MASTER

A business process harmonization approach using process mining techniques

Melissen, R.

Award date:
2013

A Business Process Harmonization Approach

Using process mining techniques

R. Melissen

Eindhoven, November 2012

A Business Process Harmonization Approach

Using process mining techniques

R. Melissen

Eindhoven, November 2012

Eindhoven University of Technology,
Faculty of Industrial Engineering & Innovation Sciences,
Information Systems Group
Supervisors: Dr. ir. R.M. Dijkman
Dr. A.J.M.M. Weijters

Capgemini Nederland B.V.
Business Analysis for Distribution Technology & Industry
Supervisor: Drs A.J. Vethman

Abstract

Business Process Harmonization (BPH) concerns the standardization of an organization's business processes while leaving room for desired variations. These variations can be desired because of, e.g., differences in legislation and market needs across countries. BPH is a relatively new field of research and so far only focused on the conceptualization of the domain and a concretization at the strategic level. At the same time, there is a need for practical pointers that can be used to do process harmonization. Therefore, this thesis focuses on the development of a concrete harmonization approach that has its practical relevance.

In this thesis we developed a BPH framework based on a survey of contemporary literature in the field of BPH. The framework consists of harmonization steps that need to be performed to arrive at a harmonized model and process model aspects which are subject to harmonization. For each harmonization step an overview of techniques has been provided which are able to execute the step with respect to one or more process modeling aspects. Next, by means of selecting a specific technique for each step in the BPH framework, a specific BPH approach has been developed. An explanation of the BPH approach is provided by means of an explanation of the in- and output of each step in the approach, as well as an explanation of the technique used to execute each step. The first step of the approach is a preparation step which concerns the development of a log file of the business process. The second step concerns the creation of process models from the log file using a process mining algorithm. The third step is a homogenization step where the process models are made comparable and the fourth step is a comparison step where a differences analysis is conducted between the process models. The fifth step is a conflict solving step which concerns solving the differences between the process models and in the sixth and final step the process models are merged into a harmonized reference model.

In order to test the practical relevance of the approach, the BPH approach is executed in a case study with a client of Capgemini Nederland BV where the approach has been performed on the purchasing process in SAP. Furthermore, the BPH approach is evaluated by means of qualitative interviews with business analysts of Capgemini Nederland BV. The business analysts at Capgemini have indicated that the BPH approach is a useful approach and describes the steps needed to develop a harmonized model which can be used to reconfigure and harmonize an organization's operational processes. Finally, by means of the evaluation of the approach, important points of improvement have been identified.

We have made a first attempt at developing a concrete BPH approach. With the help of process mining techniques the 'real' as-is process models are discovered. The as-is models are compared against each other and differences are solved in order to reflect desired behavior and become more aligned. The result of the approach is a harmonized model and an overview of performed changes on the original process models in order to comply with the harmonized model. The harmonized model and the overview of changes can be used to reconfigure the real business processes of an organization. Thus, the BPH approach assists in harmonizing an organization's operational business processes and the result will be a more standardized way of working while leaving room for desired variations.

Keywords: *business process harmonization, event log, process mining, harmonization approach*

Foreword

This master thesis finalizes my master in Operations Management and Logistics at Eindhoven University of Technology. It was a long journey which started with 3 years of secondary vocational education, 4 years of higher vocational education, a one-year pre master program and finally the two years master program. I do not regret my choices and gained much knowledge which I hope to implement in my future career.

First of all I would like to thank my supervisor from TU/e Remco Dijkman for providing me with valuable feedback and support during the master thesis project. I would also like to thank Ton Weijters from TU/e for being in the assessment committee and providing feedback on the process mining part of the thesis. I would also like to thank all my study friends from TU/e which have made the years at TU/e a real joy.

I would like to thank my supervisor at Capgemini Ard Jan Vethman for giving me this opportunity and Jef Martens, Michiel Havenith and Selda Canbaz from Capgemini who provided me with lots of feedback on the project. Anne Rozinat from Fluxicon has also helped me out on the process mining part and I really enjoyed using the Disco tool and I think it will be a great success in the future. The client of Capgemini wants to stay anonymous but the people there know that I am thankful.

My friends have always played an important part in my life and I am blessed with having such amazing friend. My mother deserves a special thank you because the door is always open, especially when I broke my leg during a soccer game at the end of the thesis. I would also like to thank my father and sister for being there whenever I needed them. Finally, my girlfriend Mattinja is the best thing that happened to me and she really talked me through this final phase of my study even do I had to miss her for half a year during my thesis due to her internship abroad.

Rens Melissen
Eindhoven, November 2012

Contents

List of Figures

List of Tables

1. Introduction

This thesis describes the development of a Business Process Harmonization (BPH) approach which is used to develop a harmonized reference model on the basis of an organization's existing business processes. BPH is closely related to Business Process Standardization (BPS), which is used to standardize an organization's business processes in order to cut costs and increase profitability. However, standardization does not always work in situation where a certain amount of variation between processes is desirable and that's where harmonization comes in play. The starting point of a BPH approach is a correct representation of the business process under investigation in the form of a business process model. By means of process mining techniques, automated discovery of a business process model is based on event data readily available in today's information systems. Process mining allows for the discovery of real business processes (instead of assumed processes) and will therefore be used in this thesis.

This first chapter introduces the research topic and performed study, starting with the research context in section 1.1. The related research and research contribution are provided in section 1.2 and section 1.3 shows the motivation for the study. Based on the related research and the research motivation the problem statement and research goals are explained in section 1.4. Subsequently, section 1.5 shows the research method and in section 1.6 the scope of the study is stated. The structure of the remainder of this thesis is provided in section 1.7.

1.1. Research context

Nowadays, many organizations try to standardize their processes to improve their performance, enable exchange of staff between departments and unify quality criteria and measures between these departments. It is shown in [1] that variation in business processes can cause customer dissatisfaction due to inconsistent business outcomes of inconsistent business processes, additional training development and maintenance costs due to multiple versions of a process, additional development and maintenance costs of IT systems that must support the variations, increased organizational complexity adding costs and risks to management, and many more costs are mentioned. Therefore, in order to overcome the costs of variations in business processes, BPS emerged to support the unification of similar business processes and the effect of BPS on the quality of performance of organizations has been analyzed, where it is clear that BPS has a positive effect on process performance [2][3].

However, a one size fits all solution for all instances of a process across the organization is not always possible and that is why standardization projects can fail. For example, a standard process for purchasing product related goods in Europe may deviate in important details from a process for purchasing product related goods in Asia and a standardized reference model for the complete organization will not support these required differences. An explanation for the failure of standardization projects is provided in [4], which classifies processes according to the scheme described in [5], which include Standard, Routine and Non-routine processes. Standard processes are set up to deal with a single variety using binary logic, designed to accept a specified type of input and to produce an ex-ante specified type of output, which means that every activity can be processed each time in an optimal way. Routine processes can distinguish a limited amount of variety using fuzzy logic and can have two or more types of inputs, and two or more types of alternative outputs. The overall aim of a routine process is usually clear, but can be achieved through different types of actions. In contrast to standard processes, routine processes show some uncertainties concerning the process execution and the assessment of a routine process cannot be reduced to a binary logic. Non-routine processes are open systems in which unrestricted variety is interpreted and assigned meaning and Non-routine processes are characterized by an unknown or vague set of inputs and outputs. The findings in [4] indicate that standard processes are suitable for BPS but that "it is more difficult to manage and standardize more complex business processes such as routines and non-routines". These findings make clear that there is a need for a different approach when standardizing processes that posses a higher degree of variation and this approach is known as Business Process Harmonization (BPH).

In line with [1], [2], [4], BPS is defined as:

> *The unification of processes based on a best in class model adapted from a reference model or other best-practice model from within or beyond the organization, and that BPS is performed to increase the performance of an organization in terms of time, costs and quality.*

In line with [6], [7], BPH is defined as:

> *Solving the differences between a set of (similar) process models to the extent possible, designing a reference model with an optimal number of process variants based on the remaining set of process models, and reconfiguring the real processes according to the harmonized reference model.*

When comparing the two definitions, it is clear that BPS uses a top down approach where processes are configured on the bases of a standard best practice or reference model which allows for a limited amount of variation in the processes. In contrast, BPH uses a bottom up approach where the existing processes are used to develop a harmonized reference model which takes into account the variations that are present in the existing processes. BPH thus focuses on the standardization of processes but leaves room for desired variations.

1.2. Related research and research contribution

In scientific literature there are only two harmonization frameworks available, namely: a conceptual framework described in [7] that tries to capture the relevant concepts in a harmonization approach when measuring the level of harmonization. Secondly, a strategic framework described in [8] is developed to support the process of harmonizing two or more reference models as companies often use multiple reference models for their processes. The strategic framework comprises all strategic elements necessary for an organization dealing with the harmonization of multiple models. Based on this existing literature, the elements of a harmonization approach are defined, e.g., different aspects of processes can be harmonized, different steps need to be performed to arrive at a harmonized reference model and techniques exist that can be used to carry out the harmonization steps. An excerpt of these techniques is given in [9] and among these techniques are, e.g., techniques to measure similarity between process models and techniques to merge process models. However, a concrete harmonization approach with practical instructions has not yet been developed. Therefore, in this thesis a practical BPH approach will be developed and executed in a case setting, which will show the usefulness of a BPH approach.

1.3. Research motivation

A BPH approach is valuable when complex organizations want to become more aligned, trying to harmonize their business processes in order to increase the performance of the organization. Large organizations often have difficulties when standardizing their processes because, e.g., due to mergers and acquisitions, these organizations have multiple versions of a process across different business units. A certain amount of variation in these processes is desired due to, e.g., different regulations and different market needs across countries. But, one standard reference model will not be sufficient in maintaining these desired variations. However, BPS has been studied in the scientific community and BPS projects have been executed in many organizations. As a result standardized reference models and standardization approaches are readily available today. In contrast to BPS, the field of BPH is relatively new and therefore little research has been conducted towards BPH and based on related research, it is not clear how a harmonized reference model can be developed. Therefore the motivation of this thesis is to develop a business process harmonization approach that can be used in a practical setting in order to design a harmonized reference model that better serves the needs of these multinational organizations. Furthermore, the development and implementation of a harmonization approach will uncover the usefulness of certain scientific techniques and pinpoint underexposed elements of a harmonization approach. Science will benefit from this because the findings will help as a guide for future research in the field of BPH.

1.4. Problem statement and research goals

Based on related research conducted in the field of BPH the following problem statement is formulated:

Problem statement
There are no business process harmonization approaches available using practical analysis techniques that can support the harmonization of business process models.

The goal of this thesis is therefore stated as:

Research goal
Develop a business process harmonization approach that can be used to create a harmonized reference model.

Sub goals
The research goal can be further subdivided into four sub goals and each of these sub goals has their results. The four sub goals and results are stated as follows:

1. *Identify existing BPH approaches and the important components of an approach, such as, process aspects that can be harmonized, steps in a BPH approach and techniques that can execute the steps in a BPH approach.*

   The result will be a BPH framework which comprises the components of a BPH approach, e.g., process aspects that can be harmonized, possible steps in a harmonization approach and harmonization techniques.

2. *Construct a BPH approach by selecting the appropriate techniques from the developed BPH framework.*

   In order to obtain the second sub goal, the second sub goal is subdivided into:
   a. *Develop selection criteria in order to select the techniques to be used in a BPH approach.*
   b. *Develop a BPH approach based on the selection criteria.*

   The result will be a set of selection criteria and a concrete BPH approach.

3. *Implement the BPH approach in a case setting by means of executing the techniques in the selected BPH approach*

   The result will be an elaboration on the execution of each step in the BPH approach.

4. *Evaluate the BPH approach.*

   The result will be an overview of possible points of improvement of the implemented BPH approach.

1.5. Research Method

The research method is schematically shown in appendix 1: research method, which consists of the goals, deliverables and the methods used. The first sub goal stated under section 1.4 concerns the development of a BPH framework and in order to obtain the first goal, a literature review will be conducted. By means of a literature review existing harmonization approaches and frameworks can be discovered, as well as steps in a harmonization approach and accompanying techniques that can support the execution of the steps. The literature findings will be used to develop a harmonization framework. Sub goal 2a concerns the development of BPH selection criteria which will be established in cooperation with Capgemini Nederland BV. Sub goal 2b contains the selection of a BPH approach which will be based on these criteria. In order to obtain the third sub goal which is the implementation of the BPH approach, a case study will be conducted in cooperation with Capgemini Nederland BV and a client of Capgemini Nederland BV. This client wishes to remain anonymous and will therefore be referred to as 'Company-Z'. Capgemini Nederland BV will facilitate in the execution of the BPH approach and Company-Z will provide the data that allows for the actual execution of the BPH approach. Company-Z wants to harmonize their purchasing processes across several business units in countries within the EMEA (Europe Middle-East and Africa) area. Company-Z needs a harmonized reference model in order to establish a shared service centre for their core business units. Their goal is to integrate the purchasing processes in a shared service centre where procurement is performed for multiple countries at ones. This will result in a more efficient purchasing organization that is able to purchase against better prices. Therefore, Company-Z will serve as a case to execute the developed harmonization approach and develop a harmonized reference model for their purchasing processes. The fourth and final goal concerns the evaluation of the executed approach. Qualitative interviews will be conducted with Business analysts of Capgemini Netherlands BV in order to assess the executed harmonization approach.

1.6. Scope

The thesis will be conduct in cooperation with Capgemini Nederland BV and Company-Z. Capgemini Nederland BV has a special interest in Process Mining which is the discovery of process models based on event data readily available in today's information systems. In order to harmonize business processes, process models must be identified that represent the business processes. The purchasing process models at Company-Z will therefore be identified with the help of process mining techniques. The SAP system used by Company-Z is based on a relational database which is suitable for the extraction of the purchasing process with process mining techniques. Due to time and data restrictions the purchasing process of 6 core countries will be used to execute the developed harmonization approach and the focus will be on the purchasing process for raw and packaging materials. The qualitative interviews, that will be used to evaluate the harmonization approach, will be conducted within Capgemini Nederland BV.

1.7. Report structure

The remainder of this thesis is structured as follows.
First some preliminary concepts are explained in chapter 2, which are used throughout this thesis. Chapter 3 discusses the development of a harmonization framework based on literature findings. Next, chapter 4 contains the development of BPH selection criteria and discusses the selection of the harmonization approach. Then, chapter 5 provides an explanation of the selected harmonization approach. Chapter 6 explains and discusses the implementation of the harmonization approach in a case study and chapter 7 shows the evaluation of the approach. Finally, chapter 8 shows the conclusions of the study.

2. Preliminaries

This chapter introduces preliminary concepts used throughout this thesis. Section 2.1 starts with the explanation of a business process, a business process model and the aspects of business process models. An explanation of SAP ERP is provided in section 2.2 and in section 2.3 we focus on Process Mining, an event log and ProM.

2.1. Business process, Business process model and Aspects

In literature many definitions exist for a business process and in the early 90's a business process is defined in [10] as:

"A business process is a set of logically-related tasks performed to achieve a defined business outcome".

And in [11] as:

"A business process involves a set of activities that are executed in some enterprise or administration according to some rules in order to achieve certain goals.

Business processes have two important characteristics. Firstly, business processes have customers; that is, processes have defined business outcomes, and there are recipients of the outcomes. Customers of business processes may be either internal or external to the firm. Secondly, business processes cross organizational boundaries, i.e., normally they occur across or between organizational subunits. Business processes that meet these characteristics are for example the development of a new product, ordering goods from a supplier, creating a marketing plan, processing and paying an insurance claim, writing a proposal for a government contract.

A business process model is a formal and details description of a business process and is used for different purposes:

- Documentation of the business process
- Better understanding of the business process
- Collaborative design of the business process
- Communication and teaching of the business process
- Analysis and verification of the business process
- Optimization and re-engineering of the business process
- Computer support and automated execution of the business process

The purpose of a business process determines which process modeling formalism and process model aspects are used. Examples of modeling formalisms are Petri-net, Event-driven Process Chain (EPC), Business Process Modeling Notation (BPMN) and many more languages exist. A company's business processes can be seen from different perspectives like, e.g., the tasks that are executed in the process (Function aspect), in which order these task are executed (Control flow aspect), the information that is used to execute the tasks (Data aspect) and the persons that are authorized to execute the tasks (Organizational aspect) and these aspects are reflected in different process modeling formalisms. For example, a simplified model of a purchasing process is shown in the EPC and BPMN notation shown in figure 1. Both models address the four before mentioned modeling aspects.

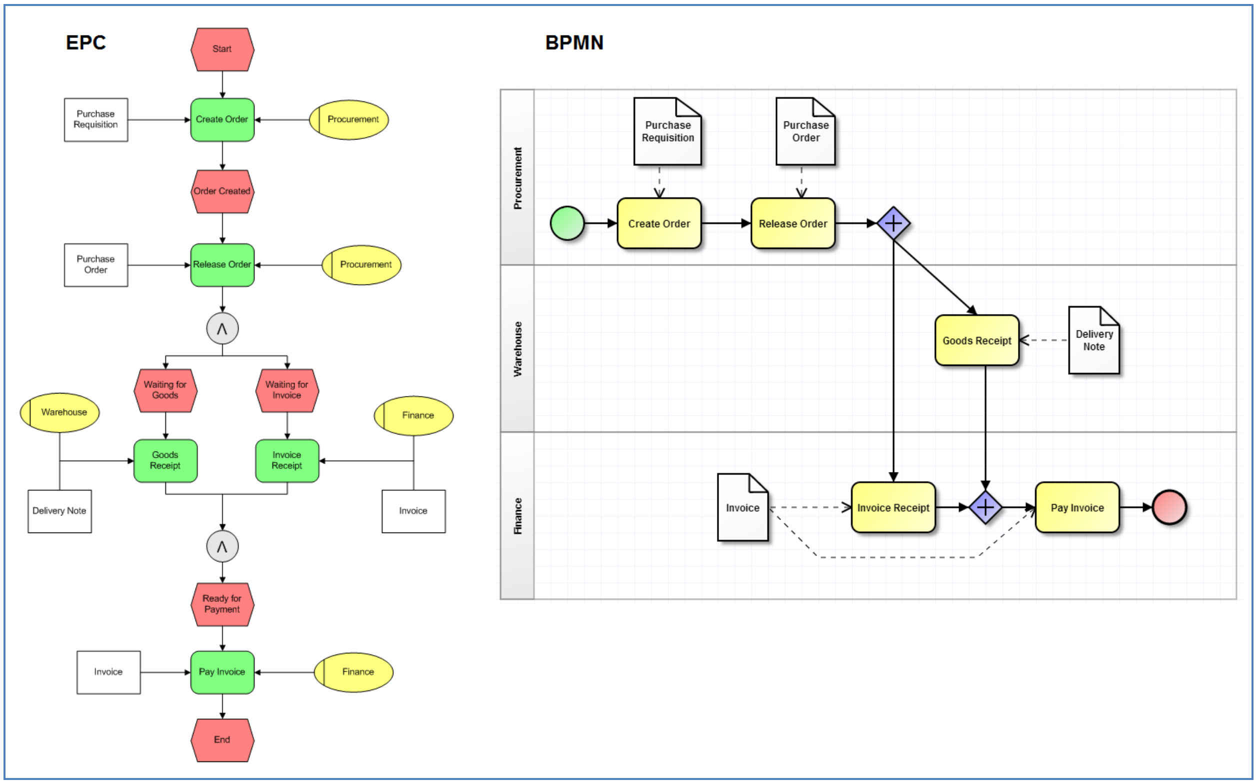


Figure 1: Purchasing process in EPC and BPMN modeling notation

2.2. SAP

SAP AG (System Analyses and Program development) is a German software company known for its ERP (Enterprise Resource Planning) software package SAP. SAP ERP supports virtually all business processes of an organization and its functions are arranged into functional modules, e.g. Finance and Controlling (FI/CO), Human Resources (HR), Production Planning (PP), Materials Management (MM) and Sales and Distribution (SD). The tasks in business processes can be executed by performing transactions in the SAP system. A transaction can be executed by entering the correct transaction code in the system or via the system's menu via the corresponding task description. SAP contains tens of thousands of tables where information is stored and every transactions causes updates in one or more of the database tables. For example when a purchase order is created, a purchase order number is created and all purchase order related data is stored in the purchase order table EKKO (Which is the Purchasing document header table, 'einaufsbeleg kopf' in German). However, when a delivery arrangement has been made during the creation of the order, this information is stored in the EKET table. In order to know which delivery arrangement is made for which order, SAP tables are linked to each other by so called Primary- and Foreign-key relationships. For example, a unique order number is created in the EBELN field of the EKKO table which serves is the Primary identifier or Key of an order (PK in the EKKO table shown in figure 2). This unique order number can also be found in the EKET table which is taken from the EKKO table and is indicated as a foreign key (FK1 in figure 2). Every executed transaction and its related data can thus be found back in the underlying structure of the database of the SAP system.

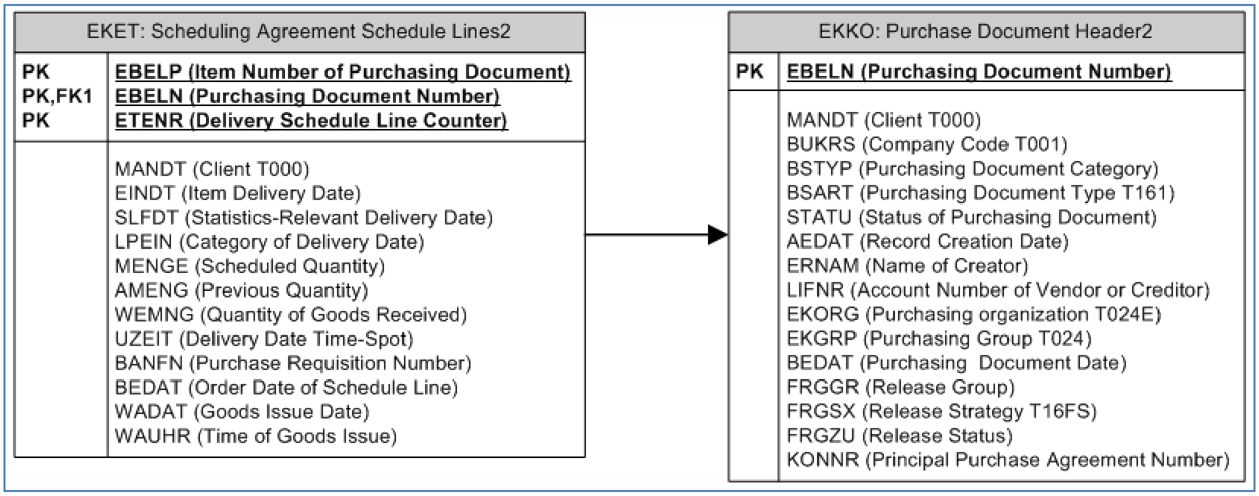


Figure 2: SAP database tables EKKO and EKET

2.3. Process Mining, Log-file and ProM

Process mining is a discipline that emerged at the end of the nineties providing comprehensive sets of tools to provide fact-based insights into processes and to support process improvement, based on event log data readily available in today's information systems [12]. A positioning of process mining is shown in figure 3, where process mining establishes links between the actual processes and their data on the one hand and process models on the other hand. Today's information systems record enormous amounts of (event) data that is captured in event logs (see figure 3). Information systems like e.g. workflow management systems (WFM), business process management systems (BPM), enterprise resource planning systems (ERP), customer relationship management systems (CRM) and all other process aware information systems (PAIS) that provide detailed information about the activities that have been executed. As shown in figure 3, event logs are used for three different types of process mining, namely: discovery, conformance and enhancement. *Conformance* compares an existing process model with an event log of the same process and shows where the real process deviates from the modeled one. *Enhancement* takes an event log and process model and extends or improves the model using the observed events. *Discovery* takes an event log and produces a process model without using any other a priori information.

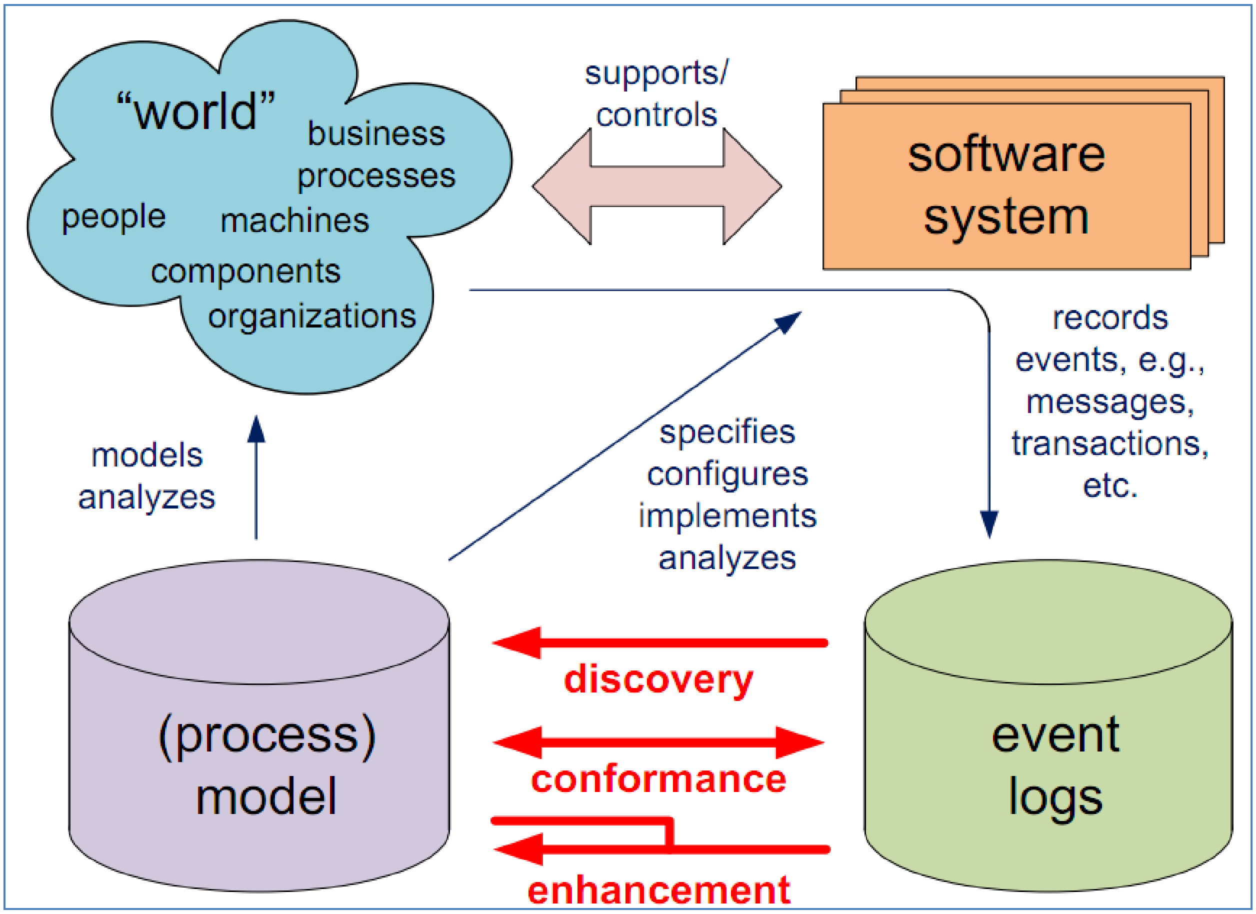
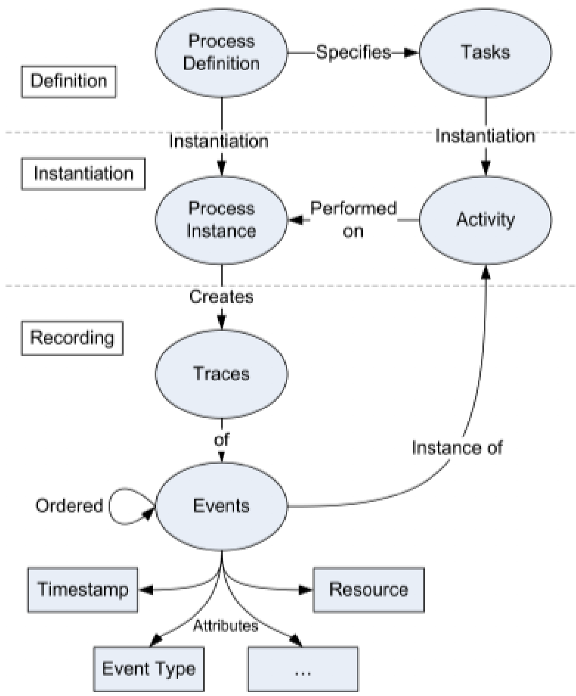


Figure 3: Positioning of process mining

When discovering process models with process mining techniques a log file has to be created that contains the execution traces of the business process under investigation. The general structure of an event log and an example event log are shown in figure 4. An explanation of the general structure is given in [13], where the *process definition* specifies which *tasks* should be executed in a process and in which structure. When a new case is started a new *process instance* of the process is generated. An example of a process instance or case are, e.g., persons, purchase orders, complaints etc. and in order to distinguish between different cases a *case ID* must be recorded in an event log. The process instance might leave a *trace* of *events* that are executed for that case in the event log. Each event is an instance of a certain *activity* as defined within the process definition. In the example log these activities are for instance named A, B, C, D or E and for case 1 the trace is for example ABCE. Furthermore, events are *ordered* to indicate in which sequence activities have occurred. In most cases this order is defined by the date and time or *timestamp* attribute of the event. Sometimes the start and stop information is recorded of a single activity. This is recorded in the *event type* attribute of the event. Another common attribute is the *resource* that executed the event which can be a user of the system, the system itself or an external system. Many other *attributes* can be stored within the event log related to the event, e.g., the data attributes added or changed in the activity. Each line in an event log thus relates to a specific event and the minimum information required to extract a process model from an event log is a case ID, activity name, event type and timestamp and additional information can be added, like the resource and other attributes. There is also a standard file format for event logs which can be used in process mining software, which is the MXML or XES event log format.



Case ID	Activity	Event Type	Resource	Timestamp	Attributes
1	A	Completed	Anne	20-7-2004 14:00	...
1	B	Completed	system	20-7-2004 15:05	...
1	C	Completed	system	24-7-2004 15:05	...
1	E	Completed	system	24-7-2004 15:06	...
2	A	Completed	Anne	20-7-2004 15:00	...
2	B	Completed	system	20-7-2004 15:07	...
2	D	Completed	Mary	20-8-2004 10:00	...
2	C	Completed	system	22-8-2004 9:05	...
2	E	Completed	system	22-8-2004 9:06	...
3	A	Completed	Anne	21-7-2004 10:00	...
3	B	Completed	system	21-7-2004 14:00	...
3	D	Completed	John	21-8-2004 10:00	...
3	D	Completed	John	21-9-2004 10:00	...
3	D	Completed	John	21-10-2004 10:00	...
3	C	Completed	system	25-10-2004 14:00	...
3	E	Completed	system	25-10-2004 14:01	...
4	A	Completed	Anne	22-7-2004 11:00	...
4	B	Completed	system	22-7-2004 11:10	...
4	D	Completed	John	22-8-2004 15:10	...
4	D	Completed	Mary	22-8-2004 17:10	...
4	C	Completed	system	29-8-2004 14:01	...
4	E	Completed	system	29-8-2004 17:30	...

Figure 4: General event log structure and event log example

Several commercial and open-source process mining software tools have been developed. ProM[1] is an extensible open-source framework developed to support a wide variety of process mining techniques. Different process mining algorithms have been implemented in ProM in the form of Plug-ins serving different process mining functionalities. ProM can read event logs in the MXML and XES event log format and supports a wide variety of filtering techniques which can be used on an event log before a mining algorithm is applied. Numerous discovery algorithms have been implemented in ProM and ProM is able to read and write a variety of file formats. A wide range of analysis techniques have been implemented in ProM and these techniques can be performed on event logs, process models or a combination of an event log and a process model. There are now more than 280 plug-ins available in ProM which makes it a comprehensive process mining tool.

[1] http://www.processmining.org/prom/start

3. Harmonization Framework

This chapter describes the development of a BPH framework based on contemporary literature in the field of BPH. As has been shown in section 1.2, two existing BPH frameworks could be identified. The framework described in [7] is a conceptual framework that tries to capture the relevant concepts in a harmonization approach when measuring the level of harmonization. The framework makes clear that different process model aspects are subject to harmonization and section 3.1 will further elaborate on these process model aspects. The second framework shown in [8] is a strategic framework developed to support the process of harmonizing two or more reference models and comprises all strategic elements necessary for an organization dealing with the harmonization of multiple models. The core of the strategic framework comprises the actual harmonization process of multiple models, which consist of steps that need to be performed to arrive at a harmonized reference model. However, the strategic framework only mentions that steps need to be performed, but the actual identification of these steps is not mentioned. Therefore, section 3.2 focuses on the identification of harmonization steps. The strategic framework also mentions that in order to execute the harmonization steps, techniques are required. But again, it is not mentioned which techniques exist and are able to support the execution of which step. Section 3.3 will therefore identify these harmonization techniques. Finally in chapter 3.4 the literature findings are combined into a harmonization framework.

3.1. Process model aspects

As described in [11], the four basic aspects of a process model are the Control, Organization, Information (Data) and Core (Function) aspect. For the harmonization framework we use the EPC modeling notation described in [14] which uses and clarifies these four aspects. The EPC notation is also used to model business processes in SAP and in chapter 6 a case study is conducted related to SAP. Therefore, EPC will be used in the harmonization framework. The EPC notation uses the 4 different aspects of a business process that together form a business process model. These 4 aspects are shown in table 1. The component model aspect shows which tasks or functions are executed within an organization and the organizational model aspect shows which organizational units and accompanying roles are involved in executing the tasks. The data model aspect shows what information is needed to execute the tasks and the interaction model aspect shows the interaction between tasks and roles. These 4 aspects together form the EPC process model which is shown in figure 5.

EPC Aspect	Explanation
Component (Function)	Business process tasks or functions.
Interaction (Control)	The order in which the tasks of a business process are instantiated.
Organization	The structure of the organization in which the business process is executed, its organization units and the relations among them; and it defines the resources and persons within these organization units.
Data	The information involved in a business process.

Table 1: The four EPC modeling aspects and explanation

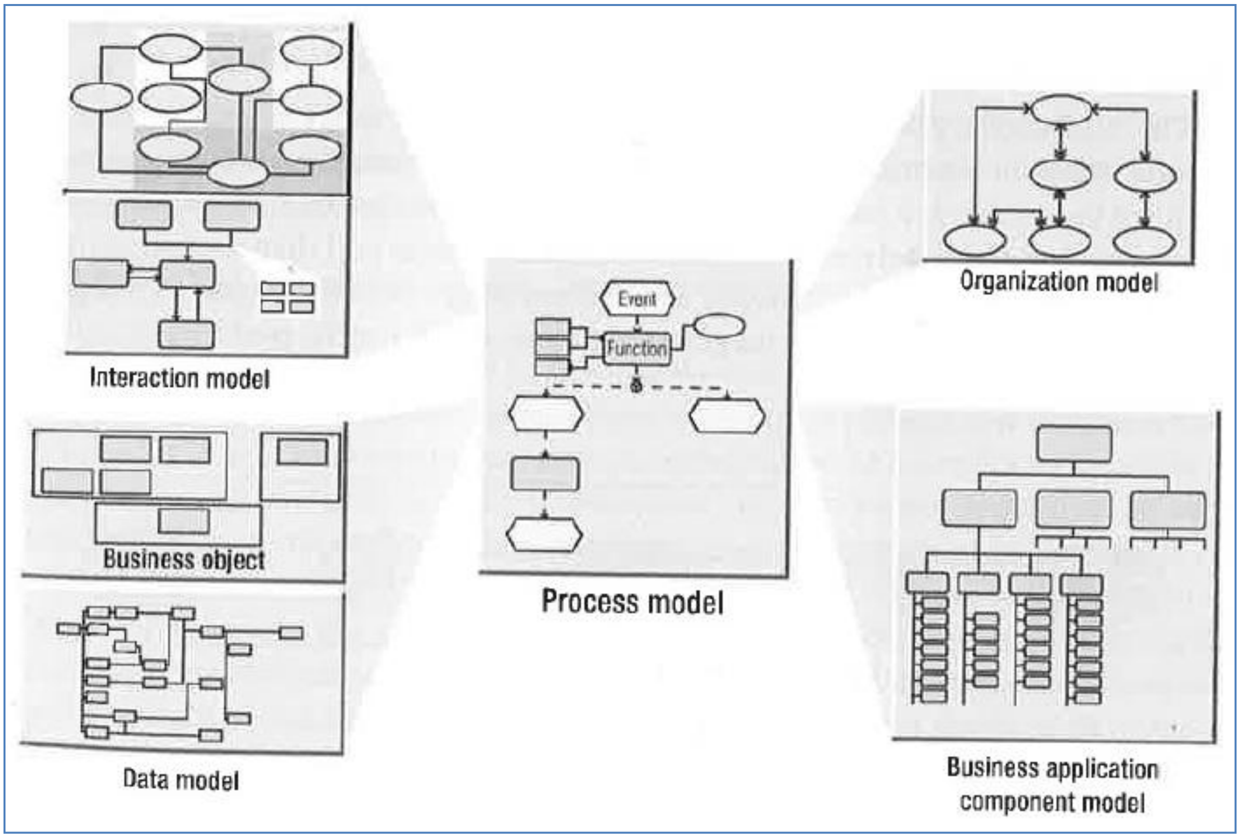


Figure 5: The aspects of the EPC modeling notation

3.2. Harmonization steps

Different steps need to be performed to design a harmonized reference model based on a set of underlying process models. This section will provide an overview of possible harmonization steps.

The harmonization steps have been discovered in [15] and the most important steps will be discussed. Several steps in a harmonization strategy are described in [16] which emerged from there harmonization framework shown in [8]. Three main steps are identified, namely:

1. Homogenization; setting in harmony the models involved, adding their information by means of a common schema or common structure of process entities.
2. Comparison; to carry out the identification of differences and similarities between multiple models.
3. Integration; combining and/or unifying the best practices of multiple models.

A number of main steps are described in [17] in their framework for business process design. The framework is used when different stakeholders model the same business process on their own, and these separate models are used to produce a more complete model. A harmonization approach also comprises process design where several business process models are used to design a harmonized reference model. Four main steps are identified, namely:

1. Identifying the processes
2. Identify conflicts with process comparison
3. Solve conflicts
4. Integrate processes

Furthermore, the scope in section 1.6 states that process models need to be discovered with process mining techniques. A preparation step is required when discovering process models with process mining techniques and in the preparation step a log file will be constructed that serves as input for the discovery step.

In conclusion, the following steps in a harmonization approach are distilled from the before mentioned steps, which are shown in table 2 and will be used in the harmonization framework.

Step	Explanation
1. Preparation	Constructing a log file of the business process
2. Discovery	Identifying the processes that are subject to harmonization.
3. Homogenization	Make the processes comparable, e.g. using the same modeling notation.
4. Comparison	Identification of differences and similarities between models.
5. Resolve conflicts	Solve conflicts between the models to the extent possible.
6. Integration	Combining the remaining set of models in a harmonized reference model.

Table 2: Harmonization steps

3.3. Harmonization techniques

Each harmonization step identified in section 3.2 can be supported by specific harmonization techniques and for each harmonization step the techniques will be discussed. First, preparation techniques will be discussed in section 3.3.1 and process discovery techniques in section 3.3.2. Then, homogenization techniques will be discussed in section 3.3.3 and process comparison techniques in section 3.3.4. Next, conflict solving techniques will be discussed in section 3.3.5 and finally an elaboration on integration techniques will be provided in section 3.3.6.

3.3.1. Preparation techniques

Nowadays, many multinational companies work with ERP systems like, e.g., SAP. SAP ERP is an integrated information and control system in which business processes can be defined and managed. The business processes are incorporated within different modules and the data in the modules can be mutually interchanged, creating a fully integrated system. Within SAP all information is stored in a hierarchical way and this structure is reflected in the structure of the relational database. However, the information recorded in the relational database, when executing task in a particular business process, is not stored in a, for process mining correct way. In other words, SAP does not store the data in an event log format that is suitable for process mining. But, it is possible to reconstruct an event log suitable for process mining by extracting the correct information from the database. This problem has been investigated and a method has been proposed in [18] to construct an event log from a SAP process. The study shows how to detect the relevant database tables in the process, but the actual construction of the event log has not been performed. The study performed in [13] has resulted in a usable application which makes it possible to create an event log based on the tables of a relational database. Finally, in [19] a method is proposed that guides the extraction of event logs from SAP, but mainly focuses on how to incrementally update a log file with only the changes from the SAP system that were registered since the original event log was created.

3.3.2. Discovery techniques

Business process discovery can be done in different ways and basically two techniques can be distinguished, namely manual discovery and automated discovery. The manual technique (also known as the traditional method of process discovery) as well as the automated technique (also known as discovery with process mining techniques) will be discussed.

Traditional method
Traditional business process discover techniques include interviewing managers and related business participants, making documentations and drawing diagrams step by step so as to incrementally identify the inputs, outputs, purposes, rules, etc., that govern a specific process. Detailed implementation of the above procedure is based on specific workshop technique that the discovery team exploits and process analysts have to join the pieces of information together to form the process flow. The traditional method typically follows three phases [20], starting with discovery workshops. Process Analysts organize workshops or meetings with the required domain expert or process owners. Domain experts provide a view of the process that they follow in their domain. Analysts take note of all the information from different domain experts so that it can be used to define the entire process and analysts can ask leading questions to understand the process handovers between different units. The next phase is modeling the As-Is process view. All the pieces of information are put together, the process flow is defined and different process models should be created and linked to provide a comprehensive As-Is process view. Finally, review workshops are conducted where process analysts organize a review of the process defined. Any review comments are treated and the process model will be modified in response. Review ensures that the process documented or modeled in line with the provided information. While conducting the process review, the analyst should ensure that all the conflicting points during the process discovery discussion sessions are reviewed and then agreed upon by the related stakeholders.

Process mining
Discovery of process models with process mining techniques takes an event log and produces a process model without using any other a priori information. Numerous algorithms have been developed that are able to produce a process model based on event data and these techniques are summarized in several literature reviews [21][22][23]. The review by [23] appears the most comprehensive and will be uses in this thesis. An overview of discovery algorithms based on [23] is provided in appendix 2: process mining discovery algorithms and reference list. The overview shows the name of the algorithm (if present), on which technique the algorithm is based, the modeling notation used, if the algorithm is implemented in a software tool and which process modeling aspect(s) the algorithm is related to.

3.3.3. Homogenization techniques

In order to make process models comparable a homogenization step is required. Activities in process models can have different labels when they actually refer to the same activity and business processes can be modeled in different modeling formalisms. In the former case a matching or mapping step is required and in the latter case a conversion of modeling formalism might be necessary.

Matching/mapping

Before a comparison can take place, [24] and [25] mention that the first step in business process model comparison is to determine which activities in one business process model correspond to which activities in the other. This step is also known as the matching or mapping step and five different ways of measuring similarity between model elements are mentioned in literature [26], namely: Syntactic similarity measure, Linguistic/Semantic similarity measure, Attribute similarity measure, Type similarity measure and Structural/Contextual similarity measure. The syntactic similarity measure uses the edit distance between two task labels, which is defined by the number of changes (addition, deletion and replacement of characters) necessary to turn one string into another, thus the greater the edit distance, the more different the strings are. The linguistic/semantic similarity measure is based on equivalence between the words that task labels consist of. The attribute similarity measure does not look at node labels but measures the similarity between the attribute values of the nodes and the type similarity measure also does not look at node labels but measures the similarity of the nodes type. Finally, when determining the similarity of two model elements, the structural/contextual similarity measure also takes the model elements that precede and succeed these elements into account. This matching or mapping step of process activities is used when comparing complete process models, where a match between model elements can be made by hand or automatically by using the before mentioned techniques.

Conversion

A conversion step might be necessary when process models are created in different notations. The desired modeling notation must be selected and the process models must be converted into the desired notation. Many conversions between models exist and conversion between models is for instance possible with the ProM tool, which support several conversions between different modeling notations.

3.3.4. Comparison techniques

Process comparison consists of identifying the differences and similarities between two business process models. As described in [9], similarity techniques mainly focus on the development of methods that, given two process models, return the similarity of those two models, on a scale from 0 to 1. In a BPH approach, similarity measuring is not very useful as we are looking for techniques which are able to point out the exact differences between two models in order to solve these differences.

Thus, when comparing business process models, similarity measures indicate the extent to which business process models are equal. But, if similar business process models are not entirely equal, the similarity metrics do not give any insight in how these process models are different. In order to provide insight in how process models are different a classification of frequently occurring differences between similar business processes is presented in [27]. The classification is independent of any modeling notation, but throughout the paper UML is used for illustrative purposes. Techniques have been developed to point out where two processes are different and the classification shown in [27] is used to explain the type of a difference in [28] and [29], where EPC [28] and BPMN [29] are used as modeling notation. Another method is proposes in [30] and [31], that tries to identify business process logic differences between a company's processes and best practice processes based on ARIS flowcharts.

3.3.5. Conflict Solving techniques

The extent to which differences between business processes occur and the extent to which these differences can be resolved is dependent on process variation factors that are present in the organization. Little research has been conducted towards an understanding of these factors, although a number of these process variation factors are mentioned in literature.

Various reasons for process variation are described by [1], including legislative requirements. Legislation is in important factor for process variation and variation in processes due to legislation is caused by, e.g., differences in financial regulations, taxation regimes and import/export regulations, which often result in mandatory and unavoidable variations in a process. Legislation is also identified by [7] and [32] as being an important factor of variations between process models. According to [1], differences in product and services may require variation in the processes that create, deliver and maintain them, and this product/service type factor has also been identified by [7] as being an important factor. IT systems are mentioned by [1] and [32] as being an important factor for process variations, where IT systems, particularly legacy systems, may force variations in business processes. In large companies many different job descriptions exist and if jobs are available in one location but not in the other, this can cause differences in business processes [1][32]. Therefore, Resources/jobs is identified as a factor causing process variation. Furthermore, [1] and [32] have a set of less obvious factors that can induce process variations. Among these factors are: personal preference, legacy process, local market imperatives and language/culture. Personal preference means that an individual with authority causes variation in processes by e.g. creating or changing the process according to its own understanding [1]. Legacy process often happens when organizations, and processes, merge after an acquisition and multiple versions of a process remain in the new organization [1]. Local market imperatives are caused by customer expectations, market maturity, competitive landscape or local market conditions and can have a significant effect on process variation [1] and Language/culture difference may also effect the amount of variation between business processes [32]. Variation in business processes are often enforced by business rules and according to [33] a business rule is a guideline to influence or guide the conduct of business. Business rules can thus be seen as a factor that causes variation in business processes. Finally, another factor is the business process type that determines the amount of variation in a business process. According to [5], processes are standard, routine or non routine and [34] classify processes as being either artistic or scientific. The scheme presented by [5] is used by [4] and [35] to clarify which factors determine business process standardization success and [4] and [35] clarify that not all process can be unified. This means that business process type is a factor that determines to what extend processes can be harmonized and thus conflicts between processes can be solved.

Although some factors have been identified in literature, a full understanding of factors determining the extent of variation between similar business processes and the extent to which differences between processes can be resolved remains unclear.

3.3.6. Integration techniques

When combining or integrating a set of process models, this is also known as process model merging. Process model merging is described by [9] as merging a collection of process variants into a consolidated process model. Several merging techniques have been proposed in literature [36][37][38][39][40][41], but there are some fundamental differences between these techniques. The techniques are classified by [9] based on three aspects. First of all, there is a distinction between merging techniques where the merged model still allows for the behavior possible in any of the original models, and techniques where it is not guaranteed that the behavior of the original models can be correctly replayed in the merged model. Second, a distinction is made between merging techniques that can only merge models which have identical task labels or techniques which can merge models that have similar task labels, which means that task labels do not have to be exactly identical. Finally there are merging techniques that can only merge models with a certain modeling notation and there are techniques that are formalism independent. Another aspect of merging techniques that can be identified is that some techniques are only able to merge pairs of process models while other techniques are able to merge multiple models at once, and some of the techniques have been implemented in a software tool. A comparison of the proposed methods is provided in table 3.

Paper	Behavior preservation	Identical task labels	Similar task labels	Formalism	Model merging quantity (At-once)	Software support
[36]	No	Yes	No	Petri-net	Two models	No
[37]	Yes	No	Yes	EPC	Two models	Yes (ProM)
[38][39]	Yes	No	yes	Directed Graph	Two models	Yes (Synergia)
[40]	Yes	Yes	No	Business process Graph	Multiple models	No
[41]	No	Yes	No	UML	Two models	No

Table 3: Comparison of process model merging techniques

In summary, it is clear that there are several merging techniques available that have their similarities and differences. The techniques by [37],[38] and [39] seem the most mature, because the behavior of the original models is maintained in the merged model, approximately equal task labels can be merged instead of identical task labels and both techniques are implemented in a software tool. The disadvantage is that it is not able to merge multiple models at once, although in order to merge a collection of process variants, it is possible to merge a first pair and then add one variant at a time to the merged process model.

3.4. Business process harmonization framework

The literature findings of sections 3.1, 3.2 and 3.3 have been aggregated into a BPH framework which is shown in figure 6 on page 19 and an explanation of the framework follows next.

BPH concerns the development of a harmonized reference model based on the underlying business processes in an organization. In order to arrive at a harmonized reference model, section 3.2 mentions six steps that need to be performed, namely: *Preparation*, *Discovery*, *Homogenization*, *Comparison*, *Resolve Conflicts* and *Integration*. These steps are shown in the framework. Each harmonization step in the framework can be executed for different aspects of a business process model and the EPC modeling aspects mentioned in [14] in section 3.1 have been used in the framework, which are the: *Function* aspect, *Control* aspect, *Organization* aspect and *Data* aspect. For each harmonization step and process aspect in the framework a certain harmonization technique is needed. These techniques have been discovered in section 3.3 and the specific paper that mentions the technique is shown in the yellow rounded rectangles in the framework. Some techniques only affect one specific modeling aspect, e.g., the technique described by [42] shows how to discover the organization model of a business process and this concerns the discovery step and the organizational aspect within the BPH framework. Other techniques have their impact on more than one modeling aspect, e.g., the integration techniques mentioned in the BPH framework all have their effect on the function aspect as well as the control aspect of business process models. The framework also makes clear that multiple techniques exist for each step and aspect in the framework. Furthermore, not all discovery techniques that are based on process mining techniques are selected for the framework and only those techniques that are implemented in a software application are mentioned (source [43–51] in the framework).

The BPH framework will serve as input for the selection of a specific approach consisting of a specific technique for each step and aspect of the framework. The selection of a specific technique for each step and process model aspect in the framework will be discussed in chapter 4.

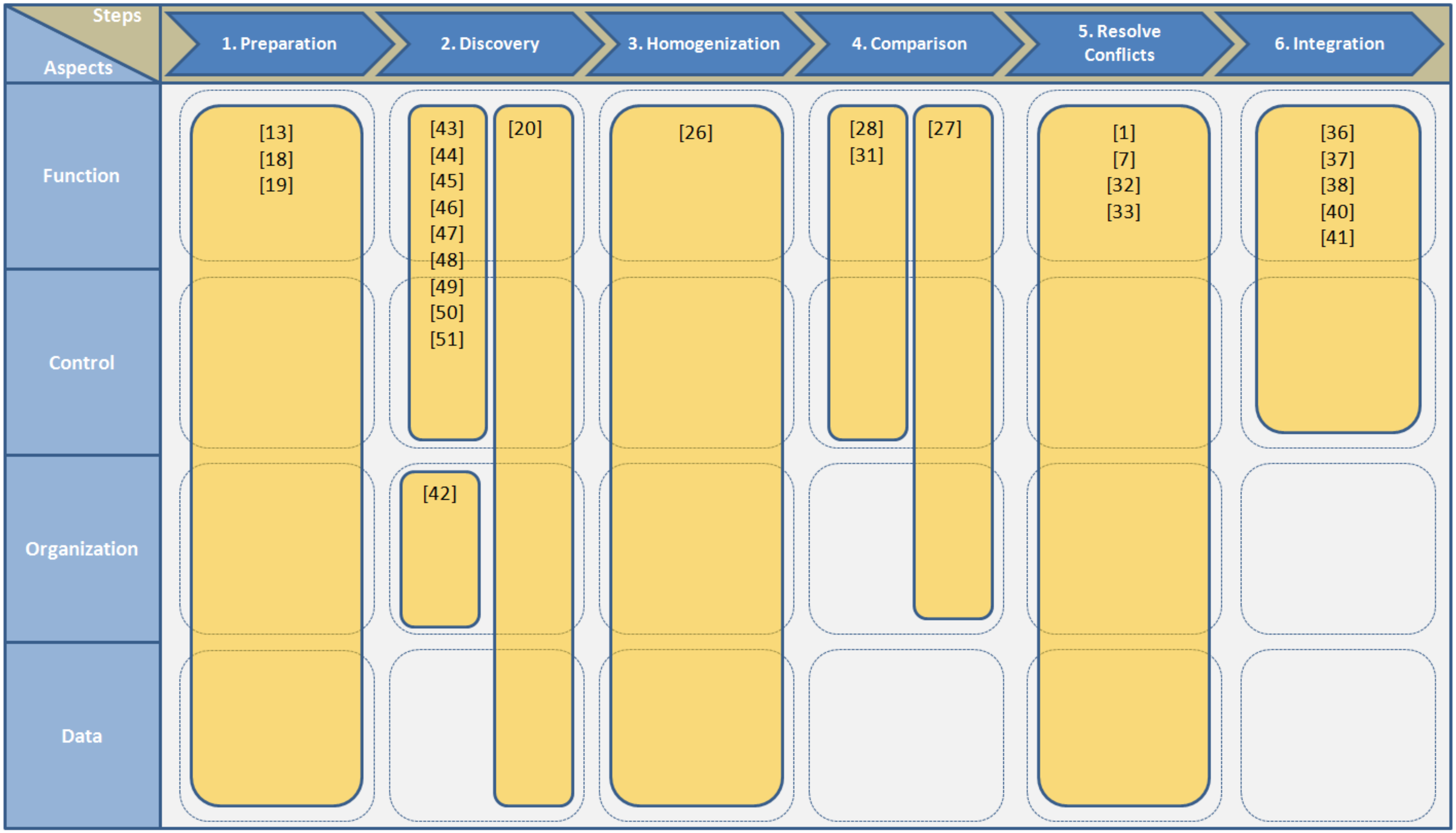


Figure 6: Business Process Harmonization Framework

4. Harmonization Criteria and Selection of Approach

In order to select a specific approach from the harmonization framework developed in chapter 3, selection criteria must be established. The development of the selection criteria will be discussed in section 4.1. The selection of the approach based on the developed selection criteria will be discussed in section 4.2.

4.1. Selection criteria

As multiple techniques exist for a part of the steps in the harmonization framework, selection criteria have been develop to select a specific technique from the framework which will form the harmonization approach. The selection criteria have been established in cooperation with a BPM/business analyst expert at Capgemini Nederland BV. General selection criteria have been established which apply to all harmonization steps. Each step in the approach should be executable by using a software tool. There are techniques that are supported by a specially developed software tool that can only be used for that technique. As different techniques will be used in a harmonization approach, it is preferred to use a software tool that incorporates many techniques and makes it possible to interchange results. An important tool which comprises many of the techniques described in the harmonization framework is the ProM tool described in [52]. The ProM tool also makes it possible to export and convert results in different file formats. Therefore, Capgemini indicates that techniques that are incorporated within the ProM tool are preferred. Furthermore, Company-Z uses SAP ERP and the EPC modeling formalism is used to document the processes in SAP. Capgemini mainly makes use of the BPMN modeling formalism and therefore, the preferred modeling formalism is the EPC or the BPMN modeling notation.

For the discovery step more specific selection criteria are required as multiple techniques exist for this step. For the discovery step a specific criteria has been established. The discovery step proposes many process mining techniques and with the before mentioned general criteria no single selection can be made. Several process mining techniques are implemented within ProM and are able to discover process models in the EPC notation. The important difference between these techniques is that they are either able to detect all behavior of a process model, which means also all exceptions, or only the main behavior of the model ignoring the exceptions. Many exceptions are possible in the SAP system used by Company-Z and including the exceptions will make the comparison of models highly complex. Therefore, in context of this thesis and in agreement with Capgemini only the main behavior of models will be discovered. However, exceptions are part of real behavior and cannot simply be ignored and when executing a full harmonization project, also the exceptions should be taken into account. With the established criteria a specific harmonization approach can be made and the general and specific selection criteria are shown in table 4.

Harmonization Step	Type	Criteria
All steps	General	1. Make use of the ProM process mining software tool where possible 2. EPC or BPMN is the preferred modeling formalism
Discovery	Specific	3. Use a discovery mining algorithm that is able to discover the main behavior of the business process (not all exceptions).

Table 4: Selection criteria

4.2. Selection of harmonization approach

For each step in the harmonization framework shown in figure 6 on page 19, a specific technique can be selected based on the selection criteria. The harmonization framework in figure 6 on page 19 has shown that for the function and control aspect, for each harmonization step, techniques are available that are able to execute the step. With respect to the remaining aspects (Organization and Data), for some steps no techniques have been discovered. Thus, for the organization and data aspect no complete approach can be developed. Therefore, a choice will be made between techniques that affect the function and control aspect. The selection of the technique for each step in the approach will be discussed next.

Step 1: Preparation

The BPH framework in figure 6 on page 19 shows that only three papers describe techniques to construct an event log using SAP, which means that no selection criteria are needed to select a specific technique. In [18] and [19] an event log extraction procedure is described. The extraction procedure described in [19] is the most complete as it shows how to construct the actual event log. Therefore, it will be used in the harmonization approach. In [13] an application is developed which makes it possible to convert SAP data to an event log and this application will be used in the selected extraction procedure.

Step 2: Discovery

The BPH framework in figure 6 on page 19 shows that there are 9 scientific papers which discuss a mining algorithm that has been implemented in a software tool and which are able to discover process models in terms of the function and control aspect. The first selection criteria discussed in section 4.1 states that the ProM tool should be used when possible and the algorithms discussed in [43][45][47][48][51] have been implemented in the ProM tool. The fuzzy miner described in [51] is not able to show the models in the EPC or BPMN modeling formalism and thus does not fulfill the second criteria in section 4.1. The third selection criteria in section 4.1 states that only main behavior should be discovered by the algorithm. The multi-phase miner explained in [47] aggregates all instances that are executed into a process model. However, the aggregated process model allows for more behavior then only the main behavior and does not meet the third criteria. The genetic algorithm described in [43] is able to discover exact process models, showing also exceptional behavior of a business process and thus does not meet the third criteria. What remains, is the α-algorithm described in [45] and the heuristic algorithm described in [48]. The α-algorithm is able to detect exceptional behavior but is not able to detect short loops in a process which can be part of the main behavior of the business process. Therefore, the most suitable algorithm is the heuristic algorithm described in [48], which is able to express the main behavior registered in an event log. The heuristics algorithm will be used in the harmonization approach.

Step 3: Homogenization

A match between process models elements must be made before a comparison can be performed. The BPH framework in figure 6 on page 19 shows that in [26] techniques are described which are able to make a match between process model elements. These matching techniques will be used in the homogenization step of the harmonization approach.

Step 4: Comparison

When two process models are different, the techniques described in [28] and [31] are able to point out where the processes are different. Solely the technique described by [28] is implemented in the ProM tool and is able to handle the EPC notation, complying with the first and second criteria in section 4.1 and will be used in the harmonization approach. Furthermore, the technique described in [28] is able to point out the differences in terms of the function and control aspect.

Step 5: Resolve Conflicts

The extent to which conflicts between models can be solved is dependent on factors that are present in the organization. For each difference between the models a decision must be made of which elements should be maintained in the harmonized model. These decisions are made by Company-Z and are made based on, among others, the factors described in [1], [5], [7], [32], and [33].

Step 6: Integration

The BPH framework in figure 6 on page 19 shows that there are 5 scientific papers that describe a process model merging technique, where the techniques described in [37][38] have been implemented in a software tool. The technique described in [37] is the only technique implemented in the ProM tool and meets the first criteria in section 4.1. The technique described in [37] uses the EPC modeling notation, thus also meets the second criteria in section 4.1 and can merge models in terms of the function and control aspect. Therefore, the technique described in [37] will be used in the harmonization approach.

An overview of the selected techniques is shown in figure 7 where for each step in the approach the reference is shown which describes the technique.

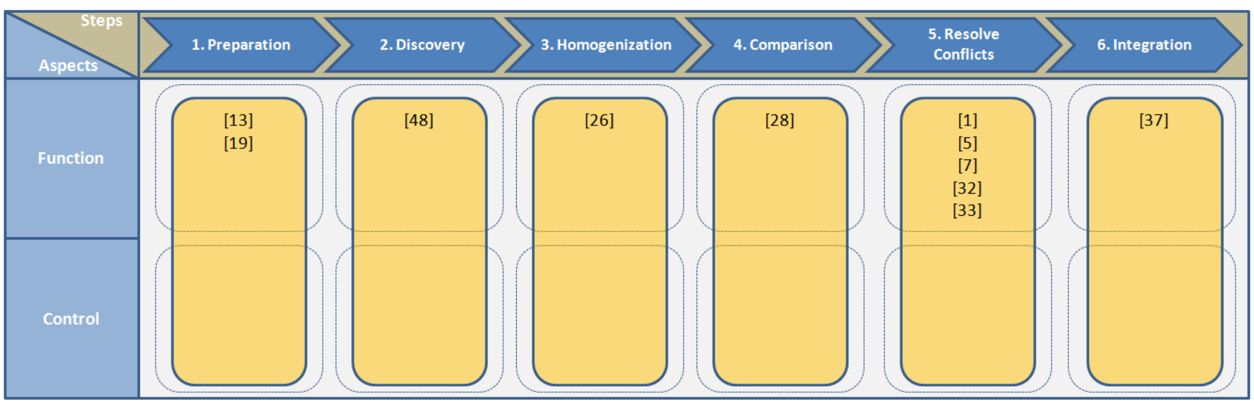


Figure 7: Selected business process harmonization techniques

5. Harmonization Approach

The harmonization approach is shown in figure 8, where for each step in the approach the input and output is provided, as well as the tool and plug-in used to execute the step. The homogenization (matching), comparison, conflict solving and integration step can only be performed on two models at once. Therefore, it is shown in figure 8 that iterations are required when more than two models need to be harmonized. The specific technique(s) used to execute the steps have been selected in section 4.2 and will be explained in sections 5.1 - 5.6.

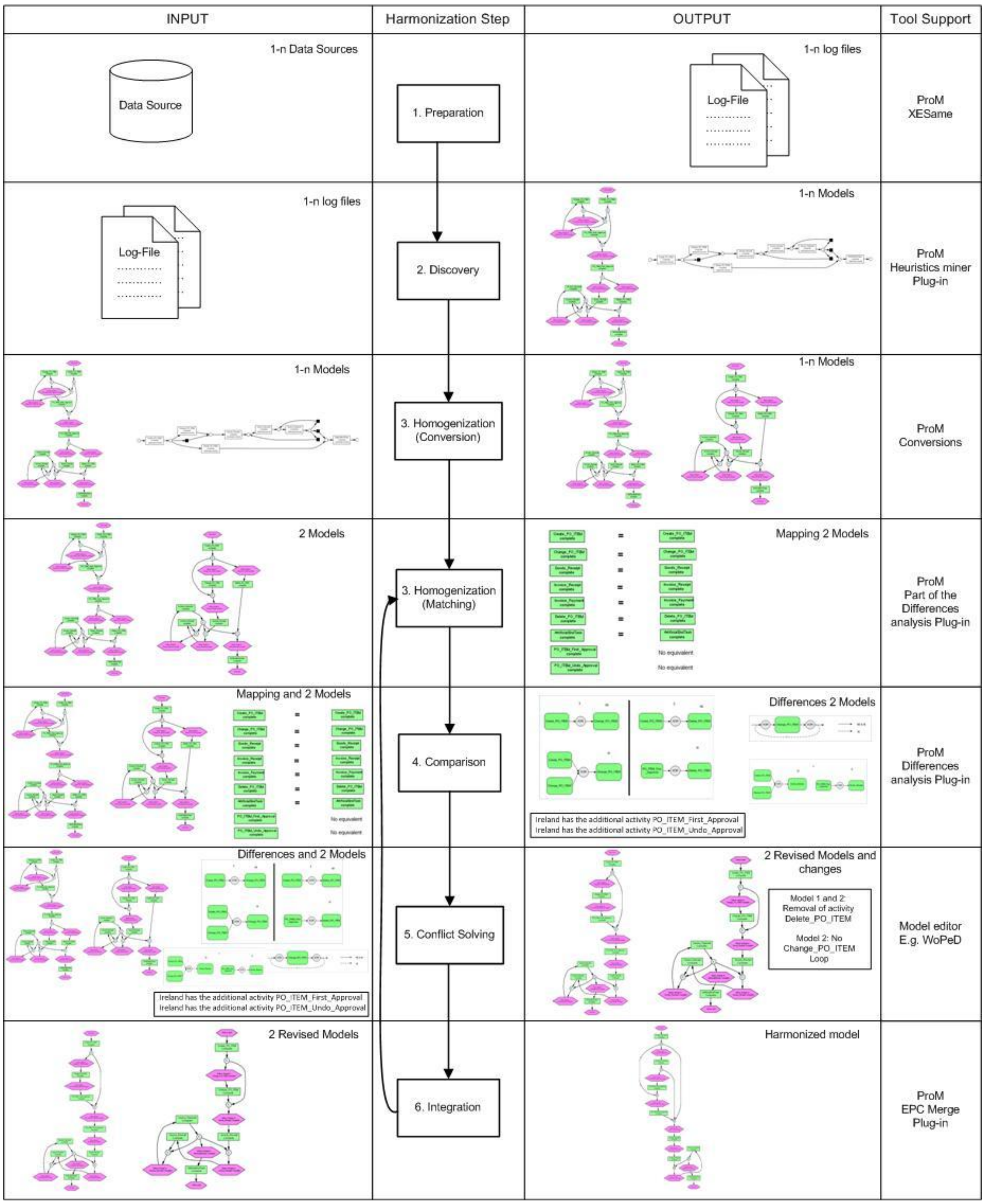


Figure 8: Business Process Harmonization Approach

5.1. Step 1: Preparation

The preparation step makes use of the event log extraction procedure described in [19] which is shown in figure 9. The procedure consists of a preparation phase and an extraction phase consisting of 6 steps. The input of the preparation step is a data source and the output is a log file in the MXML/XES event log format as shown in figure 8 on page 23.

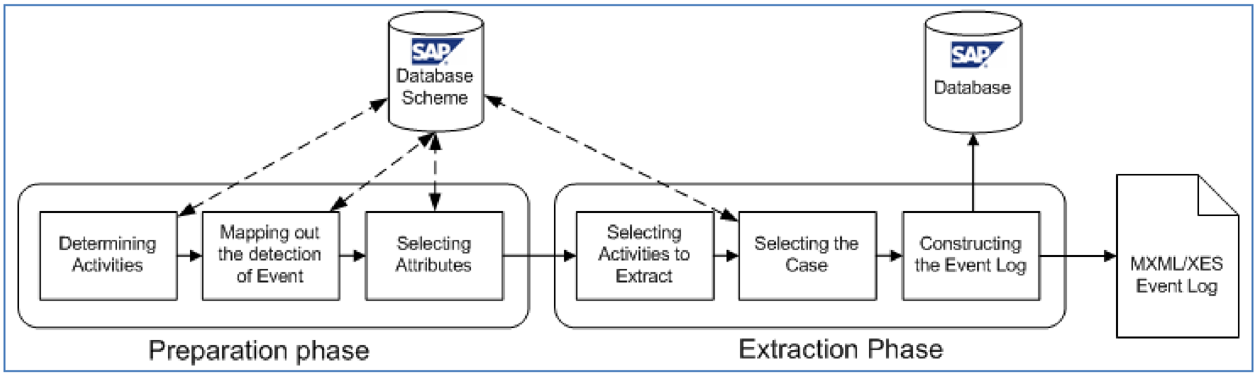


Figure 9: SAP event log extraction procedure

The first step concerns determining the process activities and according to [19] this can be done in the following ways:

1. Consult a SAP best practice reference model to identify the activities
2. Look in the SAP system itself to find the possible transactions that are related to the process
3. Consult the SAP community network on the internet which contains many literature
4. Talk to the process executor that actually executes the related activities
5. Consult a SAP expert how is specialized in the implementation of the system
6. Detailed changes are logged in, so called, change tables and activities can be identified by looking at changes that took place

The second step concerns mapping out the detection of events. In other words, for each event that took place the location of the case ID, timestamp and resource have to be detected in the relational database. This information is however stored across multiple database tables and a mapping should be made to identify the relations between the tables and the location of the relevant data fields.

An event log contains the case ID, timestamp and resource of the executed activities in a process. However, more attributes can be selected that provide information on the process. In the third step a selection of these desired attributes must be made. In the first step all process activities have been identified and in the fourth step a selection of the desired activities must be made. The selection of relevant activities is dependent on the defined project scope and goal.

The fifth step concerns the selection of an appropriate case and as described in [19], a case is a valid case for an event log if there is a way to link each event in the event log to exactly one instance of that case. This means that one case is selected and used as case throughout the process, a case is, e.g., a person/patient, a document, an application, a complaint, etc. The selection of the case is again dependent on the defined scope and goal of the project.

The final step in the procedure is the actual construction of the event log by querying the SAP database, based on the before mentioned steps. The case for the process has been identified in step 5. For the selected activities under step 4, a mapping of the events is made under step 2. This mapping should be converted to a SQL query and this can be done by using the XESame tool described in [13]. In the tool it is also possible to add the, under step 3, identified attributes to the query. The tool guides the definition of a conversion and the conversion can be defined without the need to program. The application is able to execute the conversion on a data source, producing an event log in the MXML or XES event log format. The tool makes it possible to connect to a data source and the tool can communicate with the data source via JDBC API (Java Database Connectivity Application Programming Interface). The JDBC API contains the language which enables communication between the XESame tool and the data source. For a full explanation of the tool we refer to [13] or section 6.3.1 where the preparation step is performed in a case setting.

5.2. Step 2: Discovery

The heuristics miner described in [48] is selected to discover the main behavior in an event log. As can be seen in figure 8 on page 23, the input of the heuristics miner is an event log in the MXML/XES format and the output is a heuristics net which can be converted to a desired modeling notation like EPC. The heuristics miner is implemented in the ProM tool and takes frequencies of events and sequences into account when constructing a process model. When using the algorithm, eight different settings can be entered which will be explained next and the eight settings are:

1. Dependency divisor
2. All-activities-connected heuristic
3. Dependency threshold
4. Positive observations
5. Relative-to-best-threshold
6. Length-one-loops threshold
7. Length-two-loops threshold
8. AND threshold

The starting point for the heuristics miner is a dependency graph. A frequency based metric is used in the dependency graph which indicates if there truly exists a relationship between two events a and b, shown as $a => w\ b$. This dependency measure is stated as follows:

Definition 1:
$Let\ W\ be\ an\ event\ \log over\ T\ , and\ a, b\ \in T.$
$Then\ |a > w\ b| is\ the\ number\ of\ times\ a\ is\ directly\ followed\ by\ b\ in\ W, and$

$$a => w\ b = \left(\frac{|a > w\ b| - |b > w\ a|}{|a > w\ b| + |b > w\ a| + 1}\right)$$

The "+1" in the formula is the *dependency divisor*.

The following event log will be used to construct the dependency graph:

$$W = [\langle a, e\rangle^5, \langle a, b, c, e\rangle^{10}, \langle a, c, b, e\rangle^{10}, \langle a, b, e\rangle^1, \langle a, c, e\rangle^1, \langle a, d, e\rangle^{10}, \langle a, d, d, e\rangle^2, \langle a, d, d, d, e\rangle^1]$$

First, the $=>w$ values between all activity combinations are calculated and are shown in table 5.

$=>w$	a	b	c	d	e
a	$\frac{0}{0+1}=0$	$\frac{11-0}{11+0+1}=0.92$	$\frac{11-0}{11+0+1}=0.92$	$\frac{13-0}{13+0+1}=0.93$	$\frac{5-0}{5+0+1}=0.83$
b	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0}{0+1}=0$	$\frac{10-10}{10+10+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
c	$\frac{0-11}{0+11+1}=-0.92$	$\frac{10-10}{10+10+1}=0$	$\frac{0}{0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
d	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0-0}{0+0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{4}{4+1}=0.80$	$\frac{13-0}{13+0+1}=0.93$
e	$\frac{0-5}{0+5+1}=-0.83$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0}{0+1}=0$

Table 5: Dependency measures between five activities based on event log W

The heuristic miner can work with the *all-activities-connected heuristic* which means that we know that each non-initial activity must have at least one other activity that is its cause and each non-final activity must have at least one dependent activity. Using this information in the so called *all-activities-connected heuristic*, for each activity the highest dependency measure is used to build the dependency graph. For example, the highest dependency measure of activity a is 0.93 which means that the relationship between a and b is included in the dependency graph. The dependency graph based on the *all-activities-connected heuristic* is shown in figure 10. The numbers on the arcs show the frequency of the relation and the calculated dependency measure.

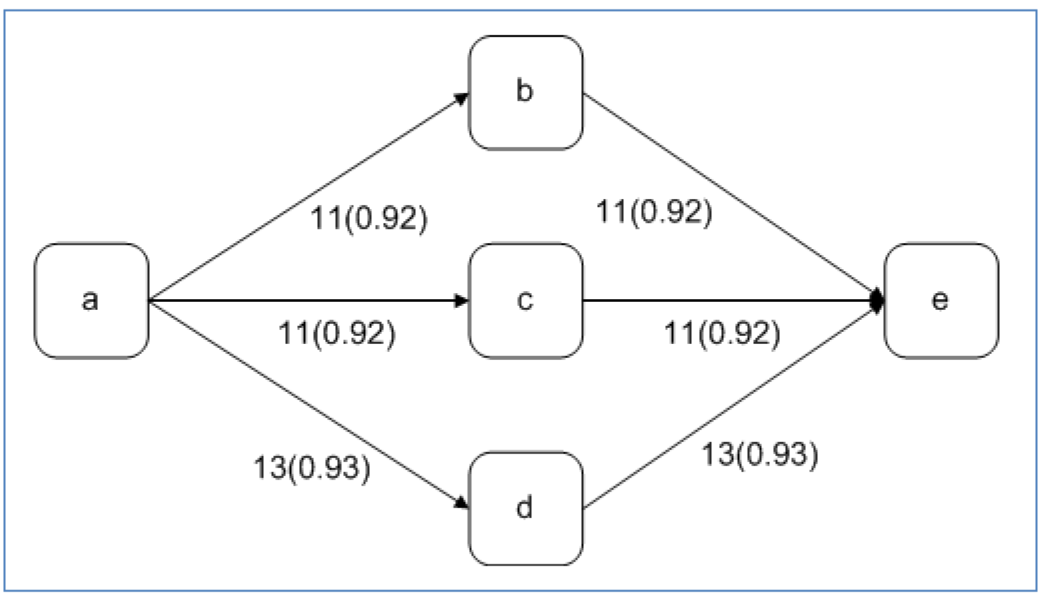


Figure 10: Dependency graph based on the all-activities-connected heuristic

From table 5 it becomes clear that there are also lower dependency values, e.g., the relationship between a and e has a dependency value of 0.83. This is less frequent behavior or maybe even noise and in case of less frequent behavior we might want to include it in the dependency graph. In order to do so, three threshold values have been developed. With these thresholds we can indicate that we will also accept dependency relations between activities that have a dependency measure above the value of the *dependency threshold*, and have a frequency higher than the value of the *positive observations threshold*, and have a dependency measure for which the difference with the 'best' dependency measure is lower than the value of *relative to best threshold*. For example, if we choose the following values: dependency threshold = 0.80, positive observations threshold = 4 and relative to best threshold = 0.15. This means that the relationship between activity a and e will also be included in the dependency graph because the dependency value is above the dependency threshold (0.83 > 0.80), the positive observation value is above the positive observations threshold (5 > 4) and the difference with the best dependency measure is lower than the value of the relative to best threshold ((0.93-0.83) < 0.15).

In a process it is also possible to execute an activity multiple times, also known as loops. Long distance loops (e.g. abc,abc,abc) can be discovered with the dependency measure stated under definition 1. However, for length one (acb,accb,acccb) and length two loops (acdb,acdcdb,acdcdcdb) the values of $c => w\ c$ and $c => w\ d$ are generally very low and cannot be discovered by the dependency measure developed under definition 1. Two new definitions are needed which are stated as follows:

Definitions 2 and 3:
$Let\ W\ be\ an\ event \log over\ T\ , and\ a, b\ \in T.$
$Then\ |a > w\ a| is\ the\ number\ of\ times\ a$
$> w\ a\ occurs\ in\ W, and\ |a \gg w\ a| is\ the\ number\ of\ times\ a \gg w\ b\ occurs\ in\ W.$

$$a => w\ a = \left(\frac{|a > w\ a|}{|a > w\ a| + 1}\right), \qquad a => 2w\ b = \left(\frac{|a \gg w\ b| + |b \gg w\ a|}{|a \gg w\ b| + |b \gg w\ a| + 1}\right)$$

The *length-one-loops threshold* and *length-two-loops threshold* are based on these definitions. For example, if the length-one-loops threshold is 0.70, the loop form activity d to d happens 4 times which means that the length one loops measure is 0.80 and the loop will be included in the dependency graph (0.80>0.70).

It is also possible that activities are performed in parallel which means that as soon as activity a is performed, activities b and c are performed in a dependent relation, shown as $a => w\ b \wedge c$. In order to detect dependent relations the following definition is developed:

Definition 4:
$Let\ W\ be\ an\ event \log over\ T\ , and\ a, b, c\ \in T,$
$and\ b\ and\ c\ are\ in\ depending\ relation\ with\ a. Then$

$$a => w\ b \wedge c = \left(\frac{|b > w\ c| + |c > w\ b|}{|a > w\ b| + |a > w\ c| + 1}\right)$$

The *AND threshold* is based on definition 4 and as an example, the AND threshold is set to 0.1. In Figure 10 on page 26 it is clear that activity a is followed by activities b,c and d. It is however not clear if these activities occur in a dependent relation. First we will look at activities b and c, which have a dependent relation score of 0.87 ((10+10)/(11+11+1)). This score is clearly above the AND threshold of 0.1 which means that activities b and c are performed in parallel. Activities c and d have a dependent relation score of 0 ((0+0)/(11+13+1)). This score is below the AND threshold which means that activities c and d do not have a dependent relation.

The eight settings of the heuristics miner have been discussed and it is dependent on the projects scope and goal if only the main behavior or also the exceptions of the process models must be discovered. The settings can then be adjusted accordingly.

5.3. Step 3: Homogenization

Homogenization consists of two steps; conversion and matching. When models are mined in different modeling formalisms a conversion step is required. In figure 8 on page 23 it is shown that the input of the conversion step are two or more models in different modeling formalisms. In ProM, models in different formalisms can be loaded and converted to a desired formalism. In table 6 a conversion overview of three commonly used notations is provided, including the file extension used. As can be seen in table 6, three conversions are possible with which each model can be converted to the other model in one or two steps.

	EPC	Petri net	Heuristics net	File Extension
EPC		X		.epml
Petri net			X	.pnml
Heuristics net	X			.hn

Table 6: ProM modeling formalism conversion table

In order to compare two process models, a match is required between the activities of the process models. This matching can be performed by hand and is integrated in the ProM tool. In ProM the matching step is integrated within the differences analysis plug-in which will be used in the comparison step and the EPC merge plug-in which will be used in the integration step. As can be seen in figure 8 on page 23, the input of the matching step is two process models in the EPC notation and the output is a match between the activity labels of the two models.

5.4. Step 4: Comparison

The differences analysis technique described in [28] is selected which is implemented in the ProM tool as the differences analysis plug-in. As shown in figure 8 on page 23, the input of the differences analysis is two process models and a match between the activity labels of the two models. The output in ProM is an indication of the type of a difference and an explanation of the exact position of a difference in the business process models. The following differences are included in ProM:

1. Different conditions
2. Additional conditions
3. Additional start condition
4. Different dependencies
5. Additional dependencies
6. Different moments
7. Iterative vs. Once-off
8. Skipped activity

For the first seven types of differences an explanation is provided in figure 11. For the skipped activity difference no visual explanation is provided, because it simply means that an activity is present in one model but not in the other.

The output of the comparison step is an overview of all types of differences between the two models and for each type of difference a visualization can be made of the position of the difference. This visualization is made based on the types of differences shown in figure 11 which is derived from [28] and the explanation of the position provided in the differences analysis in ProM.

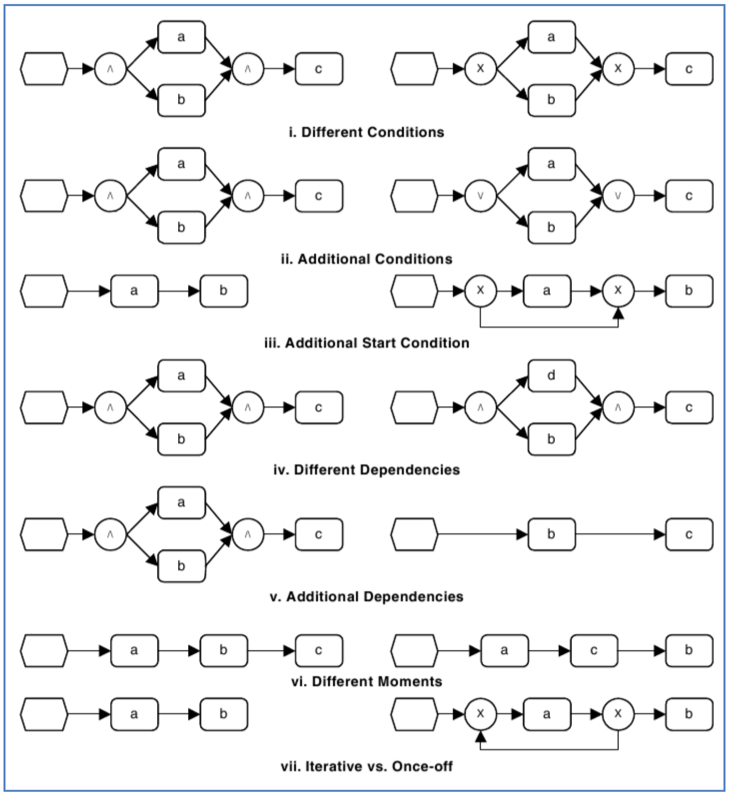


Figure 11: Types of differences

5.5. Step 5: Conflict Solving

As shown in figure 8 on page 23, the input of the conflict solving step are the two process models and differences between the two models. For each difference between the models a decision is needed if the difference is desired or undesired. In other words, if the difference is undesired one of the models must be changed to reflect the desired behavior. The decision to solve or maintain the conflicts is dependent on the requirements of the organization and for each difference, sensemaking is needed to give a meaningful explanation to the decision to maintain or solve the conflict. The decisions to solve or maintain the conflicts are controlled by factors that are present in the organization which have been explained in section 3.3.5. Figure 8 on page 23 shows that the output of the conflict solving step are two revised models (or the original models when no conflicts are solved) and a listing of the changes that have been made to the models. The models can be changed in a model editor like, e.g., WoPeD (Workflow Petri Net Designer). In ProM models can be converted to, e.g., the petri net notation and exported as a pnml file. The pnml file can be imported in WoPeD where the model can be edited. The revised model can be exported as pnml and imported in the ProM tool and converted back to the desired modeling formalism.

5.6. Step 6: Integration

As can be seen in figure 8 on page 23, the input of the integration step is two revised models. These revised models can now be integrated into one harmonized model. The EPC merge plug-in in ProM which is described in [37] can be used to merge two EPC models. The two revised models can be imported in the ProM tool and merged into one new model. The output of the merge is an integration of the two revised models which represent the harmonized model. Before performing the actual merge of two selected EPC's in ProM, the plug-in allows users to create a mapping between the functions of the two input EPC's as well as between their events. In this way, it is prevented that there are superfluous elements in the merged model. The merge algorithm works in three phases, where first the EPC's are converted to their active behavior, i.e. the functions of the EPC. This active behavior is shown in models called function graphs. The two function graphs are then merged into one function graph representing the combined behavior. Finally, the new function graph is converted back to an EPC. As stated in [37], the merge algorithm generalizes which means that the merged model may allow for more behavior than the sum of the parts behaviors'. This means that it may be possible that the business analyst must rework the merged model. In ProM, the merged model can be converted to, e.g., the petri net notation and exported as a pnml file. The pnml file can be imported in WoPeD where the model can be edited. The reworked model can be exported as pnml and imported in the ProM tool where it can be converted back to the EPC notation. A reflection of the merged model on the original models is always required to see if undesired behavior takes place in the merged model.

6. Case Study: Purchase to Pay

The harmonization approach described in chapter 5 will be executed in a case study with Company-Z. A problem description of the current situation and an explanation why Company-Z is suitable as a case for the implementation of the approach is provided in section 6.1. The process which needs to be harmonized is the purchase to pay process which will be explained in section 6.2. The scope of the purchase to pay process will also be handled in section 6.2. The execution of the harmonization approach will be discussed in section 6.3, where all steps of the approach will be addressed.

6.1. Problem description

Company-Z produces products for the building industry. Customers are, among others, resellers for the professional market and large DIY (Do It Yourself) retailers. Due to mergers and acquisitions in the history of the company multiple production sites/business units are located across the EMEA area. In order to increase profits, Company-Z tries to cut costs by optimizing the business unit's processes. One way in obtaining this goal is the standardization of processes. The processes are supported by information systems and Company-Z managed to standardize their information systems landscape and nearly all business units have SAP ERP installed. The problem is that the business units received lots of freedom during the configuration of the business processes within the ERP system. One of these processes is the purchase to pay process, concerning the acquisition of, e.g., raw materials, packaging materials, non production related goods and services. Company-Z has a special interest in the purchase to pay process because they want to further standardize these processes and work towards the implementation of a shared service centre. In the current situation each business unit arranges its own purchasing and has its own purchasing organization. In a shared service centre procurement is no longer performed for each location separately, but procurement is performed for multiple countries at once. The result will be a more efficient procurement organization which can operate with fewer employees that is able to purchase against better prices due to economies of scale. The processes in a shared service centre are often designed on the basis of a standardized (purchasing) reference model but Company-Z notices that this will not give them the desired result. The business units are located in different countries and have their own needs and regulations. A standardized model will not support these desired variations and Company-Z is looking for an alternative approach. BPH does take these variations into account and develops a harmonized reference model on the basis of the underlying processes which are currently executed in the existing business units. Therefore, Company-Z will serve as a case to execute the developed harmonization approach and develop a harmonized reference model for their purchasing processes, which can serve as a reference model for the implementation of the purchasing process in the shared service centre.

6.2. Purchase to Pay Process and scope

The purchase to pay (P2P) process is a process concerning the activities of requisitioning, purchasing, receiving and paying for goods and services. The term P2P emerged in the 1990s when organizations wanted to further optimize the process of buying. In order to bring financial rigor and process efficiency to the process of buying, with the help of IS organizations automated the purchasing process, from the way an item is ordered to the way that the final invoice is processed. The benefits of P2P are, increased financial and procurement visibility, efficiency, cost savings and control. The Automation of the purchasing process reduces processing times and the goal is that incoming invoices are handled without any (slow) manual intervention. Company-Z also followed this trend and used SAP ERP as the system to automate the P2P process.

In this case study the P2P process of Company-Z will be used to execute the harmonization approach. The scope concerns the purchasing of raw and packaging materials and only part of the P2P process will be used. The start activity of the process will be the creation of a purchase order and the end activity will be the payment of the order. Furthermore, a purchase order often consists of multiple order items and the actual receipt and payment is performed on an item level, therefore the process will be examined on an item level. Finally, the P2P process of six countries will be used during the execution of the approach and these countries are: United-Kingdom, Sweden, Ireland, Turkey, Italy and France.

6.3. Execution of BPH approach

The preparation step is described in section 6.3.1 and contains the construction of six event logs which will be used as input for the discovery step. The discovery of the six process models will then be discussed in section 6.3.2. In order to make a match between the different activity labels of the process models, a homogenization step is required which is shown in section 6.3.3. Next, section 6.3.4 discusses the differences analysis of the process models. Subsequently, section 6.3.5 concerns the conflict solving step and section 6.3.6 will show the integration of the models which will result in a harmonized model. The differences analysis, conflict solving step and integration step can only be performed on two models at the time and in total six models must be harmonized. Therefore, these three steps all consist of 5 parts. In the first part models 1 and 2 are compared, conflicts are solved and the two models are integrated into merged model 1. Then, merged model 1 is compared against model 3, conflicts are solved and merged model 1 and model 3 are merged into merged model 2. Merged model 2 is compared against model 4, conflicts are solved and the two models are integrated into merged model 3, etc. Thus, five iterations are performed in order to integrate the process models of the six countries and therefore the comparison step, conflict solving step and integration step are divided into 5 parts.

6.3.1. Step 1: Preparation

An event log is required in order to discover the P2P processes. The event log extraction procedure described in [19] is used and contains six steps, which will be explained next.

1. Determining activities

In order to mine the P2P process in SAP, we need to select the set of relevant activities for this process. The SAP reference model, a process executor at Company-Z and a purchasing expert at Capgemini have been consulted to identify the purchasing activities which are shown in table 7.

Create Purchase Requisition	Delete Purchase Order	Return Delivery
Change Purchase Requisition	Undelete Purchase Order	Invoice Receipt
Delete Purchase Requisition	Block Purchase Order	Parked Invoice
Undelete Purchase Requisition	Unblock Purchase Order	Payment
Release Purchase Requisition	Outline Agreement : Create Contract	Account Maintenance
Create Request for Quotation	Create Scheduling Agreement	Down Payment
Change Request for Quotation	Create Shipping Notification	Service Entry
Delete Request for Quotation	Change Shipping Notification	Vendor Confirmation
Undelete Request for Quotation	Subcontracting	Purchase order approval
Maintain Quotation	Issue Goods	Undo purchase order approval
Create Purchase Order	Goods Receipt	
Change Purchase Order	Delivery Note	

Table 7: Purchasing activities

2. Mapping out the detection of Events

An example of the mapping phase is provided in figure 12 which shows a part of the purchasing process consisting of two activities, namely, create purchase order and goods receipt. For these two activities the event information that is required for the construction of an event log (i.e. Case ID, Activity, Resource and Timestamp) is shown in the five tables of the relational database. We will look it this simplified process from an order item perspective. Created order items are stored in the EKPO table and each order has its own identifier, namely, the EBELP field. Orders consist of order items and for each order item there is an identifier, namely, the EBELN field. The case ID can thus be retrieved from the EKPO table and is shown as EBELP&EBELN in the activity. However, the timestamp and resource are not located in the EKPO table. Every time an order item is created or changed this is recorded in the change tables CDHDR (header) and CDPOS (item). The CDHDR table does contain the timestamp and resource information which is shown as USERNAME and UDATE&UTIME in the activity. But, the CDHDR table does not show for which order item this creation took place and this information is logged in the CDPOS table. The CDHDR and CDPOS tables are linked via a so called primary/foreign key relationship (PK and FK in the tables). This means that the CDHDR table and CDPOS table are linked to each other via the CHANGENR field. In the same way CDPOS is linked to EKPO via the TABNAME field which stores the EBELP and EBELN identifier. Thus, when EBELP and EBELN from EKPO are equal to TABNAME from CDPOS, and CHANGENR from CDHDR is equal to CHANGENR from CDPOS, we are talking about the same case. Furthermore, to make sure we are talking about an order item creation, the TABNAME field in CDPOS must be EKPO and the CHNGIND field in CDPOS must be I, which means that there was an insert (order item creation) in the EKPO table. The same logic can be applied to the goods receipt activity and every time an order item is received in stock, the receipt is logged in the MKPF (material header) and MSEG (material item) tables. The case identifier are the EBELN and EBELP fields form MSEG, which are actually the same EBELN and EBELP fields from the EKPO table, via the primary and foreign key relationship between MSEG and EKPO. MKPF contains the time and resource information, CPUDT, CPUTM and USNAM, of the goods receipt and in order to make sure we are talking about the same event; MSEG and MKPF are connected with a primary/foreign key relationship via the MBLNR field in these two tables.

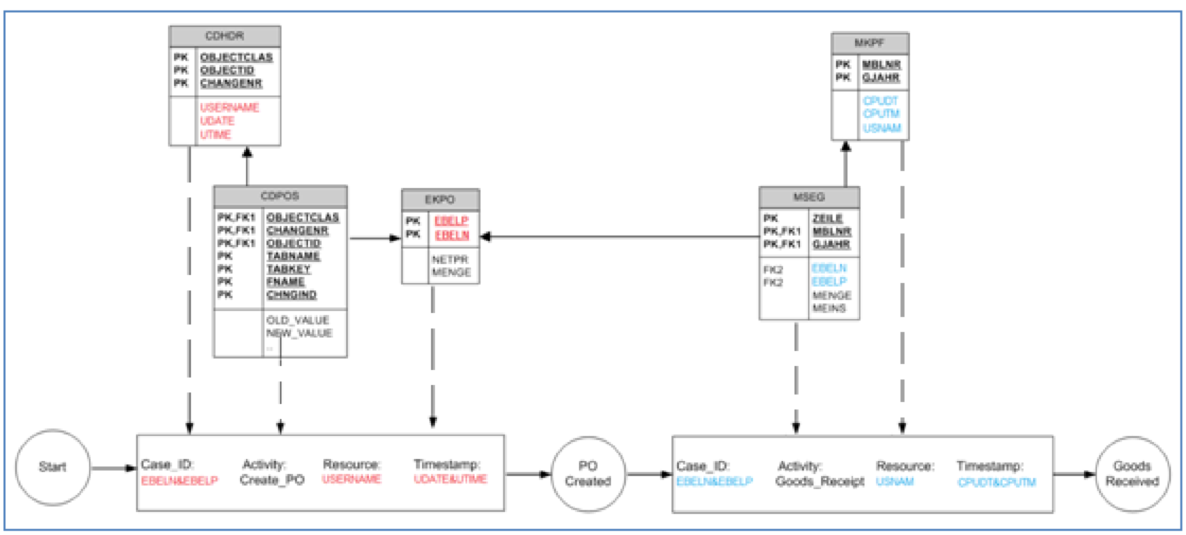


Figure 12: Mapping of SAP tables and fields on purchasing process

In conclusion one can state that, in order to select the right event log data for each of the activities of the process a mapping should be made on the relational database. This mapping will be further discussed in step 6 of the procedure, where the mapping will be used to build a SQL query to extract and convert the data into a log file.

3. Selecting attributes

Events in an event log contain information about the case identifier, activity name, resource and timestamp of the event. However, more information can be added in the form of attributes. An example of an attribute is the quantity when receiving goods or the price when creating an order. However, the harmonization approach only focuses on the function and control aspect of process models. The data aspect is out of scope and therefore, we are not interested in the data attributes of the activities in the P2P process, so this step is skipped.

4. Selecting activities to extract

Based on the scope discussed in section 6.2 the following activities are selected and shown in table 8.

Create Purchase Order	Goods Receipt	Invoice Payment
Change Purchase Order	Invoice Receipt	Undo purchase order approval
Delete Purchase Order	Undelete Purchase Order	
Block Purchase Order	Vendor Confirmation	
Unblock Purchase Order	Purchase Order Approval	

Table 8: Selected purchasing activities

5. Selecting the case

In SAP, a multitude of processes exist, which makes the selection of a correct case very difficult. With respect to the P2P process the identification of a suitable case is more obvious. An obvious choice would be the selection of the purchase order document as case and if we would do so, all activities are extracted from a purchasing document point of view. However, orders consist of items, and goods receipt and payment often takes place on an item level. Therefore, we will take the order item as the case for the P2P process.

6. Constructing the event log

In order to build the log file, the mapping of the selected activities must be converted to a SQL query and executed on the SAP database. This step will be performed by using XESame. An explanation of how to set up the connection with the SAP database in XESame and the P2P process mapping of activities on the data source in XESame is provided in Appendix 3: Constructing the event logs in XESame. As explained under step 5 of the preparation phase, we are using the purchase order item as a case and therefore the selected activities shown in table 8 have received new labels in XESame. The relation between the selected activities in table 8 and the activity labels used in XESame are shown in table 9.

Activity label	XESame label
Create Purchase Order	Create_PO_ITEM
Change Purchase Order	Change_PO_ITEM
Delete Purchase Order	Delete_PO_ITEM
Block Purchase Order	Block_PO_ITEM
Unblock Purchase Order	Unblock _PO_ITEM
Goods Receipt	Goods_Receipt
Invoice Receipt	Invoice_Receipt
Undelete Purchase Order	Undelete_PO_ITEM
Vendor Confirmation	Vendor_Confirmation
Purchase Order Approval	PO_ITEM_First_Approval
	PO_ITEM_Second_Approval
	PO_ITEM_Third_Approval
Undo purchase order approval	PO_ITEM_Undo_Approval
Invoice Payment	Invoice_Payment

Table 9: Relation between activity labels and XESame labels

The result of the preparation step is an event log for each of the six countries which will be used in the discovery step in the following section.

6.3.2 Step 2: Discovery

The heuristic algorithm discussed in [48] is used to discover the P2P process model for each of the six countries. The heuristics algorithm is implemented in the ProM tool which is used to convert the event log to a process model. The first step is loading the event log in the ProM tool and applying filters. These filters make sure that only complete cases are maintained and a single start and end event is used in the process model. The start event is 'Create_PO_Item' and the end event is 'Invoice_Payment'. But, there is also an event 'Delete_PO_Item' which can serve as an end event when order items are created and deleted somewhere in the process. This means that there are two end events and therefore an artificial end event is created.

The heuristic algorithm can now be applied and the algorithm's settings are shown in table 10. Where the relative-to-best-threshold is 0, the positive-observations are 1000 and the dependency threshold is 1 and the remaining settings are the default settings. The all-activities-connected heuristic is used which means that the values of the different parameters are ignored; simply one ingoing and outgoing connection with the highest dependency value is accepted. We assume that we have a noise free log and when using the extraordinary intolerant parameter settings, as described in [48] the right connections are made in the process model.

Relative-to-best threshold	0
Positive observations	1000
Dependency threshold	1
Length-one-loops threshold	0.9
Length-two-loops threshold	0.9
Long distance threshold	0.9
Dependency divisor	1
AND threshold	0.1

Table 10: Heuristic algorithm settings

The heuristics algorithm shows the main behavior of the process which means that not all behavior possible in the event log is represented in the mined process model. Therefore, a fitness measure shows how well the mined model represents the behavior in the event log. One way in measuring the fitness is calculating the number of correct parsed traces divided by the total number of traces in the log. However, this measure appears to naïve as partially correct traces are also handled is traces that could not be executed in the mined model. Therefore, an improved measure is used which is based on the number of successfully parsed tasks tokens instead of the number of parsed traces and the improved fitness measure is shown in table 11, where also the number of used cases is shown. The result of the heuristics miner is a heuristics net and as an example the heuristics net for Turkey is shown in figure 13. However, the preferred modeling notation is EPC and in ProM a conversion is possible from a heuristics nets to EPC. The results of the heuristics miner in the EPC notation for respectively, France, Ireland, Sweden, Italy, Turkey and the United-Kingdom, is shown in appendix 4: mined models.

Country	Improved continues semantics fitness	Cases
France	0,96	2865
Ireland	0,96	559
Sweden	0,97	3215
Italy	0,94	761
Turkey	0,89	268
United-Kingdom	0,92	11256

Table 11: Fitness measure event logs and used cases

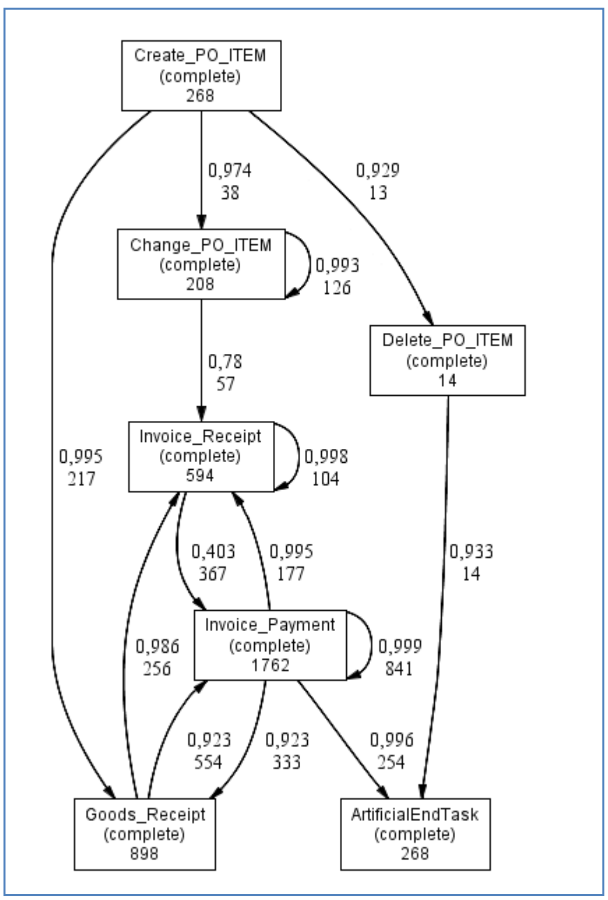


Figure 13: Heuristics net Turkey

6.3.3 Step 3: Homogenization

The discovered process models are all in the same process modeling notation, which means that no conversion of process models is needed. Before a comparison can be made, a mapping between activities of the process models is required. Due to the event log mapping that has been created in the preparation phase in section 6.3.1, the activity labels of the process models are all the same. Differences can only occur when an activity is present in one model but not in the other. A mapping of the activities between the process models is shown in table 12.

FR	IE	SE	IT	TR	UK
Create_PO_ITEM	Create_PO_ITEM	Create_PO_ITEM	Create_PO_ITEM	Create_PO_ITEM	Create_PO_ITEM
Change_PO_ITEM	Change_PO_ITEM	Change_PO_ITEM	Change_PO_ITEM	Change_PO_ITEM	Change_PO_ITEM
Goods_Receipt	Goods_Receipt	Goods_Receipt	Goods_Receipt	Goods_Receipt	Goods_Receipt
Invoice_Receipt	Invoice_Receipt	Invoice_Receipt	Invoice_Receipt	Invoice_Receipt	Invoice_Receipt
Invoice_Payment	Invoice_Payment	Invoice_Payment	Invoice_Payment	Invoice_Payment	Invoice_Payment
Delete_PO_ITEM	Delete_PO_ITEM	Delete_PO_ITEM	Delete_PO_ITEM	Delete_PO_ITEM	Delete_PO_ITEM
Undelete_PO_ITEM	-	-	-	-	Undelete_PO_ITEM
ArtificalEndTask	ArtificalEndTask	ArtificalEndTask	ArtificalEndTask	ArtificalEndTask	ArtificalEndTask
-	PO_ITEM_First_ Approval	-	-	-	PO_ITEM_First_ Approval
-	-	-	-	-	PO_ITEM_Second_ Approval
-	PO_ITEM_Undo_ Approval	-	-	-	PO_ITEM_Undo_ Approval
-	-	-	-	-	Vendor_Confirmation

Table 12: Mapping of activity labels

6.3.4 Step 4: Comparison

Because a difference analysis can only be performed on two models at once, the comparison step consists of a differences analysis which is performed in five parts. The models of the countries which are most alike in visual sense are compared first, starting with Italy versus Turkey. The results of the differences analysis of Italy vs. Turkey is shown in figure 14 and figure 15. Then the integrated model of Italy and Turkey (merged model 1) is compared against Sweden and the results of the differences analysis is shown in figure 16 and figure 17. Next, the integrated model of Italy, Turkey and Sweden (merged model 2) is compared against France and the results of the differences analysis is shown in figure 18, figure 19, figure 20 and figure 21. Then, the integrated model of Italy, Turkey, Sweden and France (merged model 3) is compared against Ireland and the results of the differences analysis is shown in figure 22, figure 23, figure 24 and figure 25. Finally, the integrated model of Italy, Turkey, Sweden France and Ireland (merged model 4) is compared against the United-Kingdom and the results of the differences analysis is shown in figure 26, figure 27, figure 28, figure 29, figure 30 and figure 31.

1. Differences between Italy and Turkey

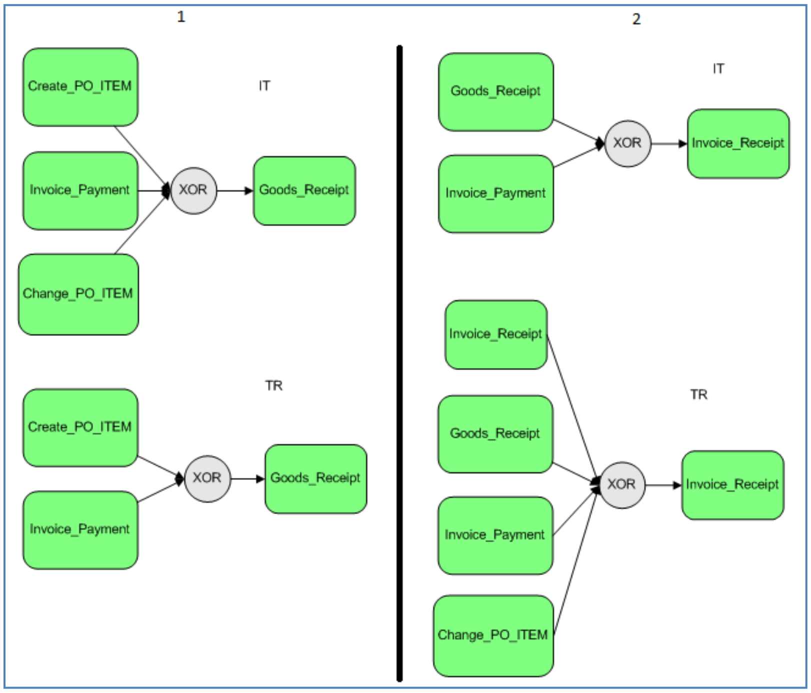


Figure 14: Additional dependencies Italy vs. Turkey

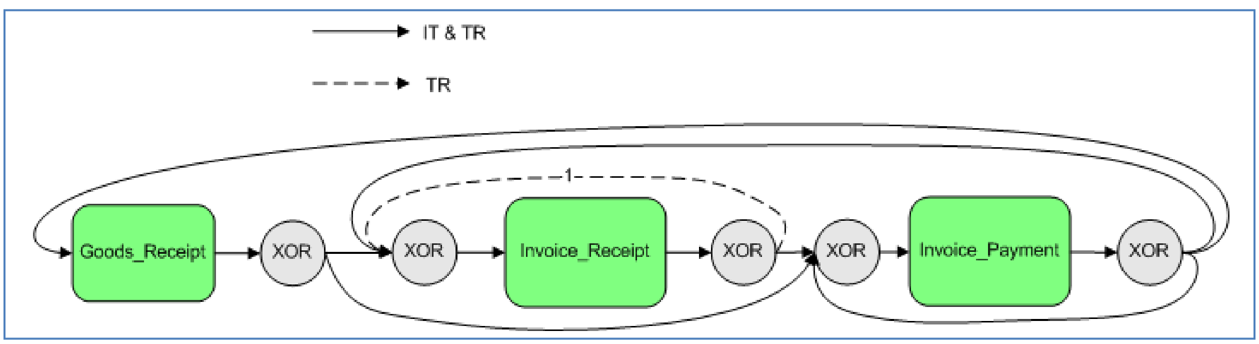


Figure 15: Iterative vs. Once-off Italy vs. Turkey

2. Differences between merged model 1 and Sweden

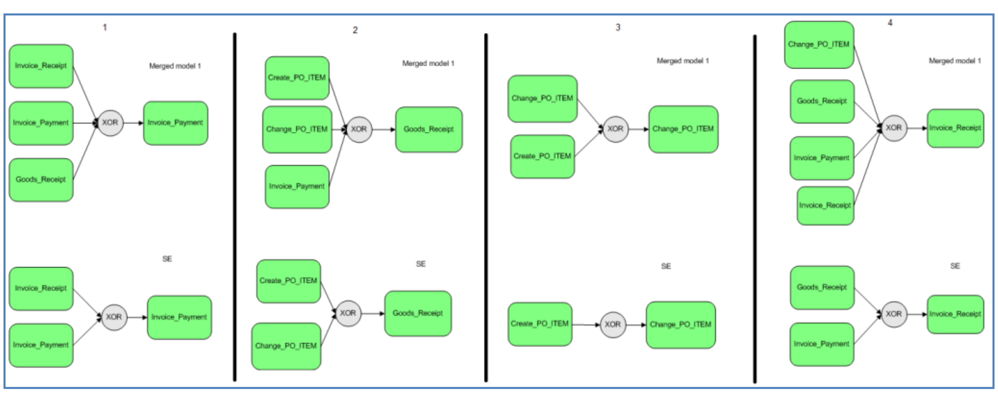


Figure 16: Additional dependencies Merged model 1 vs. Sweden

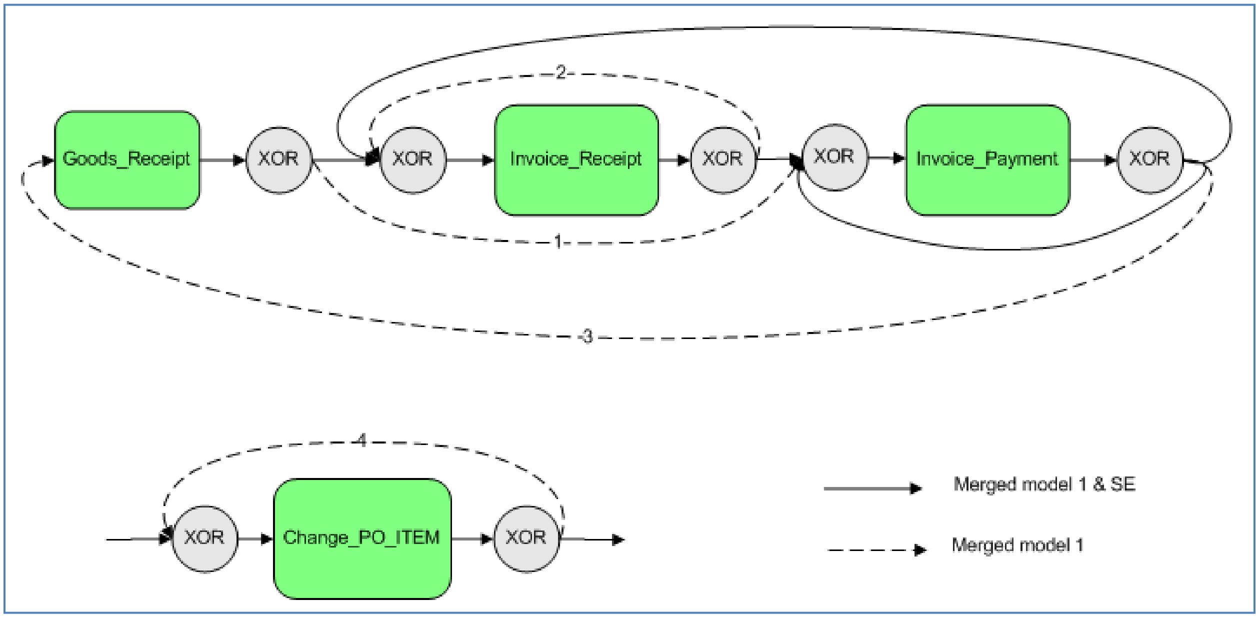


Figure 17: Iterative vs. Once-off Merged model 1 vs. Sweden

3. Differences between merged model 2 and France

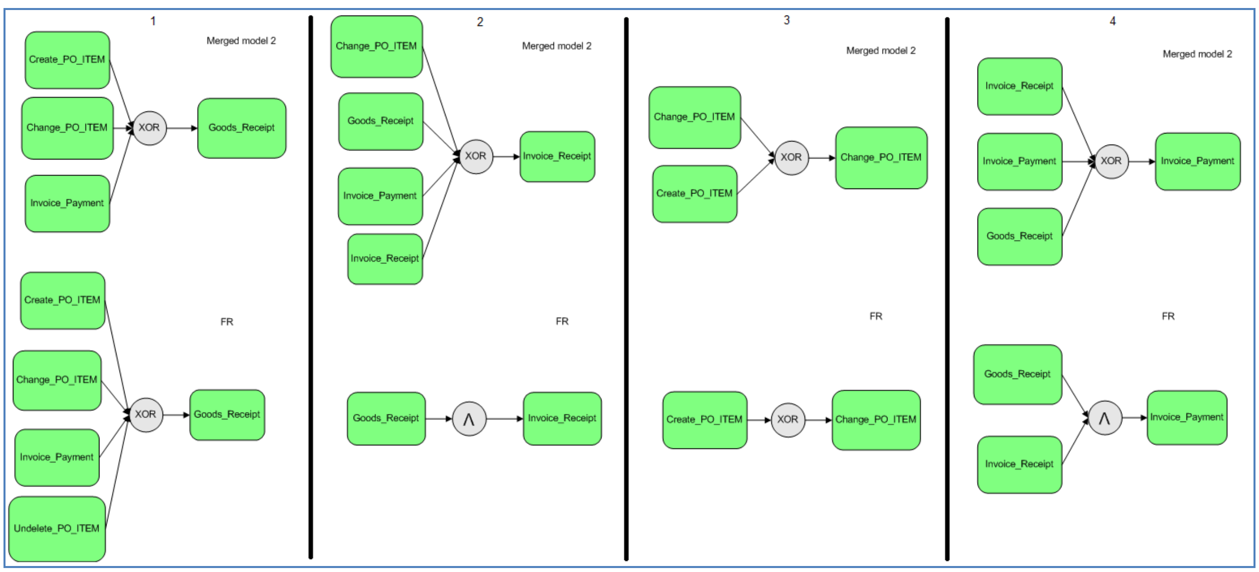


Figure 18: Additional dependencies Merged model 2 vs. France

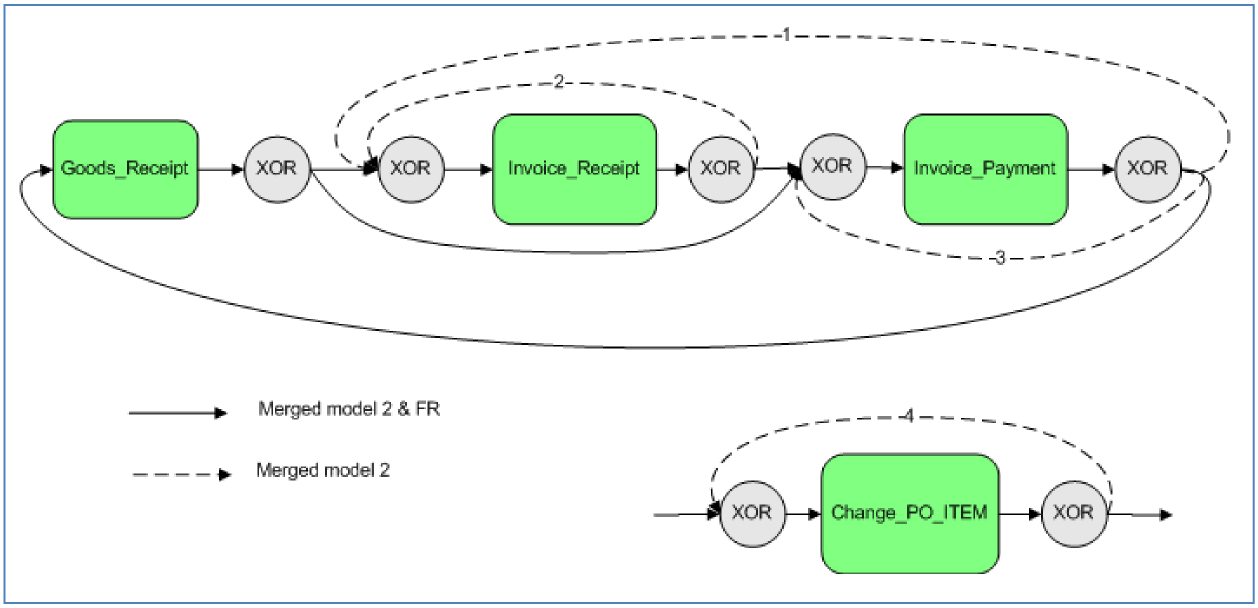


Figure 19: Iterative vs. Once-off Merged model 2 vs. France

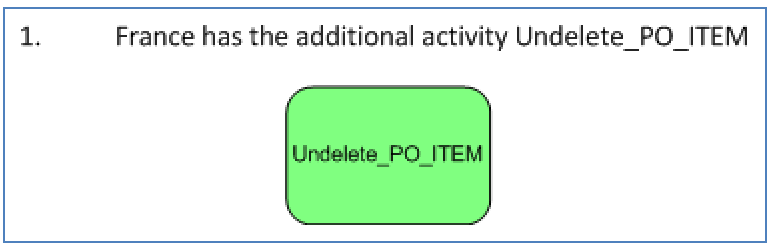


Figure 20: Skipped activity Merged model 2 vs. France

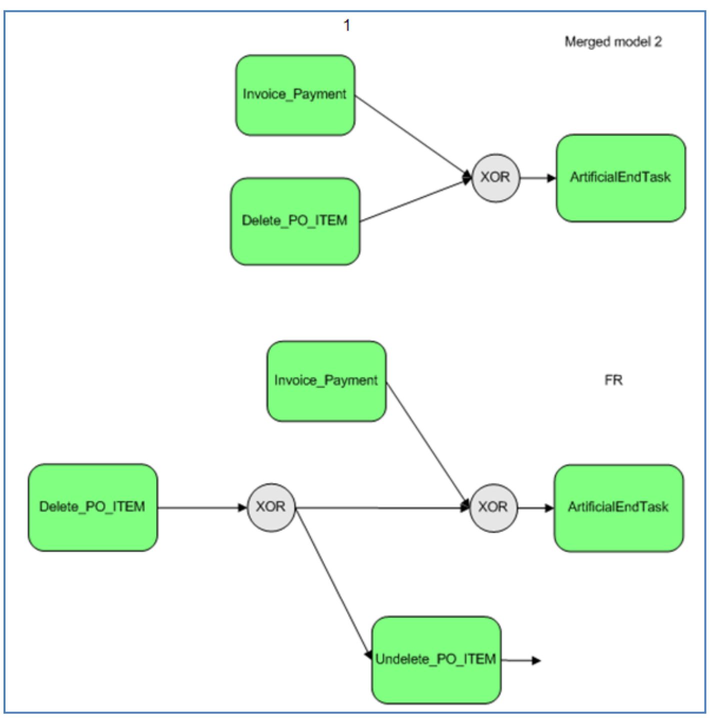


Figure 21: Additional conditions Merged model 2 vs. France

4. Differences between merged model 3 and Ireland

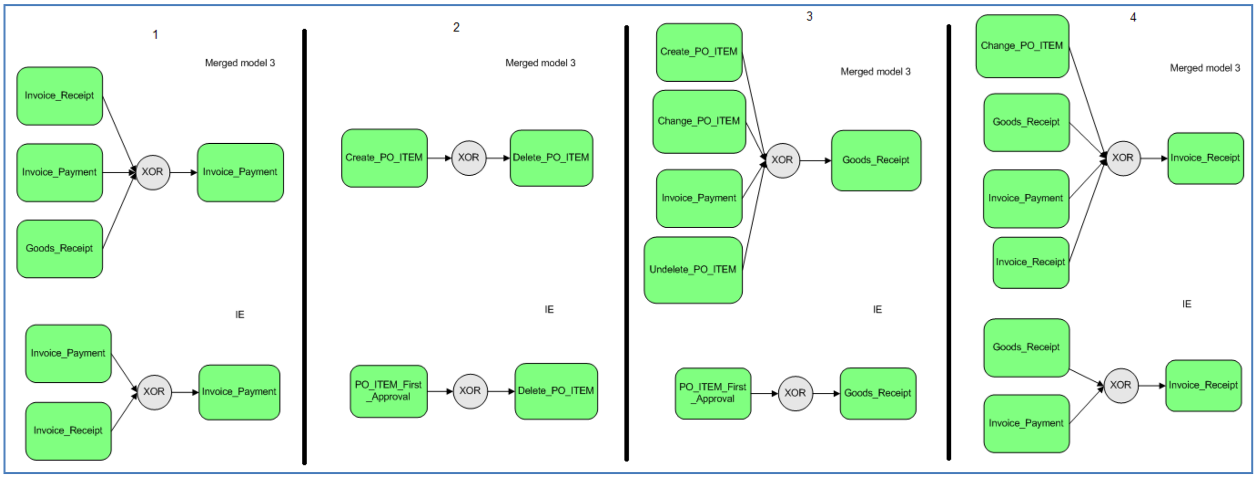


Figure 22: Additional dependencies Merged model 3 vs. Ireland

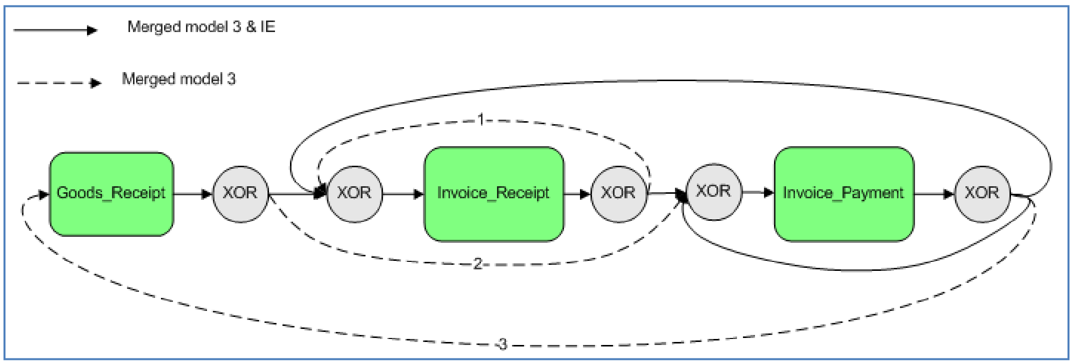


Figure 23: Iterative vs. Once-off Merged model 3 vs. Ireland

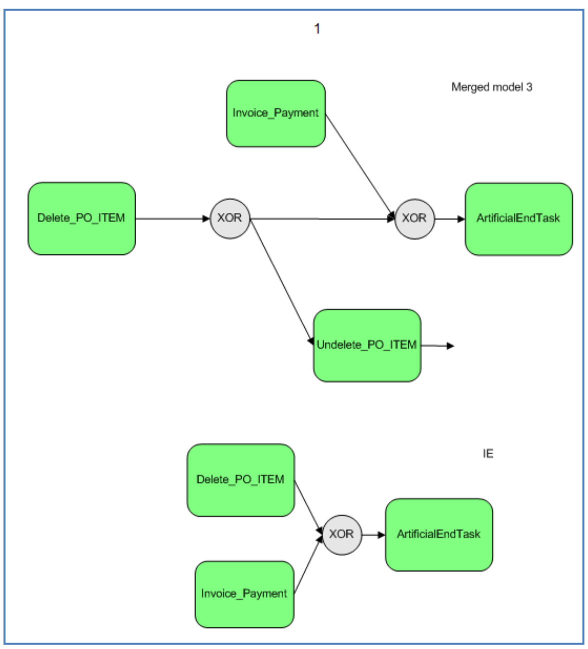


Figure 24: Additional conditions Merged model 3 vs. Ireland

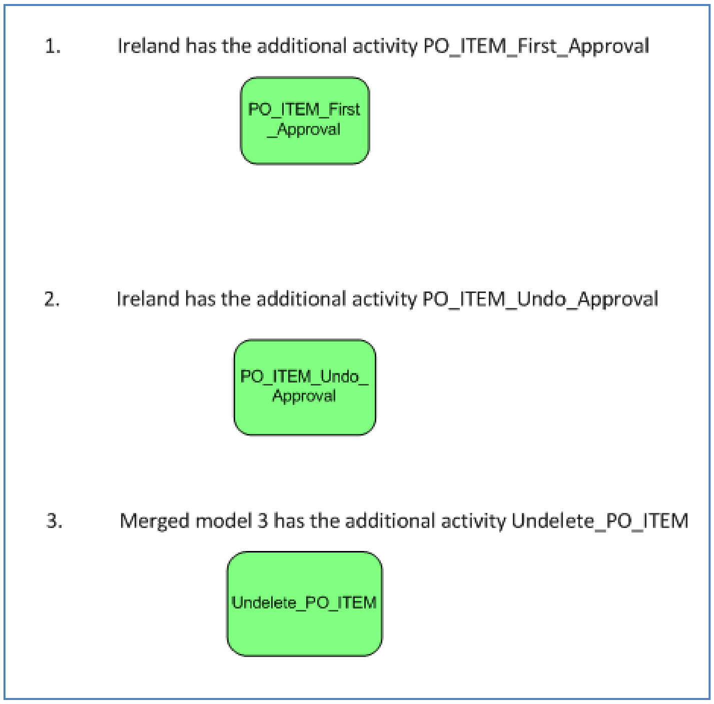


Figure 25: Skipped activity Merged model 3 vs. Ireland

5. Differences between merged model 4 and United-Kingdom

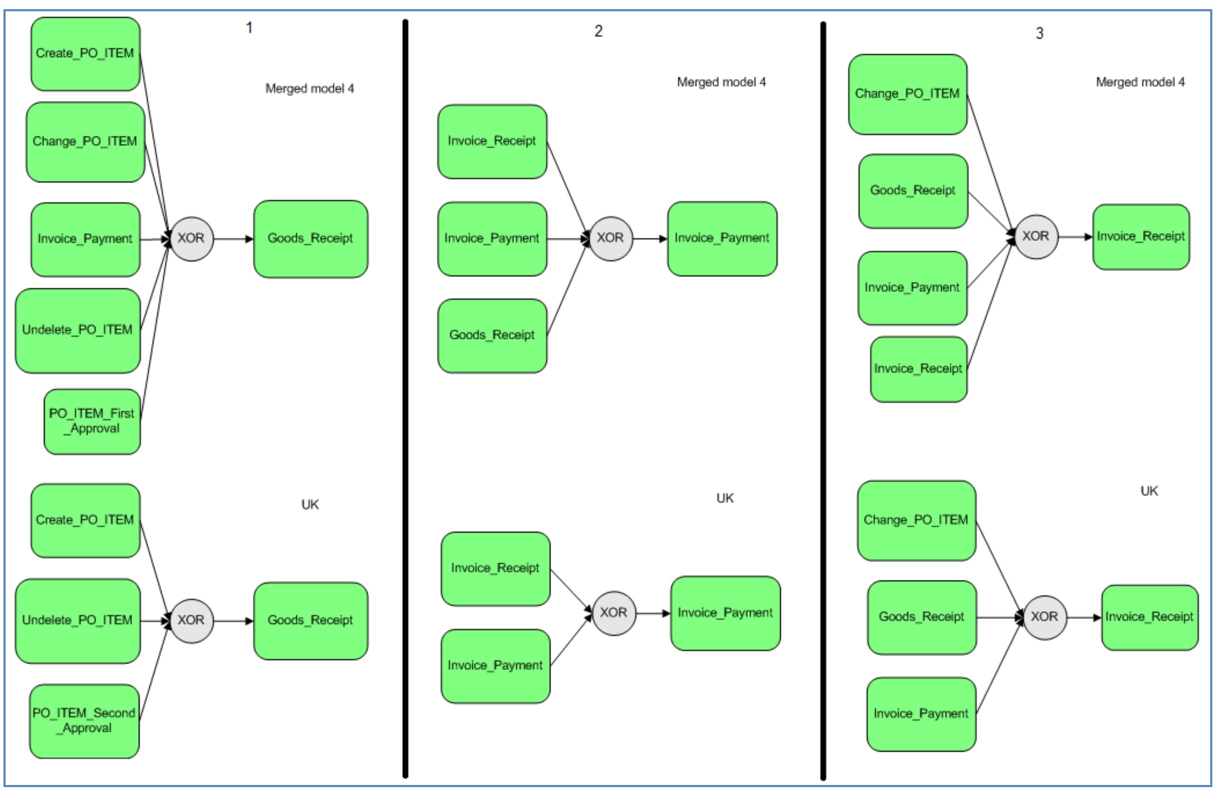


Figure 26: Additional dependencies Merged model 4 vs. United-Kingdom

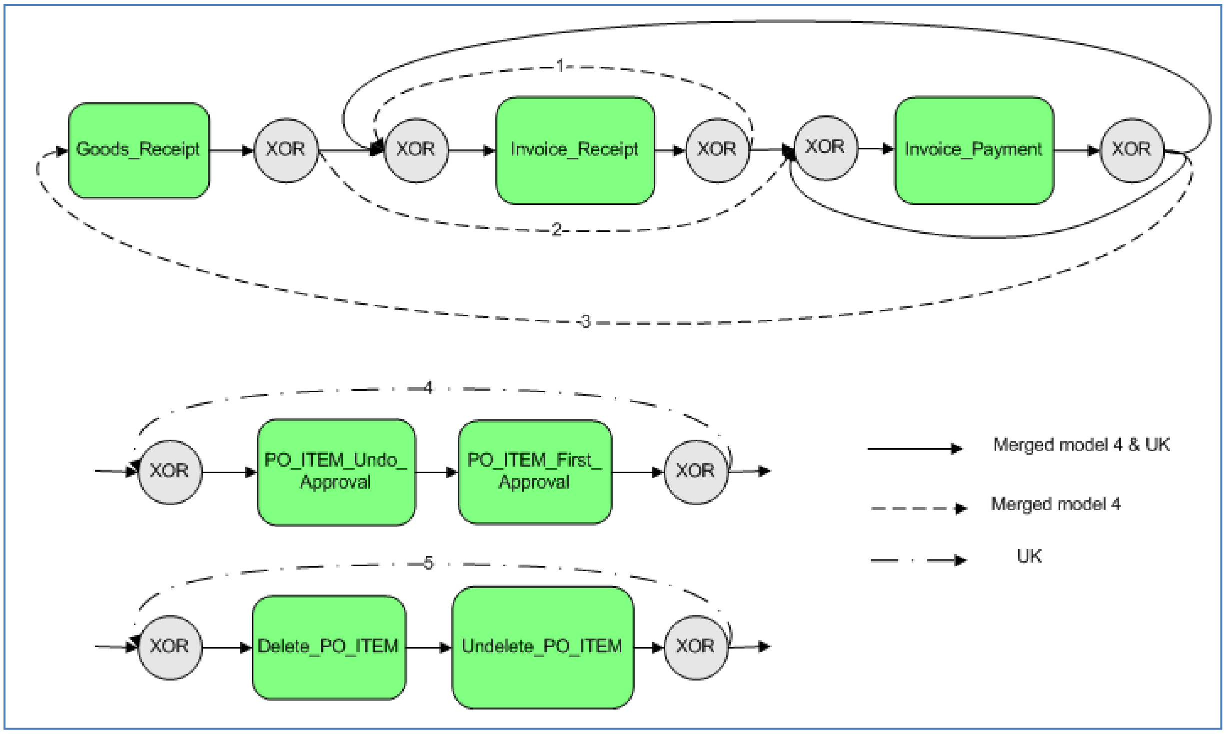


Figure 27: Iterative vs. Once-off Merged model 4 vs. United-Kingdom

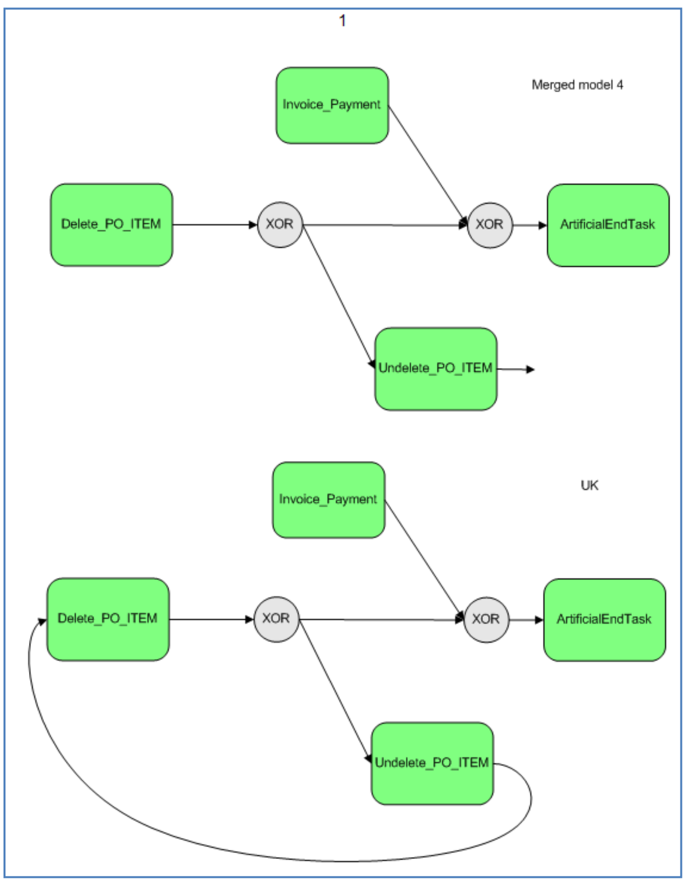


Figure 28: Additional Conditions Merged model 4 vs. United-Kingdom

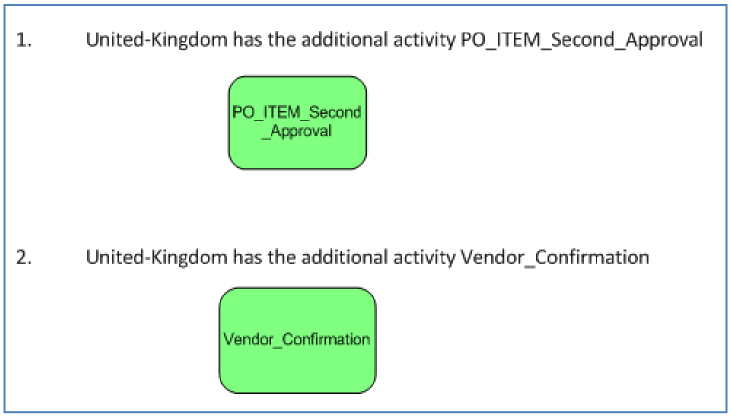


Figure 29: Skipped activity Merged Model 4 vs. United Kingdom

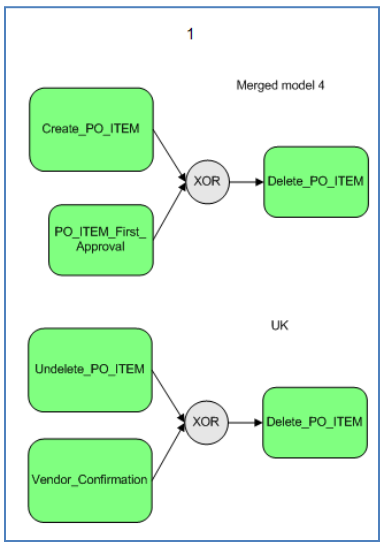


Figure 30: Different dependencies Merged model 4 vs. United-Kingdom

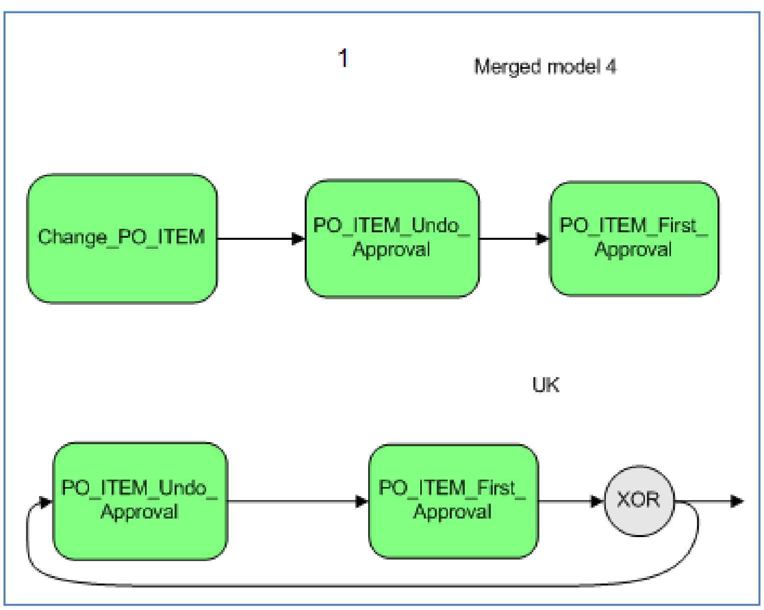


Figure 31: Different moment Merged model 4 vs. United-Kingdom

6.3.5 Step 5: Conflict Solving

For each of the five comparisons that have been made in section 6.3.4, the conflicts must be solved before integration of the models can take place. In other words, a decision should be made of which behavior is desired and which behavior is undesired. Each difference is discussed with Company-Z and Company-Z determines whether the behavior is desired or undesired. When undesired behavior exists in a model this model must be updated to reflect the desired behavior. Once conflicts are solved, integration can take place between the two models. For each of the five comparisons described in section 6.3.4 the conflicts will be discussed.

1. **Italy vs. Turkey**

Additional dependency 1
In Italy it is possible to change order items before goods receipt which is not possible in Turkey. Company-Z indicates that order items can be changed before goods receipt when, e.g., it is decided to change the order amount due to a sudden increase in demand. After the change, the order is send to the supplier and goods receipt takes place.

Additional dependency 2
In Turkey it is possible to change order items before invoice receipt. Company-Z already indicated that it is possible to change order items before goods receipt. But, in some cases it is possible that the invoice is received before the goods are received. In this case after changing, the order is send to the supplier and first invoice receipt takes place.

Additional dependency 2 and iterative versus once-off 1
Turkey has a loop from invoice receipt to invoice receipt. Company-Z indicates that it is possible to receive multiple invoices for an order item. In other words, it is possible to be billed in several installments.

In conclusion, no undesired behavior takes place and the two models can be merged as they are.

2. **Merged model 1 vs. Sweden**

Additional dependency 1 and iterative versus once-off 1
The first additional dependency states that it is possible to have goods receipt before payment in the merged model. Company-Z indicates that it is possible to have a partial goods receipt and invoice receipt for that partial delivery. Then another partial goods receipt takes place and the first invoice is paid for. This mean that invoice payment takes place after goods receipt, where this is actually the payment of the initial delivery and invoice.

Additional dependency 2 and iterative versus once-off 3
The second additional dependency states that it is possible to have invoice payment before goods receipt in the merged model. Company-Z indicates that it is possible that payment takes place for a partial goods receipt and another partial goods receipt takes place after paying for the initial partial delivery. This means that payment takes place before goods receipt.

Additional dependency 3 and iterative versus once-off 4
The merged model has a loop from change purchase order item to change purchase order item. Company-Z indicates that it is possible to perform multiple changes and that this is desired behavior.

Additional dependency 4 and iterative versus once-off 2
A loop from invoice receipt to invoice receipt is desired behavior. See additional dependency 2 and iterative versus once-off 1 from Italy versus Turkey.

Additional dependency 4
Changing order items before invoice receipt is desired behavior. See Additional dependency 2 from Italy versus Turkey.

In conclusion, no undesired behavior takes place and the two models can be merged as they are.

3. Merged model 2 vs. France

Additional dependency 1, additional condition 1 and Skipped activity 1
Un-deletion of purchase order items only occurs in France. Company-Z indicates that un-deleting purchase order items might be useful when a deletion is invalid and must be reversed. Re-creating the purchase order item is laborious and reversing the deletion gives the same result. Therefore, the activity undelete-PO-ITEM must be maintained in the harmonized model. Un-deleting a purchase order item occurs after deleting a purchase order item and un-deleting a purchase order item is followed by goods receipt. Company-Z indicates that this is desired behavior.

Additional dependency 2 and iterative versus once-off 1
It is clear that it is possible to have payment before invoice receipt in the merged model. Company-Z indicates that it is possible to be billed in several installments. This means that after receiving a partial invoice, payment takes place. After paying the partial invoice, the following partial invoice can be received. This means that payment can take place before invoice receipt and according to Company-Z this is part of desired behavior.

Additional dependency 2 and iterative versus once-off 2
A loop from invoice receipt to invoice receipt is desired behavior. See additional dependency 2 and iterative versus once-off 1 from Italy versus Turkey.

Additional dependency 2
Changing order items before invoice receipt is desired behavior. See Additional dependency 2 from Italy versus Turkey

Additional dependency 3 and iterative versus once-off 4
The merged model has a loop from change purchase order item to change purchase order item. See additional dependency 3 and iterative versus once-off 4 from merged model 1 versus Sweden.

Additional dependency 4 and iterative versus once-off 3
The merged model has a loop from invoice payment to invoice payment. Company-Z states that after goods receipt multiple partial invoices can be received for one order item and it is thus also possible to perform multiple payments in sequence.

In conclusion, no undesired behavior takes place and the two models can be merged as they are.

4. **Merged model 3 vs. Ireland**

Additional dependency 1 and iterative versus once-off 2
Goods receipt before payment is desired behavior. See additional dependency 1 and iterative versus once-off 1 from merged model 1 versus Sweden.

Additional dependency 2 and skipped activity 1
In Ireland an approval process takes place after order creation. Company-Z indicates that certain orders need approval. The level of approval (no approval, first or second approval) is dependent on the order amount, purchasing group (one or more individuals in a purchasing department) or purchasing organization (country). The approval process is different for each country and is based on decisions made by corporate management. In Ireland first approval takes place after order item creation and then goods receipt takes place. It is also possible that order items are deleted after first approval and is long as the order item has not been send to the vendor, order items are allowed be deleted.

Additional dependency 3 and skipped activity 1
As has been stated under 'additional dependency 2 and skipped activity 1' from merged model 3 versus Ireland, in Ireland first approval takes place before goods receipt. Company-Z indicates that this is desired behavior.

Additional dependency 3 and iterative versus once-off 3
Invoice payment before goods receipt is desired behavior. See additional dependency 2 and iterative versus once-off 3 from merged model 1 versus Sweden.

Additional dependency 3
In the merged model it is possible to create, change and un-delete purchase order items before goods receipt and this is part of desired behavior as indicated by Company-Z. After creation of a purchase order item, the order is send to the supplier and goods receipt takes place. For changes before goods receipt see additional dependency 1 from Italy versus Turkey. For un-deletion of purchase order items before goods receipt see additional dependency 1, additional condition 1 and Skipped activity 1 from merged model 2 versus France.

Additional dependency 4 and iterative versus once-off 1
A loop from invoice receipt to invoice receipt is desired behavior. See additional dependency 2 and iterative versus once-off 1 from Italy versus Turkey.

Additional dependency 4
Changing order items before invoice receipt is desired behavior. See additional dependency 2 from Italy versus Turkey.

Additional condition 1 and skipped activity 3
Un-deletion of purchase order items is desired behavior. See additional dependency 1, additional condition 1 and Skipped activity 1 from merged model 2 versus France.

Skipped activity 2
Ireland has the additional activity undo approval. Company-Z indicates that when orders are changed, the approval process might be performed again. Therefore, undoing the approval steps is required in order to restart the approval process and the activity PO_ITEM_Undo_Approval must be maintained in the harmonized model.

In conclusion, no undesired behavior takes place and the two models can be merged as they are.

5. Merged model 4 vs. United-Kingdom

Additional dependency 1
In the merged model it is possible to change purchase order items before goods receipt and this is part of desired behavior as indicated by Company-Z. For changes before goods receipt see additional dependency 1 from Italy versus Turkey. As has been stated under ‘additional dependency 2 and skipped activity 1’ from merged model 3 versus Ireland, in Ireland first approval takes place before goods receipt. Company-Z indicates that this is desired behavior.

Additional dependency 1 and skipped activity 1
Second approval before goods receipt takes place in the United-Kingdom and as stated under additional dependency 2 and skipped activity 1 from merged model 3 versus Ireland; this is part of desired behavior.

Additional dependency 1 and iterative versus once-off 3
Invoice payment before goods receipt is part of desired behavior as stated under additional dependency 2 and iterative versus once-off 3 from merged model 1 versus Sweden.

Additional dependency 2 and iterative versus once-off 2
Goods receipt before invoice payment is part of desired behavior is stated under additional dependency 1 and iterative versus once-off 1 from merged model 1 versus Sweden.

Additional dependency 3 and iterative versus once-off 1
The merged model has a loop from invoice receipt to invoice receipt and this is part of desired behavior is stated under additional dependency 2 and iterative versus once-off 1 from Italy versus Turkey.

Iterative versus once-off 4 and different moment 1
In the United-Kingdom it is possible to undo an approval after first approval. According to Company-Z this is not allowed as first a change has to be executed before the approval process is undone. This means that the path from first approval to undo approval in the United-Kingdom must be removed from the model. A path from first approval to changing the order item as well as a path from changing an order item to undo approval is not present in the United Kingdom and this path must be added to one of the process models.

Iterative versus once-off 5, additional condition 1 and different dependency 1
The united-Kingdom has a loop form un-deleting a purchase order item to deleting a purchase order item. Un-deletion after deletion is part of desired behavior as has been explained under additional dependency 1, additional condition 1 and Skipped activity 1 from merged model 2 versus France. However, again deleting the purchase order item after un-deletion means that a worker is deleting and

un-deleting an order item in sequence. According to Company-Z this is misguided but permitted behavior.

Different dependency 1 and skipped activity 2

The United-Kingdom has the additional activity vendor confirmation. Company-Z explains that for part of the packaging suppliers in the UK an EDI (Electronic Data Interchange) is made when orders are placed. This means that an automatic confirmation takes place when orders are received at the vendor's site. Company-Z indicates that the EDI must be preserved and therefore the activity Vendor_Confirmation must be maintained in the harmonized model. Vendor confirmation takes place after order creation. Orders are not allowed to be deleted after vendor confirmation. This is however possible in the UK and the UK model must be updated to remove this undesired behavior. Company-Z also indicates that after vendor confirmation goods receipt takes place, this is not possible in the UK model and the model must be update to represent this desired behavior.

Different dependency 1

In the merged model creation and first approval of order items is followed by deletion. Company-Z indicates that this is part of normal behavior. As long as orders are not send to the vendor (which means that there has been no goods receipt), order are allowed to be deleted.

In conclusion two paths must be removed from the UK model, namely: the path from first approval to undo approval and the path from vendor confirmation to deletion of order items. Furthermore, three paths must be added to the UK model, namely: a path from first approval to changing an order item, a path from changing an order item to undo approval and a path from vendor confirmation to goods receipt. The revised UK model is shown in figure 32 and can now be merged with merged model 4.

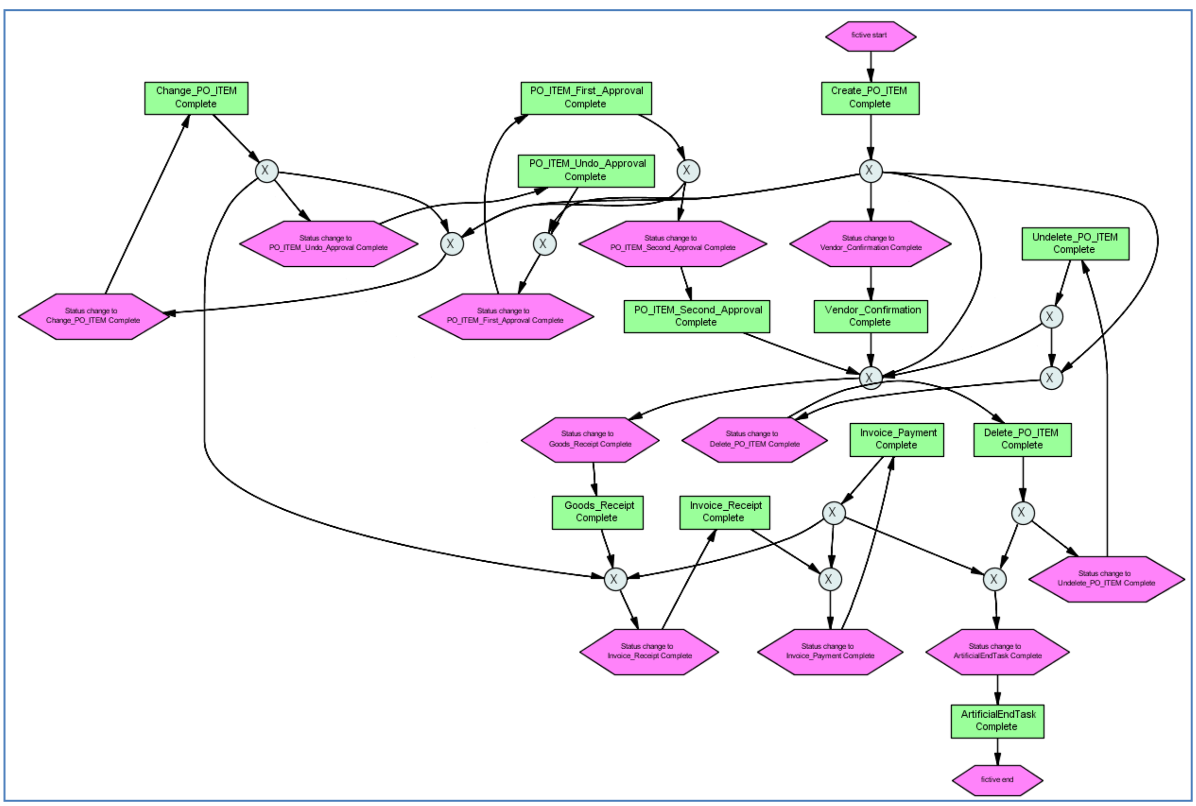


Figure 32: Revised process model United-Kingdom

6.3.6 Step 6: Integration

The integration of the models takes place in 5 steps as only two models can be merged at once. As explained in section 5.6, it may be possible that the merged model allows for more behavior than possible in the original models and therefore the merged model may need a rework to represent the correct behavior. The rework of the merged models is performed in WoPeD as has been described in section 5.6. The intermediate steps of the integration are shown in appendix 5: merged models, where the initial and reworked models are shown. In these intermediate steps Italy and Turkey are merged into merged model 1. Then, merged model 1 is merged with Sweden into merged model 2. Next, merged model 2 is merged with France into merged model 3 and merged model 3 is merged with Ireland into merged model 4. Finally, merged model 4 is merged with the United-Kingdom into the merged model of all six countries which is shown in figure 33. The merged model shown in figure 33 serves as a harmonized model of the 6 countries and together with the changes described in section 6.3.5, it can be used as a reference model for the implementation of the shared service centre.

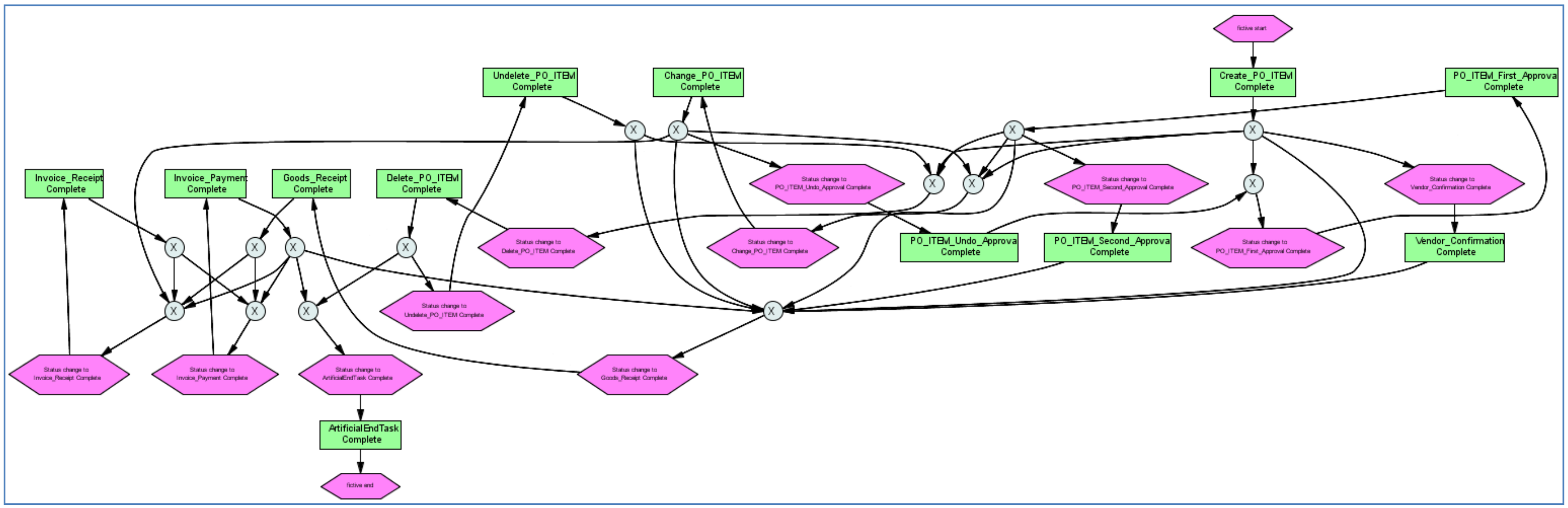


Figure 33: Harmonized model

7. Evaluation

The evaluation of the BPH approach is based on qualitative interviews and the evaluation method is described in section 7.1. The results of the evaluation are discussed in section 7.2 and a conclusion of the evaluation is provided in section 7.3.

7.1 Method

In order to evaluate the business process harmonization approach, qualitative interviews are conducted with business analysts form Capgemini Nederland BV. Capgemini is a consulting company which, among many other things, assists organizations in harmonizing their business processes. The business analysts of Capgemini are therefore suitable candidates to evaluate the BPH approach.

Semi-structured interviews are conducted with business analyst and the interviews start with the introduction of the topic. The introduction consists of the following items:

- A definition of Business Process Standardization and Business Process Harmonization
- An explanation of the Business Process Harmonization framework shown in figure 6 on page 19
- An explanation of the Business Process harmonization approach shown in figure 8 on page 23
- An explanation of the case study shown in chapter 6

Then, for each step in the harmonization approach two open questions are formulated:

1. What do you think of the practicability of the step?
2. What do you think of the output of the step?

By means of questions 1 and 2, we want know for each step if it can be performed easily and if the output of the step provides the business analyst with useful information.

Next, in order to evaluate the complete approach the following questions are formulated:

3. Are there any missing and/or superfluous steps in the approach? If yes, please explain.
4. Can the harmonized model be used as a reference model when implementing a shared service centre? If yes/no, please explain.

By means of question 3 we want to know if there are superfluous steps in the approach which make the approach unnecessarily complex and if there are missing steps which can improve the approach. By means of question 4 we want to know, with respect to the conducted case study at the client of Capgemini, if the output of the approach (i.e. a harmonized model) can be used is a reference model for the implementation of a shared service centre.

The questions have served as a guide for the semi-structured interviews and the results will be discussed in the following section.

7.2 Results

A total of 14 questions have been asked and the answers are classified and the results are shown in table 13. Two senior business analysts (BA 1 and BA 2 in table 13) and a principal consultant business analyses for the private sector (PC in table 13) have been interviewed. In table 13, each answer to a question is transcribed by a short description and for each interviewee it is indicated if a certain type of answer has been given. If two or three of the interviewees have given the same type of answer it will be stated in the results. The first 12 questions are related to the harmonization steps and question 13 and 14 are related to the approach as a whole. For each harmonization step and the approach as a whole, two questions have been answered by means of a descriptive summary of the answers of the interviewees. For each harmonization step and the harmonization approach as a whole, illustrative quotes are provided that followed from the interviews.

Question		Answer	Interviewee			Total
			BA 1	BA 2	PC	
1. Practicability preparation step	a.	The preparation step is difficult	X	X	X	3
	b.	Documentation on the system and access to the database must be available	X	X		2
	c.	Domain knowledge on SAP system and building an SQL query must be available	X	X	X	3
	d.	Building your own query is laborious and complex and not generic enough		X	X	2
2. Output preparation step	a.	No opinion, you just need it to perform process mining	X	X	X	3
3. Practicability Discovery step	a.	The discovery step is easy to perform	X	X	X	3
	b.	Education and domain knowledge on process mining is needed	X		X	2
	c.	Parallelism is hard to discover	X			1
4. Output Discovery step	a.	Useful models		X		1
	b.	Divergence and convergence in SAP system is a problem		X	X	2
5. Practicability homogenization step	a.	In the context of the case study it is an easy step as SAP activities have the same name	X	X	X	3
	b.	Conversion is not needed in this case study but it is a simple task in ProM.		X		1
	c.	When activities are not well documented (have deviating names) it can be difficult	X	X		2
	d.	When composite tasks or compared against multiple task it can be difficult	X	X		2
6. Output homogenization step	a.	No opinion, you need a mapping to perform a comparison between two models	X	X	X	3
7. Practicability comparison step	a.	The differences analysis is easy to perform	X	X	X	3
	b.	The visualization is made by hand which is a laborious task		X	X	2
8. Output comparison step	a.	The visual differences analysis is useful	X	X	X	3
	b.	Comparison against 7th desired model would be useful	X			1
	c.	The visual representation is more useful then the text based output in ProM	X	X	X	3
9. Practicability conflict solving step	a.	The conflict solving step is difficult	X	X	X	3
	b.	The conflict solving step lacks a proper methodology	X	X	X	3
	c.	Reconfiguring the models in ProM is not possible			X	1
	d.	Comparison against desired model is needed	X		X	2
10. Output conflict solving step	a.	Useful when overview of applied changes is maintained next to the revised models		X	X	2
11. Practicability integration step	a.	The integration step is an easy step in ProM	X		X	2
	b.	Does not always give the right output and some manual intervention may be needed			X	1
12. Output integration step	a.	It show a harmonized view of the underlying models			X	1
	b.	It can serve as input for configuration	X			1
	c.	A configurable models can provide more information	X	X		2
	d.	It is based on main behavior and convergence/divergence data			X	1
13. Missing/superfluous steps	a.	There are no superfluous steps	X	X	X	3
	b.	An additional step can be a comparison with a desired model	X	X	X	3
14. Can harmonized model be used as reference model	a.	No, more process model aspects are needed to use it as a blueprint		X	X	2
	b.	You must be able to turn on and off activities when dealing with a certain country	X			1

Table 13: Classification interview results

Preparation step

Question 1: What do you think of the practicability of the preparation step?

From the interviews it becomes clear that it is difficult to create a log file from SAP. Many conditions must be satisfied, e.g., a detailed design of the system must be available as well as SAP experts which are able to provide detailed explanations of the system's configuration and you must have access to the database in order to retrieve the data. As soon as a mapping has been made on the database and all data is available, one has to build its own query to construct the log file. Building the query requires domain knowledge and as the query is build for a specific system it is not flexible enough to reuse it on a different system.

Question 2: What do you think of the output of the preparation step?

The output of the preparation step is a log file which is only used to discover process models with process mining techniques and according to the interviewees it does not provide any more information.

Quotes:

> "Assuming that there exists documentation on the system in use and a specification of differences with the standard SAP system, SAP experts are available and you have access to the system's database, it is doable. However, you still need to build your own query." (Senior Business Analyst 1)

> "It is a lot work and you need domain knowledge. The mapping and resulting log file which has been built in XESame is not generic enough to reuse on a new dataset." (Principal Consultant Business Analysis Private Sector)

Discovery step

Question 3: What do you think of the practicability of the discovery step?

Mining an event log is a relatively easy task in ProM. But, domain knowledge in process mining and the algorithm used is required to get insightful results. When using the heuristic miner for example, the person performing the mining needs to know the effect of different settings of the miner.

Question 4: What do you think of the output of the discovery step?

The output is a mined model which can contain incorrect behavior due to Divergence and Convergence in the log file caused by a data-centric system like SAP. In [19] Divergence is described as: 'a divergent event log contains entries where the same activity is performed several times in one process instance'. E.g. in the case study in chapter 6, multiple payments are received for a single line item, which will result in a process model with a payment activity that has a single loop. In [19] Convergence is described as: 'a convergent event log contains entries where one activity is executed in several process instances at once', which for instance means that one payment is received for multiple line items at once.

Quotes:

> "The mining step is relatively easy to perform. But you must be educated and have domain knowledge in process mining." (Senior Business Analyst 1)
>
> "For a person without domain knowledge it is very hard to understand what happens if certain settings of the heuristics miner are changed." (Principal Consultant Business Analysis Private Sector)
>
> "Divergence and convergence is a problem as it induces behavior in the model which is not really possible in the business process." (Principal Consultant Business Analysis Private Sector)

Homogenization step

Question 5: What do you think of the practicability of the homogenization step?

With respect to the matching step, the same query is used to construct the process models for the different countries and therefore the activities of the models have the same names. Thus, with respect to the case study, matching the activities between two process models is an easy task. However, when activities have different names and composite activities have to match with multiple separate activities, making a match is more difficult.

Question 6: What do you think of the output of the homogenization step?

The output is a match between the activity labels of two process models which is needed to perform the differences analysis in the comparison step. According to the interviewees the homogenization step does not provide valuable information, it is only performed in preparation of the remaining harmonization steps.

Quotes:

> "In the context of the performed case study, matching the activities is easy to perform as the activities have the same names. But, in cases where activities are not well documented (have deviating names) and composite tasks have to match with multiple separate tasks, it is difficult." (Senior Business Analyst 2)
>
> "The matching step is part of the comparison step in ProM and when labels are similar it is an easy job." (Principal Consultant Business Analysis Private Sector)

Comparison step

Question 7: What do you think of the practicability of the comparison step?

From the interviews it becomes clear that performing the differences analysis in ProM is an easy task. The visual representation of the position and type of a difference is however made by hand and this is a laborious task.

Question 8: What do you think of the output of the comparison step?

The output of the differences analysis in ProM is a text based explanation of the position and type of a difference. This text based explanation is hard to interpret, where the visual representation provides more useful information.

Quotes:

> "The differences analysis is an easy job in ProM and the visual results are very useful" (Senior Business Analyst 2)
>
> "The output is very useful and gives an exact explanation of the position of the difference between two models. However, the output in ProM is hard to understand as it is text based. The visual representation is more useful but this step is performed by hand which is a lot of work." (Principal Consultant Business Analysis Private Sector)

Conflict solving step

Question 9: What do you think of the practicability of the conflict solving step?

During the conflict solving step all differences between two models are discussed with the client and solved in accordance with the client and the original models are then revised accordingly. The business analysts of Capgemini indicate that solving the differences between two models is very difficult and that it is only based on what customers want. A methodology is needed to solve the difference between two models which is, e.g., based on a desired model, the Lean philosophy or a Root Cause Analysis.

Question 10: What do you think of the output of the conflict solving step?

According to the interviewees the revised models are useful when an overview of applied changes is maintained next to the revised models.

Quotes:

> "The conflict solving step is very difficult as there is no real methodology available to solve the conflicts. Comparison against a desired model would be helpful." (Senior Business Analyst 1)
>
> "The conflict solving step is simply discussing the conflicts and solving where possible, where this is based on what the client wants. This step must be more rationalized by using techniques like Lean and Root Cause Analysis." (Principal Consultant Business Analysis Private Sector)
>
> "You need to maintain an overview of performed changes in order to know what must be changed in the countries to comply with the harmonized model." (Senior Business Analyst 2)

Integration step

Question10: What do you think of the practicability of the integration step?

The integration step is an easy to perform step in ProM. A disadvantage of the merge algorithm is the fact that the merged model may allow for more behavior than possible in the two underlying models. This means that the merged model must always be checked if it represents the correct behavior and must be corrected if needed.

Question 11: What do you think of the output of the integration step?

Furthermore, the merged model does not show which activities are performed by which country and a merged model in the form of a configurable reference model could solve this problem.

Quotes:

> "The harmonized model can serve as input for the configuration of the new system. However, a configurable reference model would be nice. For instance, you must be able to turn off certain activities when you are dealing with a case for country X which does not need all activities. In other words what is the consequence of turning on all activities for all countries; they have to perform tasks they never did." (Senior Business Analyst 1)

> "The integration step is an easy step in ProM. However, it does not always give the right output and some manual intervention may be needed. It is difficult to see if the merged model represents the behavior possible in the original models." (Principal Consultant Business Analysis Private Sector)

Evaluation of complete approach

Question 13: Are there any missing and/or superfluous steps in the approach? If yes, please explain.

The interviewees indicated that no superfluous steps are present in the approach. However, they did indicate that a possible additional step can be the comparison of the models against a desired/reference model. This step can be part of the conflict solving step or should be performed before the conflict solving step.

Question 14: Can the harmonized model be used as a reference model when implementing a shared service centre? If yes/no, please explain.

The final output of the complete approach is a harmonized model. In order to use the harmonized model as a reference model for the implementation of a shared service centre, the interviewees indicate that the harmonized model should be drawn up as a configurable model. Furthermore the interviewees indicate that in order to use the output of the approach as a blue print for the implementation of a shared service centre, the organizational and data aspect should also be taken into account.

> "A possible additional step would be the comparison with a desired model/behavior" (Senior Business Analyst 1 and Principal Consultant Business Analysis Private Sector)

"You must be able to turn on and off activities when dealing with a certain country." (Senior Business Analyst 2)

"More process model aspects are needed to use the harmonized model as a blue print." (Principal Consultant Business Analysis Private Sector)

7.3 Conclusion

Based on the evaluation results several points of improvement can be drawn up, which are:

- **Generic log construction**
  In the case study in chapter 6, a log file is created from scratch which means that a mapping must be made on the database to identify the process activities. This mapping is converted to a SQL query to extract the data and build the log file (where the XESame tool is used to guide this conversion and construct the log file). However, the developed query is suitable for the case study but cannot be used on a different purchase to pay case. The purchase to pay process is a standard process in SAP but subtle differences between SAP systems exist and the developed query is not able to handle these differences. The construction of a log file from SAP takes much effort and a more generic and flexible approach is needed to simplify the preparation step. The development of a log file from SAP is discussed in [13], [18] and [19] but more research is required towards the development of a generic query, which is able to construct a log file across different SAP systems.

- **Solving divergence and convergence issue**
  As described in section 7.2, Divergence and Convergence is a problem which occurs in a data-centric system like SAP and the phenomena can have an influence on the mined process models. A solution must be found for the phenomena in order to mine correct process models. The problem of Divergence and Convergence is discussed in [13] and [19], where two possible solution have been stated in [19]. The first solution is changing the representation of the process instance. For example, one could for instance change the process instance from the order to the order line. Changing the process instance could solve the problem of divergence if each payment would only relate to one order line. However, in SAP multiple payments can even be received for a single order line which means that the divergence problem cannot be solved. The second solution focuses on a new way of representing the processes executed in SAP. SAP is data-centric and focuses on objects and information and the data-centric design is reflected in the underlying database of the SAP system. A monolithic representation of a SAP process might not be the right way to model a SAP process. Therefore, a new way of modeling is developed referred to as Artifact-Centric-Process-Models. An example of an Artifact-Centric-Process-Model is a Proclet. As described in [19], 'a Proclet can be seen as a workflow process, able to interact with other Proclets that may reside at different levels of aggregation. Several distributed data objects, called artifacts, are present in such process models and are shared among several cases.' The development of this research field is ongoing and more research is needed to investigate if Artifact-Centric-Process-models can solve the Divergence and Convergence issue.

- **Comparison with desired model and development of conflict solving method**
  In the case study described in section 6, six purchase to pay process models are discovered for six different countries. In order to develop a harmonized model, a comparison is made between

the process models. The models are however not compared against a desired or best-practice (reference) model and a comparison against a desired or best-practice model could facilitate the conflict solving step. Discussing and solving the differences between two models can then be based on the desired or best practice reference model. The conflict solving step also lacks a proper methodology to solve the differences between two process models and research is needed towards a conflict solving methodology. Furthermore, differences between process models of the same family are induces by process variation factors. A better understanding of these factors is required in order to solve the differences between two process models.

- **Automation visual representation of differences**
  The visual representation of the position and type of a difference has been indicated to be very useful. The visual representation is however made by hand and automating this step can help to prevent mistakes and reduce the laboriousness of the visualization. More research is needed to investigate if this step can be automated.

- **Integration towards a configurable process model**
  The interviewees indicated that the output of the approach, i.e. a harmonized reference model, can be more useful when it is represented as a configurable process model. A configurable process model is an integrated representation of multiple variants of a same business process in a given domain and a configurable process model offers benefits over traditional process models. For example, a configurable process model enables a clear distinction between those parts that are shared by all process variants and those parts that are specific to certain process variants and keeps track of which process variant(s) each element in the configurable model originates from[2]. In the context of the conducted case study, a configurable process model would show which activities and traces are shared among the countries and which are specific for a certain country. The configurable process model can then be configured to meet specific requirements related to a country when implementing the P2P process within a shared service centre. More research is needed to find out whether configurable process models are suitable for a BPH approach.

- **Adding remaining process modeling aspects to approach**
  The harmonization approach only focuses on the function and control aspect of process models. In order to use the harmonized model is a blue print for the configuration of an information system the remaining modeling aspects should also be addressed. More research is needed to investigate how to harmonize the remaining process modeling aspect, e.g., the organizational and data aspect.

[2] http://www.processconfiguration.com

8. Conclusion

This master thesis focused on the development of a concrete Business Process Harmonization Approach (BPH) using process mining techniques. The goal of the master thesis was defined as follows: *Develop a business process harmonization approach that can be used to create a harmonized reference model.*

The first contribution was the development of a BPH framework based on contemporary literature in the field of BPH. The important elements of a BPH framework have been identified, consisting of steps that need to be performed to arrive at a harmonized reference model, process model aspects that can be harmonized and for each step and aspect, harmonization techniques have been discovered. The BPH framework can be used to select a specific BPH approach, which can be used to create a harmonized reference model.

The second contribution was the development of a concrete BPH approach from the BPH framework. A BPH approach has been developed which focuses on two process modeling aspects of the BPH framework, namely the function and control aspect. For each step in the approach a specific technique has been selected from the BPH framework. An explanation of the approach has been provided by showing the in- and output of each step, as well as an explanation of the selected technique. The BPH approach consists of six steps, namely: Preparation, Discovery, Homogenization, Comparison, Conflict Solving and Integration. A log file needs to be created in the preparation step. The log file contains the execution traces of the cases in the business process under investigation. This file is needed in the discovery step, where the log file is used to mine the process models of the business process. In the harmonization step the discovered models are made comparable by creating a match between the elements of the process models. Then, the comparison step consists of a differences analysis between the process models. During the conflict solving step, the discovered differences between the process models are solved, the models are revised and an overview of performed changes is maintained. Finally, the revised models are merged into a harmonized model in the integration step.

The third contribution was the execution and evaluation of the developed BPH approach in a case setting. The SAP data of six purchasing processes of a client of Capgemini Nederland BV has been used to execute the developed approach and qualitative interviews with business analysts from Capgemini have been conducted to evaluate the BPH approach.

The business analysts at Capgemini have indicated that the BPH approach is a useful approach and describes the steps needed to develop a harmonized model which can be used to reconfigure and harmonize an organization's operational processes. By means of the execution and evaluation of the BPH approach six important points of improvement could be drawn up which serve as a guide for future research. These points of improvement will be discussed in section 8.1.

8.1 Limitations and future work

We have made a first attempt in developing a BPH approach which can be used in a practical setting. However, improvements are required to make the BPH approach a more useful approach and six future research directions could be identified.

1. Future work should focus on the development of a generic log file creation method. In order to build a log file from SAP a SQL query must be developed which is a laborious task and requires

domain knowledge. The developed query can only be used on a single SAP system as many configurations of SAP exist. Therefore, more research is needed towards the development of a more generic and flexible method to build a log file from SAP.

2. Divergence and Convergence occurs in a data-centric system like SAP and it can have an influence on the correctness of models mined with process mining techniques. Divergence occurs in a log file when the same activity is performed several times in one process instance. E.g. when multiple payments are received for a single order item, which will result in a process model with a payment activity that has a single loop. Convergence occurs in a log file when one activity is executed in several process instances at once, which for instance means that one payment is received for multiple order items at once. Business processes in a data-centric system like SAP are often not of a monolithic nature and a different approach may be needed to model SAP process in a correct way. The Artifact-Centric modeling approach has been put forward recently which tries to explain the process with its data objects, called artifacts (e.g. an order, order item, invoice, delivered goods). In an Artifact-Centric approach a separate model is made for each artifact and the models can interact with each other via ports. The ports represent the interaction between the models (and the artifacts) and contain cardinality constraints which indicate how many instances of an activity in one artifact interacts with how many instances of another activity in another artifact. More research is needed to investigate if an Artifact-Centric process modeling approach is able to solve the Divergence and Convergence issue.

3. The conflict solving step of the BPH approach lacks a proper conflict solving methodology as conflicts between process models are solved based on the basis of client's wishes. A comparison of the process models with a desired or best practice model could facilitate the conflict solving step. More research is needed towards a conflict solving methodology and a comparison against a desired or best practice model. Furthermore, a deeper understanding is needed in factors that induce differences between process models of the same family. In order to understand why process models are different and why certain conflicts can be solved and others cannot, more research is needed on these factors.

4. The comparison step consists of a differences analysis where a visual representation of the position and type of a difference is made by hand. More research is needed to investigate if this step can be automated.

5. The output of the BPH approach is a harmonized merged model on the basis of multiple underlying models. A merged model does however not show those parts that are shared by all process variants and those parts that are specific to certain process variants. A configurable process model is able to handle this issue and keeps track of which process variant(s) each element in the configurable model originates from. More research is needed towards the usefulness of using a configurable model is output of the BPH approach.

6. The BPH approach only focused on the harmonization of the function and control aspect of process models and more research is needed to investigate how the remaining modeling aspects (e.g. the data and organizational aspect) can be harmonized.

References

[1] R. Tregear, “Business Process Standardization,” *Handbook on Business Process Management 2*, pp. 307-327, 2010.

[2] B. Münstermann, A. Eckhardt, and T. Weitzel, “The performance impact of business process standardization: An empirical evaluation of the recruitment process,” *Business Process Management Journal*, vol. 16, no. 1, pp. 29-56, 2010.

[3] T. H. Davenport, “The coming commoditization of processes,” *Harvard Business Review*, vol. 83, no. 6, pp. 100-108, 2005.

[4] M. Schafermeyer, D. Grgecic, and C. Rosenkranz, “Factors Influencing Business Process Standardization: A Multiple Case Study,” *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pp. 1-10, 2010.

[5] P. Lillrank, “The quality of standard, routine and nonroutine processes,” *Organization Studies*, vol. 24, no. 2, pp. 215-233, 2003.

[6] A. Ricken and A. Steinhorst, “Standardization or Harmonization? You need Both,” *European Health Informatics*, 2005.

[7] H. Romero, R. Dijkman, P. Grefen, and A. Weele, “Harmonization of Business Process Models,” *Business Process Management Workshops*, pp. 13-24, 2012.

[8] C. Pardo, F. J. Pino, F. Garcia, M. Piattini, and M. T. Baldassarre, “A process for driving the harmonization of models,” *Proceedings of the 11th International Conference on Product Focused Software*, pp. 51-54, 2010.

[9] R. Dijkman, M. L. Rosa, and H. A. Reijers, “Managing large collections of business process models—Current techniques and challenges,” *Computers in Industry*, vol. 63, no. 2, pp. 91-97, 2012.

[10] T. DAVENPORT, “The New Industrial Engineering: Information Technology and Business Process Redesign,” *Sloan Management Review*, pp. 11-27, 1990.

[11] E. Kindler, B. Axenath, and V. Rubin, “Amfibia: A meta-model for the integration of business process modelling aspects,” *The Role of Business Processes in Service Oriented Architectures*, vol. 6291, 2006.

[12] W. Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

[13] J. Buijs, “Mapping Data Sources to XES in a Generic Way,” 2010.

[14] T. Curran, G. Keller, and A. Ladd, *SAP R/3 business blueprint: understanding the business process reference model*. Prentice-Hall, Inc., 1997.

[15] R. Melissen, “A Business Process Harmonization Approach: Literature Review,” 2012.

[16] C. Pardo, F. Pino, F. Garc’ia, M. Piattini, M. Baldassarre, and S. Lemus, “Homogenization, comparison and integration: a harmonizing strategy for the unification of multi-models in the banking sector,” *Product-Focused Software Process Improvement*, pp. 59-72, 2011.

[17] L. Xiao and L. Zheng, “Business process design: Process comparison and integration,” *Information Systems Frontiers*, pp. 1-12, 2010.

[18] M. van Giessel, “Process Mining in SAP R/3,” 2004.

[19] D. A. M. Piessens, “Event Log Extraction from SAP ECC 6.0,” 2011.

[20] S. Jadhav, “Business Process Discovery,” *BPTrends*, 2011.

[21] D. Yue, X. Wu, H. Wang, and J. Bai, “A review of process mining algorithms,” *Business Management and Electronic Information (BMEI), 2011 International Conference on*, vol. 5, pp. 181-185, 2011.

[22] W. M. P. van der Aalst and A. Weijters, “Process mining: a research agenda,” *Computers in Industry*, vol. 53, no. 3, pp. 231-244, 2004.

[23] A. Tiwari, C. Turner, and B. Majeed, “A review of business process mining: state-of-the-art and future trends,” *Business Process Management Journal*, vol. 14, no. 1, pp. 5-22, 2008.

[24] M. Becker and R. Laue, “Analysing Differences between Business Process Similarity Measures,” *Business Process Management Workshops*, pp. 39-49, 2012.

[25] M. Weidlich, R. Dijkman, and J. Mendling, “The ICoP framework: Identification of correspondences between process models,” *Advanced Information Systems Engineering*, pp. 483-498, 2010.

[26] R. Dijkman, M. Dumas, B. Van Dongen, R. Käärik, and J. Mendling, “Similarity of business process models: Metrics and evaluation,” *Information Systems*, vol. 36, no. 2, pp. 498-516, 2011.

[27] R. Dijkman, “A classification of differences between similar Business Processes,” *Enterprise Distributed Object Computing Conference, 2007. EDOC 2007. 11th IEEE International*, pp. 37-37, 2007.

[28] R. Dijkman, “Diagnosing differences between business process models,” *Business Process Management*, pp. 261-277, 2008.

[29] R. Dijkman, *Feedback on differences between business processes*. Beta, Research School for Operations Management and Logistics, 2007.

[30] Y. Juan and C. Ou-Yang, "A process logic comparison approach to support business process benchmarking," *The International Journal of Advanced Manufacturing Technology*, vol. 26, no. 1, pp. 191-210, 2005.

[31] Y. C. Juan, "A string comparison approach to process logic differences between business process models," *Proceedings of the 9th Joint Conference on Information Sciences*, 2006.

[32] J. A. Gulla and R. Mollan, "Implementing SAP R/3 in a Multi-cultural Organization," *1 extordmasculine International Workshop on Enterprise Management Resource and Planning Systems EMRPS, Venice, Italy*, pp. 127-134, 1999.

[33] R. Crerie, F. A. Baião, and F. M. Santoro, "Discovering business rules through process mining," *Enterprise, Business-Process and Information Systems Modeling*, pp. 136-148, 2009.

[34] J. M. Hall and M. E. Johnson, "When should a process be art, not science?," *Harvard business review*, vol. 87, no. 3, pp. 58-65, 2010.

[35] C. Rosenkranz, S. Seidel, J. Mendling, M. Schaefermeyer, and J. Recker, "Towards a framework for business process standardization," *Business Process Management Workshops*, pp. 53-63, 2010.

[36] S. Sun, A. Kumar, and J. Yen, "Merging workflows: A new perspective on connecting business processes," *Decision Support Systems*, vol. 42, no. 2, pp. 844-858, 2006.

[37] F. Gottschalk, W. van der Aalst, and M. Jansen-Vullers, "Merging event-driven process chains," *On the Move to Meaningful Internet Systems: OTM 2008*, pp. 418-426, 2008.

[38] M. La Rosa, M. Dumas, R. Käärik, and R. Dijkman, "Merging business process models (extended version)," 2009.

[39] M. La Rosa, M. Dumas, R. Uba, and R. M. Dijkman, "Business process model merging: an approach to business process consolidation," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2012.

[40] W. Derguech and S. Bhiri, "Merging Business Process Variants," *Business Information Systems*, pp. 86-97, 2011.

[41] J. M. Küster, C. Gerth, A. Förster, and G. Engels, "A tool for process merging in business-driven development," *Proc. of the CAiSE*, pp. 89-92, 2008.

[42] W. van der Aalst and M. Song, “Mining Social Networks: Uncovering interaction patterns in business processes,” *Business Process Management*, pp. 244-260, 2004.

[43] A. K. A. De Medeiros and A. Weijters, “Genetic process mining,” *Applications and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science*, 2005.

[44] J. Herbst and D. Karagiannis, “Workflow mining with InWoLvE,” *Computers in Industry*, vol. 53, no. 3, pp. 245-264, 2004.

[45] W. Van der Aalst, T. Weijters, and L. Maruster, “Workflow mining: Discovering process models from event logs,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 9, pp. 1128-1142, 2004.

[46] G. Schimm, “Mining exact models of concurrent workflows,” *Computers in Industry*, vol. 53, no. 3, pp. 265-281, 2004.

[47] B. F. van Dongen and W. M. P. van der Aalst, “Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets,” *PNCWB 2005 workshop*, pp. 35-58, 2005.

[48] A. Weijters, W. M. P. van der Aalst, and A. K. A. de Medeiros, “Process mining with the heuristics miner-algorithm,” *Technische Universiteit Eindhoven, Tech. Rep. WP*, vol. 166, 2006.

[49] A. K. A. de Medeiros, B. F. van Dongen, W. M. P. van der Aalst, and A. Weijters, “Process mining: Extending the alpha-algorithm to mine short loops,” *Eindhoven University of Technology, Eindhoven*, vol. 19, 2004.

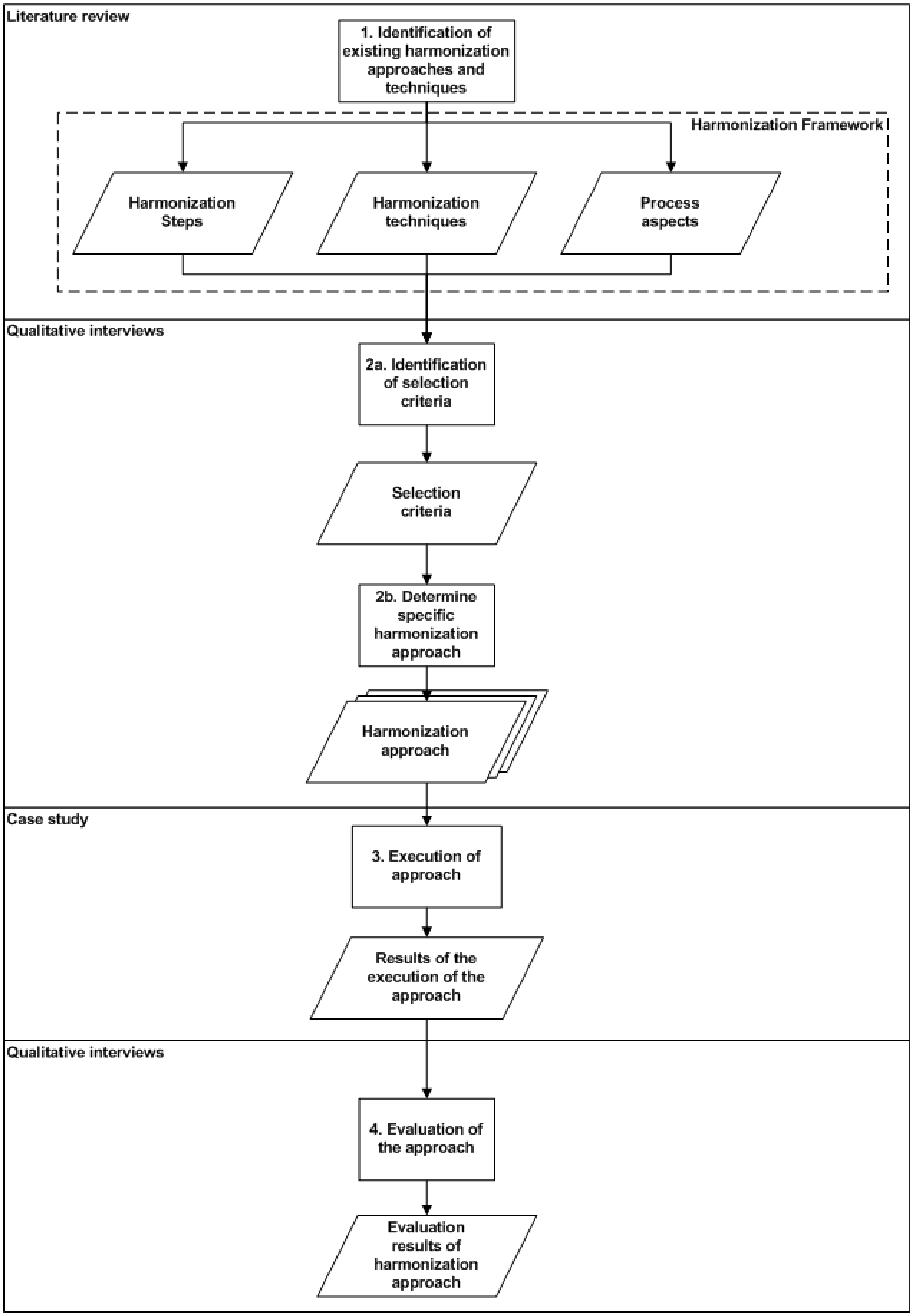
[50] J. E. Cook and A. L. Wolf, “Discovering models of software processes from event-based data,” *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 7, no. 3, pp. 215-249, 1998.

[51] C. Günther and W. van der Aalst, “Fuzzy mining-adaptive process simplification based on multi-perspective metrics,” *Business Process Management*, pp. 328-343, 2007.

[52] B. Van Dongen, A. de Medeiros, H. Verbeek, A. Weijters, and W. Van der Aalst, “The ProM framework: A new era in process mining tool support,” *Applications and Theory of Petri Nets 2005*, pp. 1105-1116, 2005.

Appendix

Appendix 1: Research method



Appendix 2: Process mining discovery algorithms and reference list

		Process mining techniques						Modeling Notation			
Papers	Algorithm (Name)	Data Mining techniques	Genetic Algorithms	Markovian Approach	Cluster Analysis	Neural Networks	Other Algorithmic Approaches	EPC	Petri-Net	Software/tool Support	Process Aspect
(1)(2)(3)	Genetic		X					X	X	Yes (ProM)	Function & Control
(4)	Inductive						X			No	Function & Control
(5)	induceUnique-NodeSAG	X								Yes (InWoLvE)	Function & Control
(6)(7)(8)(9)	Alpha						X		X	Yes (ProM)	Function & Control
(10)(11)		X			X					Yes (Process Miner)	Function & Control
(12)		X								No	Function & Control
(13)							X		X	No	Function & Control
(14)(15)	Multi-phase mining						X	X	X	Yes (ProM & ARIS PPM)	Function & Control
(16)(17)	Heuristic						X		X	Yes (Little Thumb & ProM)	Function & Control
(18)							X		X	Yes (MiSoN & ProM)	Organization
(19)							X			No	Function & Control
(20)		X								No	Function & Control
(21)	Greedy						X		X	No	Function & Control
(22)							X			No	Function & Control
(23)(24)	Alpha+						X		X	Yes (EMiT)	Function & Control
(25)							X			No	Function & Control
(26)	Ktail/Markov/RNet			X		X				Yes (DaGama)	Function & Control
(27)	Fuzzy Miner						X		X	Yes (ProM)	Function & Control

1. De Medeiros AKA, Weijters A. Genetic process mining. Applications and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science 2005;

2. De Medeiros AKA, Weijters A, Van Der Aalst W. Using genetic algorithms to mine process models: representation, operators and results. 2004;

3. de Medeiros A, Weijters A, Van Der Aalst W. Genetic process mining: a basic approach and its challenges. Business Process Management Workshops 2006;203-215.

4. Zhang S, Gu N, Lian JX, Li SH. Workflow process mining based on machine learning. Machine Learning and Cybernetics, 2003 International Conference on 2003;4:2319-2323.

5. Herbst J, Karagiannis D. Workflow mining with InWoLvE. Computers in Industry 2004;53(3):245-264.

6. Van der Aalst W, Weijters T, Maruster L. Workflow mining: Discovering process models from event logs. Knowledge and Data Engineering, IEEE Transactions on 2004;16(9):1128-1142.

7. van Dongen B, van der Aalst W. EMiT: A process mining tool. Applications and Theory of Petri Nets 2004 2004;454-463.

8. de Medeiros A, van der Aalst W, Weijters A. Workflow mining: Current status and future directions. On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE 2003;389-406.

9. Van der Aalst W, Weijters A, Maruster L. Workflow mining: Which processes can be rediscovered. 2002;

10. Schimm G. Mining exact models of concurrent workflows. Computers in Industry 2004;53(3):265-281.

11. Schimm G. Mining most specific workflow models from event-based data. Business process management 2003;1021-1021.

12. Greco G, Guzzo A, Manco G, Sacca D. Mining and reasoning on workflows. Knowledge and Data Engineering, IEEE Transactions on 2005;17(4):519-534.

13. Weijters A, Van der Aalst W. Process mining: discovering workflow models from event-based data. Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001) 2001;283-290.

14. van Dongen BF, van der Aalst WMP. Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets. PNCWB 2005 workshop 2005;35-58.

15. Van Dongen B, van der Aalst W. Multi-phase process mining: Building instance graphs. Conceptual Modeling-ER 2004 2004;362-376.

16. Weijters A, van der Aalst WMP. Rediscovering workflow models from event-based data using little thumb. Integrated Computer Aided Engineering 2003;10(2):151-162.

17. Weijters A, van der Aalst WMP, de Medeiros AKA. Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP 2006;166

18. van der Aalst W, Song M. Mining Social Networks: Uncovering interaction patterns in business processes. Business Process Management 2004;244-260.

19. Golani M, Pinter S. Generating a process model from a process audit log. Business Process Management 2003;1020-1020.

20. Hwang SY, Yang WS. On the discovery of process models from their instances. Decision Support Systems 2002;34(1):41-57.

21. Cook JE, Du Z, Liu C, Wolf AL. Discovering models of behavior for concurrent workflows. Computers in Industry 2004;53(3):297-319.

22. Hwang SY, Wei CP, Yang WS. Discovery of temporal patterns from process instances. Computers in Industry 2004;53(3):345-364.

23. De Medeiros A, Van Dongen B, Van Der Aalst W, Weijters A. Process mining for ubiquitous mobile systems: an overview and a concrete algorithm. Ubiquitous Mobile Information and Collaboration Systems 2005;151-165.

24. de Medeiros AKA, van Dongen BF, van der Aalst WMP, Weijters A. Process mining: Extending the alpha-algorithm to mine short loops. Eindhoven University of Technology, Eindhoven 2004;19

25. Chen K, Yun D. Discovering process models from execution history by graph matching. Intelligent Data Engineering and Automated Learning 2003;887-892.

26. Cook JE, Wolf AL. Discovering models of software processes from event-based data. ACM Transactions on Software Engineering and Methodology (TOSEM) 1998;7(3):215-249.

27. Günther C, van der Aalst W. Fuzzy mining-adaptive process simplification based on multi-perspective metrics. Business Process Management 2007;328-343.

Appendix 3: Constructing the event logs in XESame

This appendix describes and visualizes how the raw log file belonging to the purchase to pay process is extracted by using the XESame tool. The exported data from the SAP system is stored in a folder containing 11 database tables in the csv file format, namely:

BKPF	Accounting Document Header
BSEG	Accounting Document Segment
CDHDR	Change Document Header
CDPOS	Change Document Items
EKES	Vendor Confirmations
EKKO	Purchasing Document Header
EKPO	Purchasing Document Item
MKPF	Header Material Document
MSEG	Document Segment Material
RBKP	Document Header Invoice Receipt
RSEG	Document Item Incoming Invoice

First a connection must be made to the data folder via the ODBC (Open DataBase Connectivity) Data Source Administrator in Windows which is shown in Figure A. 1. A new data source must be added in the Administrator and a driver must be selected (in our case the Microsoft Access Text Driver) as shown in Figure A. 1. Next, a name must be created for the data source and the directory must be selected where the csv files are stored as shown in Figure A. 2.

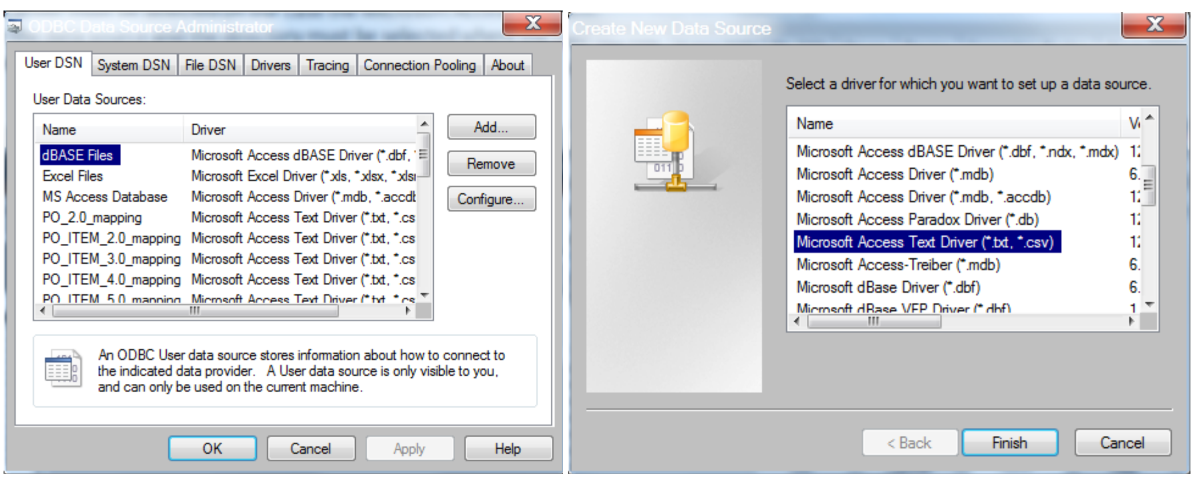


Figure A. 1: ODBC Data Source Administrator and selection of driver

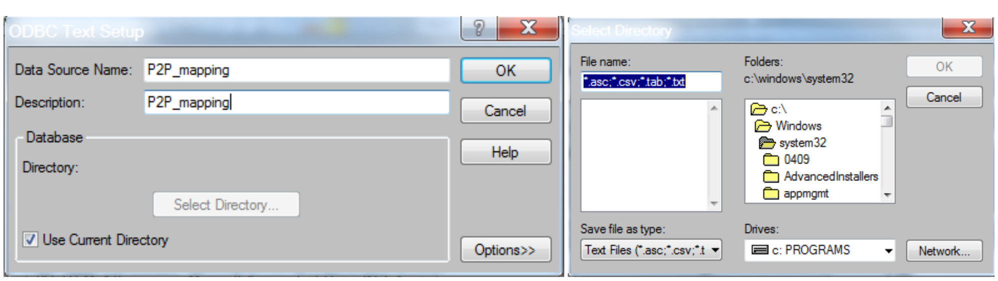


Figure A. 2: ODBC Name creation for data source and directory selection

Then via the options (shown in Figure A. 2) and define format controls a new format must be defined for all column headers that are present in each of the database files, as shown in Figure A. 3. This step will create a schema file in the directory where the database files are stored. This schema file contains the format definition of the database files which are stored in the directory.

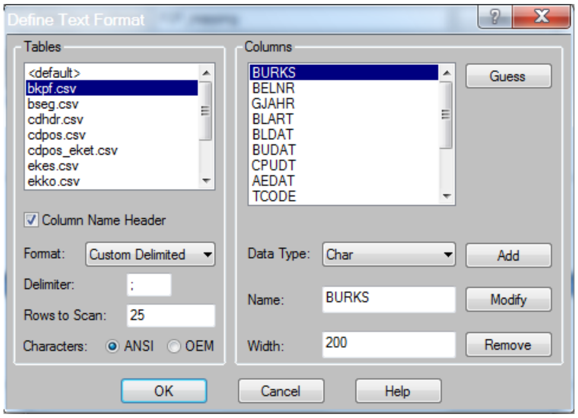


Figure A. 3: ODBC Define format window

In XESame a connection must be made to the database. In Figure A. 4 the connection configuration is shown where it is important that the part behind the jdbc:odbc: part in the URL to database section has the same name as the name that has been created in the ODBC setup mentioned previously. Via the test connection control shown in Figure A. 4 the connection can be tested.

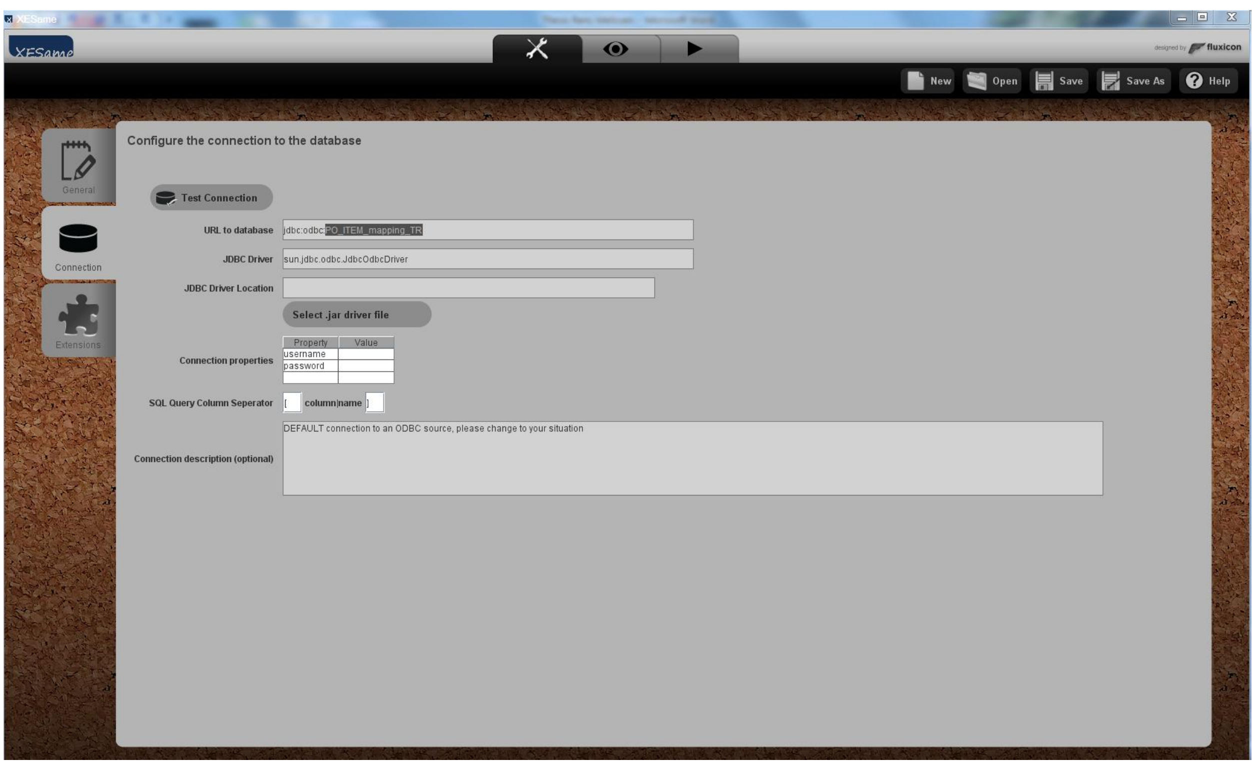


Figure A. 4: XESame database connection

Now the actual mapping can be configured in XESame where the mapping of Turkey is used as an example. In XESame a tree structure is shown which consists of a Log tab, Trace tab and Event tab, where for each event a separate Event tab is shown. For each of these three tabs an Attributes and Properties window must be specified and these specifications are shown in Figure A. 5 to Figure A. 36.

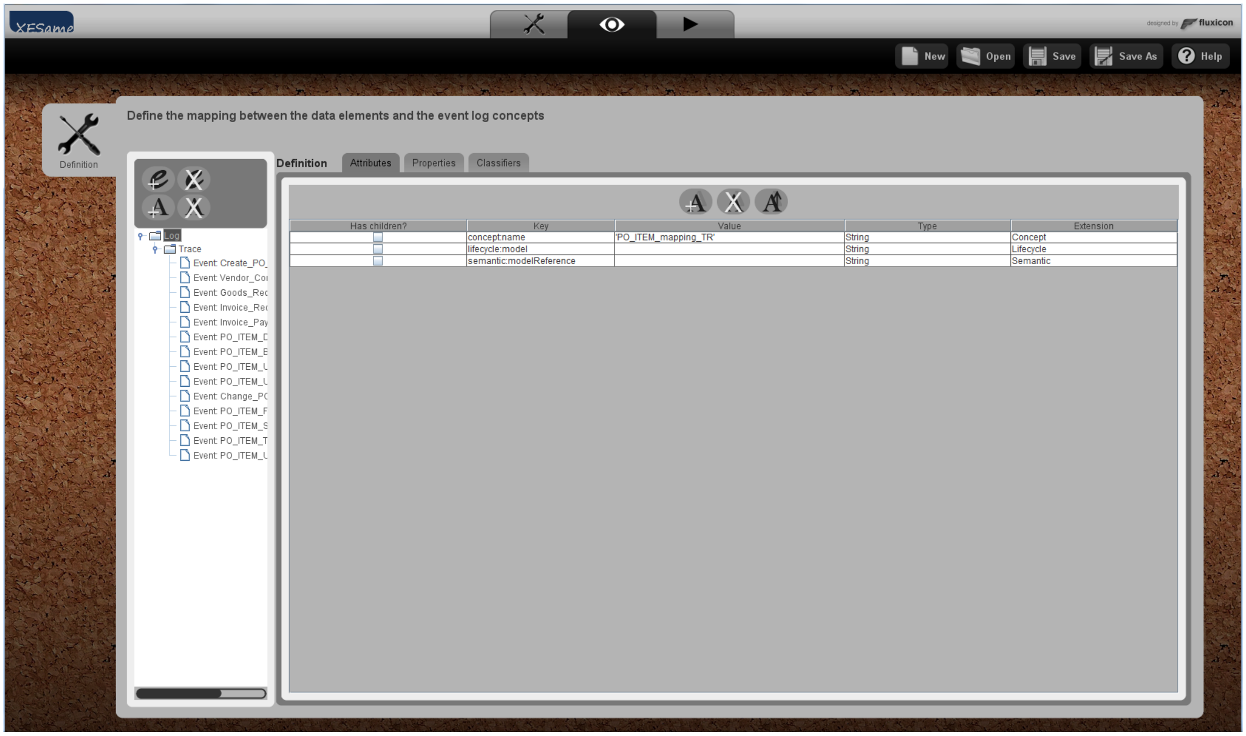


Figure A. 5: Log Attributes specification

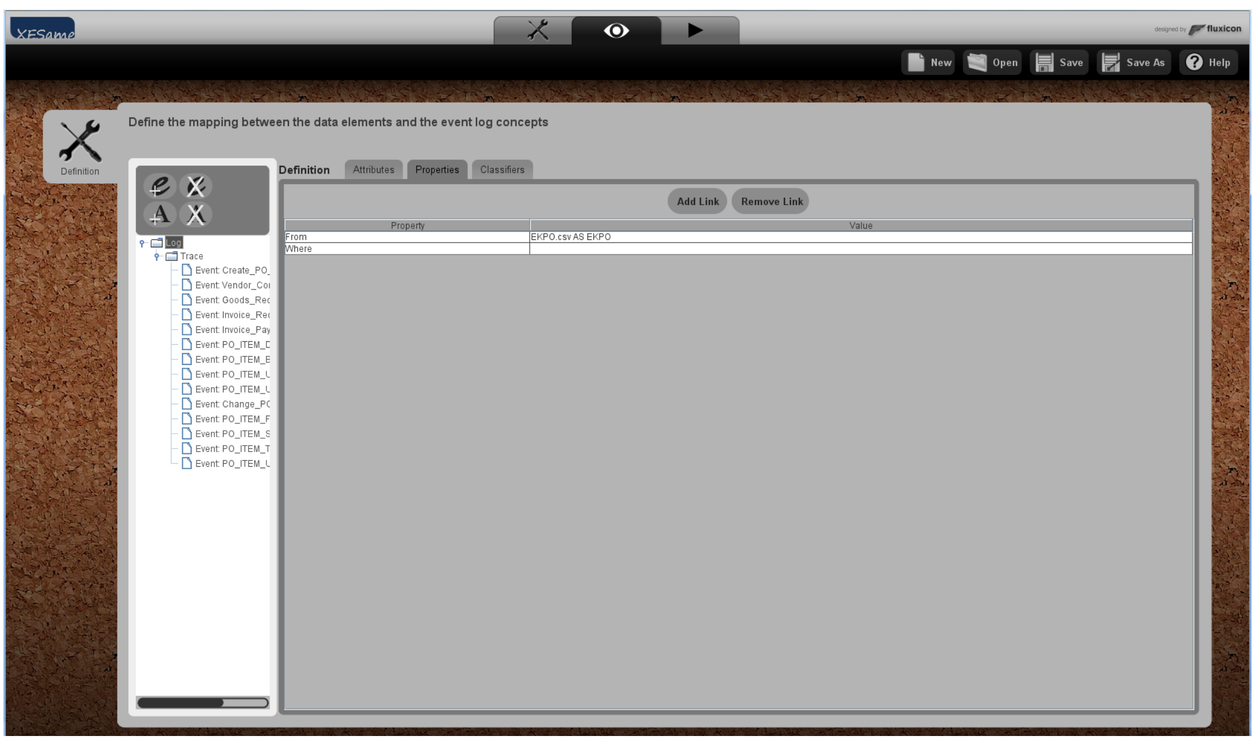


Figure A. 6: Log Properties specification

Figure A. 7: Trace Attributes specification

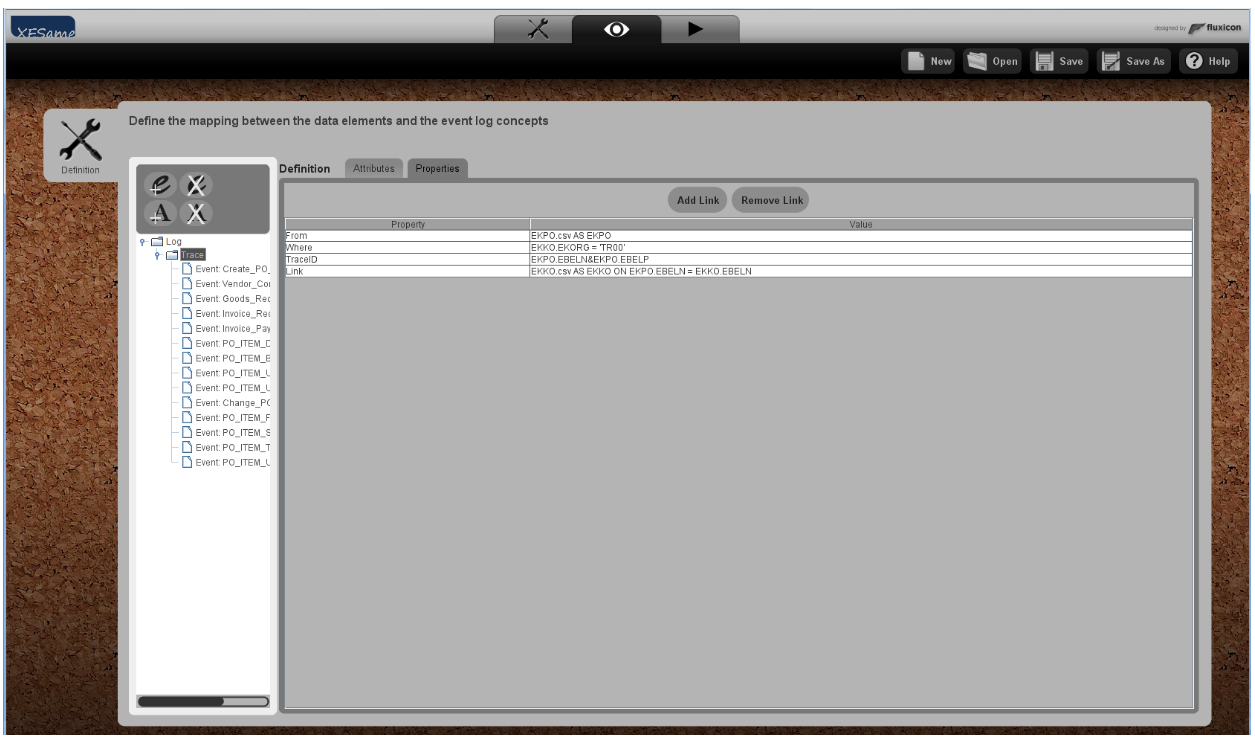


Figure A. 8: Trace Properties specification

Figure A. 9: Event Create_PO_ITEM Attributes specification

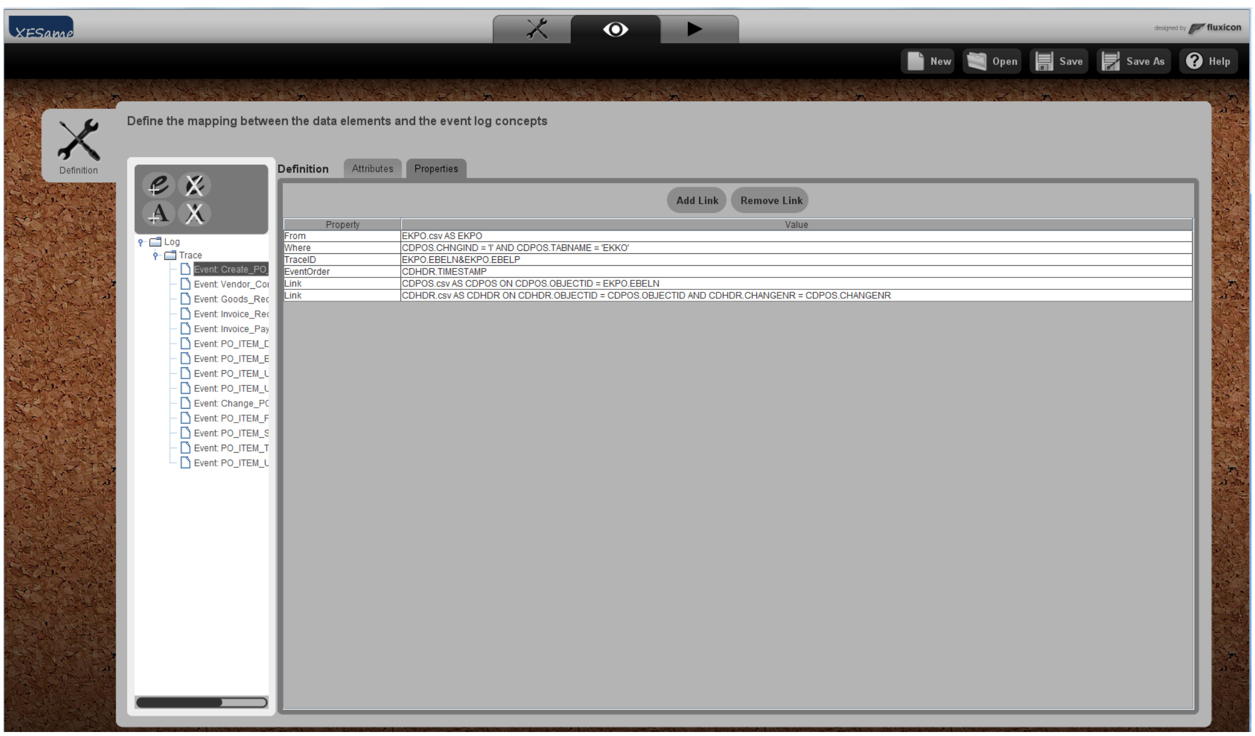


Figure A. 10: Event Create_PO_ITEM Properties specification

Figure A. 11: Event Vendor_Confirmation Attributes specification

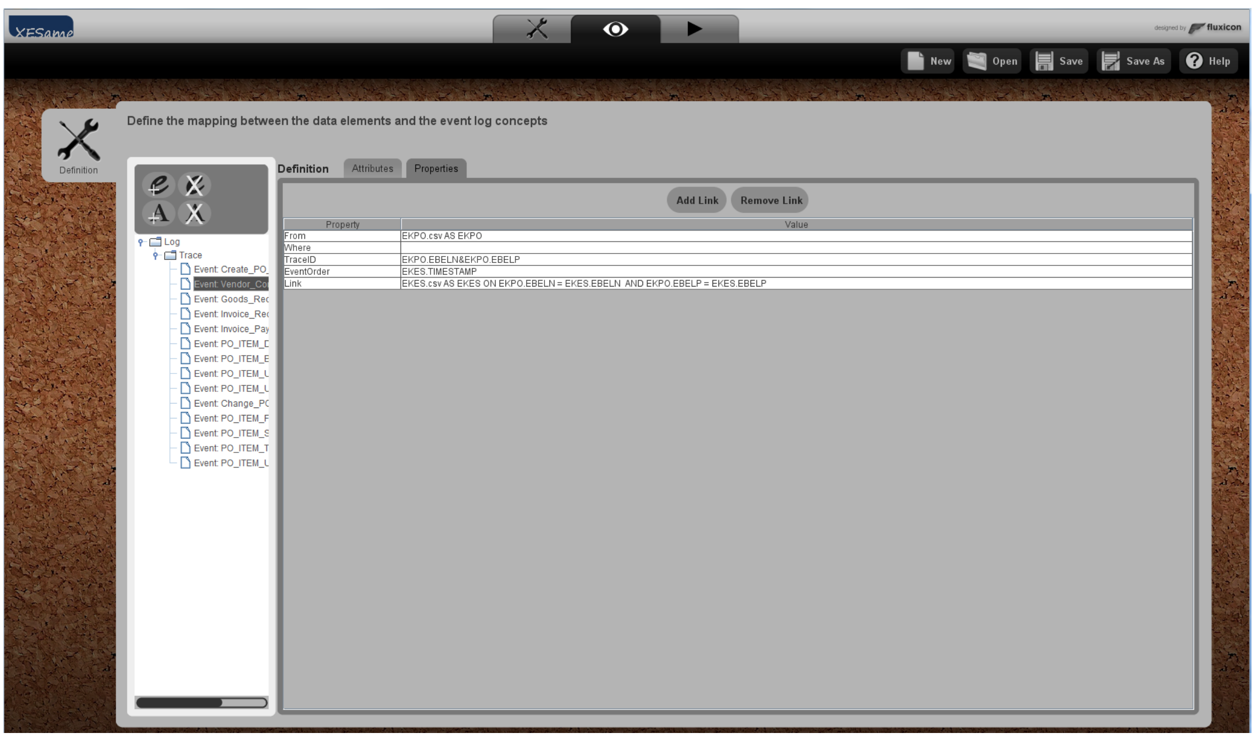


Figure A. 12: Event Vendor_Confirmation Properties specification

Figure A. 13: Event Goods_Receipt Attributes specification

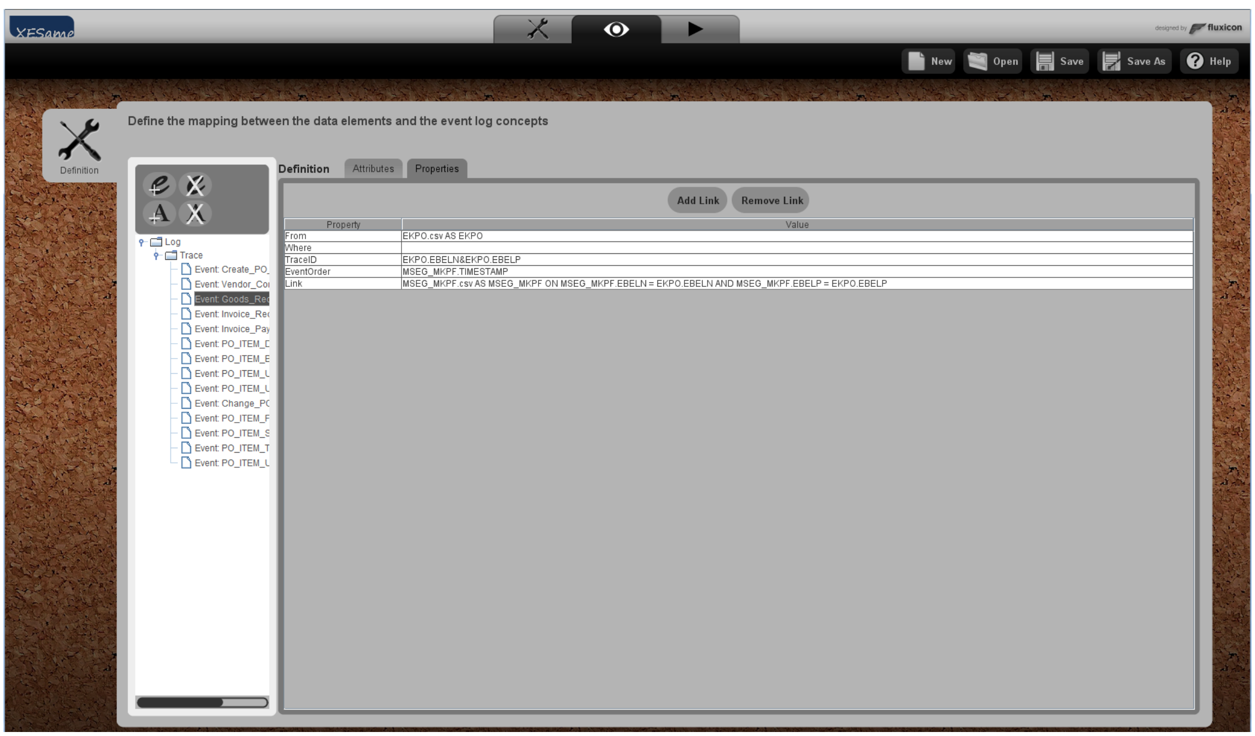


Figure A. 14: Event Goods_Receipt Properties specification

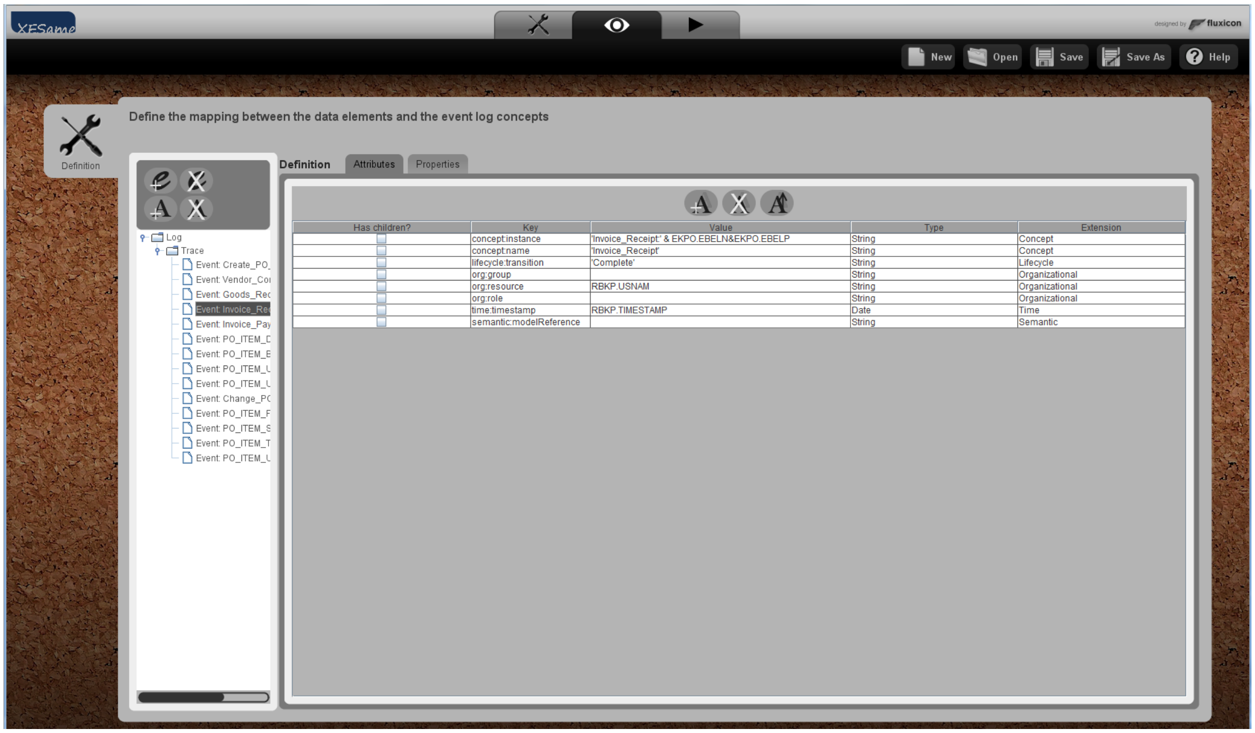


Figure A. 15: Event Invoice_Receipt Attributes specification

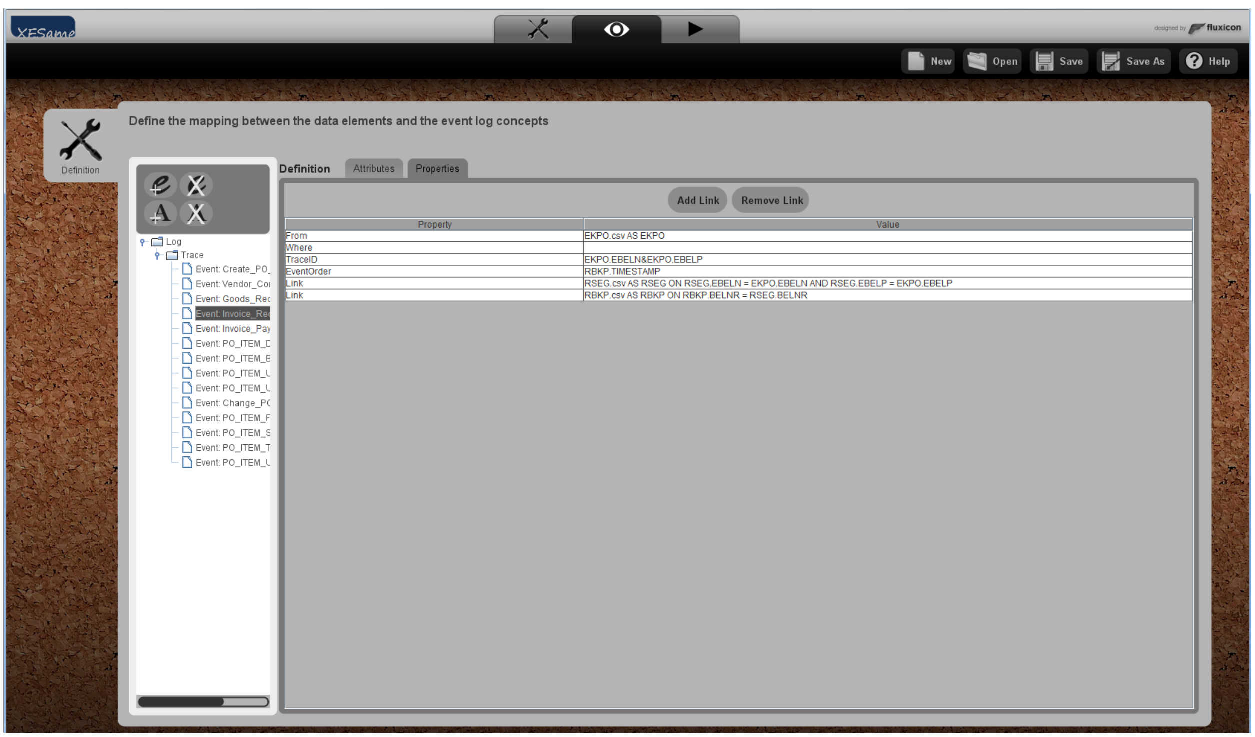


Figure A. 16: Event Invoice_Receipt Properties specification

Figure A. 17: Event Invoice_Payment Attributes specification

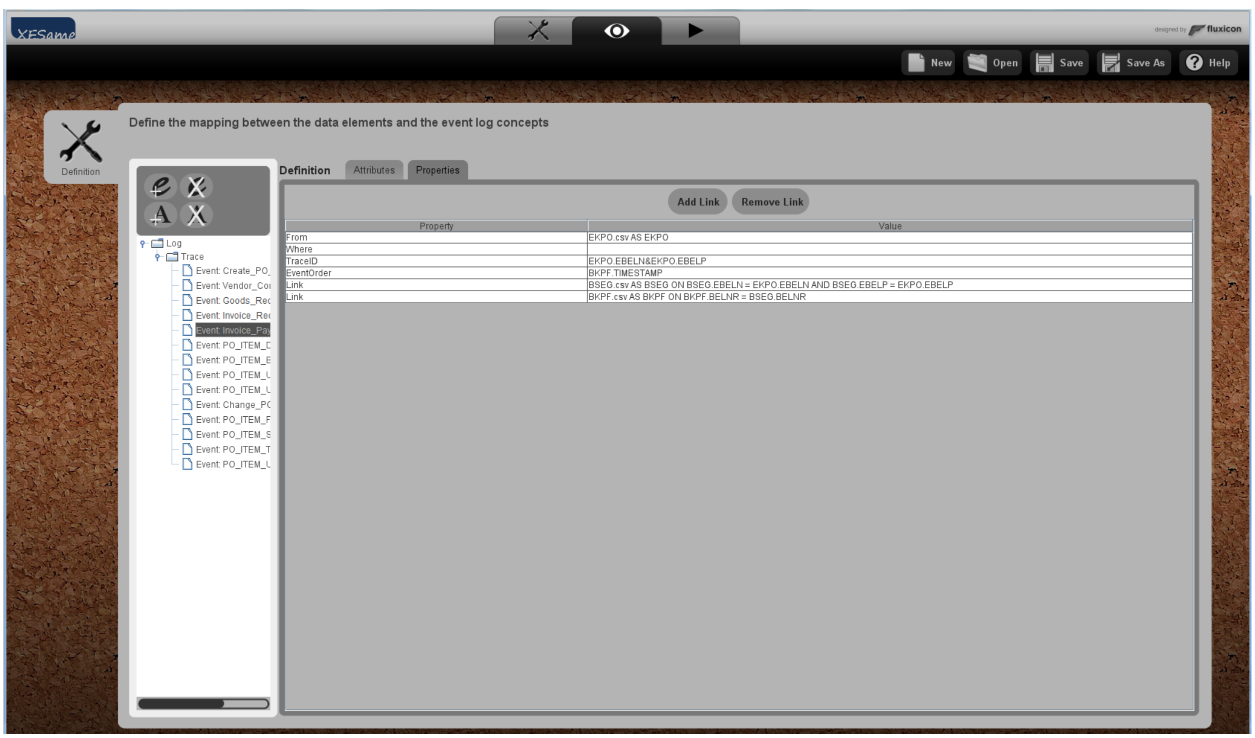


Figure A. 18: Event Invoice_Payment Properties specification

Figure A. 19: Event Delete_PO_ITEM Attributes specification

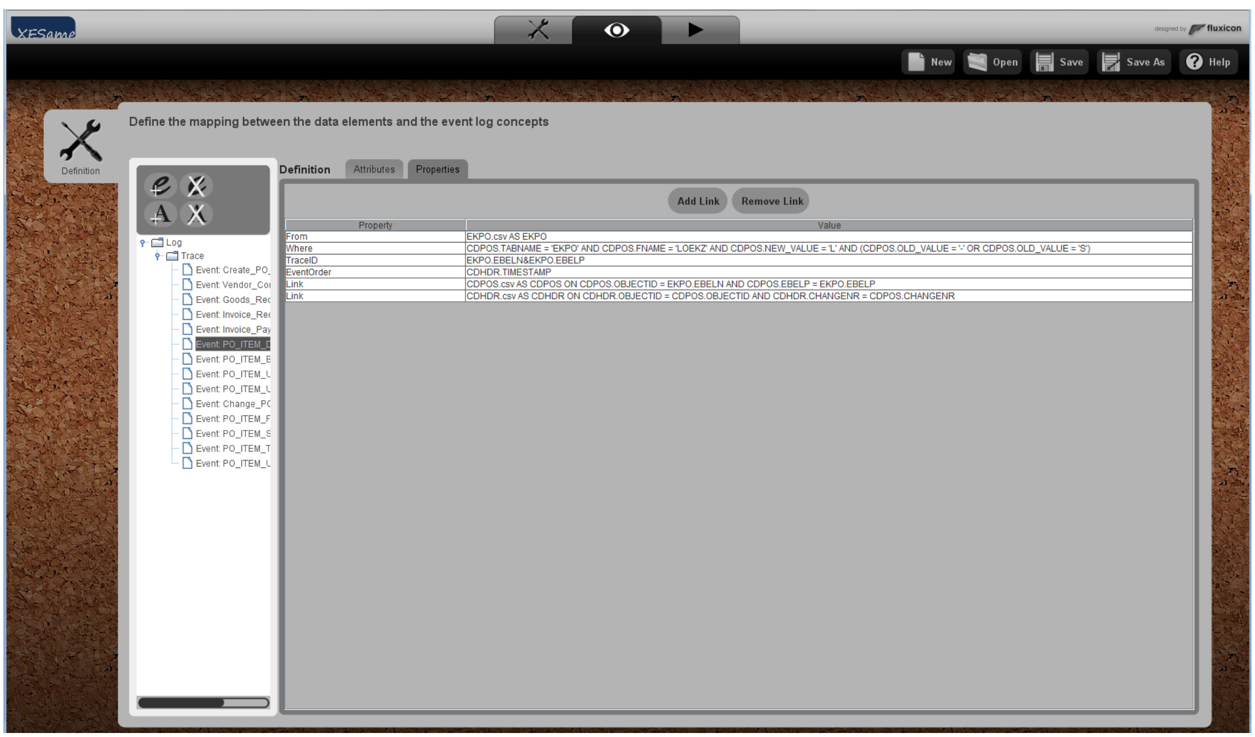


Figure A. 20: Event Delete_PO_ITEM Properties specification

Figure A. 21: Event Block_PO_ITEM Attributes specification

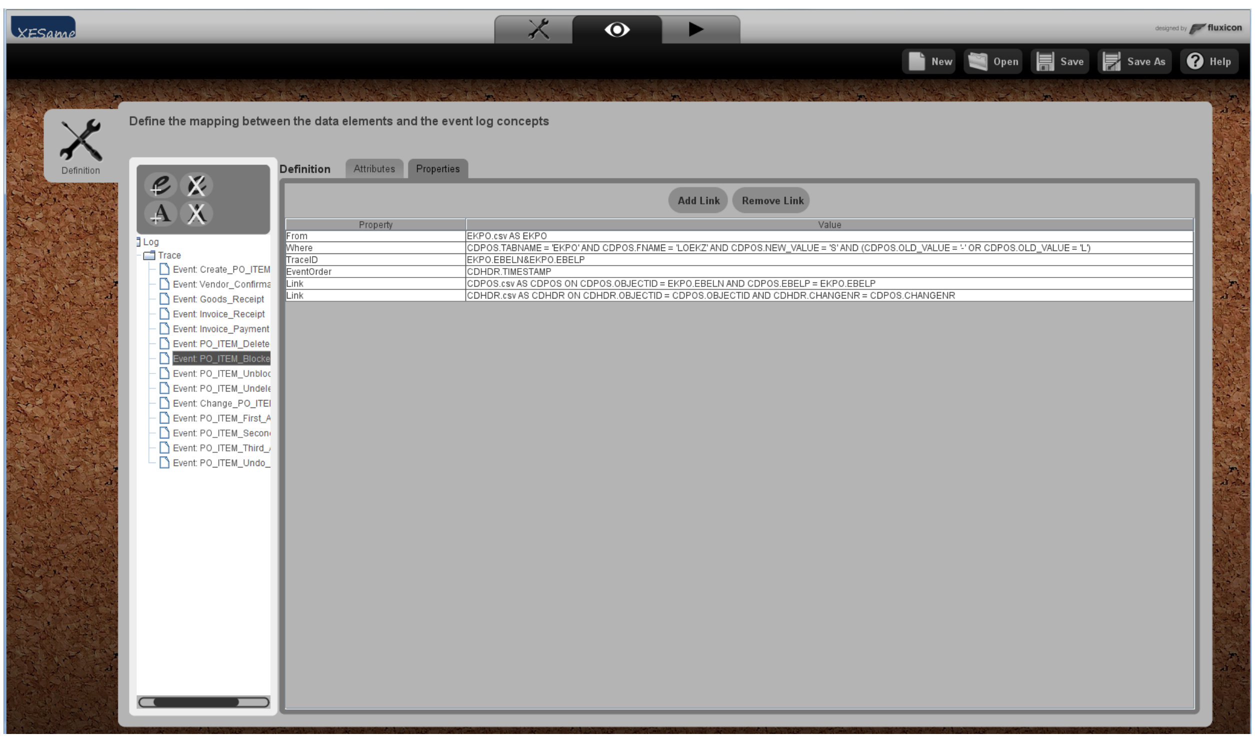


Figure A. 22: Event Block_PO_ITEM Properties specification

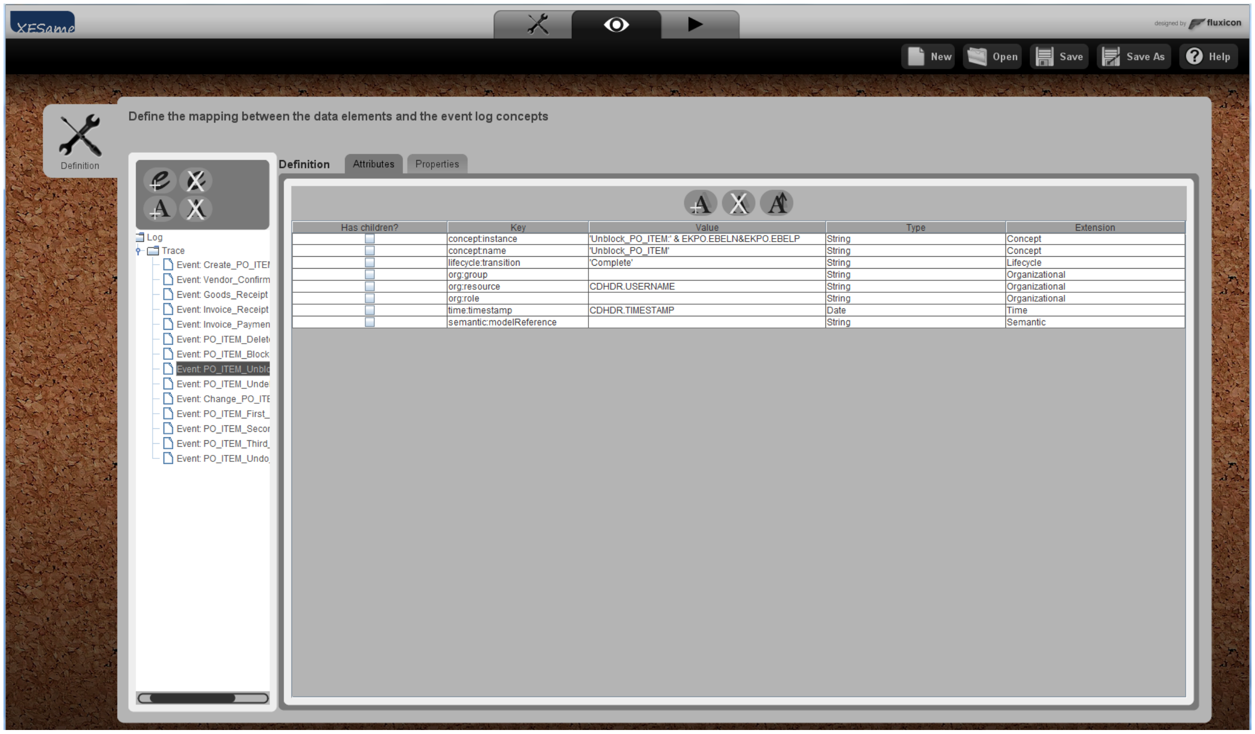


Figure A. 23: Event Unblock_PO_ITEM Attributes specification

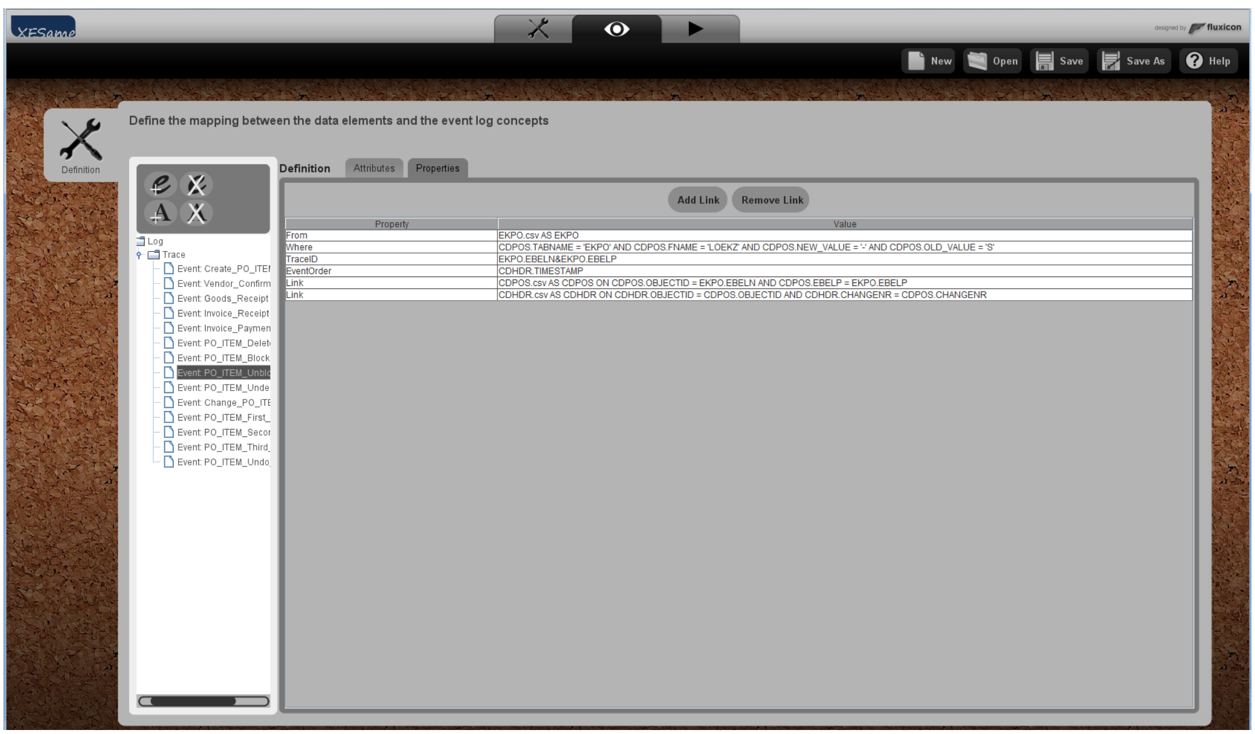


Figure A. 24: Event Unblock_PO_ITEM Properties specification

Figure A. 25: Event Undelete_PO_ITEM Attributes specification

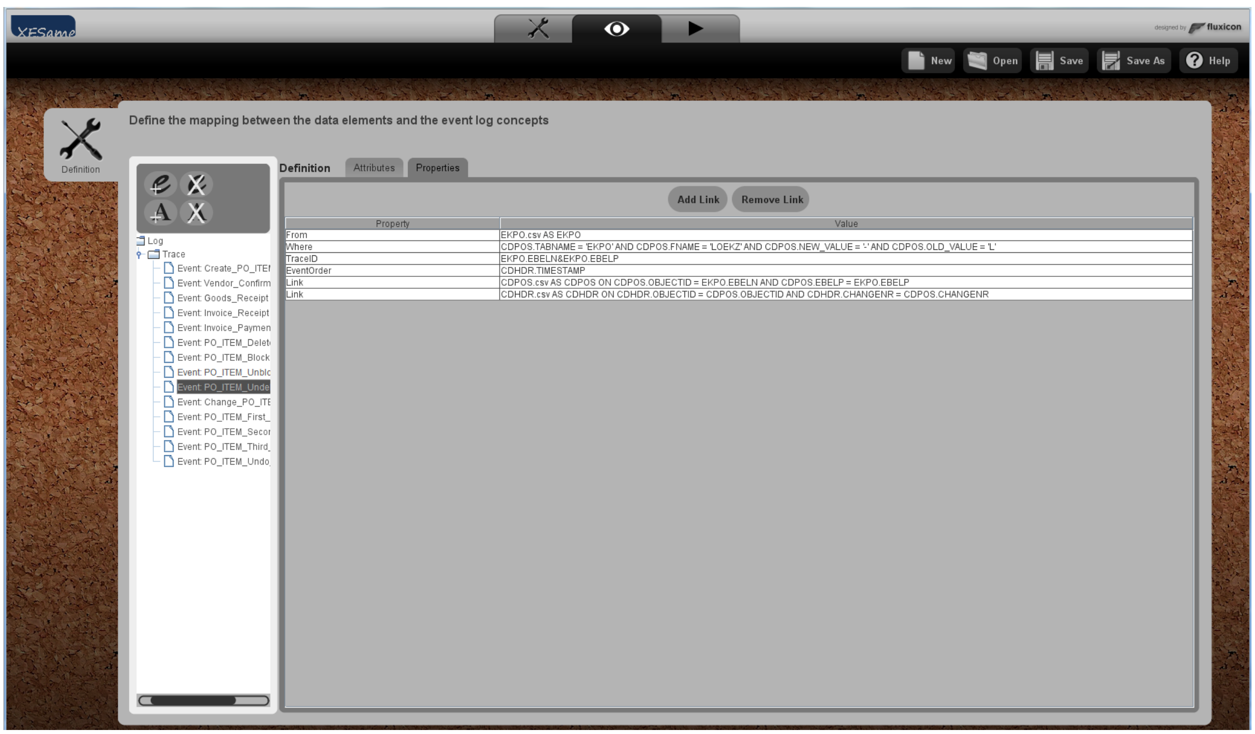


Figure A. 26: Event Undelete_PO_ITEM Properties specification

Figure A. 27: Event Change_PO_ITEM Attributes specification

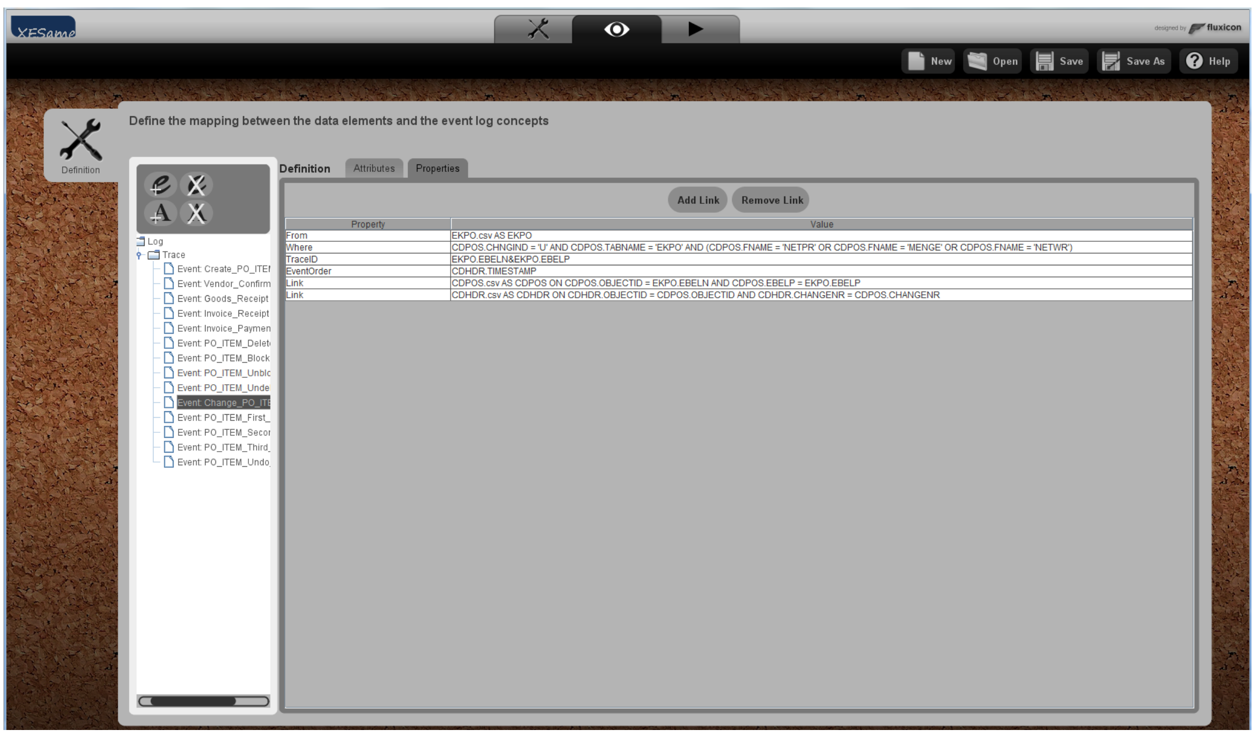


Figure A. 28: Event Change_PO_ITEM Properties specification

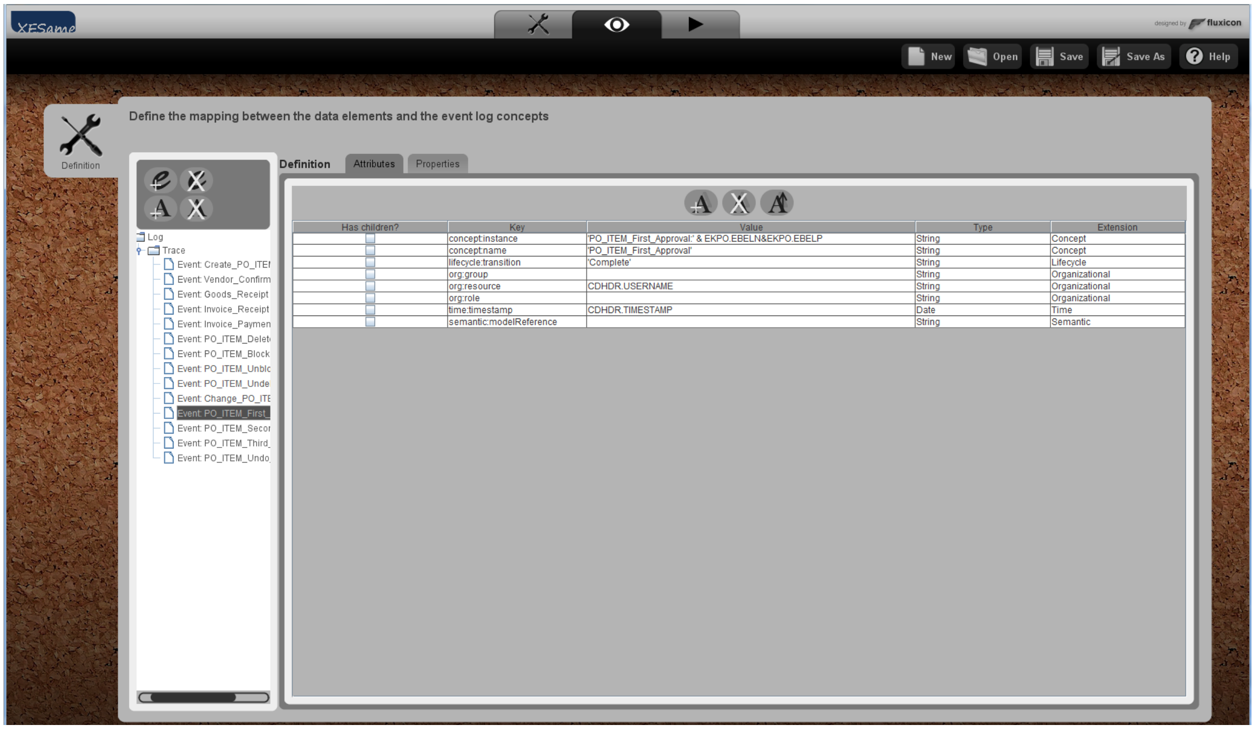


Figure A. 29: PO_ITEM_First_Approval Attributes specification

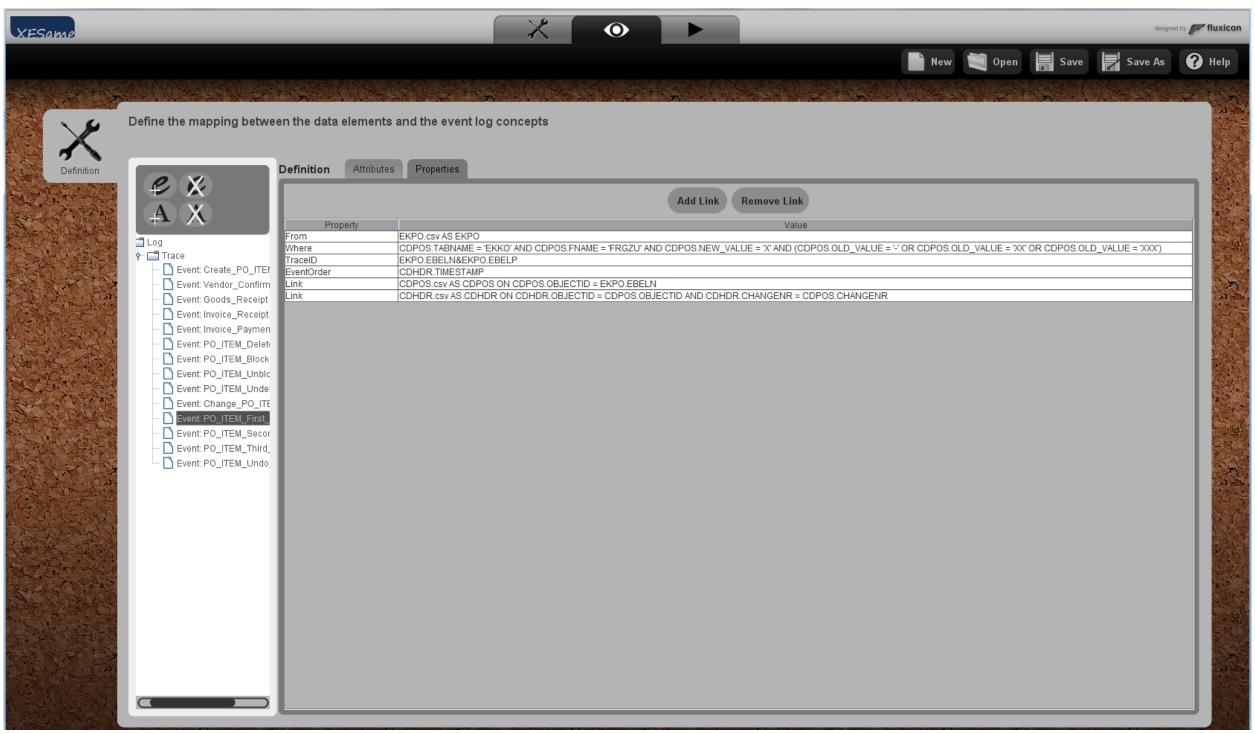


Figure A. 30: PO_ITEM_First_Approval Properties specification

Figure A. 31: PO_ITEM_Second_Approval Attributes specification

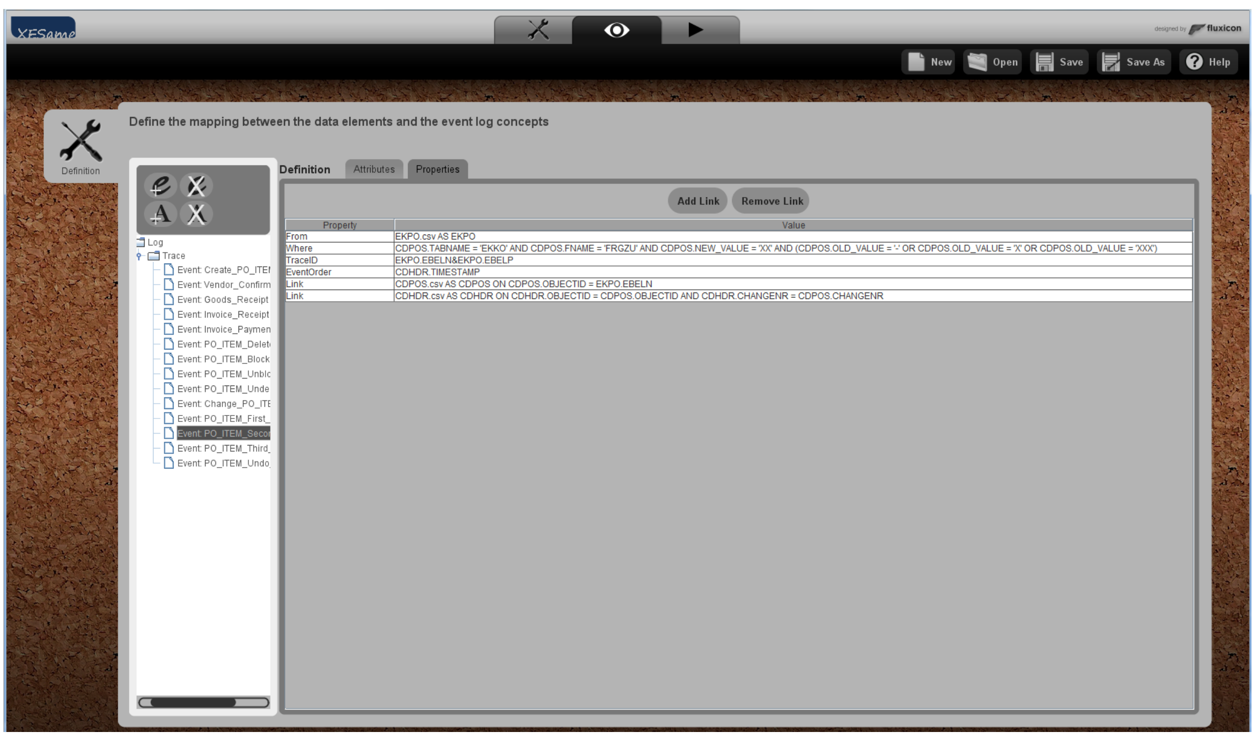


Figure A. 32: PO_ITEM_Second_Approval Properties specification

Figure A. 33: PO_ITEM_Third_Approval Attributes specification

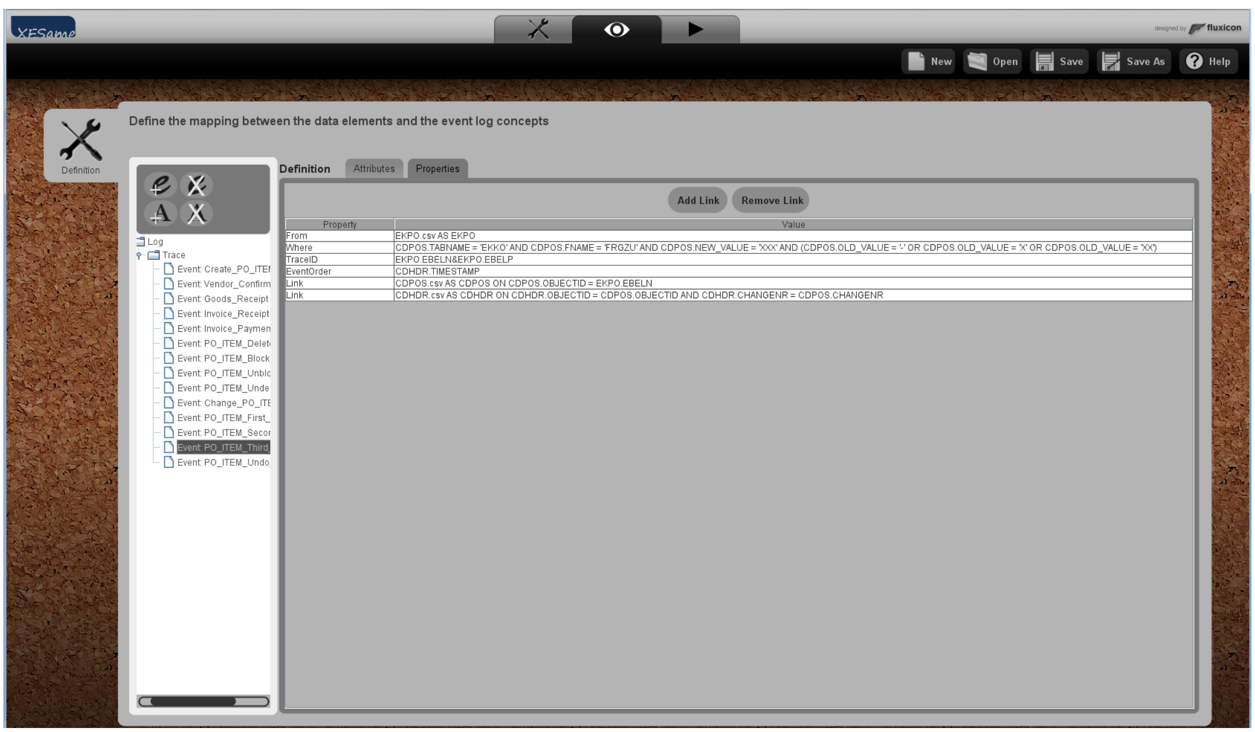


Figure A. 34: PO_ITEM_Third_Approval Properties specification

Figure A. 35: PO_ITEM_Undo_Approval Attributes specification

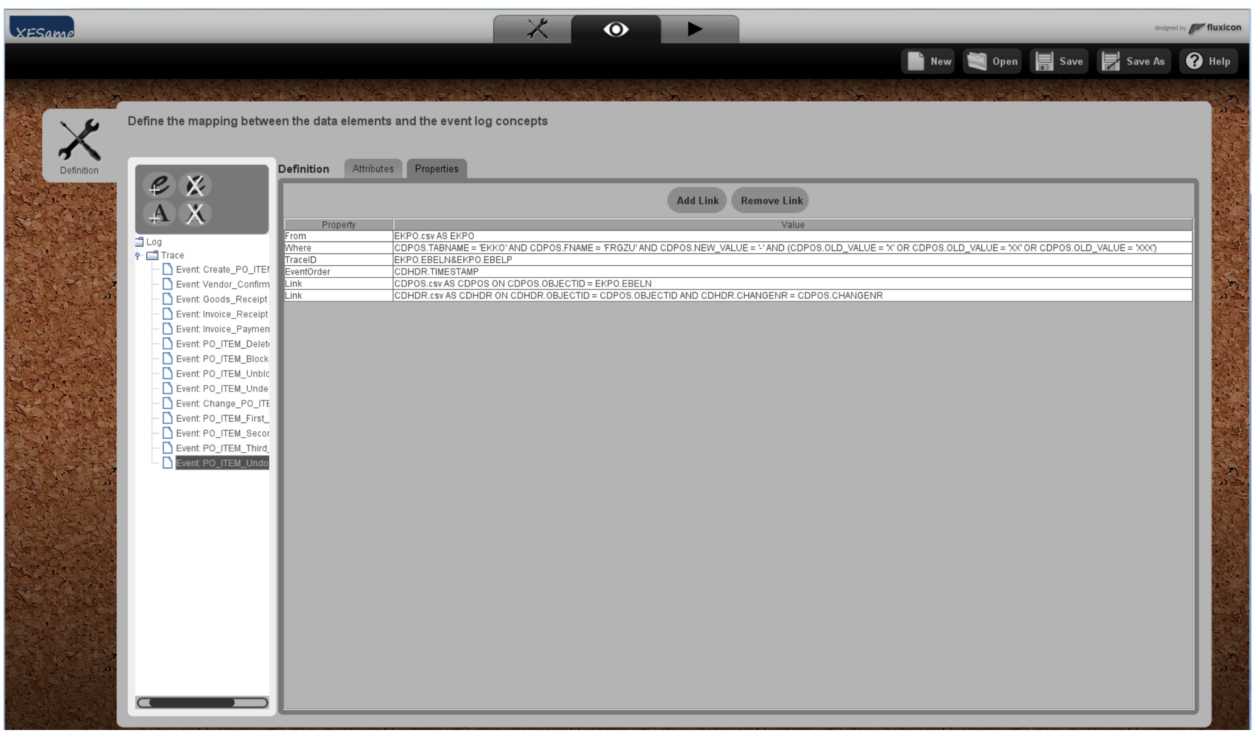


Figure A. 36: PO_ITEM_Undo_Approval Properties specification

Finally the mapping can be executed via the execute conversion control and the MXML/XES event log can be found in the specified output event log location shown in Figure A. 37.

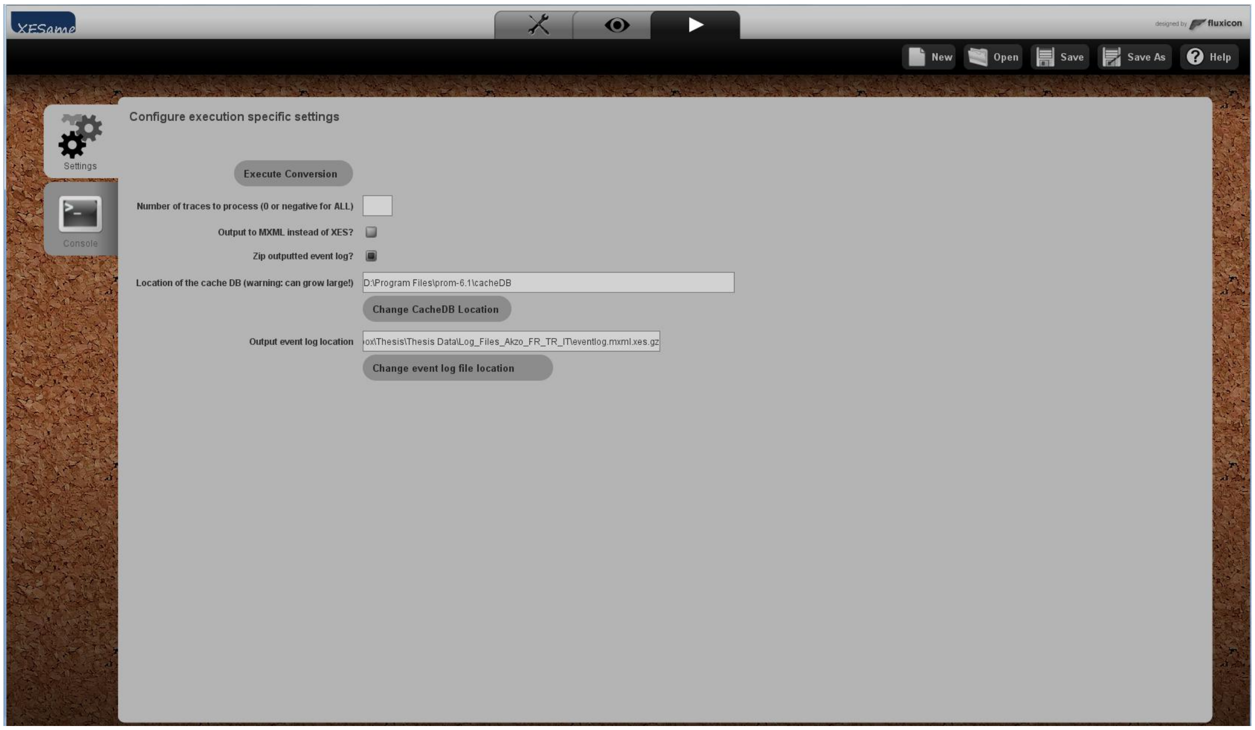
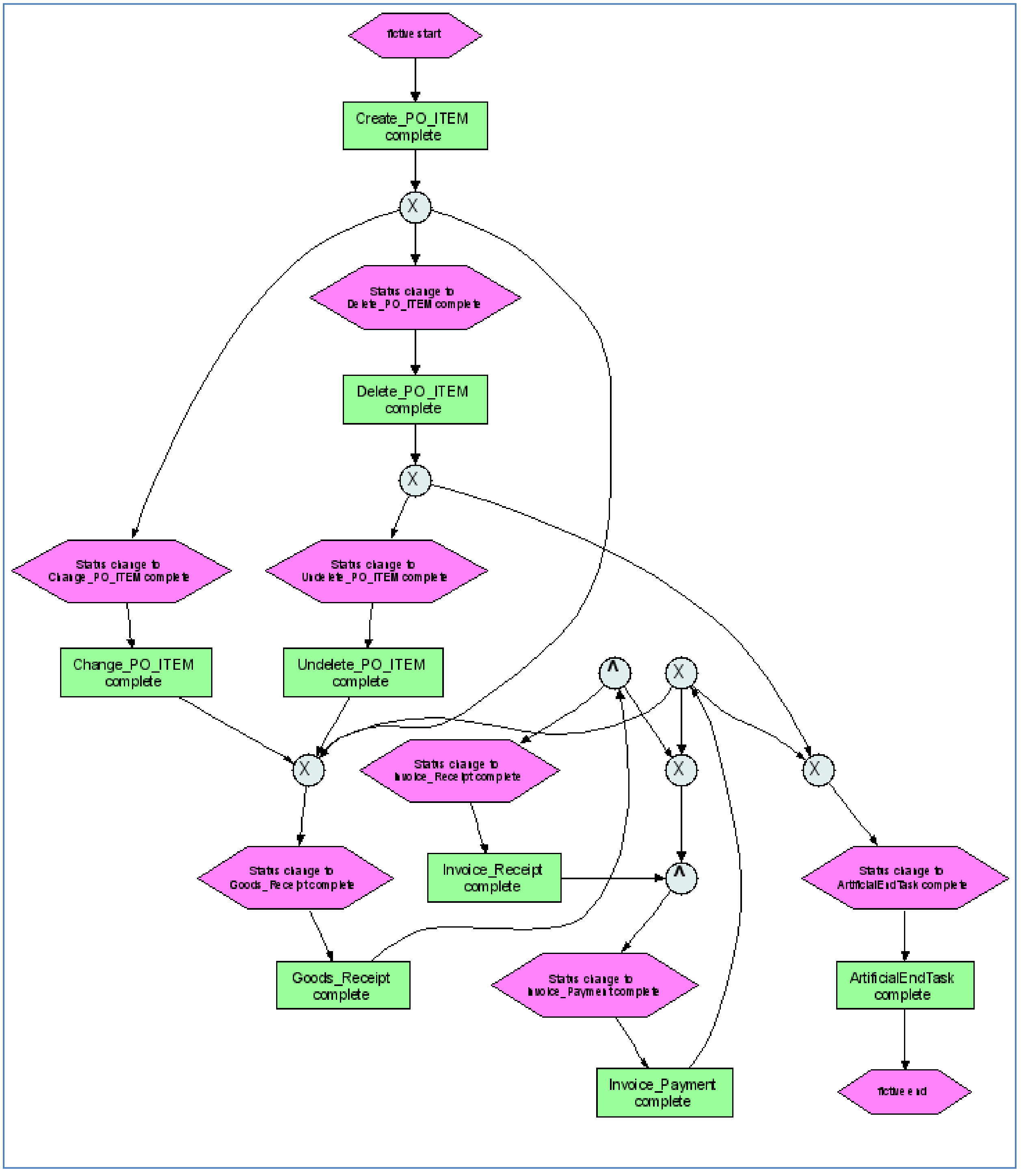


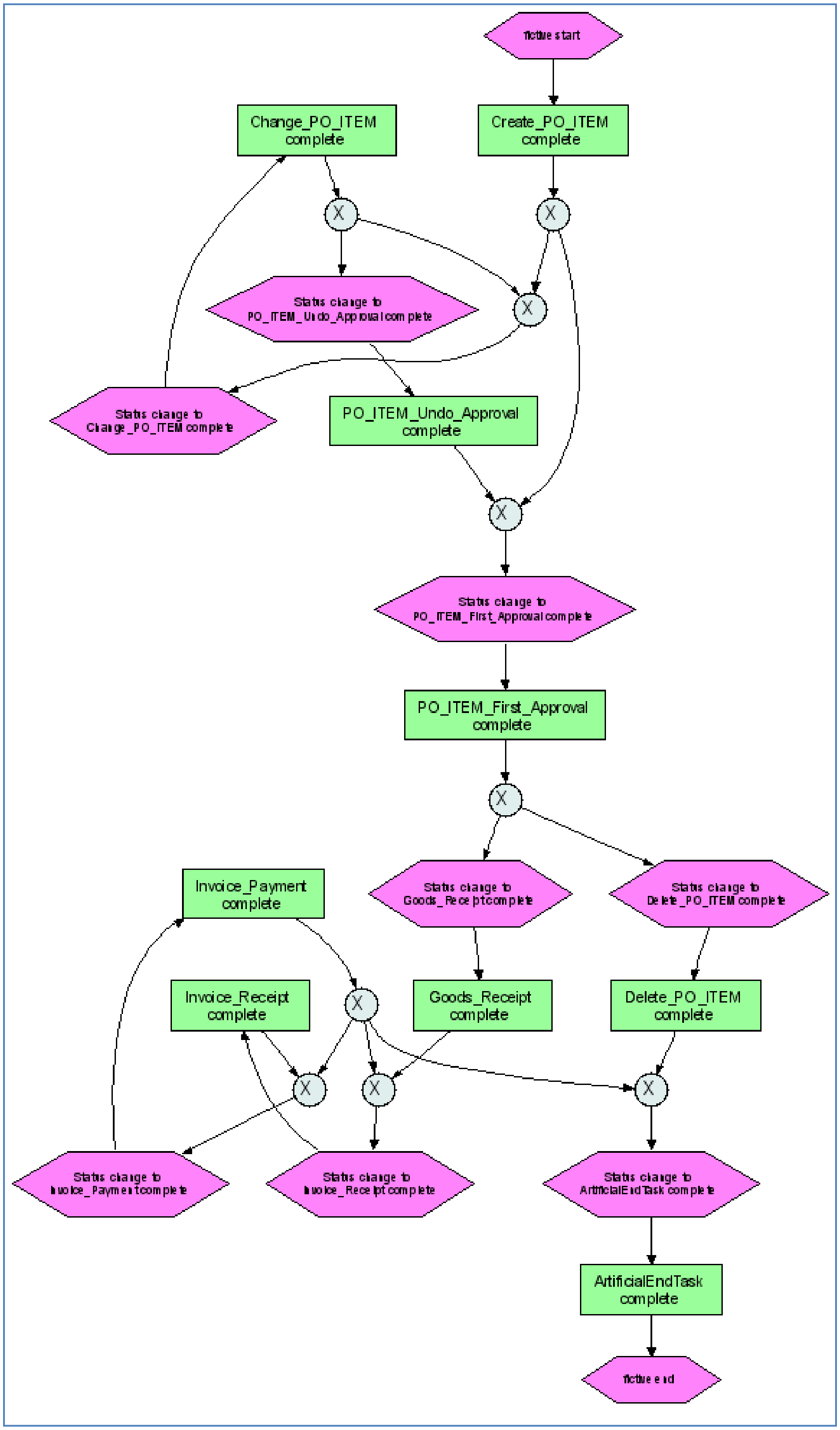
Figure A. 37: XESame execution window

Appendix 4: Mined models

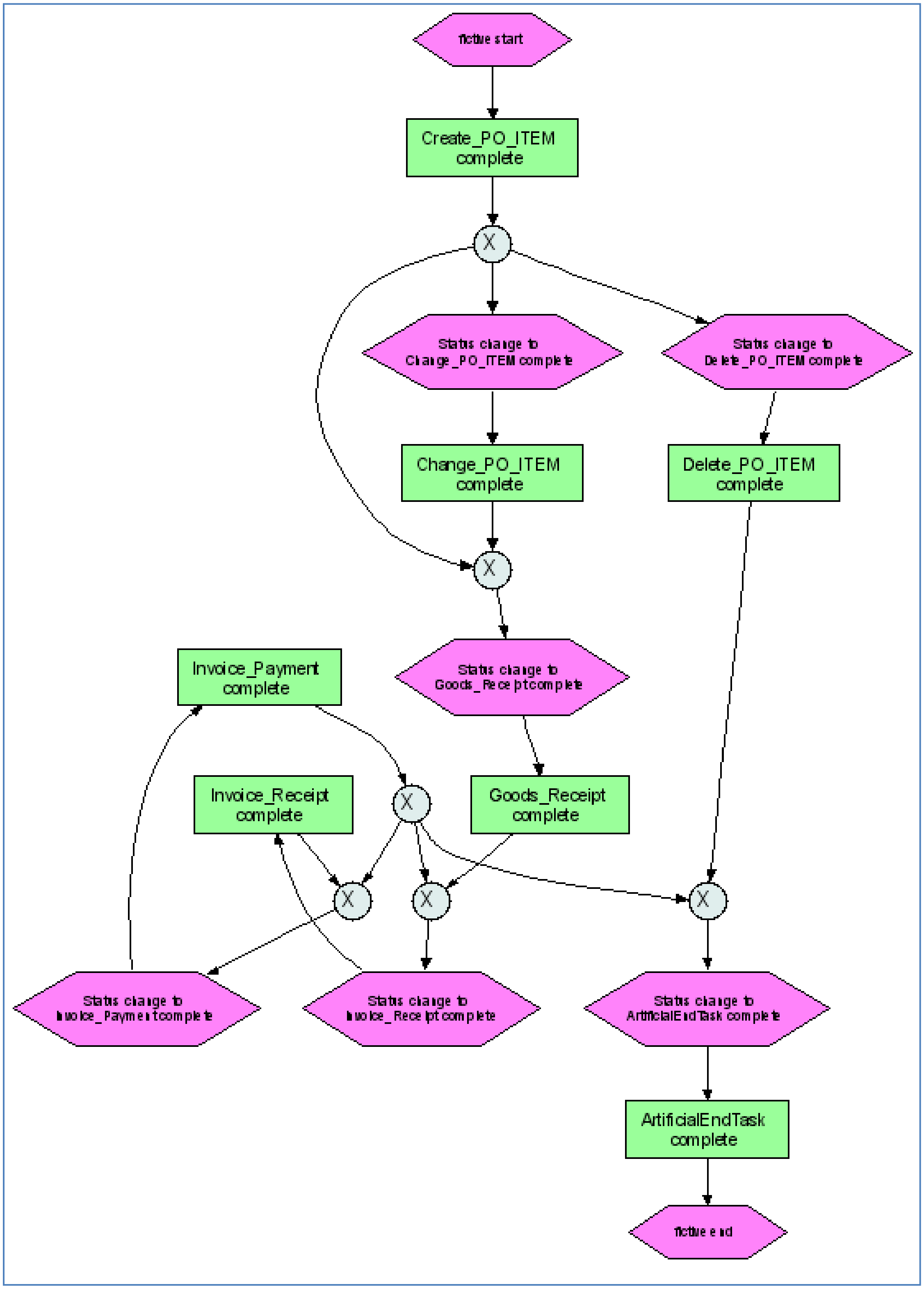
EPC model France



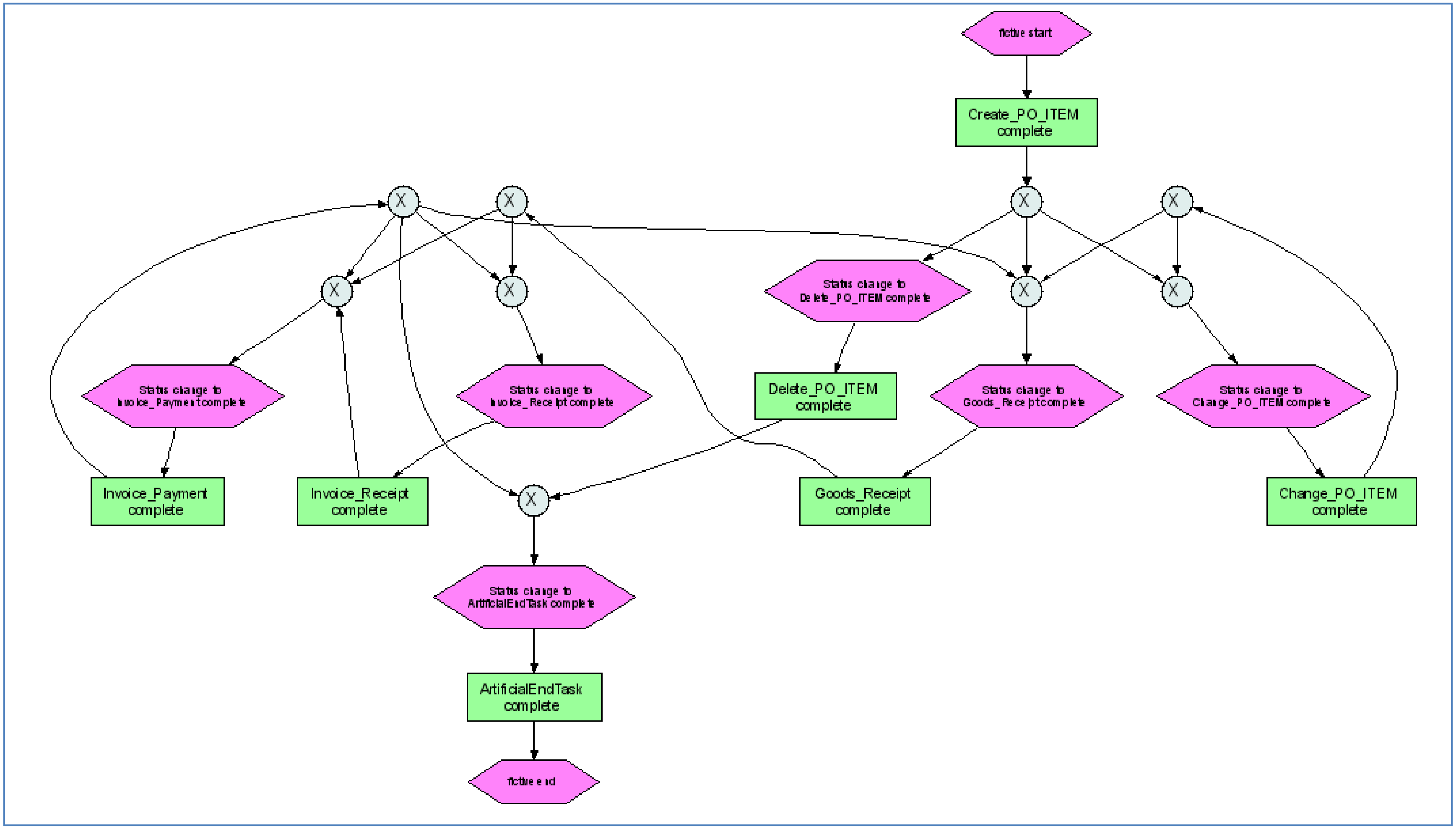
EPC model Ireland



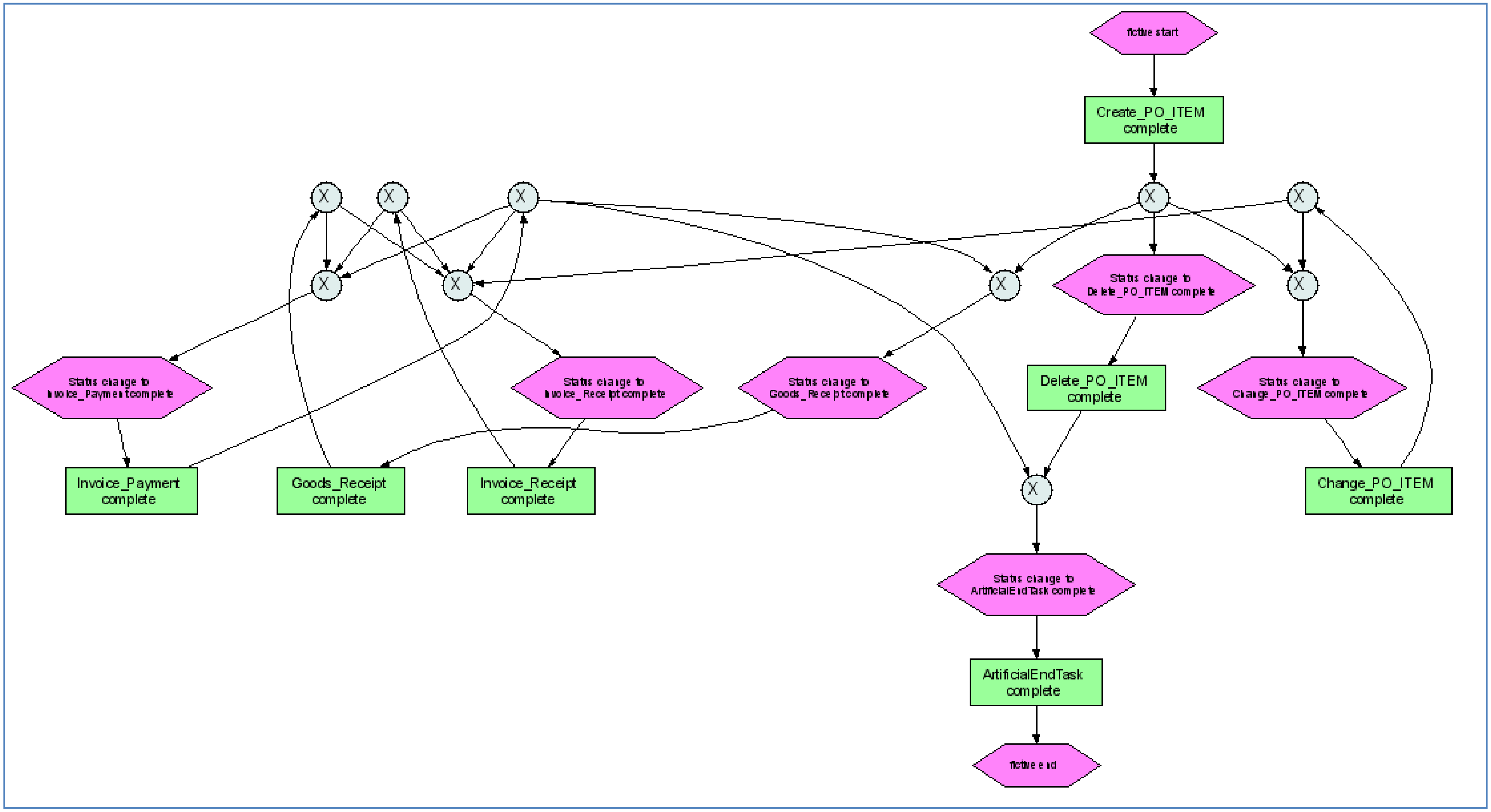
EPC model Sweden



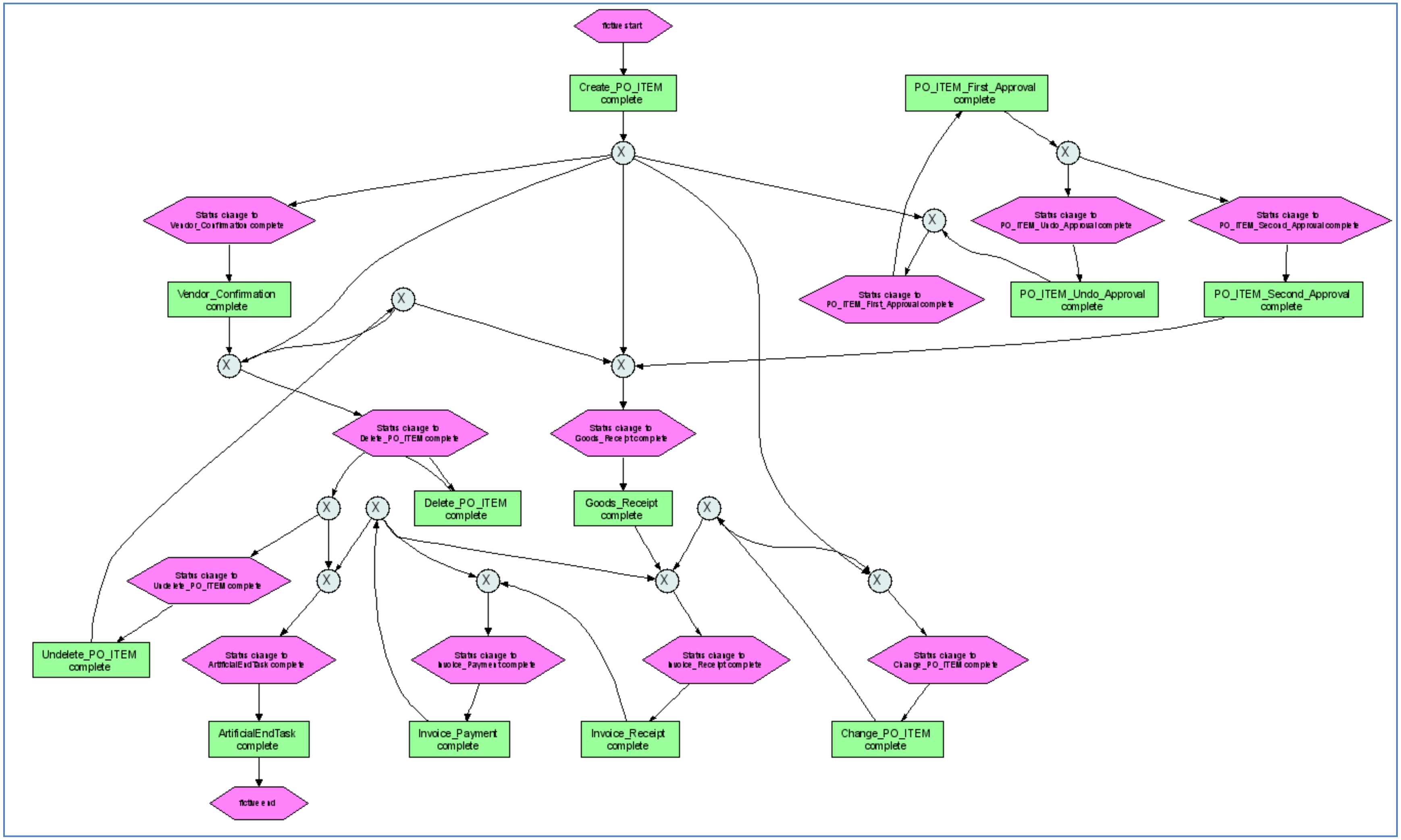
EPC model Italy



EPC model Turkey

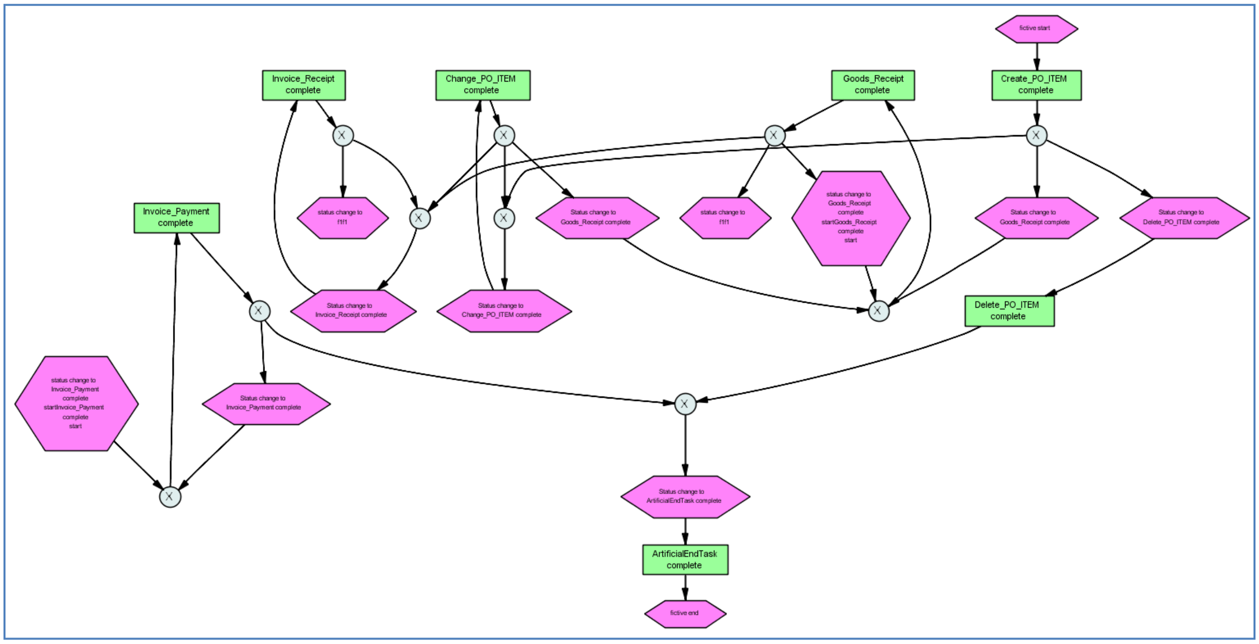


EPC model United-Kingdom

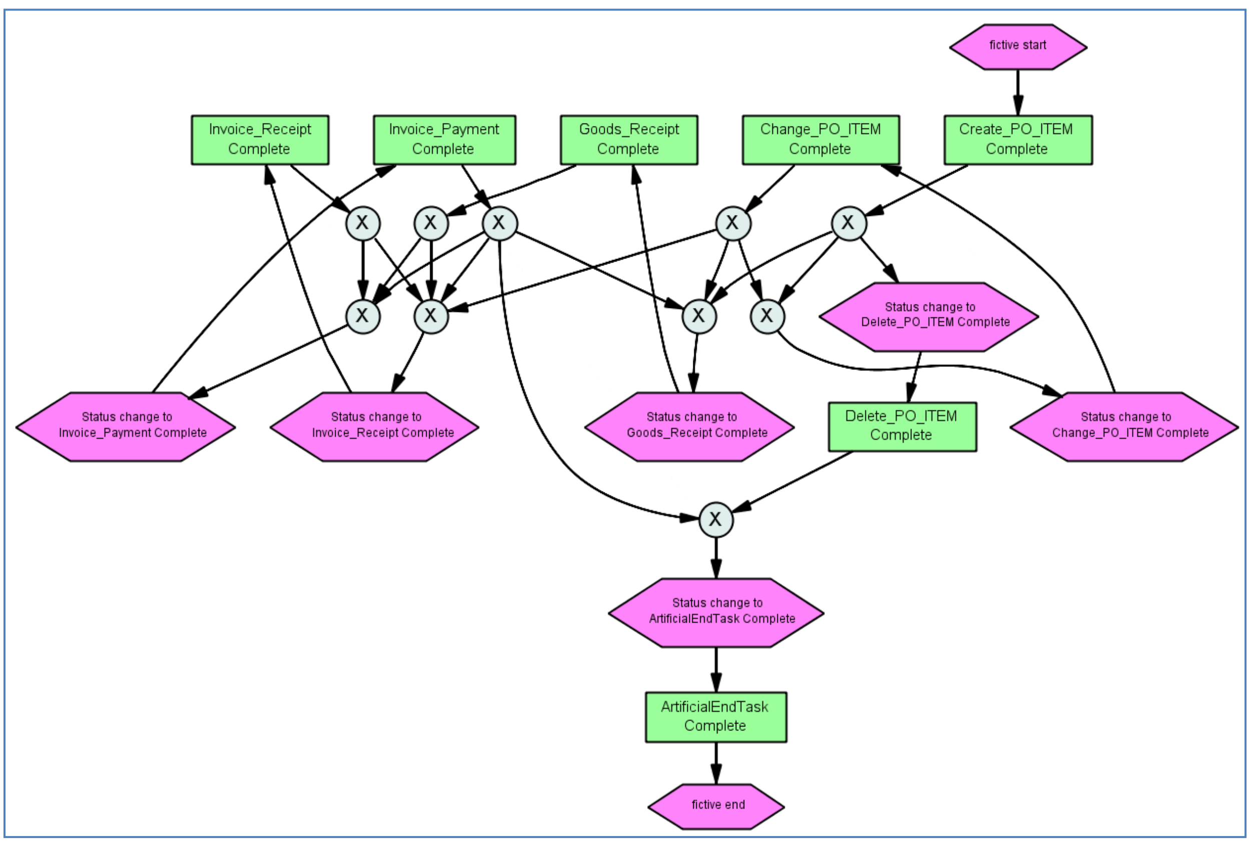


Appendix 5: Merged Models

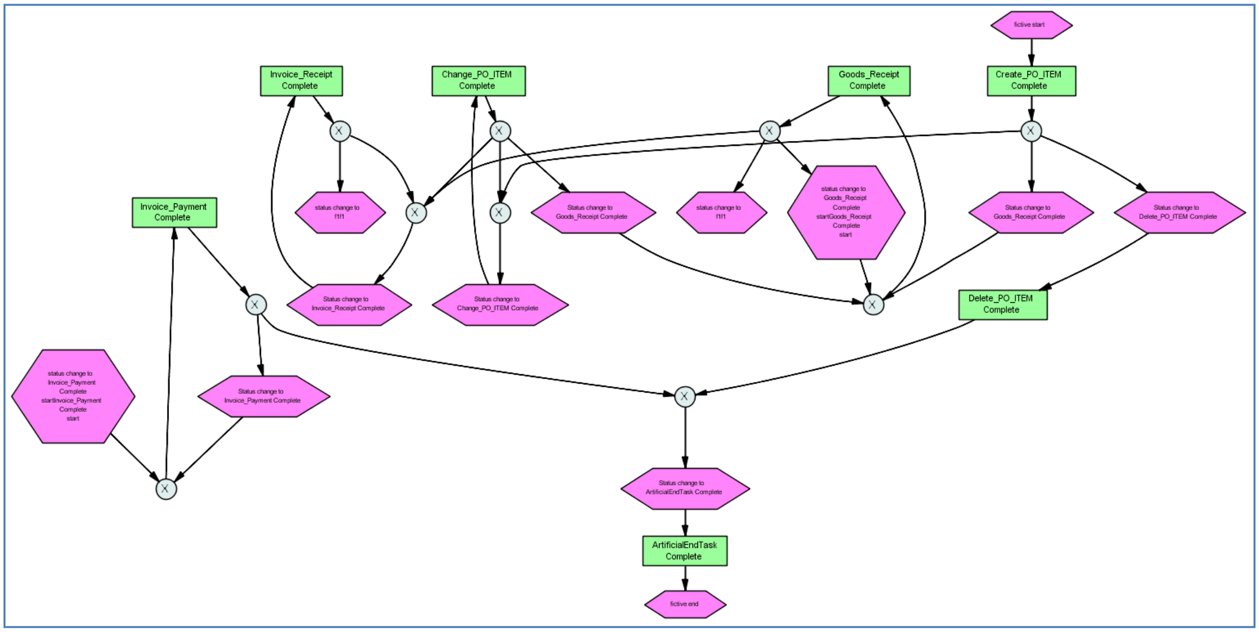
Merged model 1: Merge between Italy and Turkey "Initial model"



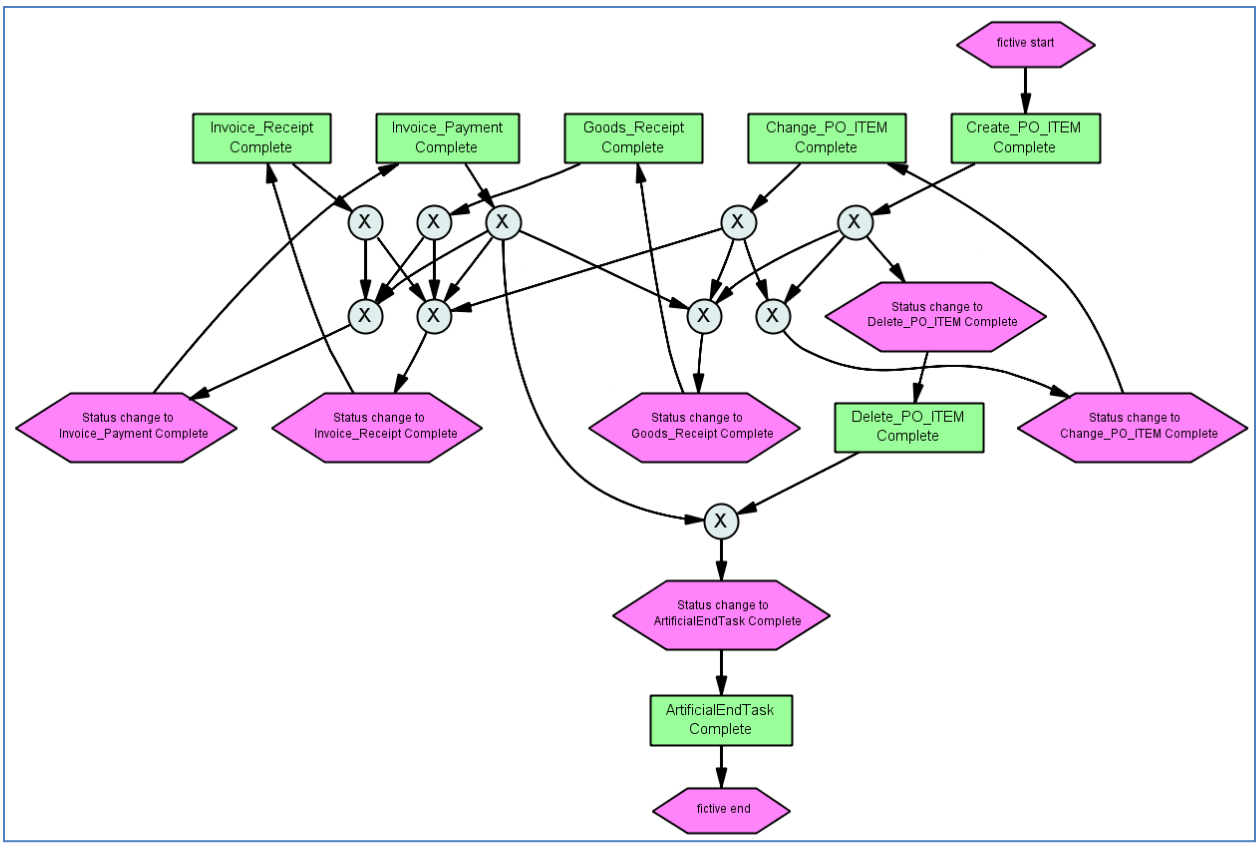
Merged model 1: Merge between Italy and Turkey "Reworked model"



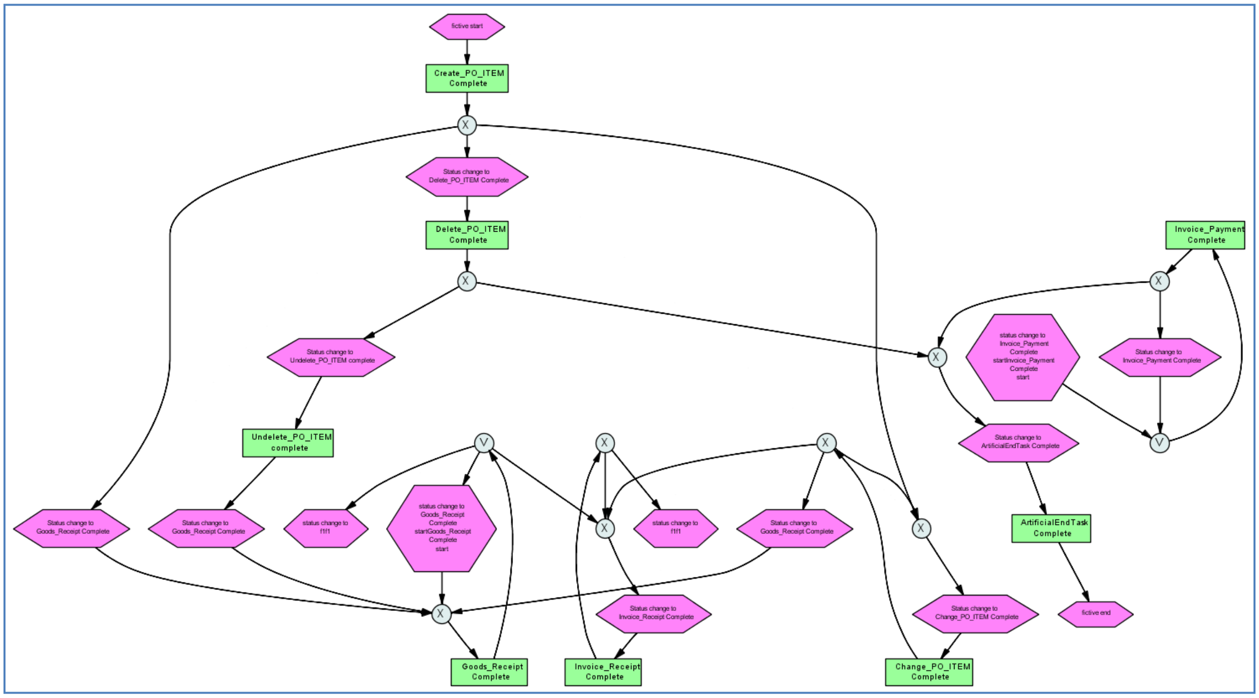
Merged model 2: Merge between Italy, Turkey and Sweden "Initial model"



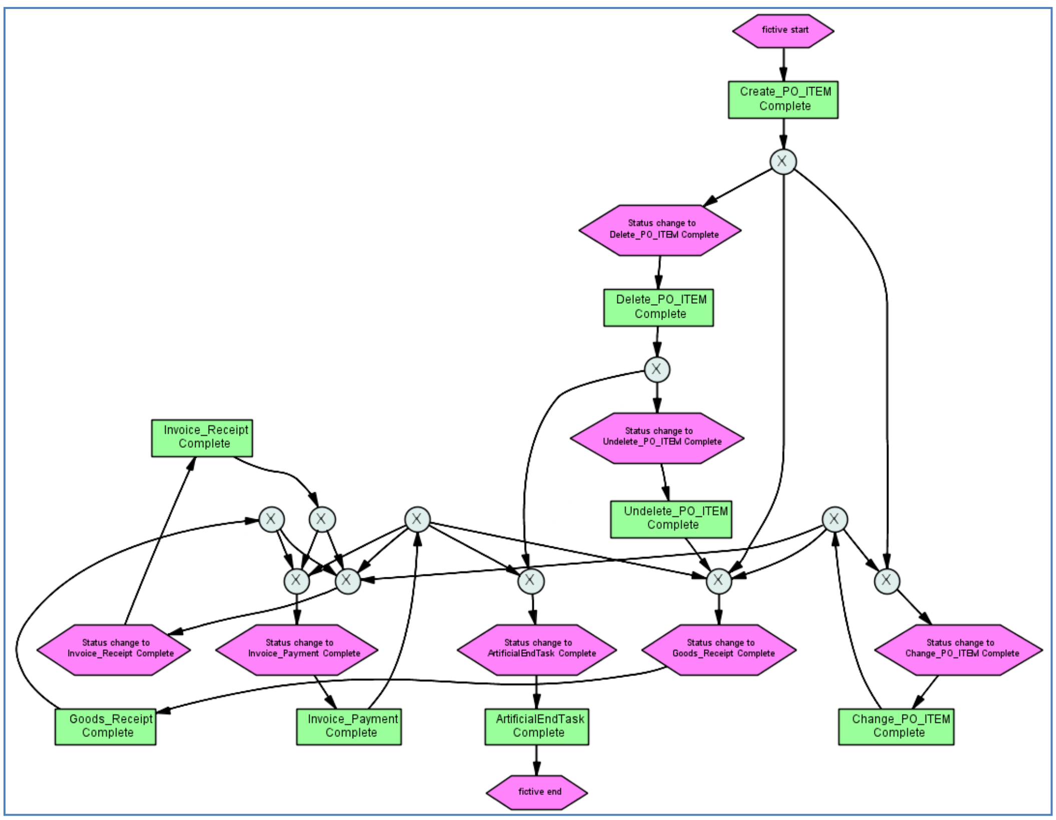
Merged model 2: Merge between Italy, Turkey and Sweden "Reworked model"



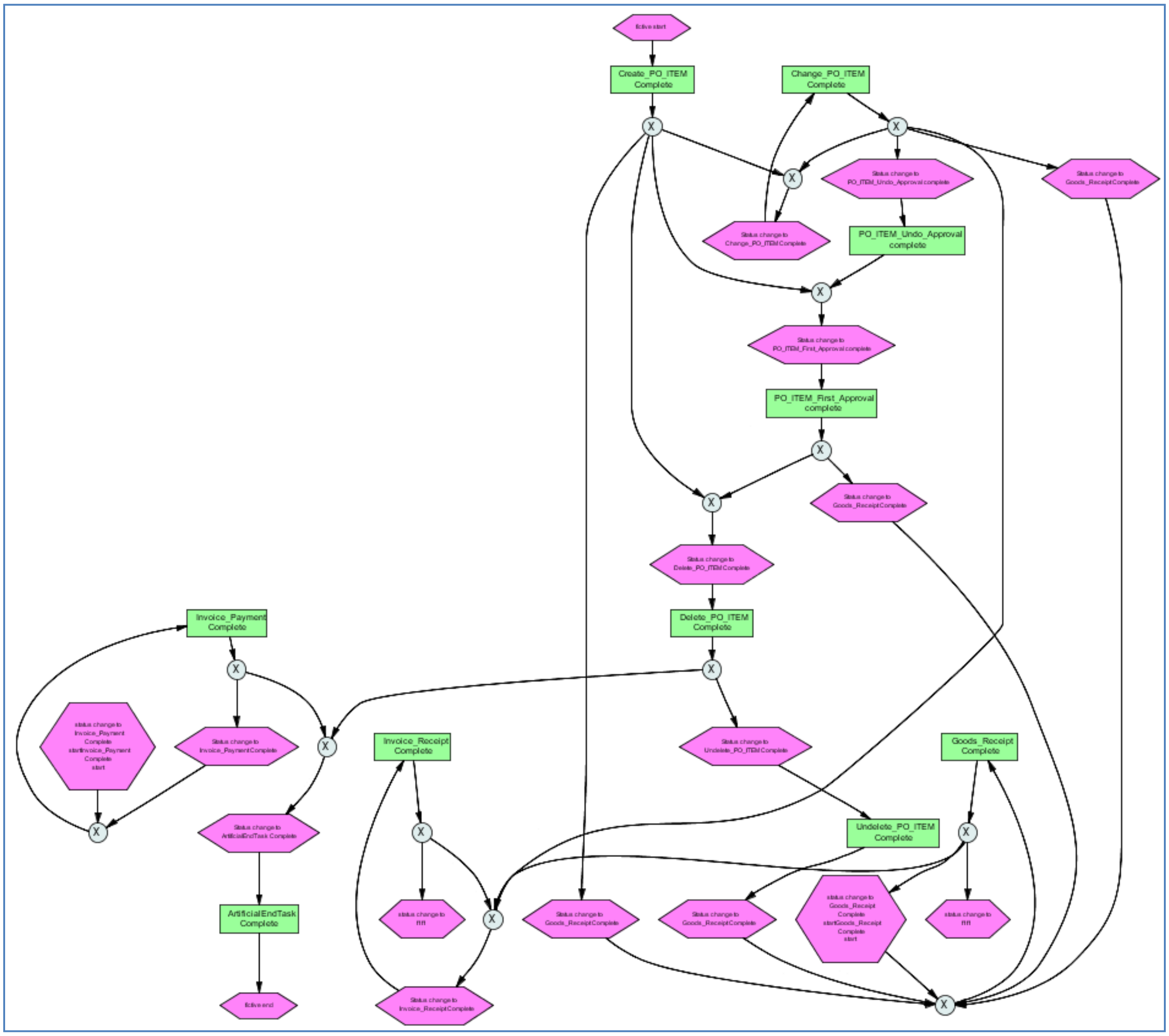
Merged model 3: Merge between Italy, Turkey, Sweden and France "Initial model"



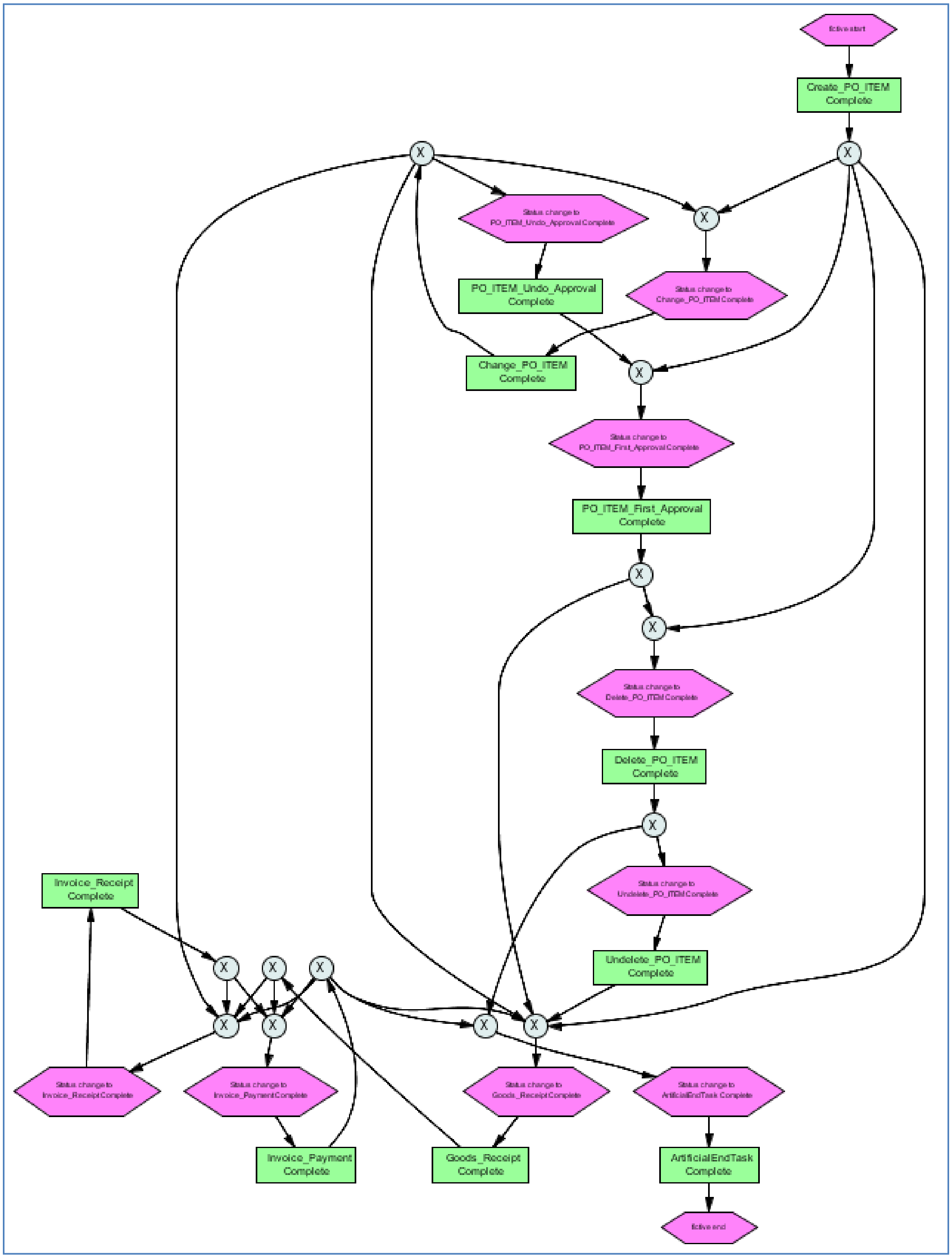
Merged model 3: Merge between Italy, Turkey, Sweden and France "Reworked model"



Merged model 4: Merge between Italy, Turkey, Sweden, France and Ireland “Initial model”



Merged model 4: Merge between Italy, Turkey, Sweden, France and Ireland "Reworked model"



Merged model 5: Merge between Italy, Turkey, Sweden, France, Ireland and United-Kingdom “Initial model”