ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE

TESI DI LAUREA

in

Cambiamento Organizzativo
e Progettazione dei Processi Aziendali

BUSINESS PROCESS MANAGEMENT AND PROCESS MINING WITHIN A REAL BUSINESS ENVIRONMENT: AN EMPIRICAL ANALYSIS OF EVENT LOGS DATA IN A CONSULTING PROJECT

CANDIDATO:
Philippe Gonella

RELATORE:
Chiar.mo Prof. Alessandro Grandi

CORRELATORE:
Ing. Stefano Aiello

Anno Accademico 2015/16

Sessione II

Contents

List of Figures

List of Tables

Introduction

Everything around us is characterized by *processes*. The reality, as we know it, is build up by pre-constructed set or activities or tasks, interacting to achieve a result. These, are grouped together into processes. To take the car to go to the supermarket, to schedule free time in order to study for an exam, as well as the construction of a chair, are instances of processes. We used to think about our objectives or expectations in terms of processes, and companies do the same. With the continuous evolving of markets, technologies and customers' needs, corporate are been facing issues that are more and more challenging. To do business in a context characterized by dynamism, multitude of drivers, financial exposure, pressing on outcomes, new threats and opportunities, it makes mandatory for organizations to pursue continuous changes and evolution, as well as to improve their internal efficiency. These recommendations are more than actual in these years, characterized by the run through the so-called *industry 4.0*.

Not only companies, but human beings in general have always felt the necessity to fell confident with everyday processes, in order to reach their desired achievements. From firsts rural organizations, to the Ford's model factories, and finally to the recent management theories that define processes as the central element of the corporate environment. Over the last decades, many fields have been evolving in order to support executives and decision makers enhancing their firms' performances, and lots of these insights and updating have been possible thank to the role of *information and communication technology*.

Every company today rests on technology for executing a huge diversity of tasks, no matter the specific industry and the business context considered. As a consequence, technologic devices produce data that, if exploited properly, will be an outstanding source of value for businesses. Even though business intelligence and data-driven relate disciplines, introduced in last decade, has released opportunities previously unimaginable, they still seem to be not adequate within the very complexity of real business processes. Hence the birth of a new technique, process-driven and oriented, able to manage today's variety of digital information present in the form of event logs, which are stored in the organizations' computer devices: *process mining*.

The presented dissertation is aimed to explore this new subfield of computer science discipline, and to verify if it may provide a tangible and practical benefit in a real business scenario. Hence, the description of a *consulting pilot project* in which I participated, involving a big application infrastructure provider, which has been conducted with the support of process mining technique and software. The reason for carrying out this case study in a real business context, is to strive to put in evidence if this methodology could be considered as the definitive tool for achieving the desired outcomes, within a consulting project or a change plan.
Moreover, a *database* I have created, containing the most complete list of the adoptions of process mining techniques in real corporate scenarios, is provided in the final appendix of the research. Since the list represents the most complete census of the applications of the subject existing worldwide, it has been presented at the IEEE Task Force on Process Mining annual meeting in September, during the 14th International BPM 2016 Conference in Rio de Janeiro. The purpose was to create a list of practical cases, no matter the specific industries and the final results, with the only aim of contextualizing the spread of the phenomena in the last decade.
Nevertheless, the endeavour is structured into four chapters, each exploring a relevant aspect of the subject, in order to provide readers with all the notions necessary to appreciate the context as well as the basic technical knowledge for understanding recent process modeling formalisms and the way process mining algorithms actually work.

The first chapter offers a brief overview of the business organization structures over time, considering the way human beings had grouped together in the past, in a way to achieve their common goals. Then, the fundamental influence of *Fordism* and the birth of mass production is taken into account, together with the first management theory which defined process as being the central aspect of an organization: *Business Process Reengineering*. From BPR limitations and issues, the last and more recent theory of *Business Process Management*, whose main characteristics will be discussed in the second chapter, is introduced. Finally, last part of the chapter provides readers with an overview of the main *process modeling formalism* used in industrial and business situations: Petri net, YAWL, EPC and especially BPMN are presented.

The second chapter precisely describes *Business Process Management* and how this theory is able to overcome the limits highlighted for the previous discipline. Firstly, a framework for the subject is presented, in order to collocate it within the fundamental BPM driver of transparency and between the internal attributes of efficiency, compliance, integration and the external ones of quality, flexibility and networking, that characterizes a generic business environment.

Secondly, the *BPM lifecycle* is taken into account, and each of the phases that contribute to the successful of this kind of management theory are explored. These stages are process identification, process discovery, process analysis, process re-design, process implementation, process controlling and monitoring.

Then, third chapter finally guides readers into the phenomena of *process mining*, the innovative science that allows to analyse business processes in a deeper way and that enables next-level manipulation possibilities, permitting to outstanding insights, previously unimaginable, to emerge. The subject is presented through the requirements of event logs for its application, formalized by several scientific publications conducted during the last years over the world, whereas additional guidelines and principles theorized in other studies are citied. Then, the three fundamental techniques of process mining are deeply described: process discovery, process conformance and process enhancement. In particular, during the process discovery paragraph the most important *process mining algorithms,* both from an academic point of view and for their applicability in real business context, are briefly discussed. The objective of this part is not to provide a full mathematical and logical explanation of the formulations, but just to let readers understand the basics of the process mining algorithm theory.

Last chapter presents the description of the *pilot project* I was involved in, working within a management consulting group as analyst. The customer, object of the study, was an application infrastructure provider based in Italy, with more than 100 customers and resources involved in the specific process analysed. In particular, the process taken into account is a ticket handling process, counting more than 75000 data logs within a time scope of nine months, between 2015 and 2016. The chapter contains a description of the analysis conducted within the consulting firm, which have carried out the project with the support of the process mining software Disco, developed by Fluxicon. After a data cleaning phase, process mining application let several *evidences* to emerge that were then presented to the customer' executives involved, in the form of projects deliveries and with a final work package at the end the project.
Evidently, due to not disclosure agreement, company's and its customers' names are not citied within the study, and some data have been modified before the publication of the research. Nevertheless, project's outcomes, as described in the dissertation, and the evidences that process mining application has enabled to emerge, remain the authentic ones of the actual business case.

Chapter 1 The Origin of Business Process Management

1.1 The Functional Organization

Since Prehistoric Age, human beings have felt the need to join with other people in order to carry out their activities and to achieve their common goals.

In the very first forms of social organizations, people used to execute their activities independently and just for themselves, with the aim of meeting all their needs: they were both producers and consumers of the same outputs of their work. In this sense, the human being had the capacity to produce absolutely heterogeneous products, and to create all the tools needed to satisfy his basic necessities. Therefore, it is possible to define these first labour organizations composed by workers who were *generalists* (Dumas *et al.* 2013).

However, in Ancient Times, thank to the progressive technological discoveries and the improvement of production and processing techniques, which had allowed people to use new material resources, organizations moved to an intermediate level of specialization. Individuals began to focus on creating just a single product type: as a consequence, they became *intermediate specialists*. This trend took shape in a more evident way in the Middle Ages, when the various professions of artisans began to spread. These figures knew the whole process of shaping the outcomes of their work and they were the responsible for selling them too: just think, for instance, the professions of shoemaker, blacksmith, etc.

During the Second Industrial Revolution, between the second half of the nineteenth century and the beginning of First World War, this trend of specialization finally concretized and organizations reached a *pure specialization* level. This was unquestionably related to the contribution of Frederick W. Taylor (1856-1915), and his theory of *Scientific Organization of Labour*.

The central element of this doctrine was the searching for the maximum division of labour through a "scientific" selection of the workforce, characterized by the splitting and scanning of work cycles within the production system. Taylor, a mechanical engineer and entrepreneur, spread the idea that from the meticulous measurement of the times of all work steps, it could be possible to develop a model list of instructions, specific to each resource. This from the assumption that only if a worker is committed in just one stage of the production process, he could reach the higher degree of specialization possible in that specific step. According to Taylor, this would lead to the maximum productivity and efficiency of the entire production plant. Then, this new philosophy was applied not only to the industrial and production companies in the strict sense, but was also extended to the government and administrative aspects, as well to many other fields (Kanigel, 2005, Dawson, 2003, and Head, 2005). As a consequence, the scientific division of labour became the most common form of organization.

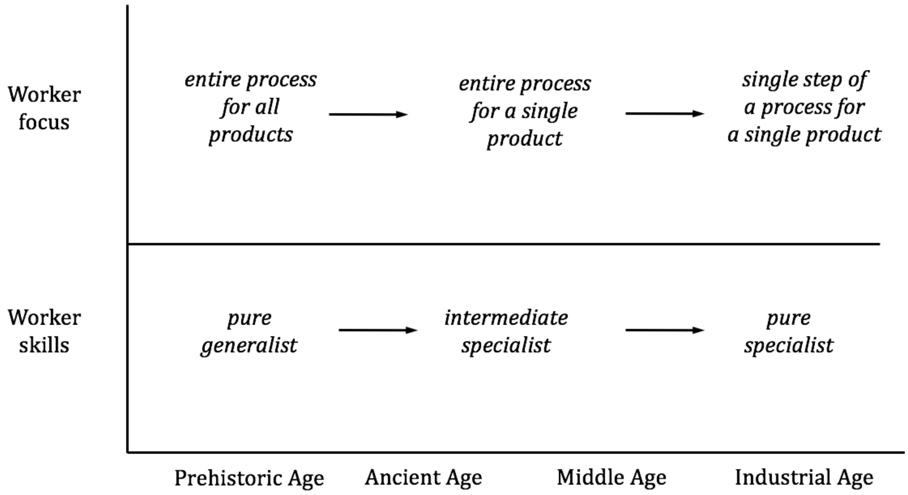


Figure 1.1: The Timeline of Business Organizations
(from Dumas, La Rosa, Mendling, Reijers, 2013)

The maximum application of the principles of the rhythm and scientific organization of work came up as a result of the contribute of the automotive entrepreneur Henry Ford (1863-1945), who had introduced the concept of assembly line, which became later the central element of the modern manufacturing industry (Settis, 2016). With the use of a conveyor belt inside the factory, a superior fragmentation of the work was obtained. With the Ford factory in Detroit the concept of *mass production* came to life and spread from the United States to all around the world. This innovative production system allowed firms to produce on a large scale a series of standardized products, at significantly lower costs than the average industrial plant of that times. Beside the philosophy of the scientific organization of work, the role of new professionals, namely managers, emerged. The managers were responsible for the achievement of individual objectives of the resource under their supervision, though they were not specialized on the particular production process they controlled. Moreover, their prior goal was to achieve the highest possible degree of optimization, since this was seen as an essential condition for gaining the desired business results. This is how the principles of labour division and internal fragmentation of roles has come to be used in a number of firms.

Therefore, it was necessary to differentiate among the responsibilities of the various managers. Hence the birth of *functional units*, which consist of resources committed to the implementation of the same output - even intermediate - grouped together.

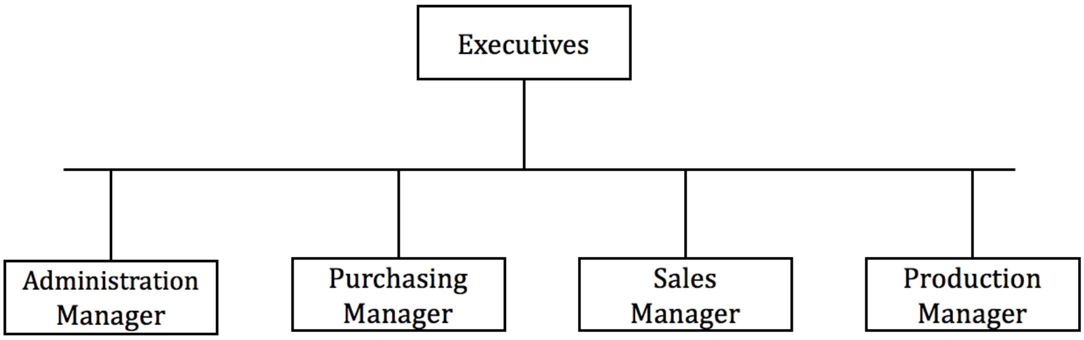


Figure 1.2: Functional Organization Structure

These units were supervised by managers with different responsibilities and hierarchically organized, for instance in clusters within departments, or departments within a business unit, etc. The organizational functions, emerged during the Second Industrial Revolution, have been dominating the business landscape for the majority of the nineteenth and twentieth centuries. Only at the end of the 80's some of the big US companies, such as IBM, Ford and Bell, realized that their excessive emphasis on functional optimization was actually creating inefficiencies in their operations. In that years, American companies were losing their competitiveness, and were suffering the increasing pressure of Japanese competitors (Womack, Jones, Roos, 1990).

1.2 The Process Thinking

Now it is clear that processes are everywhere around us. A process could be defined as a distinctive set of *tasks* or actions that happen over a precise period of *time*, which are *interrelated* to each other with the aim of a *shared objective.*
The beginning of a new philosophy, characterized by a deeper focus on production processes and on the way they take place within the organizations, is committed to be the famous real case - described for the first time by Michael Hammer and then repeatedly taken up and celebrated by the literature - of the acquisition of Mazda by Ford (Hammer, 1990), through a financial operation begun in 1979.

In Ford's accounting department, in North America alone, had more than 500 employees. Specifically, the payment process of suppliers was absolutely complex and expensive.
The triggers of the process were the *purchase order document* and the relative copy for the accounting, prepared by the procurement office and sent to the supplier. Then, the supplier delivered the goods requested to the warehouse of the production plant, where an employee filled out the form with the description of items (*receiving module*) and then sent the documentation to the accounting department. Finally, the supplier sent the *invoice document* to the accounting department. In that situation, the department operated with three documents related to the same batch of goods: the purchase order, the receiving module and the invoice form. In addiction each document was composed of 14 data items (type of product, quantity, price, etc.) and this contributed to make the control and the management of the process even more hard.

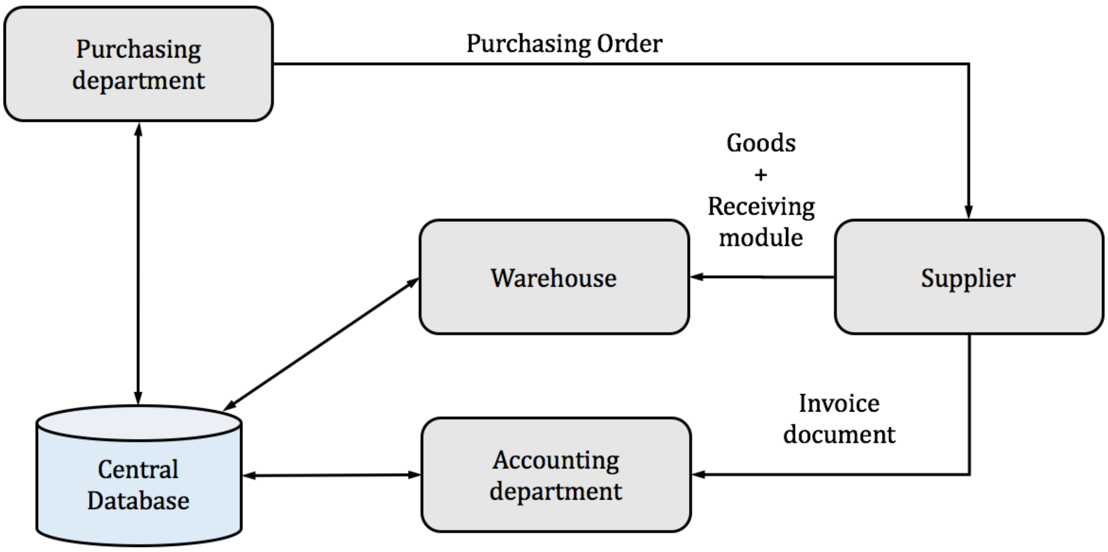


Figure 1.3: The reengineered purchasing process at Ford
(from Dumas, La Rosa, Mendling, Reijers, 2013)

Even so, if any discrepancies between the documents were present, the accounting department could proceed to the payment. However, the process often contained several anomalies: the employees spent most of the time to solve those few cases in which documents showed differences, rather than focused on their routine activities. This way, many days and even weeks were needed to reconstruct the orders and to find out the source of the discrepancy.
After the acquisition of 25% of Mazda's shares, Ford's management questioned about how the Japanese company could actually handle the same billing process with a staff of just 5

resources. This gap was too large to be justifiable just by the different volume of the two companies. Basically, Ford was taking into accounting problems (precisely ‘the differences among documents) as they occurred, while Mazda was making sure that these problems would not occur at all, just acting upstream in order to eliminate risks from their roots (Hammer, 1990).
After an analysis and a comparison between the two payment processes, Ford management realized that the way in which the process had always been done internally has to be changed. First, it was developed a central database where all the information regarding the purchasing process were stored, and finally it was interrupted the exclusive use of paper forms for data collection. In addition, some terminals were installed in the warehouse in such a way that, once arrived the goods, the resource could immediately verify the correctness of the order in terms of the initial agreement, by checking all the 14 features expected by the document. On the one hand, in case of mismatching, the order was rejected and came back to the vendor, which was committed to pay for the additional assurance of the correctness of the goods, as determined in the purchase order. On the other hand, in case of conformity, the goods were accepted and the order was recorded by means of the terminal. Therefore, the only form required to manage the billing process became the original order of purchase.
The redesign of the process allowed Ford to cut 120 employees involved in the accounting, from an initial number of 500 resources (a reduction of the 76%) (Hammer, 1990).
According to Hammer, it was precisely this episode, because of its characteristics, methods applied and outcomes, the first aware case of *process re-engineering*.

1.3 Business Process Reengineering

Business Process Reengineering (BPR) refers to a "a fundamental rethinking and a radical redesign of processes, in order to achieve dramatic performance improvements" (Hammer, Champy, 1993). As this definition from Hammer and Champy suggests, the re-engineering refers to an action of deep "rupture" from the organization's past, conducted through a careful analysis about why the single activities should be conducted as always, and with the precise aim of exploring the root of the phenomena taken into account (Grandi, 2015).
Using a metaphor, one can say that in the Business Process Reengineering companies sit down at the table in order to design the new organization, starting from a *blank sheet of paper* (Dence,

1995), in such a way that they would not be influenced by the constraints of the current situation. In this way firms could rethink and redesign existing processes in order to create even more value for their final customer. The surgery is almost never localized at a single business unit or even at a single business function: re-engineering is often a cross-organization action, which runs through different functions. Such a changes become possible if they are carefully planned and executed with a deep commitment of the top management. The main idea is that, by increasing the emphasis on the customer, companies would be able to streamline the value activities, and to offer new services - or to improve existing ones. Therefore, this would lead to a double benefit of improving internal efficiency and gaining more competitive advantage.

Besides the focus on processes, one of the pillars of the BPR is the link between innovation and the introduction - or the redesign - of the information system, seen as a fundamental enabler for changes and a driver for business decisions (Grandi, 2015). However, the reengineering should not be seen as mere automation: the use of IT technologies, although a preliminary condition for the success of the project, should not be considered as the only instrument.
Firstly, a generic BPR intervention is characterized by the redesign of the classic functional unit in cross-functional teams. Secondly, the role of technology become essential in order to improve the dissemination of data and information, the activities organization, the decision-making structure and the overall management of the process.
Other basic steps common to this type of project also are (Rigby, 2015):

- Rethinking of the value delivered, shaping it according to the real customer needs;
- Redesign the core processes, often through the contribution of ICT;
- Reorganization of the organization, defining the responsibility for processes (namely *process ownership*);
- Rethinking of the organizational aspects and the management of resources;
- Improving business processes through the creation and monitoring of specific performance indicators (*KPIs, key performance indicators*).

Companies that execute a successful BPR intervention for the purpose of enhancing the performance of their key processes, thus can achieve considerable benefits (Rigby, 2015):

- Reducing costs and cycle times by eliminating unproductive activities;

- Reorganization of the resources into teams and reducing the need for a layered management within the organizational structure;
- Quick circulation of information flows, preventing errors and rework;
- Improving quality by reducing the fragmentation of work and establishing a clear governance process. The workers may benefit in terms of increased responsibility and capability to measure their performances.

As defined by Hammer and Champy (Hammer, Champy, 1993), Business Process Reengineering involves a profound rethinking of the organization and a radical redesign of processes and activities. However, it was especially this tendency toward a radical redesign, aimed at a drastic change, that became often the cause of the failures of BPR plans (Hussein *et al.*, 2013). Indeed:

- When the change is far-reaching, it should not be conducted omitting the micro-organizational structure and the actual decision-making and coordination mechanisms;
- The complexity of the program may increase its durability: the expected benefits could arrive too late, when the context had already changed, negating all the efforts;
- The sought improvements are often local and not repeatable: too much focus devoted to just one individual project can cause negative effects above the entire plan;
- Too often organizational change resulted in a blind automation project. As introduced, it is pivot that firms understand the prior role of technology.
- Resources lack of awareness about their contribution and their insufficient involvement (through appropriate change management programs) is very often the cause of mismatching the objectives initially identified.

Besides reasons related to the characteristics of the projects, sometimes the problems took place because of the way they were managed. The causes are:

- Misalignment between the objectives of each project with the company strategy;
- Non-effective use of resources;
- No measurable objectives.

Without carefully designing the KPIs, all the efforts may result in a huge waste of resources. As introduced, key performance indicators are essential because allow companies to take a

snapshot of the initial situation (*as-is*) in such a way that they could properly design the intervention and measure the improvements that lead to the desired situation (*to-be*).
Finally, in terms of governance major errors are:

- Lack of communication and involvement. Without an accurate starting strategy (pilot projects, "quick-win", creation of the sense of urgency), the change can be perceived negatively and generate great resistance among the resources involved;
- Inconsistency with reward systems, incentives, or total lack of talent management programs as a career path for the employees;
- Poor commitment and interest by the top management;
- Lack of change agents, which act as adequate leading to the change plan.

Functional Organization		Business Process Reengineering	
Advantages	Disadvantages	Advantages	Disadvantages
• Economies of scale and reduced overhead costs • Local Efficiency • Developing expertise • Defined responsibilities	• Integration and cooperation difficulties • Local efficiency instead of global improvement • Resistance to innovation	• Reduced costs and inefficiencies • More proximity to customer needs • Better quality of the final output • Better definition of the responsibilities • Improved government processes and performance	• One-shot and non recurring improvements • Difficulties in managing "soft" aspects • Large amounts of capital and maximum commitment of the management required • Probable great resistance to innovation

Table 1.1: Advantages and disadvantages of Functional Organization and BPR

1.4 Process Modeling

With the continuous increasing of the complexity of the business activities and the number of resources involved, companies has faced the need to formalize the processes by means of a graphic representation.

Indeed, a representation allows firms to create a unified view of the activities and their methods of execution, essential in order to involve all the stakeholders. Hence, the necessity of a graphical formalization as understandable as possible by all the levels of the organization, able to communicate and share knowledge about the processes and to define responsibilities in the most transparent way possible. This necessity is even more urgent when problem areas have to be divided in a clear and structured business process, and it becomes essential to carry out any kind of change project (such as the BPR). In addiction, for organizations which do their business in specific sectors, a particular graphic representation of the process with a well defined standard may be expressly required by the regulations.

It must nonetheless pointed out that these needs were partly resolved by companies through ad-hoc modeling, adapted to the specific process to represent. However, this non-replicability of use, due to the lack of a formalized method, and the impossibility of a unique interpretation, have prompted organizations to seek for solutions as objective as possible and easily comprehensible by actors coming from different corporate levels. Anyway, process modeling in general is based on a structured approach, which has multiple objectives:

- The determination of phases and activities that make up a process and the interdependencies existing between them;
- The timely and transparent identification of process responsibility;
- The structuring of the decision-making process so as to increase the organizational efficiency;
- The optimization and rationalization of processes.

These objectives should be achieved by carefully managing the trade-off between the right level of *abstraction* and the correct *representation* of situations even more complex and structured. Before starting to model a process, every tasks with their order of representation, the events that characterize it (which act as triggers for the transition from one activity to another), and the personnel involved (whether they are internal or external to the organizational structure, human, material or information), should be defined.

The most common and easily understandable method for the representation of processes is the *flow chart*. This type of graphical formalism provides an effective representation and generally a low degree of ambiguity, ensuring an adequate balance in the trade-off abstraction-representation, at least for the less complex processes.
A flow chart is constituted by elementary entities, defined as basic blocks, which represent the operations and the control constructs. These are the lowest level of abstraction possible in the modeling. The various blocks are related to each other by means of arrows, which link any action to the next within the diagram.
Therefore, a flow chart is composed of two types of entities: *nodes* and *directed arcs*. The nodes represent the real operations and the various states of starting and ending; while the directed arcs describe the flow of the activities or the information and how they are sequenced. A structure of this type, composed of nodes connected via directed arcs is said *graph*.
Thus, using a graphic formalism, starting and ending events are represented through rectangles with rounded angles, while general executions of the activities are represented as rectangles, linked through oriented lines. Finally, decisions and conditions are described by lozenges.

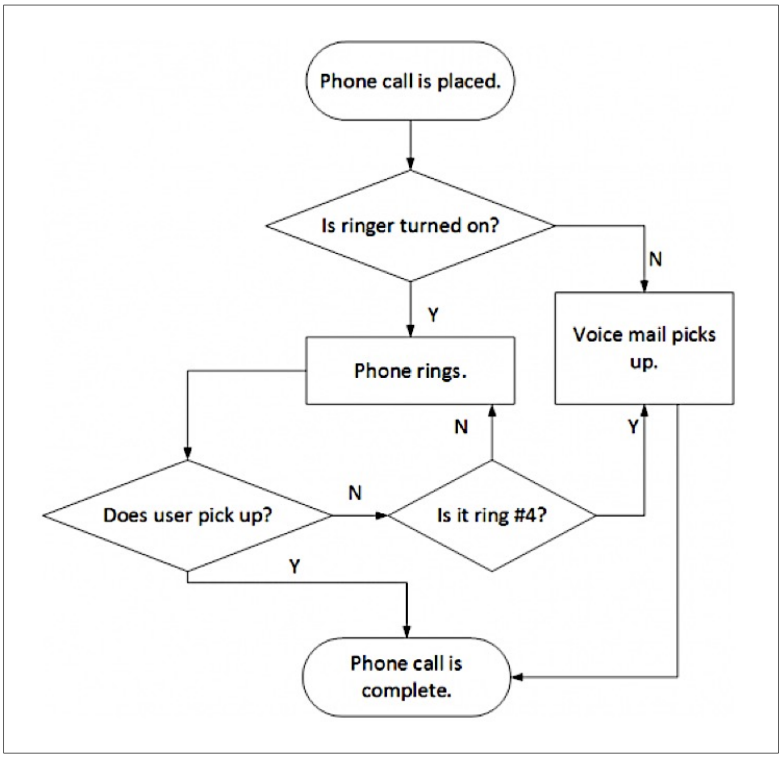


Figure 1.4: Example of flow diagram

Although very clear, the use of this method of modeling certainly involves some non-trivial limits (Carli, Grandi, 2015). First of all, without any graphic formalism for the representation of the employees involved, it may not be easy to identify which are the actors of the processes, their role and responsibilities. Another limitation arises from the difficulty of structuring very complex processes, since there is no possibility of separating the micro and macro levels. In addition, other problems occur in case of loops and uncertain decision-making processes: very often the processes and decisions do not originate a unique and direct consequence, and this can be difficult to visually represent.

1.4.1 Petri Net

A Petri net is a process model notation made of a simple set of objects. In particular, it consists of *places*, *transition* and *direct arcs* which link the objects with each other in a logical way.

- *Transitions:* they indicate a specific business task or an activity that might be performed, and are to graphically represented trough boxes or vertical black bars;
- *Places:* depicted by circles which can contain one or more tokens, represented through black dots;
- *Direct arcs:* the simple connectors of the other objects, described with oriented arrows indicating the direction of execution.

Moreover, the presence of *token* indicates the possibility for the related transition to kick off. In other words, they might be seen as the trigger for the execution of the business activities. Indeed, a transition is enabled exclusively if all the places connected to it with an ingoing direct arc (which are its input) contain a token. Then, when the transition is enabled it could fire, and the task related could be executed. This consumes the tokens contained into the inputs places and creates a token for each of the output places (connected with it by outgoing direct arcs). In this way, it is precisely the firing of the transitions that would change the marking over the net and provides for its progress. Moreover, the distribution of tokens shows the current state of a process, which activities are being performed and which ones might be performed in the future. In addiction, thank to the presence of tokens, Petri nets seems to be suitable for modeling decision constructs, such as XOR/AND gateways (see Figure 1.5).
Since firing is nondeterministic and tokens might be present within a place in several distributed systems, this notation is useful for describing real life situation. In this sense, Petri net seems to be able to guarantee a good balance in the trade off between *abstraction* and *representation* (Murata, 1989).

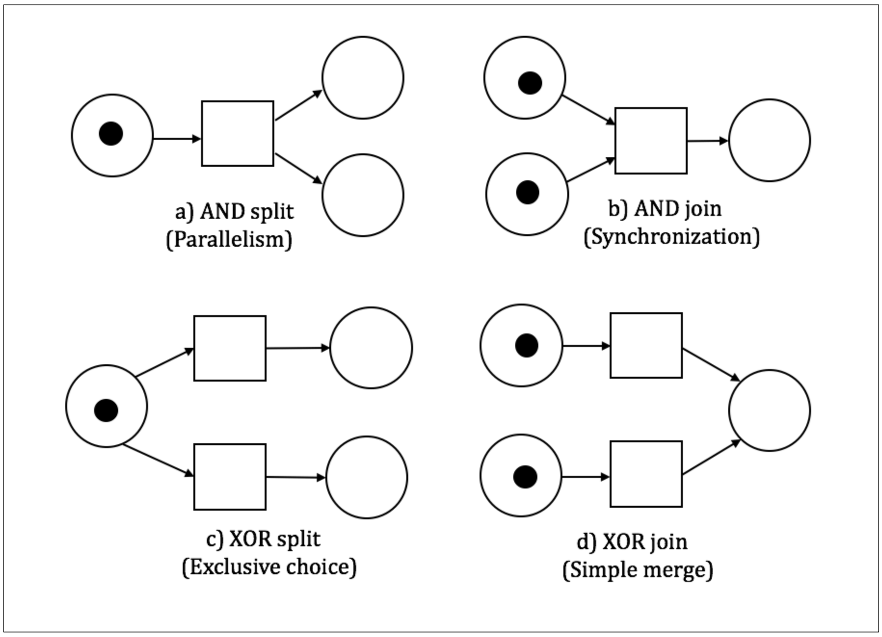


Figure 1.5: Examples of constructs via Petri nets

Unless many others business process notations, Petri net execution semantics can be described with an exactly mathematical theory. For this reason, this formalism is used in various areas, such as computer science, mechanics, economics, logic, etc. (Petri, Reisig, 2008). Nevertheless, it is not the aim of the present research to describe the mathematical theory behind this formalism. Rather, the objective here is just to give reader very basic information in order to be able to understand the diffusion of very different process model notations.

1.4.2 YAWL

YAWL (yet another workflow language), is a workflow management language, developed in an academic environment by Wil van der Aalst and Arthur ter Hofstede in 2002 with the specific aim of supporting the workflow patterns (van der Aalst, Adams, Hofstede, Russel, 2009). This language was designed in order to overcome the limitations of Petri nets in the description of workflow patterns. Indeed, situation characterized by multiple instances, very complex synchronisations and cancellation patterns are essential to express properly business process behaviour (Rozinat, 2010). Therefore, even if Petri nets could be seen as the starting point of this notation, YAWL is characterized by new mechanism and formalisms which enable analysts to perform a more direct and intuitive identification of the workflow pattern (Van Der Aalst, Hofstede, 2005).

Therefore, in YAWL new constructs are introduced, such as OR joints, removal of tokens and multiple instances activities, that make the language easier and more expressive. In particular, the OR is one of the most problematic patterns and very often other notations struggle with its semantic (Rozinat, 2010). Despite other languages, in YAWL the OR split/merge is projected to guarantee the desired synchronization. On the one hand, the OR-split triggers some, but not necessarily all the outgoing flows and it is appropriate in those situation when it is unknown until runtime what concurrent resultant work can lead from the completion of activities. On the other hand, the OR-join ensures that an activity waits until all the incoming flows have either finished, only if there is something necessary to wait. Moreover, the formalism offers several new syntactical elements which intuitively describes other workflow patters. For instance, the notation enables the description of simple choice (graphically depicted via an XOR split), simple merge (indicated as a XOR join). Clearly, the possibility of designing common situation, such as parallelism of activities, are still satisfied with the present notation (via AND split). Furthermore, in YAWL, transitions are assumed to be no atomic: in this sense they do not fire immediately, and it may require some time to do the task. For this reason, one transition here is equal to two transitions in a Petri net, with one place within them.

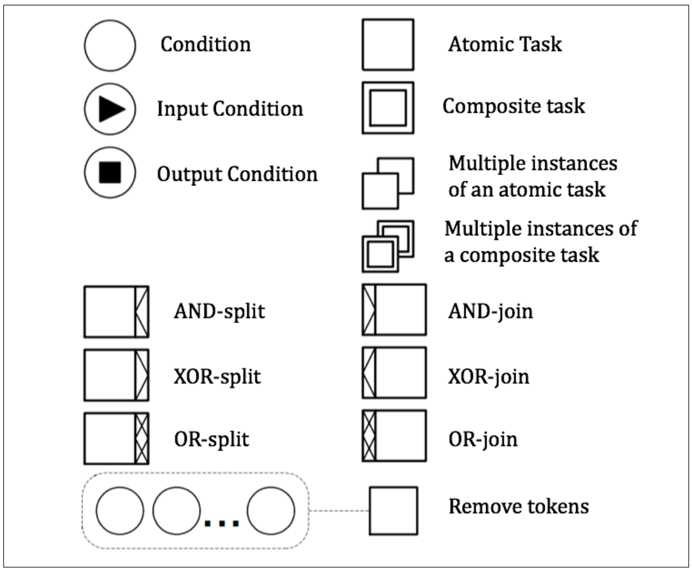


Figure1.6: New symbols used in YAWL

Finally, the language allows to create additional workflow dimension, adding new rules or prescriptions. For instance, with YAWL it is possible to link the process model with the resources involved, by defining which roles are allowed to execute particular activities or to create new routing rules based on other kind of information (Rozinat, 2010).

1.4.3 Event-driven Process Chain (EPC)

The Event-driven Process Chain is a special kind of flowchart, which is used to configure enterprise resources planning implementation and to model, analyse, and redesign other types of business process (Hommes, 2004). It was developed in the Nineties within a framework by AIRS (Architecture of Integrated Information Systems) by August-Wilhelm Scheer. Basically, the EPC could be defined as an ordered graph made of two fundamental objects: events and function. In particular:

- *Functions:* the basic blocks of the EPC language, corresponding to a generic activity that has to be performed. Then, a function may also be refined into another EPC, creating a hierarchical structure;
- *Events*: which are linked with functions, describing the pre and post-conditions of their execution. Indeed, events tell reader under which circumstances a function works or in what state the function actually is.

In this sense, a function may be seen as a place within a Petri net, while events may be considered as transitions. Moreover, the notation is made of several kind of connectors characterized by logical operator (AND, OR and XOR) (Van der Aalst, 1999). Nevertheless, the translation of connectors into a Petri net is more complex, because each single operator in EPC may correspond to a group of arcs or to a network of places and transitions. Finally, the EPC approach may be extended in a way it would become suitable for particular process view, such as data view, organizational view and functional view (Nüttgens, M., Feld, Zimmermann, 1998). For instance, is it possible to put in evidence which organizational unit is involved for a process or the owner responsible of the specific function.

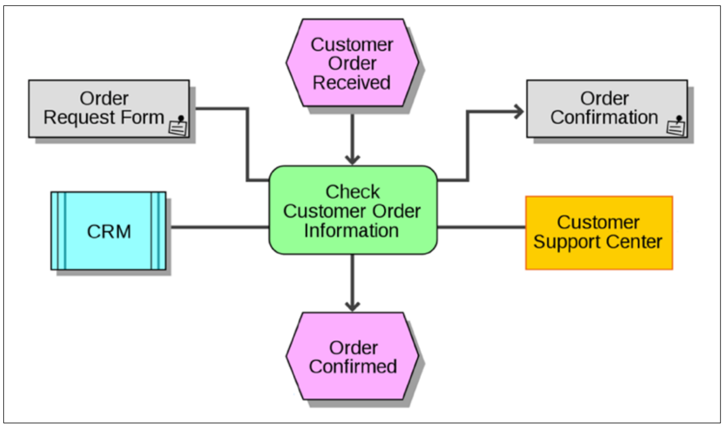


Figure 1.6: An example of EPC

1.4.4 Business Process Modeling Notation (BPMN)

Business Process Modeling (BPM) can be defined as a set of standards and techniques for the design, measurement and control of business processes. Because of its utility within real business situations, this modeling notation will be discussed more in depth in the present research. Therefore, BPM has two important functions:

- Representing the existing processes (activities or other elements);
- Representing new processes for evaluating their performance.

BPMN (*Business Process Modeling Notation*) is a structured method, standardized, coherent and consistent that allows to understand, model, analyse, simulate, execute, and continuously updating business processes (Carli, Grandi, 2015). The notation is the result of a long research on modeling languages by the BPMI working group, made up by a large number of notations developers of business processes (White, Bock, 2011).

The BPMN methodology aims to reduce fragmentation originated during the last years spread of a large breadth of business processes modeling tools. Starting from these divergent notations, the BPMI (Business Process Management Initiative) working group extracted best practices and combined them into a new standard notation. The use of a single standard allows a more rapid adoption of tools for modeling business processes, reducing the existing gap between the analysis and the implementation phase of the process (Freund, Rücker, 2012). Indeed, the two stages are often executed by actors who came from very from different backgrounds, provided with different skills. Moreover, these resources often have a substantial distinct view of the problem from one another. Indeed, the analysis is usually carried out by the management, whose focus is on the strategic aspects of the process. However, these aspects do not necessarily have the same importance at the implementation level, where the designers have to consider other constraints and logical issues. Finally, the end user - often the process owner - requires just a simple and clear tool to administer and control the process. So, this type of standard puts emphasis on all the resources involved, highlighting the responsibilities and the particular contribution on the overall performance of the process.

Before approaching to the BPMN, it is necessary to define a *Business Process Diagram* (BPD), a graphical representation describing the order to carry out the actions. The objects that made up the diagram allow to develop a simple flow chart which have the advantage of being familiar and easily understandable to most of the users.

One of the main goals of BPMN formalism is the search for a mechanism as intuitive as possible, which enables resources to create a representation of business processes ensuring the necessary expressiveness and carefully managing the complexity existing. Therefore, also in this situation a trade-off between the level of abstraction and representativeness is present. To ensure a balance between the two requirements, the notation offers a variety of graphical aspects, organized into specific *categories*. The membership of a symbol to a category allows the reader to easily recognize the behaviour and the specificity of the diagram, leading to a more intuitive understanding of the overall BPD.
Within each category some variations are existing, in order to increase the expressiveness of the notation.

Flow Object

A BPMN diagram is based on three elementary objects, which identify three different categories of *flow object* (White, Bock, 2011). As highlighted in previous paragraph, the purpose of the categories is to confer to the reader an easier way for understanding the meaning of slightly different forms, which clearly belong to a same family. The three flow object are:

- *Event*: a generic event, which occurs during a process. It is graphically represented through a circular element. There are three different types of event which vary in function of the moment they are used within the BPD: *Start*, *Intermediate*, and *End*. Besides according to the reference instant, these objects can vary for their content (a message, error, etc.).

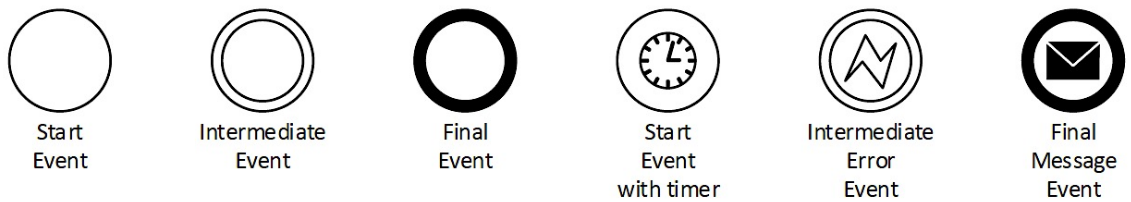


Figure 1.7: Examples of *event*

- *Activity*: a generic action carried out within a process. It is graphically represented by a rectangle with rounded corners. There are two types of activity: atomic (*Task*) or non-atomic (*Subprocess*). In the latter case, another process is present inside the activity, made up of other activities and events. In this case a '+' symbol is drawn inside the rectangle, in order to indicate the presence of a subprocess within the activity. Thank to this property is it possible to generate multi-level diagrams.

Figure 1.8: Examples of *activity*

- *Gateway*: a decision or a logical operation, which allows the process to take different developments. It is graphically represented by a rhombus. In a gateway a passage condition, or a logical symbol may be present. In addiction, both a division gateway (split) and a reunification gateway (merge) are required.

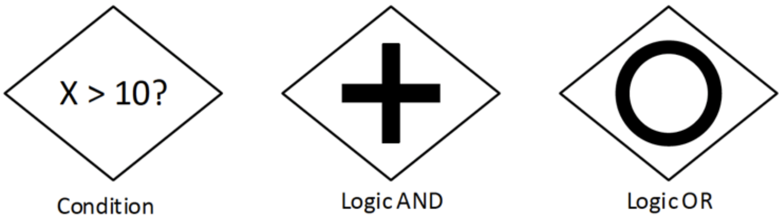


Figure 1.9: Examples of *gateway*

Connection Objects

The *connection objects* refer to those elements which link different flow objects within a diagram. In this sense, they shape the basic structure of a business process. Many types of connecting object are defined by BPMN (White, Bock, 2011):

- *Sequence flow*: it is represented by a solid line with an arrow at one end of the line, indicating the traveling direction. These objects give the reader the evidence on the order of the activities that are carried out in a diagram.

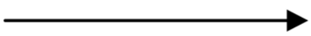


Figure 1.10: Example of *sequence flow*

- *Association flow*: it is graphically represented by a dotted line with an arrow at its extreme. Unlike the sequence flow, it is not used to connect two activities, but it has the task of associating *flow object* of the other objects, such as data, information, text or other entities.

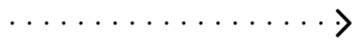


Figure 1.11: Example of *association flow*

- *Message flow*: it is designed as a dashed line with an open circle at one extreme and an empty arrow to the other. It is used to show the message flow between the participants of the business process.

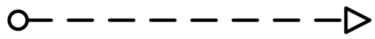


Figure 1.12: Example of *message flow*

Swimlane

In BPMN, the *swimlane* are those constructs intended to graphically organize activities and to divide them into different groups, in order to put in evidence the different competence and / or responsibilities. The two types of elements that make up the category of swimlanes are (White, Bock, 2011):

- *Pool*: it represents a specific actor within a business process. A pool acts as a graphical container that groups together the activities performed by the same resource, function, or business unit. Therefore, the identifier name of each pool indicates the resource associated with the process, which can be both an internal actor (business unit or function) or external (suppliers or customers). In that sense, the formalism allows to simple describe a large number of processes, both B2B and B2C.

Figure 1.13: Example of *pool*

- *Lane*: a set of sub-pool partitions, useful to provide an additional level of detail of activities.

Figure 1.14: Example of *lane*

Artifact

As highlighted before, the purpose of BPMN is to ensure the greatest flexibility possible in the development of a diagram. To obtain this result other elements are present in the notation, called *artifacts*. There is no limit to the number of artifacts in a diagram, because they allow to describe as accurately as possible the business processes that are to be mapped. The notation defines three objects for this category (White, Bock, 2011):

- *Date object*: they are used to show which are the required data or the information generated by an activity. They are connected to the activity through an association-type connection.

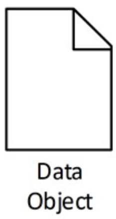


Figure 1.15: Example of *data object*

- *Group*: it is graphically represented through a rectangle with rounded corners, drawn with a dashed line. It can be used to generate documentation with various meanings, but it does not have any influence on the *sequence flow*.

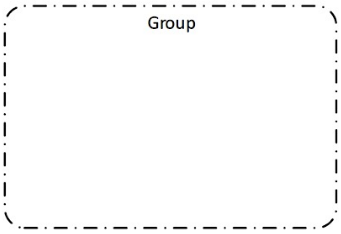


Figure 1.16: Example of *group*

- *Annotation*: useful to allow the insertion of text comments that will further guide the reader in understanding the process represented.

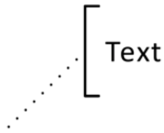


Figure 1.17: Example of *annotation*

Finally, a simple example of the elements described is reported in the next figure.
The case refers to the process related to a pizza delivery, triggered from the order by a potential customer, both online or via phone call.

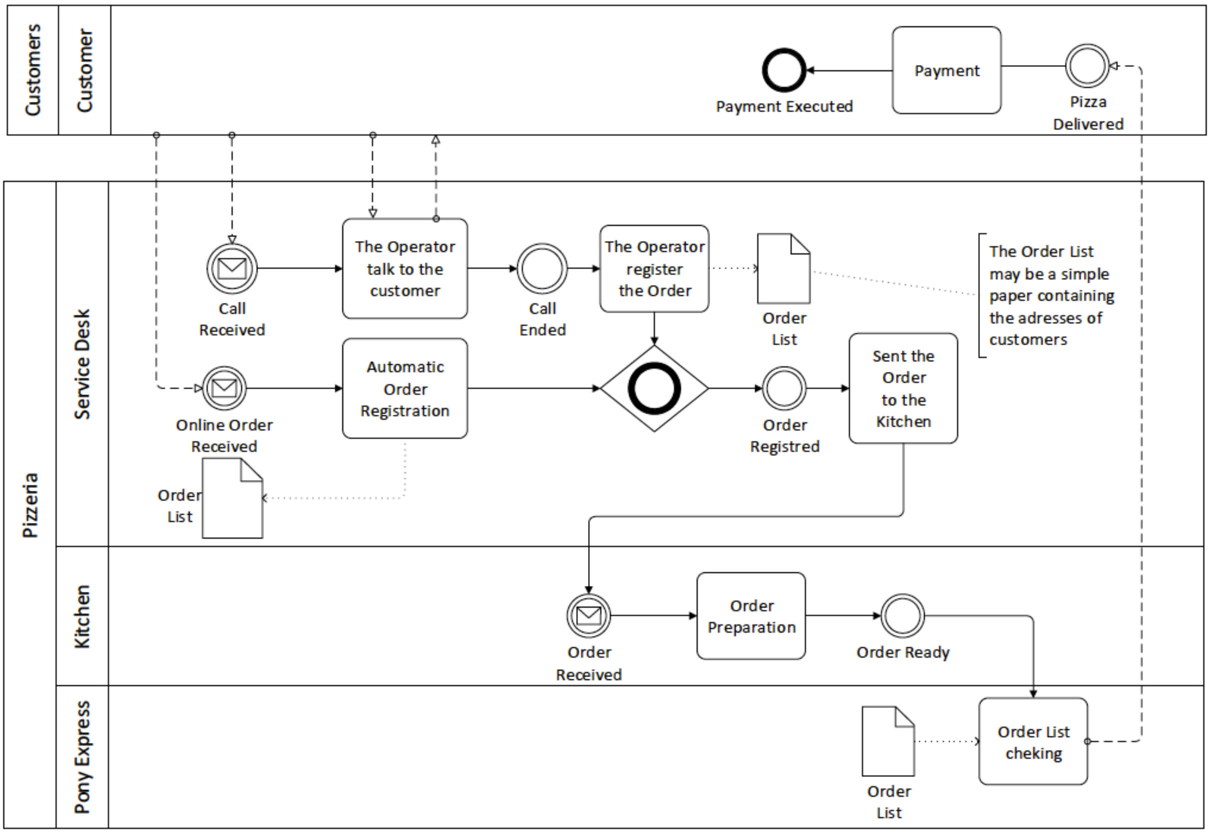


Figure 1.18: Example of BPMN diagram by using Microsoft Visio

Chapter 2 Business Process Management

2.1 A Business Process Management Framework

Business Process Management is a management theory which defines clearly the structure of the activities performed inside an organization. It can be defined as a holistic discipline, focused on the technology-oriented processes of modeling and analysis, in addition to other subjects as the strategic alignment, governance, human resources and corporate culture (Vom Brocke, Rosemann, 2010).

BPM is a science applied for monitoring how business processes are executed inside the firm, with the purpose of reaching all the predetermined goals and pursuing on performances improvement. These kind of improvements are evidently depended on the specific objectives of the organization and are influenced by its internal structure and the external context in which it operates. The principle of BPM is not only finalized towards a unique and local improvement, rather towards a better and distributed knowledge about the events chain, activities and decision flows. The role of carrying out a BPM activity is clarified by taking into account a generic supply chain. In a very simple case, just the following main actors are present:

- *Partner and supplies*, who make provision for the raw material and all the relevant resources needed to operate (e.g. financial resources, human resources, technology or material resources);
- *Customers*, who guarantee revenues.

BPM links these two different prospective with the prior aim of generating value for the customer: from various inputs a product or a service is shaped. This outcome is then delivered to the customer through various activities, arranged in processes (La Rosa, 2015).

Expanding the previous supply chain other relevant players must be considered:

- *Competitors*: the other firms which operate on the same industry, or which serve their outcomes to the same target market, starting from the same or similar inputs;

- *Business environment*: the context in which the firm do its businesses. Each business environment is characterized by its specific economical, tax, political or cultural aspects which influence both suppliers and customers, and therefore the transactions with them;

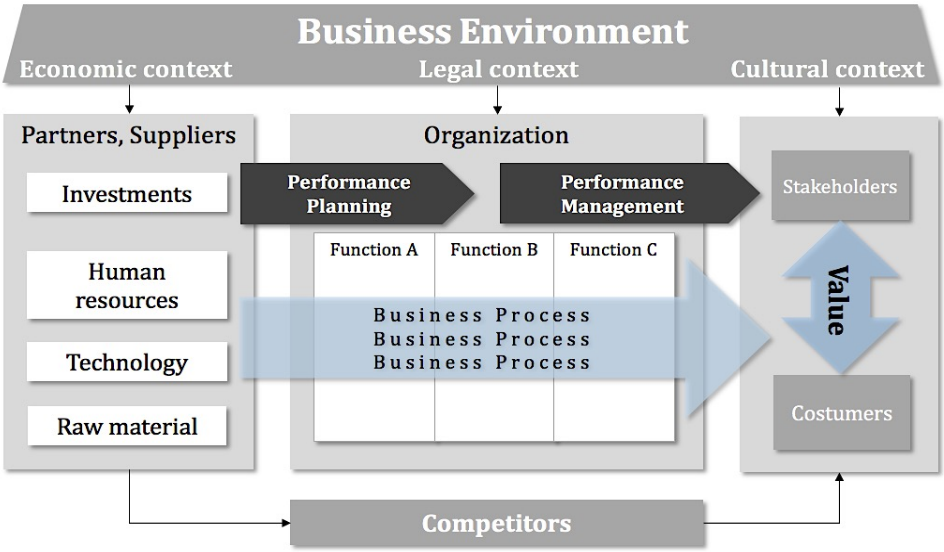


Figure 2.1: A proposed framework for BPM

- *Shareholders*, individual or institution (including other corporation) that legally owns a share of stock in the firm. Therefore, the value that a company should to generate to them is a reasonable guarantee of a future economical return in the form of more stocks or exit possibilities. In order to pursuit these objectives and to correctly serve this type of actors, organizations should correctly plan their processes and performances in accordance to their strategic goals.

However, firms that carry out a BPM initiative seem to be moved by similar objectives and necessities, synthetized through the following model (Franz, Kirchmer, Rosemann, 2011 and La Rosa, 2015). Taking into account a generic business environment, it is possible to divide every single factor impacting in someway to the organization through the belonging to the organization itself, or not. From this hypothesis it is possible to identify *internal attributes* or *external ones*. The model considers the presence of one core-value, and three pairs of peer ones. The core-value is the **transparency**, necessary to obtain all the other values. Only if the organization is able to share the knowledge relative to their processes could introduce itself correctly to a BPM plan. In this sense, transparency is considered to be the preliminary condition for the entire initiative (Franz, Kirchmer, Rosemann, 2011).
To gain transparency means to obtain visibility on the way the firm operates, on how data are

collected and used, on the various users and on the risks interrelated. This essential driver needs others specific preliminary conditions:

- A scalable internal process model (or map), necessary to support the management in their decisions related to the operational processes, which could be also personalized if the situation requires changes;
- An accessible process map, easy to understand and to consult, without compromising firm's security or causing any undesired sharing of sensible data;
- A process map containing all the relevant information, in order to obtain the specific strategic objectives, such as a risk evaluation for risk management, job descriptions for human resources management, etc.
- A process model based on roles and responsibilities, in such a way that every actor could find which map is suitable for his own position;
- Correlation with other process maps, in a way to let management to identify how a single change in an activity or in an area could impact somewhere else.

As introduced, the other objectives are in relation each other by pairs, following a *win-lose logic*: basically every internal - or external - attribute tend to has a negative effect of its external - or internal - pair factor (La Rosa, 2015). The *internal* values are:

- **Efficiency**: the goal is to improve business performances and this is often related with the managing of internal costs. Therefore, a consistent indicators system, able to measure all the relevant factors, like idle-times, time-to-market, allocation index, is needed;
- **Conformance**: the ability of making processes predictable and in certain way controllable, if possible;
- **Integration**: to gain a better assimilation of the internal structure.

By contrast, the *external* dimensions are:

- **Quality**: in this case the indicators should be related to both quantitative aspects, such as number of defection, and intangible ones, like costumer satisfaction;
- **Flexibility**: the capacity of the organization to change itself, if necessary, and to better respond to external stimuli and new business environment needs;
- **Networking**: the improvement of the relationships with other external actors, both customers or suppliers.

As mentioned, these six aspects are related according to a *win-lose logic*: for instance, the increasing of the quality (external variable) assumes an expected increasing in costs, but this may result in a reduction of the efficiency (internal variable).
The purpose of BPM technique is to improve the value of a variable without causing any marginal loss in the values of the related one (Franz, Kirchmer, Rosemann, 2011).

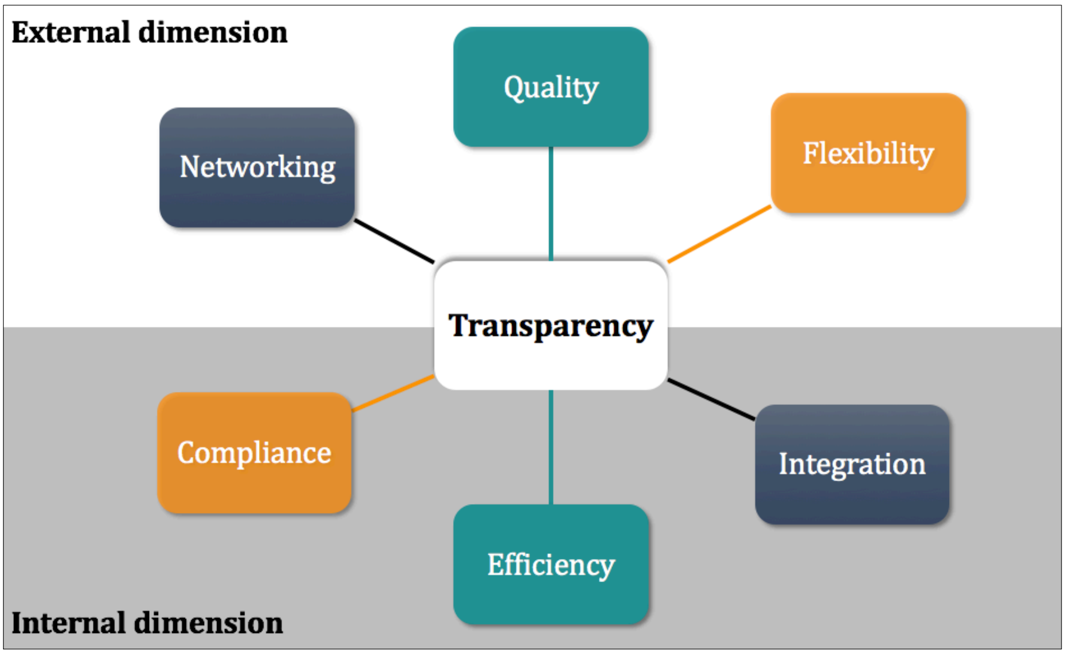


Figure 2.2: BPM external and internal dimensions (from Franz, Kirchmer, Rosemann, 2011)

The difference in pursuing a cost leadership strategy or a quality leadership one, has been widely discussed and analyzed from Porter (Porter, 1980) and form the related models. The BPM allows, instead, to balance the two variables, contributing to pursuit both strategies.

The search for internal *efficiency,* through the elimination of waste and rework and through a better allocation of resources, is committed to be one of the main drivers of a BPM initiative. Inded, the drive towards a progressive and increasing internal optimization is a goal - or compulsion - more relevant than ever, especially at this historical post-crisis period. In this sense, a successful implementation of BPM, addressed to achieve a higher efficiency degree, provides a deep economic focus in the analysis and in process design, a reclassification of assets based on the value of the outcomes producted, the adoption of costing philosophies, activity-based accouting and conscious mastery of complexity and cost drivers.
On the other hand, to pursue a strategy aimed to *quality* reflects an external focus, as this theme is strictly related to customers who receive the products or services created by the organization,

through its business processes. In BPM, this attribute integrates the views of external stakeholders, who benefit from the outcomes, with the internal analysis and design of the processes. Companies adopt customer-driven BPM to increase control over their performances, to find the root of how the activities are performed, without neglecting the metrics of customer satisfaction and all those theories - such as Six Sigma or TQM - with quality as key object. The development of awareness on the correlation between quality and outcome of the process, as well as the need of a correct involvement of the relevant external actors, are another goals pursued by this type of initiative. Therefore, BPM appears to be a bridge between the two dimensions and it seems able to fill the existing gap, which has been initially theorized by Porter.

Nevertheless, by considering the current market dynamics, organizations are increasingly exposed to the turmoil caused by the entry of new competitors, the customers' needs and the technological drivers. In this situation, companies are called to be dynamic and adaptable to changes. For the BPM, the *flexibility* can only be achieved through a process design able to react promptly to external pressure, without being slowed down by the excessive internal structure, and thought the possibility of directly monitoring the context. Companies characterized by such flexibility operate through a light and decentralized structure, not overly prescriptive processes, but they can adapt themselves in response to the needs of the moment. This is achieved through a continuous analysis of the business environment, able to quickly detect changes and to quantify how they impact the process performances.
Moreover, the need of *compliance* concerns to the use of activities as standardized and defined as possible. In this case, the priority resides in the design of processes the most predictable possible, as well as controllable. As seen with regard to the adoption of specific notations of process mapping (Chapter 1.4: Process Modeling), this aspect may be specifically required by law or regulations, or during an audit activity. BPM provides for compliance with the necessary restrictions, only if these standards are strictly related to the real business needs and already specified in the initial process design phase. Rather than on compliance, the focus lies in the careful allocation of responsibilities and ownership of the processes, as well as in the definition of milestones and releases during the first stages.

Furthermore, firms tend to have a focus either internal or external, especially in terms of how processes are seen from the employees and the perception of the outer environmental impact of these processes. *Integration* is focused on the internal balance of resources, both human or assets, and captures the implications of current process design. These implications include issues such as whether the processes are related with precise job description and operating procedures, the level of internal acceptance, and the process metrics used. A firm which is searching for integration requires close involvement of its human resources in process design in order to ensure high acceptance levels, the using of indicators which could put into relation business processes with employees' satisfaction and motivation, and a bottom-up and decentralized approach to processes.

By contrast, organizations that emphasize *networking* believe that processes have to be designed taking into account external actors and resources first. Networking as as driving value for BPM requires that the companies clearly identify the role of these external partners and resources, and complements its traditional focus on time, cost and quality with the environmental impact. All must be carefully developed to ensure maximum positive synergies and sustainable partner relationships. Finally, organizations have to explore all the potential opportunities of social technologies, such as social networks like LinkedIn or Twitter. This, with the specific aim of keeping external audience informed about changes, in order to engage even more strong and effective relationships with new stakeholders (Franz, Kirchmer, Rosemann, 2011).

2.2 The BPM Lifecycle

In general, a BPM initiative is characterized by some activities, each with specific actors and goals, that contribute together for the success of the overall plan. It is useful to view BPM as continuous cycle, by descripting the following six phases (La Rosa, 2015):

1. Process identification: defining priorities and the purpose of the process. This first stage has two fundamental outputs:

a) *Architecture*: namely the inventory processes;
b) *Process Portfolio*: the vision of the processes, sorted according to different values that indicate the execution priority;

2. Process discovery: get the representation of processes done, through interviews addressed to the resources directly involved in the processes, or through other techniques. The outcome of this stage is the representation of the *as-is process model*.

3. Process analysis: starting from the as-is process model, the purpose is to identify and to quantify the critical aspects and the weaknesses of the business process, which may result to unsatisfactory performances. The techniques used can be both *qualitative* (for instance, the famous Ishikawa diagram for the root/cause analysis), both *quantitative* statistical models (such as simulation of the processes in different scenarios).

4. Process re-design: in this phase the picture of the *to-be process model* is shaped, with the aim of filling the gaps identified and improving performances. Usually, those areas where an intervention is needed are the same as listed above, such as the value-pairs quality-efficiency, flexibility-compliance and networking-integration. However, it remains essential to improve a dimension without going to affect and worsen another one.

5. Process implementation: implementing the configuration previously designed. The two phases that made up the stage are:

a) *Process Automation*: concerning automated or computerized activities;

b) *Change management*: carefully managing all the soft aspects related to the involvement of human resources.

6. Process monitoring & controlling: to pursue the philosophy of continuous improvement, using the feedbacks and all the relevant information obtained from the system. The methods used during this phase could be, for instance, database logs and event stream for the preparation of reports and dashboards, and many other techniques.

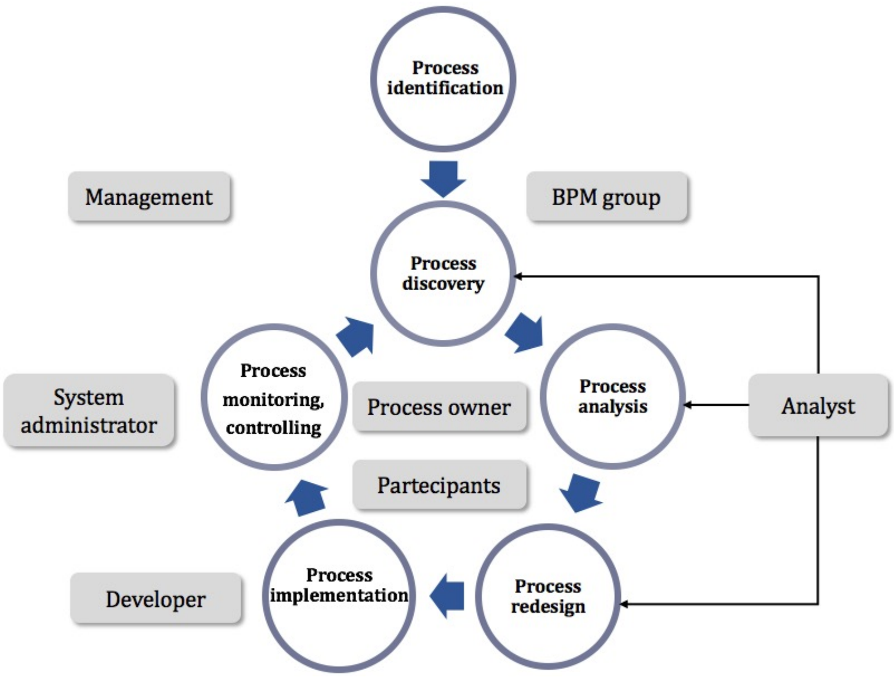


Figure 2.3: The BPM lifecycle and the stakeholders involved (from La Rosa, 2015)

As introduced, during the BPM lifecycle many different stakeholders are involved. Therefore, before exploring BPM lifecycle, is it useful to distinguish the following main participants:

- *Management Team*: the group responsible for overseeing and coordinating the processes, initiating process redesign initiatives, providing the resources needed and the strategic guidance to all the stakeholders involved within other phases;
- *Process Owner*: the responsible for the effective operation of the process, responsible both for planning and organizing, and both for monitoring and controlling the process. This stakeholder is also responsible for the process participants involved;
- *Process participants*: the stakeholders that perform the activities of a business process and that conduct routine tasks, according to the standards and the guidelines of the company. They are mainly involved during the process discovery and analysis phases. Anyway, they may support the redesign activities and the implementation stages too.
- Process *Analysts*: resources that execute the process identification, discovery, analysis and redesign activities. They may coordinate process implementation as well as process monitoring and controlling.
- *Developers*: the engineers typically involved in process redesign and implementation stages. They translate requirements of the *to-be process model* into a system design;
- *System Administrator*: the resource responsible for the operational system, which supports the monitoring and controlling of the processes;
- *The BPM Group*: a group of process management experts, that are responsible for maintaining the process architecture, prioritizing process redesign projects and for ensuring that all the process documentation is maintained in a right way, and finally that the process monitoring system works as expected.

2.1.1 Process Identification

During the very first phase the business problem is posed, processes related to the problem are identified and put into relation to each other. The outcomes of process identification are a new or updated process *architecture* and a *portfolio of processes* that provide an overall view of the processes and their relationships. Process identification is determined by some activities, with the purpose of systematically identifying the set of business processes of the firm, and establishing clear ways for prioritizing them. But first, firms have to define some criteria in order to put key processes into relation with their strategic importance and with the value they generate. It's essential, at this point, to determine which processes have stronger priority, and for this reason a map of all the processes performed inside the organization is required. More specifically, process identification is concerned with two different sub-phases: (I) *designation* and (II) *evaluation* (La Rosa, 2015).

The *designation* phase is conducted in order to categorize and to understand all the processes the firm is involved in, and the inter-relationship existing between them. The decision about how many processes have to be taken into account is the most critic step of this sub-phase: it is essential that companies consider the existing trade-off between *impact* and *manageability*. In this sense, if only few processes are considered, inevitably each of these will cover a huge quantity of operations and activities. On the other hand, many processes related with a smaller and more focused group of activities, could be hard to manage because too many specific drivers and characteristics arise. To find the right balance to this relation, seems to be useful to separate *broad* processes from *narrow* ones (Dumas *et al*., 2013). A broad process is a group of activities related to company's critic factors of success, characterizing those areas where the firm has to monitor and preside the operations in a way to guarantee quick response and fine-tuning, whereas a narrow one is a process in which there is no need of an active monitoring and of a continuous updating.

After the identification phase, the *evaluation* stage is executed in order to quantify the importance of the processes listed, and then to prioritize the amount of attention each of these required. One of the most used criteria concerns to find the importance of a process related to the strategic value it creates, therefore it is sufficient that the considerations are defined just at a very abstract level. Another criteria takes into account the potential dysfunction already existing in the as-is processes, using more qualitative approaches. The last method is related to the process susceptibility to another initiative, and it strives to quantify the potential cooperation - or contention - between them.

However, the first fundamental output of the process identification phase is the so-called *process architecture*, which is a conceptual model showing all the processes of the organization, and the relationships existing between them.

Process architecture defines different levels of detail (Joosten,2000):

1. *Permanent processes*, which have a long life cycle inside the organization and represent the processes on a very abstract level;
2. *Case processes*, which guide the execution of the process and directly identify it. In this level processes are listed at a finer degree of granularity but still in a certain abstract level;
3. *Detailed processes level*, where processes are carefully described in a detailed way, including all the inputs needed and outputs generated, the relevant actors involved and their responsibility and relationships.

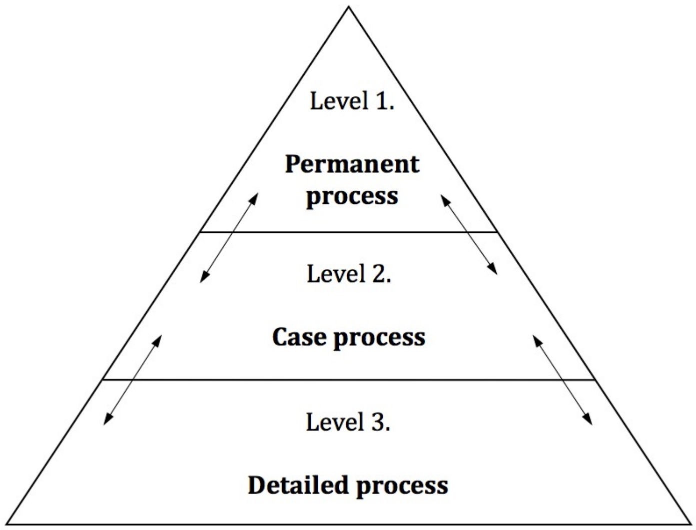


Figure 2.4: Levels of process architecture

Furthermore, the second fundamental output is the creation of a *process portfolio*, where each process is identified in different classes according to the specific dimension considered. The main purpose is to determine the ways in which groups of similar processes are managed by the organization. The classification of the group types is conducted through some properties, according to the firm and the scenario in which it operates.
The most commons are (Dumas *et al*., 2013):

- Product (or service) type: the different products (or services) are identified;
- Channel: the way in which the organization delivers the outcome to their costumers;
- Costumer type: according to the specific profile of the buyers, often identified connected to the strategy and to the market segmentation.

Then, a classification of the business processes that are performed, related to their distinctive properties, is created. This progression requires that each of the case type is inspected in detail, and that for each one both the business functions and the roles involved are identified. The classification of these characteristics and the functional units (or managers and resources) identified originates a matrix 2x2 that help to visually identify the main aspects related to the process and the relationship between all the resources involved.
Then, the architecture and the process portfolio are used during the consecutive phase for the purpose of designing the organization's process map, describing the actual situation in which activities and tasks are performed. Therefore, the techniques used here are those already described in the first chapter, such as BPMN (Chapter 1.4.4: BPMN).

2.2.2 Process Discovery

Process discovery is defined as the demonstration of gathering information around current process and the organization of data in terms of an *as-is process model*. This emphasizes assembling and organizing different sort of information and data.
In general, three types of discovery methods have been identified: (I) *evidence based discovery*, (II) *interview-based discovery* and (III) *workshop-based discovery* (Dumas *et al*., 2013).

The *evidence based discovery* takes into account the various pieces of evidence available, just to make it is possible to study how the existing processes really work. There are three methods that allow to achieve this goal: document analysis, observation, and automatic process discovery. Firstly, document analysis is surely the most direct way to collect facts about the existing processes, because every firm used to register relevant information into documentation material. However, there are some potential issues with this kind of analysis related to the subjectivity of the resource who has registered that particular information, or related to the multitude of standards used. These complications may arise not only among the various business units operating in the firm, but inside the same unit too.
Secondly, the observation method of discovery directly studies the processing of individual cases, in order to get an understanding of how the process is actually executed. This kind of analysis can either be conducted by the customer - internal or external - of the process, which has an active role in the process, or by a passive observer.
The last third option of automatic process discovery emerges especially thank to the extensive

operational support of business processes offered by a huge number of information systems. Just think about all the different enterprise resource planning (ERPs) existing and used daily in every companies, such as SAP, Oracle, etc. Automatic process discovery could be executed through the examination of the so-called *event logs*, stored within these information systems. Such event data have to be recorded in a way that each event can be exactly related to the case of the process, the activity performed and a precise time considered (van der Aalst, 2015). If these three pieces of information are available in the logs, then automatic process discovery techniques can be used to reconstruct the process model flow starting from event logs. Anyway, if the outcome of the automatic process discovery doesn't give an entire process map, or a major functionality is not present in the discovered process, to increase the data set should be a good solution. This can originate an iterative trial and error process (Jadhav, 2011).
Since this approach has in common some characteristics with data mining, where meaningful information is extracted from fine-granular data, these techniques of automatic process discovery are associated to one application of *process mining*. Process discovery method through the discipline of process mining, together with the connected methodologies, will be largely explored in the following chapter (Chapter 3: Process Mining).

Interview-based discovery pertains to methods characterized by various interviews, conducted upon domain experts, about how processes are executed and organized. Before using this method of discovery it is essential to explicitly take into account some challenges, related to the fact that process knowledge is often divided across different business units and specialists. Moreover, these experts typically think in terms of individual cases and act with a local focus, and they are often not familiar with any kind of business process modeling languages. All these considerations have relevant implications for how the interviews can be conducted, and which iterations are required. The interview-based discovery challenges emphasize the role of the process analyst, who is required for abstracting information on how individual cases are executed, and for enabling to construct meaningful process models. Firstly, the analyst has to obtain all the sensible information about the process through interviews and secondly has to organize all the material gathered and to develop an initial process model. As a consequence, interviews are often conducted in distinct iterations. After a general interview, the process analysts prepare a sort of draft of the process model, which is then discussed with the professionals directly involved in terms of correctness and completeness. The main objective of interview-based discovery is to provide a complete and detailed picture of the processes and the resources involved. The results could also reveal different perceptions on how the processes operate among the domain experts (Jadhav, 2011). The gathered evidences may help the analyst to understand the details and the specificities of the processes too. However, it is a very labour-intensive discovery method, both in term of time required and economical resources needed. In addiction, many iterations are necessary for arriving at a point where specialists feel comfortable with how a process is described in the model, especially for those experts who have the role of knowledge-owner, concerning that particular field.

Finally, another discovery method useful to let a rich set of information to arise, is the *workshop-based* one. In particular circumstances, the setting could be designed in such a way that the participants to the discussion are immediately used to model the process. Rather than interviews, workshops involve more participants together at the same time, but also need a defined set of roles. Indeed, additional roles are required for make the discussion well structured, such as the sponsor, typically a business executive who decide the scope and the goals. The sponsor identifies what is needed to be discovered, and establish the timetable of the phases. This figure is critical to ensure the success of the overall project, because a strong sponsor would be able to make the required resources and business information available to the analyst (Verner, 2004). Other essential roles are the facilitator, who takes care of organizing the verbal contributions, and the tool operator, who is the responsible for directly registering the discussion results into the modeling tool. As in the interviews, participants often came from different domains and backgrounds: both the process owner and the analyst interact at the same time.
Nevertheless, this process discovery method requires the involvement and the commitment of the resources involved as well as diligent preparation and scheduling pre-phases. Furthermore, the process will not be discovered in detail in just one session. It can be expected that numerous day sessions would be necessary and this would inevitably make costs higher.

2.2.3 Process Analysis

The analysis of the business process, previously identified and discovered, can be conducted through two main different approach: in a (I) *qualitative* or (II) *quantitative* way (van Hee, Reijers, 2000).

Generally, the first methodology concerns various techniques grouped as *value-added analysis* or *root case analysis*. The *value-added analysis* typically consists of two steps which are value classification and waste elimination. Value-added analysis is a technique aimed at identifying the unnecessary steps in a process - or waste - in order to eliminate them. The first thing an organization which wants to carry out a value-added analysis have to do, is to create a decomposed view of the process, identifying the different steps, activities and resources related. Then, the second prerequisite is to precisely identify who is the customer of the process and what are the specific outcomes that this stakeholder seeks for. These results of the process are said to add value to the customer, in the sense that fulfil all the requirements which are in the interest or for the benefit - and satisfaction - of this stakeholder. Having decomposed the process into stages and having defined the positive outcomes of a process, the firm have then to analyse each step identified in terms of the value it confers. In that sense, the stages that directly create positive outcomes or that contribute to the goods the customer wishes to see, are called *value-adding*. Anyway, some other steps do not directly add value to the customer but are committed to be essential for the business. Finally, others are pure wastes. So, according to the Lean Six Sigma management literature, value-added analysis is a technique whereby the analyst decorticates the process model, extracts every step in the process and classifies these into one of the following categories (George, 2010):

- *Customer Value-adding* (CVA): A step that straight produces value or satisfaction of the customer. A value-added activity is that task costumers are willing to pay for;
- *Business value-adding* (BVA): A step that is necessary for the business to run, or that is required due to the business environment in which the organization operates;
- *Non-value adding* (NVA): A step that does not fall into any of the previous two categories.

Having characterized the steps of the process as discussed, the following stage is about determining how to remove wastes. The general rule is that organization should strive to minimize - or to eliminate, if possible - NVA steps. Some of these NVA steps can be directly eliminated by means of automation, putting in place an information system that allows to detect and cancel them. A more radical approach is to completely remove the resources associated with the process and to aggregate tasks somewhere else in the company. This means moving some additional work to the another worker or unit, in a way there are less handovers in the process. Anyway, the consequences of this option, in terms of added workload to the resource, need to be definitely considered (Dumas *et al.*, 2013). Another approach to delete NVA steps is to directly eliminate the need for approval of requests, in the cases where the estimated costs are

below a predetermined limit value. Again, this action should be carefully weighted against the possible undesired consequences of having less control steps in place.
The removal of NVA steps is generally considered the prior desirable goal, whereas the elimination of BVA ones should be treat considering the trade-off existing between these steps and their role in the business. Therefore, an organization should first put down all the BVA steps and then link them with the business goals or requirements, such as the special regulations that have to be respected or the risks that are sought to reduce. The company has to identify what is the minimum amount of work required in order to ensure the customer satisfaction, while fulfilling the business goals and requirements.
Another useful qualitative discovery method is the *root cause analysis,* which is a family of tools that support company to identify and to understand the root cause of problems. This methodology is commonly used in the context of accident or incident analysis, as well as in manufacturing processes where it is applied to detect the cause of non conformance in a product, or the origin of undesirable events. In the context of business process analysis, root-cause analysis is useful to identify the issues that prevent a process from gaining a better performance (Andersen, Fagerhaug, 2006).
Cause-effect diagram is one of the technique belongs to root-case analysis family, and it is basically used to identify the relationship between a given negative effect and its causes. In this regard, a negative effect can be for instance a frequent problem or a lower level of performance. The causes detected are then divided in main and contributing factors. Each factors are grouped into categories and, if possible, in sub-categories too. Because of their appearance, these kind of diagrams are also known as Fishbone diagrams (or Ishikawa diagrams, from the name its creator Kaoru Ishikawa) (WBI Evaluation Group, 2007).

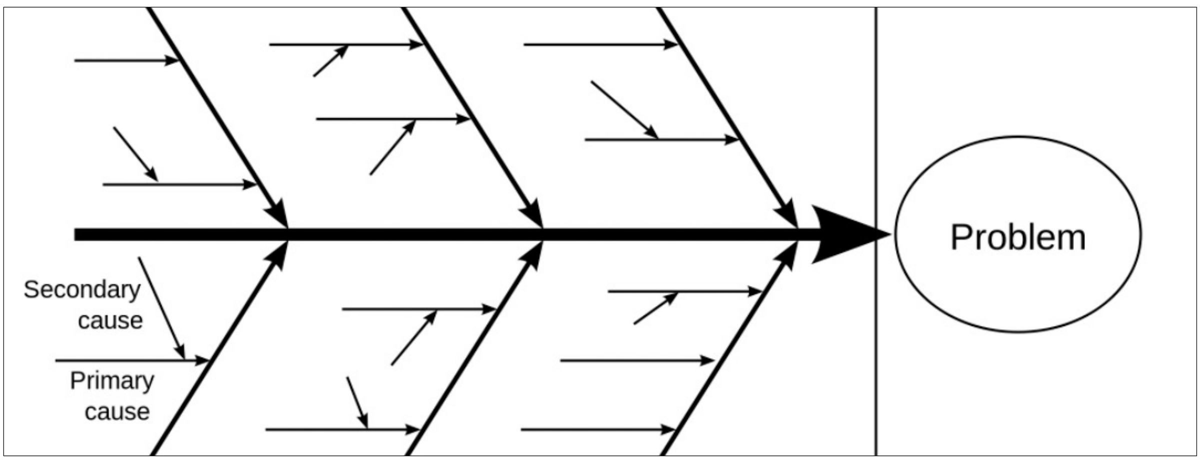


Figure 2.5: A root-cause diagram (Ishikawa diagram)

Finally*, tree diagrams (or why-why diagrams*) constitute another technique useful to analyse the cause of undesired effects in a business process. The aim of the methodology is to recursively ask why a specific issue has happened, until stakeholders would be able to identify the root-case under it.

The second main way in which the process analysis phase can be performed is through a *quantitative* approach. This allows to detect processes quantitatively, in terms of specific performance measures. Normally, three distinctive techniques are identified, namely *flow analysis, queueing analysis* and *simulation* (Dumas *et al*., 2013). All these have in common the goal of calculating performances of a process, starting from given data about the individual activities and resources involved. Generally, every quantitative analysis is based on the precise identification of the three main process performance dimensions, which constitute the so-called triple constraints of project: time, cost and quality (Atkinson, 1999). Another way of classifying and defining performance measures is given by the concept of balanced scorecard which is an approach that allows to align the objectives and facts based on four performance dimensions: financial, internal business, innovation and customer measures (Kaplan, Norton, 1992).
Flow analysis is a family of techniques that allows to quantify the performances of a complex process from the evidence about the single performance of its activities. With flow analysis it is possible to calculate the average cycle time of an entire process from the average cycle time of the single activity, quantify the error rate of a process knowing the error rate of each activity, or calculate the average cost of a process from the cost per-execution of the single activity.
Queueing theory is another one quantitative analysis method, based on the collection of mathematical techniques that enable to analyse systems characterized by resource contention, which inevitably leads to queues. This theory provides companies with tools calculating specific parameters of a queue, such as its expected length, or idle times.
Finally, p*rocess simulation* is one of the most supported methodology for quantitative analysis of the business processes. In essence, a process simulator generates some hypothetical cases of a process, executes them, and then records each result. The outcome here typically includes some statistics related to average waiting times, resource allocation or production planning.

2.2.4 Process Re-design

The process re-design phase deals with rethinking the business processes, and re-organizing them with the specific objective of making them more efficient and to reach higher level of

performances. There are two basic methods for achieve this goal: the (I) *heuristic process redesign* and the (II) *product-based design*.
Before start any project based on the re-design approach it is pivot to identify some elements of the process. They are (Dumas *et al*., 2013):

- *Customers* - internal or external - of the business process considered;
- *Business process operation view*, which reflects how the business process is implemented, the number of activities that made up it and their characteristics;
- *Business process behaviour view*, related to the way the business process is executed, the order of activities, how these are scheduled and assigned for the execution;
- *Organization and the participants* involved, in terms of the company structure (roles, users, groups, unit, etc.) and its population (human resources and those activities assigned to them);
- *Information* needed to the business process to perform, and data created from its execution;
- *Technology* used in the business process;
- *External environment* (economical, tax, political or cultural aspects) in which the process is situated in.

The first methodology used to conduct a process re-design activity is the *heuristic process redesign*. This method makes use of a fixed list of redesign techniques - often simple actions - in order to create potential improvement on the existing process. For each of the heuristics considered, it is necessary to identify not only which one suits best with the situation, but also what desirable outcome it is able to produce. Therefore, a redesign heuristic should to be applied whether it helps to acquire the desired performance improvement in the process analysed. After consideration of each of the redesign heuristics, it is a good practise to explore which clusters of applicable and desirable heuristics can be developed. Indeed, some of the heuristics have to be applied together, in order to obtain mutual benefit from the synergies originated, while others should be applied separately because they are quite contrary or because they may give rise to overall negative results (Grandi, 2015).

Heuristics can be classified in different groups, formed by a list of actions, according to a similar aim or situation (Dumas *et al*., 2013): *customer heuristics*, *business process operation heuristics*, *business process behaviour heuristics, organization heuristics*, *information heuristics*, *technology heuristics*, and *external environment heuristics*.

The following table point out the differences among the groups and shows the main actions that made up them.

Customer Heuristics	
Control relocation	To move controls towards the customer
Contact reduction	To reduce the number of contacts with customers and external stakeholders
Integration	To purse the integration with a business process of the customer or the supplier
Business Process Operation Heuristics	
Case types	To determine whether activities are related to the same type of case and, if necessary, identify new type of business processes
Activity elimination	To eliminate unnecessary activities from a business process
Case-based work	To remove batch-processing and periodic activities from a business process
Triage	To split a general activity into two or more alternative activities
Activity composition	To combine small activities into composite activities and to divide large activities into workable smaller ones
Business Process Behaviour Heuristics	
Re-sequencing	To move activities to more appropriate places and to change the scheduling
Parallelism	To consider whether activities may be executed in parallel
Knock-out	Order knock-outs in an increasing order of effort and in a decreasing order of termination probability
Exception	To design business processes for typical cases and to isolate abnormal cases from the normal flow
Organization Heuristics	
Case assignment	To assign workers perform as many steps as possible for single cases
Flexible assignment	To allocate work in such a way that maximal flexibility is preserved for the near future
Centralization	Treat geographically dispersed resources as if they are centralized
Split responsibilities	To avoid shared responsibilities for tasks by people from different functional units
Customer teams	To compose work teams of resources from different departments able to take care of the complete experience of specific sorts of customers
Numerical involvement	To minimize the number of departments, units and people involved in the same business process
Case manager	To defined one person to be the case manager, responsible for the handling of each type of case
Extra resources	To increase the available number of human resources
Specialist-generalist	Consider to let resources to increase or decrease their skills
Empower	To empower workers with most of the decision-making authority instead of relying on middle management

Information Heuristics	
Control addition	To conduct additional control about the completeness and correctness of incoming materials and check the output before it is sent to customers
Buffering	Instead of requesting information from an external source, buffer it and subscribe to updates
Technology Heuristics	
Activity automation	To automate repetitive or routine activities
Integral technology	To elevate physical constraints in a business process by applying new technology
External Environment Heuristics	
Trusted party	Instead of determining information oneself, use the results of a trusted party
Outsourcing	To outsource a business process completely or partially
Interfacing	To establish a standardized interface with customers and external partners

Table 2.1: Process redesign heuristics

The second methodology in order to perform the re-design phase is the *product-based design*, which is very different from heuristic process redesign. Basically, it promotes a radically re-thinking about how a particular product or service can be developed, rather than pursuing an incremental approach as the previous method. Furthermore, the existing as-is process is not even more the starting point of the phase. Rather, the key point that guides the whole project is the identification of the characteristics of the particular product or service that the to-be process is expected to deliver to the customer (Reijers, Limam, Van Der Aalst, 2003). The fundamental idea behind product-based design is that, by ignoring the way actual processes work and how a particular product is shaped, it becomes easy to develop the leanest, most efficient process possible. In general, without being blocked by the status quo situation, it is more feasible to design and to conduct a change project with the purpose of gaining the higher results possible. Nevertheless, product-based design method is more ambitious than the previous technique: as a consequence, it is more limited in its application scope. Anyway, the most important stages of this design technique are (Dumas *et al.*, 2013):

- *Scoping*: The initial phase in which the business process that need to be redesigned is selected. Then, both the specific performance targets and the related limitations have to be carefully identified and to be taken into consideration for the planning of the final design.
- *Analysis*: The careful study of the product characteristics and specification have to be

carried by the product decomposition into information pieces and their logical relationships, in the form of a product data model. The information that have to be considered are significant for both designing the new business process and for conduct any kind of evaluation of the ongoing processes.

- *Design*: Starting from the redesign performance objectives, the product data model previously identified and the estimated performance figures, the present phase is conducted in order to develop new processes that could match the designed goals.
- *Evaluation*: The designed process has to be verified and validated with the support of the end-users, in a way that the estimated performance could be discussed in more detail. The most promising option can be presented to the top management to assess the degree in which objectives can be realized and to select the most opportune design to be implemented.

Even if these phases are presented in a sequential order, in practice it is often preferable that some iterations will take place. For instance, the last evaluation phase is directly aimed at identifying previous design errors, which may result in the necessity for reworking on the same phase again.

2.2.5 Process Implementation

During the process implementation phase the changes required to move from the as-is situation to the to-be situation, are prepared and executed. As introduced in the preface of the chapter, process implementation stage involves two different but complementary stages, which are (I) *process automation* and (II) *organizational change management* (La Rosa, 2015).

On the one hand, *process automation* concerns to the development and deployment of new IT systems - or to the reconfiguration of an existing one - able to support all the necessary tasks for the to-be process designed. Therefore, the system should guide the resources in the execution of process activities in which they are involved. This may include, for instance, assigning tasks to the participants, helping them to prioritize their time and to schedule their work, providing them with the information they need, and performing automated crosschecks or some automated tasks, if needed and possible. There are several ways to implement such an IT system, aimed at extending the to-be process model, obtained from the previous process redesign phase.

One specific kind of technology that is particularly suitable to achieve process automation is the Business Process Management System (BPMS). With this expression, we refer to a generic system aimed at the explicit design of the processes, which realizes applications able to manage business processes (Weske, van der Aalst, Verbeek, 2004). It is a process-aware information system, able to derive an explicit representation of a business in the form of a process model, useful to coordinate the overall execution of the activities. In short, the main purpose of an BPMS is to organize an automated process in such a way that all the activities are conducted at the right time and by the right resource.

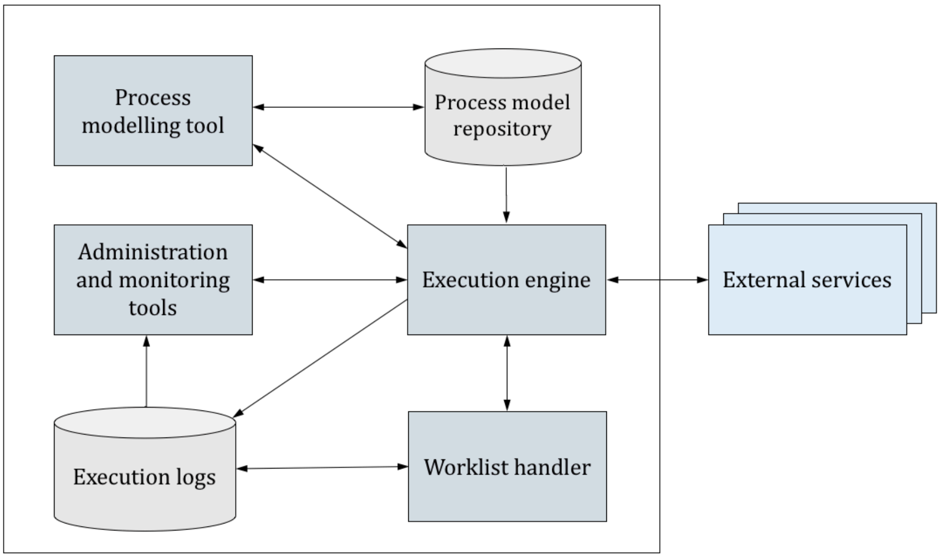


Figure 2.6: Business Process Management System (BPMS) representation
(from Dumas, La Rosa, Mendling, Reijers, 2013)

The BPMS is also a standard type of software system. Today, BPMS software industry is in a certain way mature, so the technologies used by different vendors are very similar to each other (Sinur, Hill, 2010). In that sense, it becomes feasible to support the execution of a specific business process by using the standard facilities, offered by the system.

The BPMS is made of various components with specific characteristics and objectives, which interact with each other. These components are the *execution engine*, the *process modeling tool*, the *worklist handler* and the *administrating and monitoring tools* (Dumas *et al*., 2013).

- *Execution Engine:* It is the central component of the BPMS, which provides various functionalities such as the possibility to create executable process instances, the ability to distribute tasks and activities to the right participants, the capacity to automatically store relevant data, required for the execution of the processes, and to automatically delegate tasks to the various software applications across the organization. Moreover, the engine constantly monitors the progress of the different instances, and schedules which activities have to work on next, by generating work items for each specific cases;

- *Process modeling tool:* The main functionalities this component provides are the possibility for users to create and customize process models. The tool lets the users add additional data, process roles or specific performance indicators, and enables them to store and share models from a *process model repository*. The execution engine makes use of the process modeling tool, in order to discover the temporal and logical order in which the activities of a process have to be executed: it determines which work item have to be generated and to whom it should be allocated;
- *Work list handler:* The component that enables participants to see which items are ready for them to be executed. It is the execution engine that keeps track of what items are available and makes them visible, through a sort of list, for the resources involved;
- *External services:* They are all the external applications required to execute the business processes. These applications have to expose the service interface the engine can interact with, and all the necessary data for performing the activity of a specific case. Once completed the request, the service will return the outcome to the engine. For those business processes that are not completely automated, the execution engine will invoke the appropriate services, with the right parameters, in order to support the process participants involved.
- *Administration and monitoring tools***.** This group refers to the tools applied for supporting the management above all the operational and administration matters of the BPMS. These tools are required to monitor the way processes are working, or to deal with exceptional situations, by aggregating data from different cases. In this way, the execution-related events stored, can be exported in the form of execution logs. As in the automatic process discovery (Chapter 2.2.2: Process Discovery), this theme will be discussed in depth in the following chapter. (Chapter 3: Process Mining).

On the other hand, *organizational change management* concerns with the activities required to move from the as-is process to the to-be process, especially to those related with the people directly involved. In this phase are executed all the actions required to change the way in which the company and the internal stakeholders works.

These activities include (Grandi, 2015):

- Explaining the *vision* of the changes to the participants and involving them in such a way that they perceive the tangible benefits that these change are expected to create for the organization;
- Creating a *sense of urgency* inside the functional unit, or among resources involved in the change project, in order to ensure their full acceptance and commitment, making them ready to react quick;
- Building up the *team* that will guide the change, and carefully projecting roles, responsibilities and procedures;
- Designing *quick-win projects* in order to immediately convince resources about the utility of the change project and about which benefits are expected;
- Putting in place a *change management plan,* so that stakeholders could be prepared about the scheduling of the activities and the arrangements needed to address problems during the transition to the to-be process;
- *Training resources* about the new way of working, and monitoring the plan in order to ensure a smooth transition to the to-be process.

2.2.6 Process Controlling and Monitoring

Finally, some adjustments might be required in order to match the changing of the business environment or because the implemented process does not meet the expectations.
Therefore, the process needs to be cautiously observed through the analysis of the data collected by a process monitoring tool, in a way to identify which adjustments are required for gaining an exceptional control on the overall execution of the activities. During the processes executions, relevant data are collected and studied in order to determine how well the processes are performing, in relation to the grade of expected performance measures and objectives. In this way, all the reprocesses, bottlenecks, recurrent errors or deviations are identified, and corrective actions can be prepared. Anyway, new issues may then arise within the same process or somewhere else, requiring the cycle to be repeated again. All of these activities are included in the process monitoring and controlling phase.

Even if this phase is essential, it requires a continuous great effort: the lack of a continuous commitment on the monitoring, inevitably leads to performances degradation. The process cannot improve without an adapted approach that keeps up with the ever-changing landscape of customers' needs, technology and competitors. This is why the phases in the BPM lifecycle should be seen as being circular (La Rosa, 2015): the output of the monitoring and controlling stage feeds back into the discovery, analysis and redesign phases, creating a continuous flow of data and corrective actions.

Chapter 3 Process Mining

3.1 The Importance of the Log Event for Process Mining

3.1.1 Data Everywhere

Nowadays, more and more companies use Information Technology (IT) systems in order to support their daily business processes. Evidently, organization are becoming more dependent on information systems to conduct their business operations. Starting from the late 1980s and the beginning of the 1990s, new software systems have entered the market, known as *enterprise resource planning* (ERP) systems.

The ERP is a software system for business management, characterized by different modules, each supporting specific functional units such as planning, manufacturing, sales, marketing, distribution, accounting, financial, human resource management, project management, inventory management, service and maintenance, transportation (Rashid, Hossain, Patrick, 2002). Therefore, the IT started to play an essential role in the execution of any kind of business process inside an organization, no matter the specific business or the external scenario in which it operates. These kind of systems enables companies to maintain the control over their business and the resources involved, and guarantees an extraordinary improvement in the way information are registered, collected and used. In fact, the structure of the ERP facilitates the transparent integration of modules, providing flow of information between all the functions in a consistently visible manner. The enterprise resource planning system connects the two poles of the supply chain, supporting the necessary back office activities (supplier-oriented) such as manufacturing, financial or inventory management, and the front-office ones (customer-oriented) such as sales, distribution and service application.

Nowadays, the majority of the big firms are supported by an enterprise system, in order to run their activities, and this trend seems to be in increase also among the small medium enterprises (SMEs). The increasing shifting from being predominantly "analog" to "digital" provides the basis for new ways to collect, categorize and store data. Therefore, the integration capability of todays IT systems opens new possibilities of data access and analysis. An ERP system, when used for conducting or supporting a process or a transaction, generates data. If properly managed, this digital information may be transformed into an incredible sources of value (Accorsi, Damiani, van der Aalst, 2014). The natural reference is to the phenomena of the "Big Data", which is based on the extraordinary growth of data in recent years.

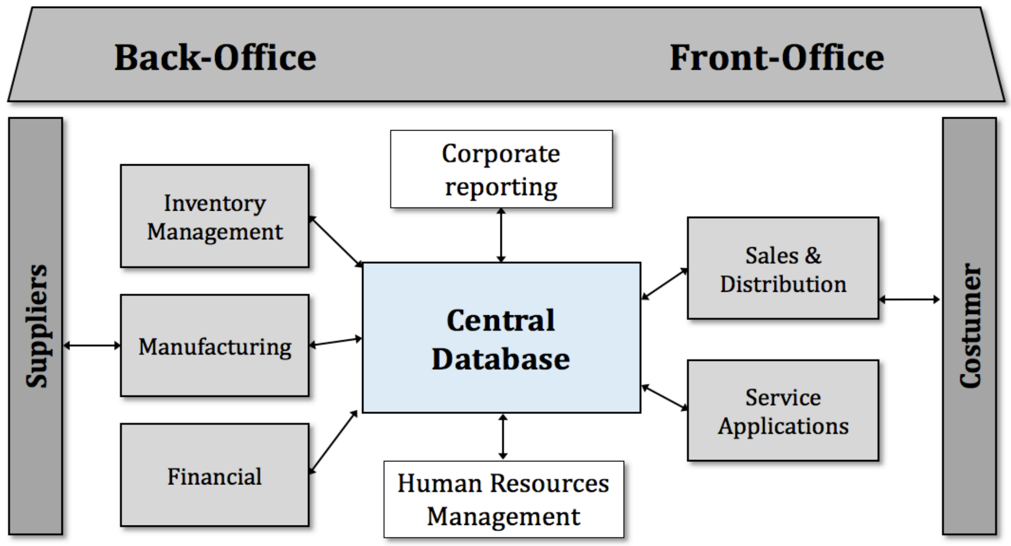


Figure 3.1: A framework for an ERP system

However, to make the concept clearer, let's take into account SAP, the most worldwide used ERP system vendor. SAP gains the 26% of total market share of the industry (Panorama Consulting Solutions, 2016), so it is an extremely representative instance to consider in order to understand which are the main functionalities of an ERP system. SAP is an integrate suite made of different modules, each specifically design to support a particular business process. Moreover, the suite is flexible and mouldable: a company rely on SAP according to their necessities, buying the specific tools that it needs. Even if SAP may ideally support a firm in almost every business tasks, the most important modules are those referred to accounting, production organization, analytics, logistic, purchasing and procurement.

Nevertheless, the prior objective in not to collect even more data, but to transform this information into real and tangible value for companies. This to the extend of improving existing processes and the services/products they support, or for enabling new ways to perform them. In general, the application of methodologies and tools which extrapolate insights from digital data, is the so-called *business intelligence* (BI), characterized by several approaches such as analytical processing (OLAP) and data mining (Kemper *et al.,* 2010). The first group refers to those tools which allows analysts to study multidimensional data using different operations of slice, dice or slip, while the second approach is primarily used for discovering new patterns among the large data sets.

Figure 3.2: An easy framework for SAP enterprise architecture solution (form Mesprosoft site)

Therefore, *event-related data* can be considered one of the most important source of information at a corporate level. Indeed, an event takes place every single time a generic task is performed with the support of an enterprise system. Just think, for instance, the creation of a purchasing order, or the goods checking-in at the warehouse (like in the case of Ford reengineered process in Chapter 1.3: Business Process Reengineering). Each of these tasks, once registered, automatically originates a data log. In this context, the techniques under the name of *process mining* may be extremely useful. Process mining aims to support decision making processes, by providing methods and tools for discovering, monitoring and improving real processes, extracting knowledge from this kind of event data, stored into companies' information systems.

3.1.2 The Required Data for Process Mining

According to the IEEE (Institute of Electrical and Electronics Engineers) Task Force, a group of professionals and researchers grouped with the specific aim of tracing the fundamentals of this new methodology, process mining may be defined as follows:

“Process mining is a relatively young research discipline, that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations.”

According to the quote (Van Der Aalst *et al.*, 2011), the source of the methodology is the *data log*, stored into companies’ IT systems. Moreover, the processes that process mining enables are defined as *automated* because they are supported by software applications that receive an event data as input and design, using different algorithms, a representation of the business process. This process model refers to the visual and structured representation of any process considered, such as the registration of a purchasing order by the resource involved in the process, the goods checking-in at the warehouse, etc. Indeed, according to IEEE Task Force, the prior output of process mining is a *structured view* of the process. Then, the discovered process models can be used for a variety of purposes, such as the *process analysis, process implementation*, and *process monitoring and controlling*, discussed in the previous chapter (Chapter 2.2: BPM Lifecycle).

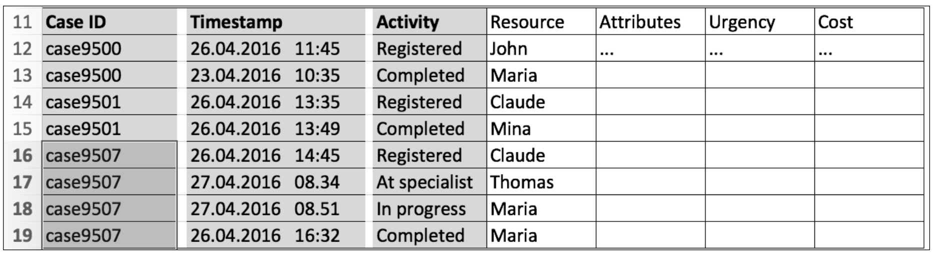
Even if this output is clearly an abstraction and a simplification of the real business process, it seems to enable insights and knowledge gathering for decision making that the complex reality of internal corporate environment often obscures. Therefore, process mining refers to all those techniques distilling structured process descriptions from a set of real executions, using event log data. (van der Aalst, Weijters, Maruster, 2004). In this sense, process mining can be considered as a new application in business intelligence whose objective is to learn processes and to gain knowledge from previously recorded actions. More specifically, process mining can be seen as the “bridge" between the data-driven approach of data mining (whose information are not process-centric) and the model-driven approach of BPM (where models aren’t constructed from real data and events stored inside IT systems) (da Silva, 2014).

Even if the idea of mining the set of activities in a context of workflow processes was firstly introduced during the late Nineties (Agrawal *et al*., 1998, Cook and Wolf, 1998), it is over the last decade that the research in this field has expanded greatly through very different disciplines and contexts. (see van der Aalst *et al.* 2003; van Dongen *et al.* 2005; de Medeiros, Weijters, Aalst, 2006; Greco et al. 2006; Gunther and van der Aalst 2007).

However, in order to perform process mining techniques on the event data, some mandatory requirements have to be contained into the traces used as input. These elements are essential because they enable algorithms mining the processes considered and discovering the real process model. Anyway, many other information may be stored into the log and could be exploited in order to make the analysis deeper.

The fundamental information required to perform process mining (Van der Aalst et al., 2003) are described as follows:

1. **Activity Trace**: each trace must contain an explicit reference to the activity or the task, related to a well defined step within the business process (such as “confirm order”);
2. **Case ID**: each event must be related to a specific instance - or case - made by one or more different activities, grouped together (such as “purchasing process”, or just an identification number like “ID12345”);
3. **Timestamp**: each event must be linked with the precise time in which it is executed, that is actually the time of registration into the information system. This requirement, with the initial assumption that events are totally ordered.



11	Case ID	Timestamp	Activity	Resource	Attributes	Urgency	Cost
12	case9500	26.04.2016 11:45	Registered	John	...	...	...
13	case9500	23.04.2016 10:35	Completed	Maria			
14	case9501	26.04.2016 13:35	Registered	Claude			
15	case9501	26.04.2016 13:49	Completed	Mina			
16	case9507	26.04.2016 14:45	Registered	Claude			
17	case9507	27.04.2016 08.34	At specialist	Thomas			
18	case9507	27.04.2016 08.51	In progress	Maria			
19	case9507	26.04.2016 16:32	Completed	Maria			

Figure 3.3: The required information to perform process mining

Sadly, especially among SMEs, it is not common to execute business processes according to a declared and explicit model; instead these companies perform their tasks without any formal structure, in a way that seems to be near to a pure and implicit sorting. So, the main problems that may occur are (La Rosa, 2015):

- *Lack of correlation* between the variables within the same process;
- *Absence of the timestamp* or inconsistency between the moment in which the event occurs and the activity registration by the resource;
- *Snapshots*, when event duration are not representative of the entire process, because it is too short or too wide;
- *Scoping difficulties*, because the relevant information may spread in more than one log events. It is essential to identify the purpose of the process first, in order to understand which systems have to be analysed and which logs have to be taken into account;
- *Granularity problem*, lack of the level of abstraction needed to obtain a useful view, from the conceptual point of view (for instance, there can be too many logs which are not significant for a process which is actually very simple and intuitive).

Even if such a premeditated model for data registration does not exist, the presence of a log of the tasks is very frequent. This because any information system that uses a transactional system will offer the three basic information in some form (Hornix, 2007). However, event log often contains also additional information. These could be, for instance, the name of the resource who physically executes the activity, the priority related to the process considered, information about costs or other attributes.

3.1.3 Guide Principles and Guidelines for Logging

At this point, it is useful to present a short framework for the log activity, proposed by the Process Mining Manifesto itself (Van Der Aalst *et al.*, 2011) in order to prevent users or analysts from making the most common mistakes. Following these *guide principles,* companies enable the opportunity to exploit process mining techniques by correctly extracting data logs stored in their ERP systems. The proposed guiding principles are:

GP1: *Event data should be treated as first-class citizens;*
GP2: *Log extraction should be driven by questions;*
GP3: *Concurrency, choice and other basic control-flow constructs should be supported;*
GP4: *Events should be related to model elements;*
GP5: *Models should be treated as purposeful abstractions of reality;*
GP6: *Process mining should be a continuous process.*

The first guiding principle aims to create consciousness, among users, that more focus should be pointed on the quality of event logs, rather than on their storage format. In this sense, event data should be considered as first-class citizen inside the IT systems supporting the processes that need to be analysed. The criteria to judge event data quality are various, such as the *trustworthy*, the *completeness* and the use of a well-defined *semantics.* Nevertheless, event logs should ensure the proper level of *privacy* and *security*.
The second principle concerns to spread the message that any process mining activities should be driven by questions. For instance, event logs should contain many information about the product/service case types: hence, analysts have to choose the type of cases to be taken into account before starting any kind of extraction and mining.
Then, the third principle refers to the fact that process mining techniques should support the main modeling elements such as parallel routing (AND-split/joint gateways), choice (XOR-split/joint gateways), or loops, provided by the other mainstream modeling languages such as Petri nets or BPMN (see Chapter 1.4: Process Modeling).
The fourth principle tells users that events have to be strongly related with the models. According to the authors, all the process mining techniques rely on the relationship between events and elements in the process model. Therefore, ambiguities about which event is related to a specific activity, need to be removed so that results can be interpreted properly.
Fifth guiding principle indicates that the derived process model obtained from log, should provide a transparent representation of the real situation, with a proper level of abstraction. Therefore, each map should be suitable according to the intended stakeholders and their back-grounds. In accordance to these assumptions, maps should be focused on different perspectives (strategic, operational, tactical, etc.), showing different levels of granularity and abstraction.
Finally, the last principle emphasises the need for a continuous focus on process mining, rather than seen it as a one-time activity. Giving the dynamic nature of processes, the objective is not to create a fixed model, but to encourage users to look at process models on a daily basis. Nevertheless, often historical event data may not be sufficient to trace real process model. Indeed, process mining techniques should be performed at runtime in order to provide actionable information related to different time scales (minutes, hours, days, etc.).

Moreover, besides the guiding principles listed before, other researches have theorized *guidelines for logging* specifically, in order to guarantee the most efficiency usage of process mining. Therefore, these advices are listed as follows (van der Aalst, 2015):

GL1: *Reference and attribute names should have clear semantics*, in such a way that the stakeholders could interpret data in exactly the same way;

GL2: *There should be a structured and a managed collection of reference and attribute names*, according to functional units, groups or hierarchic sets;

GL3: *Reference has to be stable*, and not to change according to time, location, language, scenario, or due to the resource considered;

GL4: *Attribute values should be as precise as possible*, and if some value is missing, it should be stated in an explicit way;

GL5: *Uncertainty should be described* through appropriate qualifiers;

GL6: *Events should be at least partially ordered*, explicitly or implicitly;

GL7: *Transactional information about events should be provided*, such as start event, suspend-resume events, end events, etc.);

GL8: *Perform regularly automated consistency and correctness checks,* for the purpose of ensuring the syntactical correctness of the event data;

GL9: *Ensure comparability of event logs* over time, among different groups of cases or process variants;

GL10: *Events used as input should not be aggregated before the analysis*, because data should be as raw as possible;

GL11: *Events should not be removed*, even if they mislead expected results;

GL12: *Guarantee privacy* without losing any meaningful correlations.

Surely these guidelines are very general and just want to improve the logging activity itself. The prior purpose of all these advices is to encourage companies and their resources to be conscious about the problems related to the input of process mining, and to support them in order to achieve the most value possible from the desired outcomes of their activities.

3.2 Process Mining Techniques

As mentioned above, there are various application of process mining techniques that, starting from the logs stored into IT systems, supports organization in different projects. In general, there are three main uses which allows to group all the applications into different categories. According to the Process Mining Manifesto (Van Der Aalst *et al.*, 2011) the fundamental applications of process mining are:

1. ***Automated process discovery***: a technique aimed to discover the structure of the process, in order to design a comprehensible process model (in the form of Petri nets or BPMN, social networks or organization charts, etc.);
2. ***Conformance checking***: highlighting process compliance and checking if reality, as recorded into the logs, conforms to the model and vice versa (via checklist deviations between the process originally designed and the real one);
3. ***Enhancement***: the aim is to extend or to improve a process, using the existing information about it, contained into logs. While the conformance checking aims to measure how the process is aligned with the model, the enhancement technique wants to change or extend the pre-existing model.

Once introduced the three different process mining techniques and their differences in terms of their application, it is necessary to point out their differences in terms of inputs and outcomes.

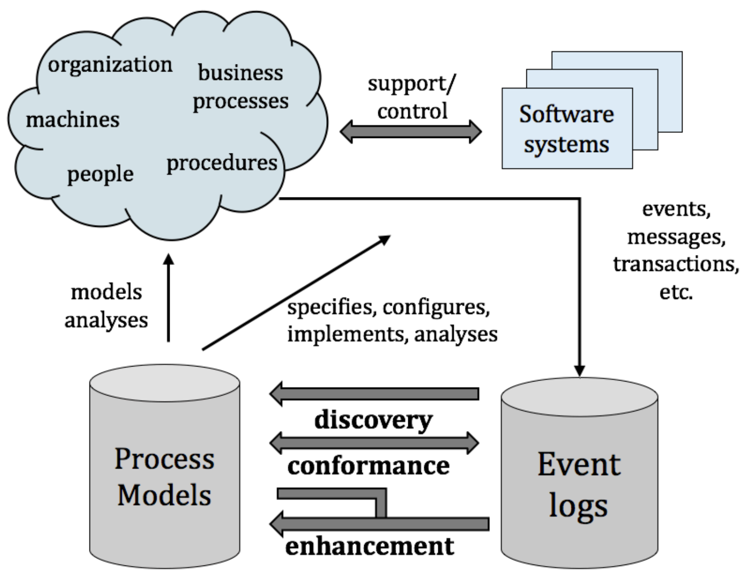


Figure 3.4: The process mining techniques (from Process Mining Manifesto, 2011)

Therefore, process discovery method takes an event log as input and generate a model exclusively based on that information. On the other hand, process conformance technique takes as input an event log or the related process model, in order to diagnose if they conform to each other. Finally, process enhancement also need an event log and a process model as input, but it give as outcome an improved or an extended model (Van Der Aalst *et al.*, 2011).

Moreover, the results of these techniques could be applied within four different perspectives: (I) *process perspective*, (II) *organization perspective*, (III) *case perspective and (IV) time perspective*. The *process perspective* is mainly focused on the control flow of the processes performed and related issues, like finding the order of tasks and obtaining a good characterization of paths in the form of a process model (though Petri-nets, BPMN, etc.). Instead, *organizational perspective* is interested on the human resources viewpoint and strive to understand which actually are the relevant stakeholders of the process, what their role and their responsibility are, and how they are related to each other (in the form of a social network). Finally, the *case perspective* is concerned with the properties of a single process case, while *time perspective* is focused on the respect of time and frequency parameters (Van Der Aalst *et al.*, 2011).

Nevertheless, the majority of the researches have confirmed the categorization in the three main techniques. On the other hand, other studies present some slightly differences. For instance, La Rosa (La Rosa, 2015) has identified four techniques. Even if the first two methodologies (automated process discovery and conformance checking) remained almost the same, two different types are proposed:

- *Performance Mining*: analysing process performance (by preparing reports or dashboards). This type of analysis can be conducted either in a static manner (by taking into account, for instance, a fixed idle time between two tasks), or in dynamic way (by repeating the analysis many times);
- *Variants and deviance mining*: tracing the differences, or discovering the root causes of process variation (with the support of statistical models or patterns of analysis).

Even if this categorization present different technique names, they are actually all comparable in terms of context. In particular, conformance checking could be seen as an extension of variants and deviance mining, while performance mining may be considered a technique employed

in order to gain enhancement.
According to this consideration, the present dissertation follows the original partition proposed by the Process Mining Manifesto into three fundamentals applications.

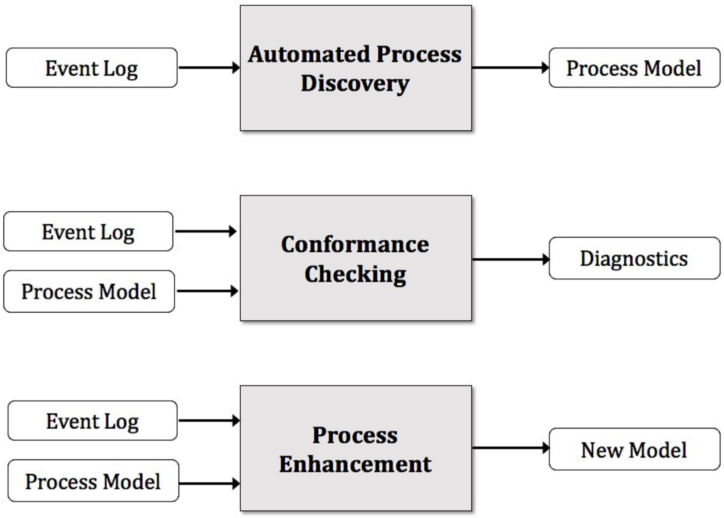


Figure 3.5: The three process mining techniques and their inputs and outputs

3.2.1 Automated Process Discovery

The prior objective of the automated process discovery algorithms is to construct a comprehensible process model which reflects the behaviour that has been observed in the event log stored inside the IT system. Control-flow discovery is often seen as the most exiting application of process mining techniques. Indeed, the real situations in which process discovery importance might be emerge are various:

- The entrance of a new CEO or a new General Manager within the organization. In the present scenario, board of directors often puts high expectation on the new entrant and demands to challenging results in the medium-short time. Thank to the automated process discovery, the manager is able to objectively understand the operational model of the company, with no need of numerous and time-consuming interviews;
- Merge and acquisition. Process mining automated discovery technique serves as a fundamental starting point to guide the re-organization and the change management plan. The possibility of enhancing data from heterogeneous systems, promotes the quick understanding of how processes are executed in all the companies involved;

- Consulting. As introduced, the relevant stakeholders often don't feel confident with the way processes are designed. Moreover, the trade-off between abstraction and representation might weaken the comprehensiveness of the solution (see Chapter 1.4: Process Modeling). Otherwise, automated process discovery methods can overcome classic modeling notations' limits, because the process designed is obtained exclusively from real data.

In this sense, automated process discovery can be extremely useful, especially during the BPM discovery phase (Chapter 2.2.2: Process Discovery). However, it is very important to remind that process mining discipline is not limited to process discovery, rather this is just one of the three main applications. Although the utility of this practice is not in doubt, it is important to notice that this technique may be very expensive and sometimes cannot justify such an investment for gaining a constant monitoring of all the company's processes (La Rosa, 2015).
Nevertheless, different algorithms have been introduced during the last 15 years (Rozinat *et al.*, 2007). Therefore, in the present research just the most important ones from an academic point of view and for their applicability in real organizational context, are briefly discussed. The aim of the present endeavour is not to depict a detailed description - in a mathematical and logical formalization - but exclusively to let audience understand the basic idea and structure of the process mining algorithm theory. If further consideration or additional notions are needed, readers might find the full list of papers about this topic in the final bibliography.
Anyway, the algorithms that will be taken into account are the *α-algorithm*, the *heuristic algorithm*, the *fuzzy miner algorithm* and the *multi-phase miner algorithm.* Each of these algorithms will be briefly described in the following paragraphs.

Alpha-Algorithm

The *α -algorithm* is the first process mining algorithm, introduced for the first time by Van der Aalst, Weijters and Măruşter in 2003 (Van der Aalst, Weijters, Maruster, 2003). Despite the interesting properties which can be proven around it and the great importance and influence from a scientific perspective, this algorithm seems to no work properly with real-life logs. (Rozinat, 2010). Anyway, an easy description within the present chapter is necessary to contextualize the overall phenomena of process mining techniques' spread. So, the aim of the alpha-algorithm is to examine the causal relationships existing between different tasks.

For instance, one particular activity may always be preceded or followed by another particular activity in every log trace, and this would be a very precious source of information. Basically, the α -algorithm is based on four fundamental relations, that can be easily derived from any kind of workflow log.
Assuming both A and B be events stored into a workflow log W, then:

1. $A>B$ if and only if there is a trace in W, in which the event A is directly followed by the event B;
2. $A \rightarrow B$ if and only if $(A>B) \wedge !(B>A)$;
3. $A \| B$ if and only if $(A>B) \wedge (B>A)$;
4. $A \# B$ if and only if $!(A>B) \wedge !(B>A)$.

The first relation A>B just points out the *direct ordered link* between the two events A and B. Then, the second relation A→B is the so-called *dependency relation*, in the sense that B depends directly on A in the workflow log W. Finally, the third relation A||B is the *parallel relation* that suggests potential parallelism among events, while the last A#B relation is the so-called *non-parallel relation,* which indicates that neither direct dependency neither parallelism between A and B are present.
In the process model based on the α-algorithm, the dependency relations are employed to connect events, while the parallel/non-parallel relations are used to detect the kind of splits/joins.

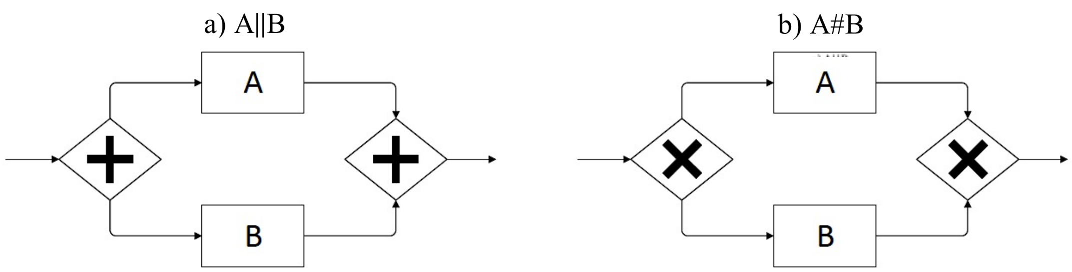


Figure 3.6: The representation of A||B and A#B relations in α -algorithm

However, this approach presupposes perfect information: in particular, (I) the log must be complete (in the sense that it should contain an example of every behaviour existing) and (II) it is assumed that there is no noise in the log (so that everything that has been stored in the log is completely correct and trustworthy).
As introduced in the previous paragraph (3.1.2: The required data for Process Mining), in real-

life situations logs are rarely complete and free from noise or data corruption. This represent the fundamental limit of process mining discovery methods based on this algorithm. Therefore, if an activity A is expected to be in the definitive model directly by B, it is mandatory that the relation A→B is observed at least one time in the data log. In addiction, it becomes very difficult, in practice, to decide if two events A and B are in dependency relation A→B or in non parallel relation A#B. For instance, the dependency relation between two tasks A and B (A→B) only holds if in the log there is actually a trace in which A is directly followed by B (A>B) and there is no trace in which B is directly followed by A (!(B>A)). Otherwise, in real-life situation characterized by abundance of data and noise, just one corrupt trace can completely mess up the derivation of a right relation (Weijters, Van der Aalst, 2003). Finally, it was proven that the algorithm does not allow any statistical analysis of the frequency of the activities (Burattin, 2013). For this reason, researchers and scientists tried to develop new heuristic mining techniques which are less sensitive for noise and incompleteness of logs.

Heuristic Miner Algorithm

The *heuristic miner* was the second process mining algorithm, closely related to the α-algorithm. It was developed by Ton Weijters (Weijters, Van der Aalst, 2003), with the aim of finding a new technique which is less sensitive for noise and the incompleteness of information. Basically, as the name suggests, the algorithm employs heuristics in order to limit the set of precedence relations included to the model (Günther, Van der Aalst, 2007).

Even if this algorithm maps processes in a notation specifically dedicated to process mining, its outcomes may be easily converted into other languages (Van der Aalst, Gunther, 2007). The heuristic miner derives XOR and AND gateways from the existing dependency relations among logs. Since it can be used also with exceptional behaviour and noise, the heuristic miner is suitable for many real-life situations (Rozinat, 2010).

In the heuristic mining approach three mining steps are identified: (I) the construction of *a dependency/frequency table*, (II) the induction of a *dependency/frequency graph* out of the dependency/frequency table, and (III) the *reconstruction of the workflow net,* out of the dependency/frequency table and graph previously identified (Weijters, Van der Aalst, 2003). As has been said, the objective of the research here is just to present a very high and basic description of the process mining algorithm's theory. Therefore, each phase will be briefly discussed above.

The first step of heuristic miner algorithm is the construction of a *dependency/frequency-table* containing different metric values. So, for each event A, the following metrics are identified:

M1. #A: the overall frequency of event A;

M2. #B<A: the frequency of event A directly preceded by event B;

M3. #A>B: the frequency of A directly followed by event B;

M4. $\$A \rightarrow^{L} B$: the strength of the dependency between A and B (local metric);

M5. $\$A \rightarrow B$: the strength of the dependency between A and B (global metric).

Even if metrics M1, M2 and M3 seem to be very clear, the last two metrics need additional explanation. Therefore, the M4 is more precisely defined as follows:

M4. $$\$A \rightarrow^{L} B = \frac{(\#A>B - \#B>A)}{(\#A>B + \#B>A + 1)}$$

Remarking that in this metric exclusively local information is used (i.e. the A>B relation) the result of this definition is that, if for instance event A is 9 times directly followed by event B, but the reverse relation never occurs, the value of $\$A \rightarrow^{L} B = 9/10 = 0{,}9$ indicates that even if we are quite sure about the dependency relation, noise could have effected the result anyway (for the remaining 0,1).

Finally, M5 is more global in the sense that not only the direct following events are involved. For instance, the metric takes into account the situation in which task B occurs shortly later then task A has occurred. The metric states it is plausible that it was actually the occurrence of task A, that causes later the occurrence of task B. So, if A occurs before B, and *n* is the number of the intermediary events between A and B, the $\$A \rightarrow B$ metric is incremented through a specific factor δ ($0{,}1 \leq \delta \leq 1{,}0$), raised to the *n* power. In this way, the factor's contribution is maximal when A and B are directly followed (*n*=0).

The second step is the induction of *dependency or frequency graphs*. This is based on the following rule, which formalizes the dependency relation A→B:

- IF $((\#A > B > \sigma)$ AND $(\#B < A \leq \sigma)$ AND $(\$A \rightarrow^{L} B \geq N_1)$ AND $(\$A \rightarrow B \geq N_2))$
  THEN A→B

The dependency relation A→B (B depends directly on A) occurs only when the 5 metrics are in relations with certain limit values (σ, N_1, N_2), which are sensitive for noise and parallelism in the workflow. However, considering that it is not mandatory to formulate a rule for each pair of event (because every not-first event must have at least one previous event, and every not-last event must have at least one dependent one), the so-called heuristic rule 1 is formalized thought the definition of a *dependency score* DS. In particular, assuming X and Y are events, the dependency score DS(X,Y) is formalized as follows:

- $\mathrm{DS}\,(X, Y) = \dfrac{(\$X \rightarrow^{L} Y)^2 + (\$X \rightarrow Y)^2}{2}$

Therefore, from the definition of the dependency score, the mining rule 1 is obtained as follows. Assuming a given task A, and X the event for which DS(A,X) = M is maximal. Then:

- A→Y if and only if DS(A,Y)< 0,95*M
- Y→A if and only if DS(Y,A)< 0,95*M

Remark that the new heuristic rule does not contain any parameter except for the threshold value of 0,95, which is appropriate and enough robust for noise and concurrent processes (Weijters, Van der Aalst, 2003). Then, heuristic rule 1 is extended with two simple heuristic rules for cases which present recursion (rule 2) and short loops (rule 3). Details are omitted.

Finally, the last step is aimed to apply the heuristic rules on a workflow log, in order to find the corresponding *workflow-graphs*. Anyway, the types of the gateways (splits and joins) are not yet represented in obtained the dependency/frequency- graph.

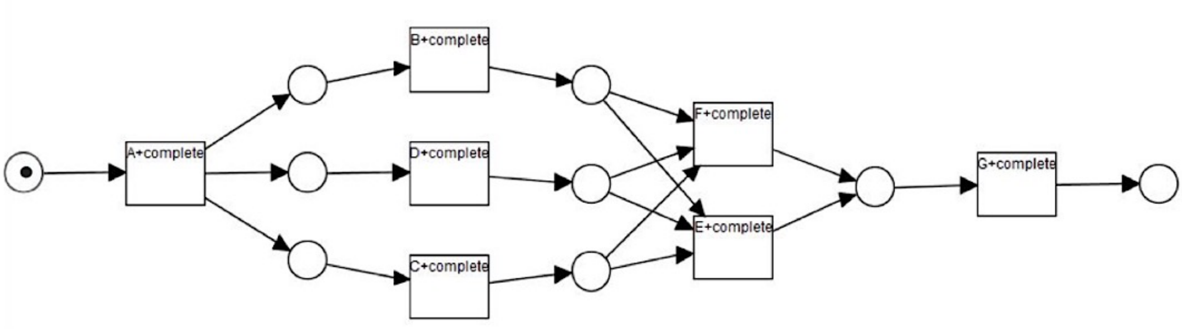


Figure 3.7: Petri net mined using the Heuristic Miner algorithm
(from TU/e Eindhoven University of Technology site)

Therefore, if two dependencies A→B and A→C are identified, it is necessary to find out if we are in the case of an AND or an XOR split. In this sense, if the resulting values are above a given limit value, the relation is considered to be an AND split, otherwise a XOR (Burattin, 2013). Indeed, information contained in the the *dependency/frequency table*, and the frequency of the nodes in the *dependency/frequency-graph,* provide the information needed to indicate the gateways' type (Weijters, Van der Aalst, 2003).

Multi-phase Miner

The *multi-phase miner* is the first algorithm which explicitly accepts the OR split/join semantics. This algorithm, developed by Boudewijn van Dongen (Van Dongen, Van der Aalst, 2004) uses EPCs (previously introduced at Chapter 1.4.3: Even-driven Process Chain) as a default representation, permitting to design complex behaviours in relatively high-structured models (Rozinat, 2010). Anyway, the EPCs representation can be converted in the other notations, such as Petri nets and its different typologies. One of the most important advantages of the algorithm is that it always produces a model that can reply the log. Indeed, it does not require any notion of completeness of a data log intended to work (Van Dongen, Van der Aalst, 2005). However, it is often hard to deal with process logs containing noise. Moreover, its application is rarely useful for mining excessive complex processes, because the model becomes difficult to read and to understand (Rozinat, 2010). In general, as its name suggests, the multi-phase miner algorithm is based on a multi step approach where (I) an *instance graph* is generated and (II) then all these models are *aggregated* into an overall process model.
Once the clustering is complete, a sort of tree containing all the aggregations is developed: every node is seen as the abstraction of its children, in such a way that different processes are abstracted into a common parent.

The first stage is based on the idea that every process case can be just described by a directed graph, called *instance graph*, rather than a sequence of events. These graphs can be seen as the process models of the single execution of a case (Van Dongen, Van der Aalst, 2004) and they may be very useful in order to identify parallelism within process execution.
Therefore, even if every process contains a set of specific tasks which have to be performed, or other information regarding rules (such as decision or frequency of execution), there is no need implementing every single kind of choice (such as every split or join in the representation).

Indeed, it is sufficient to use the instance graph, where each node indicates the execution of an activity.

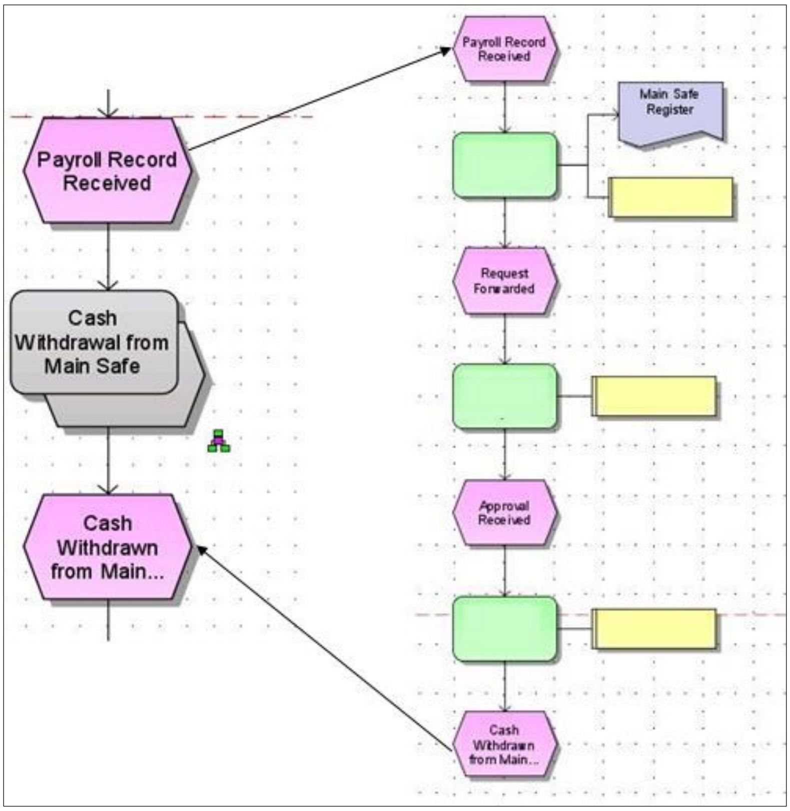


Figure 3.8: EPC Instance graph using the multi-phase miner algorithm
(from ARIS Community site)

Secondly, the final stage consists in the *aggregation* of the instance graphs obtained during the first step. This, because executives usually prefer to quantify performance measures in a more aggregated level. In addiction, looking at just one single instance, it may lead to an excessive fine-grained view. The second step makes it possible to aggregate cases just to describe a set of instances, rather than to identify just one instance. Basically, this result may be useful for monitoring performances, by grouping all the cases involving a particular characteristic, and to find out if this feature impacts on the overall indicators. For instance, the cases carried out by a predetermined amount of resources can be isolated, in a way to discover if these people influence in some ways the overall process. Therefore, the instance graph reveals causal relations existing among execution of tasks, while the aggregation graph shows the relation existing between the tasks themselves.

In conclusion, this approach may enable to generate a description of all the instances considered, that is the prior objective of process mining discovery techniques.

Fuzzy Miner

The *fuzzy miner* is another process discovery algorithm, developed by Christian W. Günther in 2007 (Günther, Van Der Aalst, 2007). Fuzzy miner is the very first algorithm directly addressed to solve problems of large groups of activities, in a context of highly unstructured behaviour. Indeed, this algorithm is particularly suitable for mining complex and unstructured data logs because is based on correlation metrics, which interactively simplify the process model at the desired level of abstraction (Rozinat, 2010). In addiction, with fuzzy miner it is possible to leave out insignificant activities from the intended analysis, or to hide them into clusters. This enables to properly manage the abundance of observed behaviours, which remained a fundamental weak point in most of the early process mining algorithms. Indeed, when logs intended to mine came from very less-structure processes, the final result is often just an unstructured and difficult to read representation. Even if this outcome, commonly defined as "spaghetti" process model, is not incorrect from a logical perspective, it does not provide any meaningful reflection of the real situation (van der Aalst, Gunther, 2007).
Firstly, two fundamental metrics are proposed for guiding discovery decision: (I) *significance* and (II) *correlation* (Günther, Van Der Aalst, 2007).

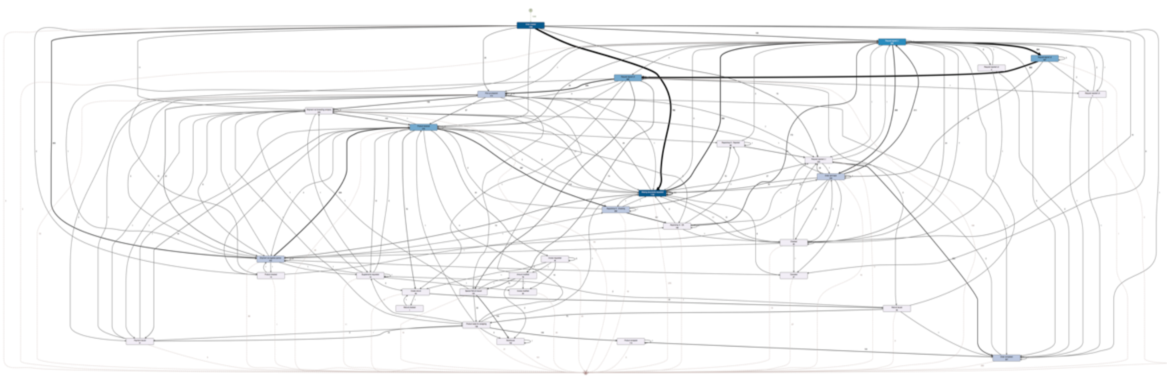


Figure 3.9: A representation of the "spaghetti" process model mined with Disco

On the one hand, *significance* measures the relative importance of a behaviour, both in terms of activities or precedence relations existing among them. In this sense, the most common and easy way to quantify significance is by measuring frequency.
On the other hand, *correlation* is relevant for precedence relations among different events, as it conveys how closely related two events, following one another, actually are. For instance, the identification of the amount of data attribute shared by events (in terms of timestamp or similarity among their recorded names), could be a method to quantify correlation.

Based on these two metrics, the simplification process is derived following this rules:

a. Highly significant behaviours are preserved;
b. Less significant but strongly correlated behaviours are aggregated in clusters;
c. Less significant and lowly correlated behaviours are abstracted or removed from the model.

Secondly, the approach is based on a framework for quantifying the two fundamentals metrics introduced before using new three types of metrics: 1) *unary significance*, 2) *binary significance* and 3) *binary correlation* (Günther, Van Der Aalst, 2007).
Each metric could be measured directly from the data logs or through the manipulation of other log-based metrics, in a derivative manner.
The first metrics group is the one of *unary significance*, which indicates the relative importance of an event class (represented as a node in the process model). Due to the fact that by removing a node, it implies deleting all the relative linked arcs, the unary significance could be seen as the prior driver for simplification. The metrics which made up the group are:

1.1 *Frequency significance*, according to the number of times that a certain event class is present in the log;
1.2 *Routing significance*, which is related to the number of successors (outgoing arcs) of a node;
1.3 *Amplifier metric*, useful for separating significant nodes from less important ones.

The second metrics family is characterized by *binary significance*, which represents the relative importance of a precedence relation among different event classes. The objective of these metrics is to isolate the behaviours which are supposed to be the most interesting. The metrics of this group are:

2.1 *Frequency significance*, related to how many times two nodes are observed after one another;
2.2 *Distance significance*, expressed by how the significance of a relation differs from its predecessors and successors.

Finally, the third group is related to the concept of *binary correlation.* The last family seeks for the distance of events within a relation. In this sense, the activities characterized by similar context (for instance, if they are executed the same day, or by the same resource, within the organization) are supposed to be higher correlated. Therefore, binary correlation is the fundamental driver for the decision between aggregation or abstraction of less significant behaviours (Günther, Van Der Aalst, 2007). The metrics that made of this family are:

3.1 *Proximity correlation*, which quantifies event classes occurring shortly one after another;
3.2 *Originator correlation*, determined by the similarity of resource names who have triggered subsequent events;
3.3 *Endpoint correlation*, expressed by the similarity of the activity names of two subsequent events.

Once the desired level of significance and correlation is obtained, the following step is the implementation of the results into a comprehensive graph.

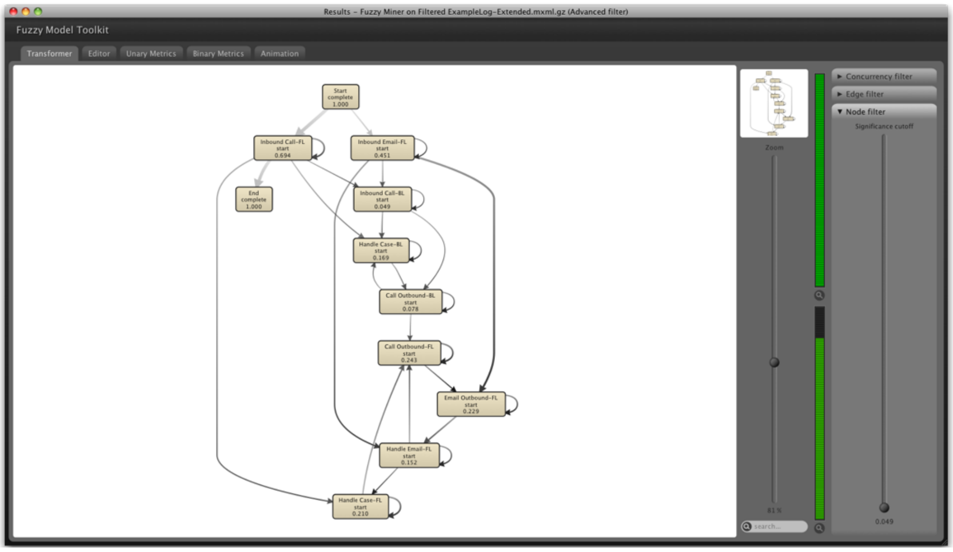


Figure 3.10: Process map mined using Fuzzy Miner algorithm in ProM
(from Fluxicon site)

Instead of following an interpretative approach like other process mining techniques (which aimed to map the behaviour using the classic AND or XOR gateways), the focus here is to propose a high-level mapping of data logs. Therefore, all the event classes are represented by activity nodes, while the precedence relations are implemented through directed edges, described by the binary significance and the correlation of the ordering relation. The importance of both the nodes and the edges is expressed by unary significance.

3.2.2 Conformance Checking

The second fundamental utilization of process mining techniques is the conformance checking, where an already existing process model is placed in juxtaposing with an event log representing the same process. The prior objective of this methodology is to compare an existing model with the data log extracted from the companies' IT systems, gaining the visibility on the discrepancies between modelled behaviours (from process models) and real-observed behaviours (from the data log) (Adriansyah, van Dongen, van der Aalst, 2011). In this sense, the information needed for the conformance checking phase are a (I) data log and (II) a process model, in a way that enable the comparison between the observed-real behaviour and what it is expected, discovering differences and discrepancies (Burattin, 2013). Therefore, the initial assumption in order to perform such a conformance is the existence of a process model able to depict the desired process, no matter which notation is used to represent it (Gehrke, Werner, 2013).
Nevertheless, conformance checking may be executed also to evaluate the quality of a mined model that has been automatically designed through process discovery algorithms, previously described (Adriansyah, van Dongen, van der Aalst, 2010).

In general, conformance checking phase is executed both for local diagnostics of specific processes, and related activities, and for assuring global conformance of the overall model.
Anyway, it is important to keep in mind that not every deviation identified should be seen as negative and must be eliminated. Sometimes deviations from the desired model may only mean that the model does not properly describe the real situations requirements and needs (Gehrke, Werner, 2013). Anyway, there are several real organization scenarios in which process mining conformance checking might be used:

- Standardization/harmonization programs. The possibility of working on end-to-end processes, promotes the rapid development of a cross-functional consciousness, needed to define effective and reliable harmonization and standardization programs;
- Compliance auditing. Conformance checking techniques might be extremely useful in order to gain a complete visibility on how processes are performed, highlighting the activities compliance (from the data log) with the procedures and policies imposed (from the process model that have to be followed).

Effectively, also taking into account process mining literature, it is possible to state that conformance checking is committed to be applied for two different scopes within a corporate context: *business alignment* (Rozinat, 2010) and *auditing* (Ghose, Koliadis, 2007). Each of these application will be discussed in the following paragraphs.

Business Alignment

The focus of *business alignment* is to measure the matching between the traces recorded in the data logs and the way processes are actually described within the process model. The prior aim here is to verify if the process model and the log are "well aligned" (Burattin, 2013). In general, the process model may be characterized by two distinctive natures (Rozinat, 2010): *descriptive* models, whose aim is just to extrapolate real processes characteristics without being mandatory, and *prescriptive* models which describe the way actual processes should be performed. In this situation, for both the natures, the focus is pointed on the identification and localization of deviations between logs and the process representation. There are three different quality dimensions that have to be identified in order to conduct the process conformance phase (Van der Aalst, 2008). Indeed, the dimensions needed for obtaining the desired business alignment are:

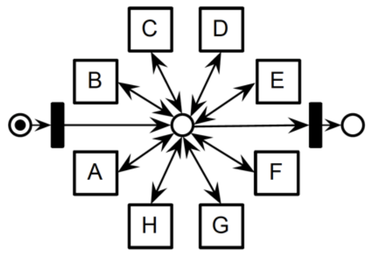
1. *Fitness*, which quantifies how much the observed behaviour is actually captured by the process model considered;
2. *Precision/Generalization*, which indicates how precise is the model in the sense of how is able to describe the situation;
3. *Structure*, which is addressed on the modeling language used, the richness of its vocabulary, the presence of duplications and the overall difficultness of understanding the way processes are represented.

In general, the identified quality dimensions seem to be orthogonal to each other. Indeed, every single characteristics is independent by others (Rozinat, 2010).

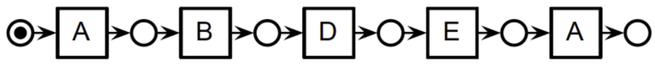
Fitness is committed to be the most important requirement for the process conformance phase (Rozinat, 2010). Basically, a generic event log and a process model fit if the model can represent each trace in the log, describing every sequence entirely. In other words, fitness evaluates whether every single log is an acceptable execution of the model. According to this perspective, a model has a perfect fitness only if all the traces in the log can be replayed by the model, from the beginning to the end (De Leoni, van der Aalst, 2013). Therefore, the easier way to quantify fitness is to measure how many times process model is able to replay the log and the mismatch between them, if present.

Log Traces	Instances
ABDEA	*1789*
ACDGHFA	*986*
ACGDHFA	*453*
ACHDFA	*54*
ACDHFA	*24*

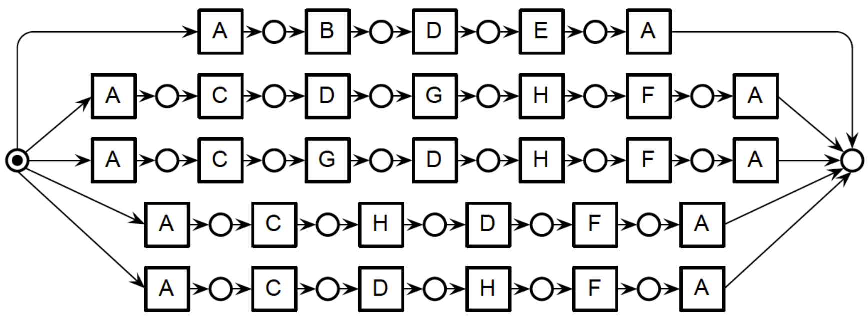
a) Table of Log traces



b) Model M1



c) Model M2



d) Model M3

Figure 3.11: Examples of log traces and process models in Petri net

Then, it become essential to localize the errors in order to provide some useful insights to the analyst. For instance, in reference to the examples in Figure 3.11, the process model M2 is just capable of reproducing the sequence ABDEA and no other of the traces present in the log. For this reason, its fitness is poor. However, a process model characterized by a good fitness often does not imply conformance, and that is the reason why the other two quality dimensions are required (Rozinat, 2010).

If the first quality dimension quantifies the capability of a process model to describe all the traces within the data log, *precision/generalization* strive to calculate the portion of behaviours which are acceptable but never executed. Indeed, when the model is too general and allows for more executions than necessary, it is not representative of the real processes. Excessive generalization will overload the process model with superfluous results and will allow for too much behaviours. Anyway, the first simple method to measure precision is to calculate the relation between the amount of possible execution cases and the ones actually present in the data. Considering again the instances in Figure 3.11, the model M1 is able to represent the execution of all the activities in any order. In this sense, even if its fitness is good (it correctly depicts all the traces contained in the log), its precision is poor because it allows all the other unnecessary behaviours.
Finally, the last quality dimension is the *structure*, which reflect the syntactic way in which the behaviour is expressed. In this sense, it might be useful to look for duplicate tasks, invisible tasks, and implicit actions. Taking into account once again the example in Figure 3.11, even if both precision and fitness are good for the model M3, its structure is low because of the presence of duplicated tasks which lead to a less comprehensiveness of the representation.

Audit

As introduced many times during the present research, especially during the last 20 years, companies have been starting to document their internal processes in some ways. There are several reason for this practice: to enable a better communication between manager and resources, to maintain some sustainable relationships with external stakeholders, to gain an ISO 9000 certification, to respect legal regulations, etc. All these applications can be grouped under the *audit* category. In other words, the conformance checking phase performed with the purpose of an audit tries to evaluate the current processes' executions with respect to *boundaries*, both *internal* and *external* (Burattin, 2013).

Thus, taking into account the *internal perceptive* it is important to notice that often processes are not entirely known by all the resources involved, especially for those cross-organizational processes which are made of different sub-steps. In particular, the compliance rules relevant for single steps necessary for running cross-organizational processes, imply different levels of responsibilities. For instance, the overall perspective of the communications among the owners of a cross-organization process (*interaction model*) is relevant for the management level, while the *local view,* derived from it, is addressed to a specific step involving the lower level of resources. The *public process model* takes into account the local view from the the interaction model, in order to depict the behaviour of the phase and to be representative from the viewpoint of the particular owner (Knuplesch, Reichert, Fdhila, Rinderle-Ma, 2013). Finally, each manager keeps his own *private process model* which include additional tasks not relevant for the the interactions with their pairs. Furthermore, these are not revealed to the other managers due to privacy matters. Anyway, all the presented model should conform with others. According to this representation, conformance between the different perspectives may be obtained through these two rules:

- *Local compliance rules*, imposed on the private process model;
- *Global compliance rules*, between the interaction model and the public process model.

To sum up this concept, it is useful to divide (I) public part from (II) private parts. Therefore, public parts include the interaction model, the public process models and the global compliance rules. On the other hand, private parts are characterized by private process models and the local compliance rules (Knuplesch, Reichert, Fdhila, Rinderle-Ma, 2013).

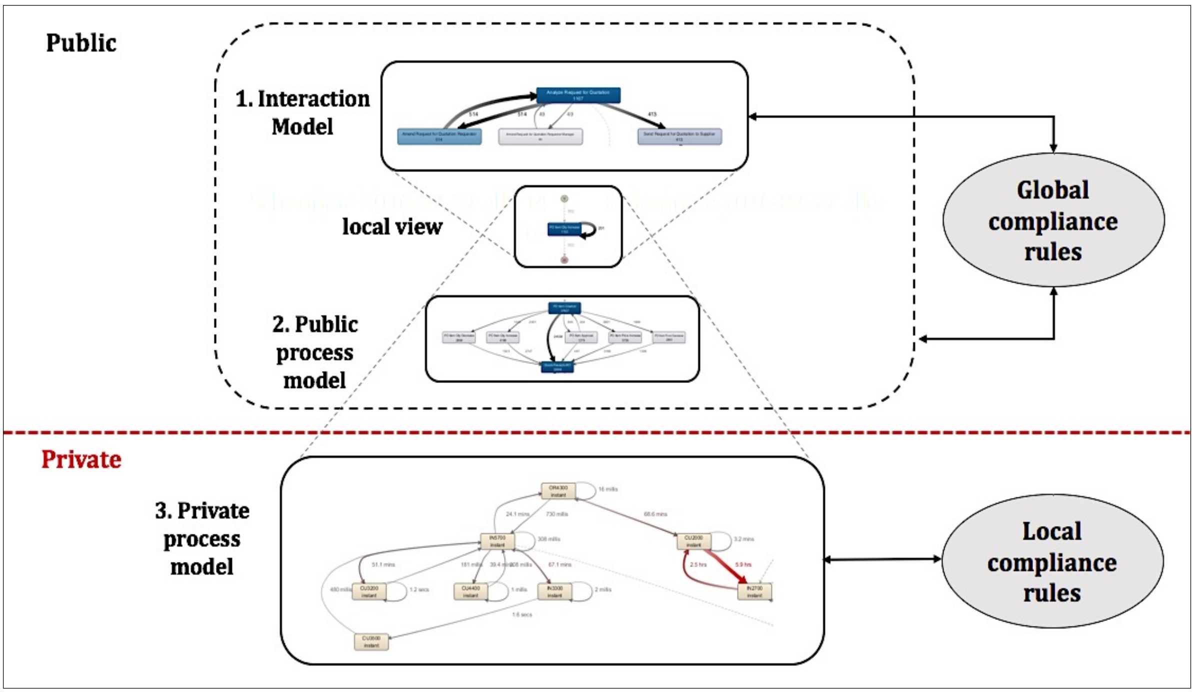


Figure 3.12: Compliance rules within a cross-organizational environment

Moreover, the second principal perspective of the process conformance phase is the *external* one. As introduced, compliance management has become an important matter for companies due to the increase of the number of regulations and laws. The regulatory environment imposes onerous and strictly requirements, so companies are forced to constantly increase investments in order to comply with them. In this scenario, firms not only have to manage problems related to the respect of legislative and regulatory frameworks, but they also have to consider the need to modify and to correct processes found to be non-compliant with the standards.

In order to obtain the desired level of compliance with these restriction, it might be useful to follow some heuristics for supporting the resolution of non-compliance in the process models (Ghose, Koliadis, 2007). The compliance framework, characterized by the mainly recurrent violations and the action needed to restore compliance, is summarized in the following Table 3.1, divided into structural patterns and semantic patterns.

Structural Patterns	
Activity/Event/Decision Inclusion: a generic activity, an event or a business decision inclusion may be defined without obligation or permission modalities ("permitted", "mandatory", "prohibited") or just based on route expressions.	**Resolution**: add or remove the activity. This may require coordination and assignment change to maintain the meaningfulness of the processes.
Activity/Event/Decision Coordination: a generic activity coordination can be serial, conditional, parallel, or repetitive.	**Resolution**: add or remove the activity. Otherwise, re-order existing activities.
Activity/Event/Decision Assignment: the assignment of an activity to a role is defined using logic operators. A statement that makes an action mandatory for some role may not preclude its assignment to other roles.	**Resolution**: add or remove the activity. Otherwise, re-assign the activity to another role.
Stakeholder/Resource Inclusion: the involvement of a stakeholder or the availability of a resource within a process can also be described using logic operators. A stakeholder's presence in a process model (for instance in the BPMN lanes) will indicate their involvement too.	**Resolution**: add or remove the process participant or the resource. This may require the addition or the removal of some others activities or interactions.
Stakeholder/Resource Interaction: the interaction between stakeholders and the transfer of resources may be governed by security or privacy matters.	**Resolution**: add or remove the stakeholder or the resource. Otherwise, add or remove the interaction or the transfer.

<table>
<tr><th colspan="2">Semantic Patterns</th></tr>
<tr><td>Effect Inclusion: an effect may be defined through obligation or permission modalities, to link it with a set of final or intermediate states of the process.</td><td rowspan="3">Resolution: to resolve the following compliance issues, these actions (or a combination of them) may be required: add or remove an action, add or remove an effect, re-assign an action, add or remove an actor, and add or remove an interaction.</td></tr>
<tr><td>Effect Coordination: the temporal relationship among effects of a process may be constrained.</td></tr>
<tr><td>Effect Modification: temporal rules may also refer to allowable changes upon intermediate effects within a process.</td></tr>
</table>

Table 3.1: Heuristics for compliance resolution (from Ghose, Koliadis, 2007)

3.2.3 Process Enhancement

The last process mining technique is the so-called *process enhancement*. Here, the prior aim is to extend, to enrich and to improve existing process models with information obtained from the event logs. From the previously discovered process models, it is possible to conduct strong analysis for obtaining performance indicators such as average idle times, cost of improving or changing a process, percentage of items that need to be reworked, etc. Again, the main advantage is related to the fact that this kind of analysis uses reliable, tangible and credible data as only sources. Therefore, this kind of optimization is extended to support processes both post-runtime inspections, using the precious insights abstracted form the manipulation of historical data, and processes under execution for the behaviour prediction of traces, by comparing actual instances with similar already executed. Finally, process enhancement techniques can be applied to derive information for redesign of processes, before they are implemented (Gehrke, Werner, 2013). Evidently, process mining enhancement techniques can be used in various real business applications:

- Identifying inefficiencies in the process mining processes. Here, process enhancement allows analysts to extract, through statistical models, particular patterns and to identify the root causes of recycles, bottlenecks or exceptional waiting times. In addiction, the possibility of exploit data at runtime through process mining methods enable firms to gain a better visibility of the execution of their reengineering or changing projects, becoming one of the main drivers to measure and to monitor performances;

- Improve the system of indicators used. Sometimes the *key performance indicators* (*KPI*s) used for measuring processes are not properly designed to highlight the company's performances. For instance, just think the case where the result in terms of provision of an office (and the consequent rewarding system) is exclusively based on the number of orders executed. In this situation, the resources would be more likely to increase their productivity in order to meet this objective, neglecting the other parameters not taken into account by the *KPIs* system used, such as the completeness of the document. Process enhancement techniques can correlate different performance metrics (in the specific case, the number of orders processed and the percentage of reworked ones, due the lack of completeness), in a way to maximize the effectiveness of the internal measurement system.

According to this instances, process enhancement is particular suitable during the process redesign, implementation, monitoring and controlling phases of the BPM lifecycle described in Chapter 2 (Chapter 2.2: The BPM lifecycle).
Nevertheless, due to the fact that process enhancement is strictly related with the concept of performance improving, it makes sense at this point of the research to examine more in depth which are the main performance dimensions of a company and how they gain advantages from process mining methodology.
In general, there are committed to be four prior process performance measurement, also called *key performance indicators* or *KPIs* (also mentioned in Chapter 1.2: Business Process Reengineering), which are (I) *time*, (II) *cost*, (III) *quality* and (IV) *flexibility.* By determining these values, it is possible to measure - or to gain knowledge from - how good business processes are carried within the company (Jansen-Vullers *et al*., 2007).

Time is the first process performance dimension type, and it is often related to the strive for reducing organization's throughput time (also know as cycle time). Besides cycle time, there are several other time-related dimensions, such as waiting time (related to the fact that no resource is available at the moment or because of synchronization), processing time (precisely the amount of time resources actually spent on handling the task), queueing time (caused by the waiting of items among machines, set-up time, etc.).
The second key performance indicator identified is *cost*. Even if this concept might be related also with increasing revenues or turnovers, it is generally addressed with the strength for reducing costs. It is essential here to distinguish between fixed cost and variable cost because

they are related to different issues. Fixed costs are costs which are independent from the production volume, whereas variable ones vary with the level of the output. Due to that, a fixed cost seems to be not possible to put into a direct correlation with a business process while a variable cost is directly correlated with some variable process quantities.
Then, the third performance dimension is *quality*, which can be observed from two different angles: from the internal side (process participant viewpoint) and from the external side (customer viewpoint). From the one hand, internal quality can be expressed with how process participants fell in control with the business processes they handle. Indeed, it can be seen as the quality of a workflow from an operator's perspective. On the other hand, external quality is often related to customer's satisfaction, which reflected if their expectations are met by the delivered product or service.
Lastly, *flexibility* reflects company's ability to react quickly to changes and to the challenges invoked by the business scenario. Accordingly, flexibility metrics quantifies the capacity of a process participants to execute various tasks and to handle different cases, organization's ability to change its internal structure, and process responsiveness to react against external stimuli.

As with the internal and external BPM dimensions (Chapter 2: A Business Process Management Framework), also this four dimensions may be described within a framework in which their mutual properties can emerge. In this situation, the framework to be considered is the so-called *Devil's Quadrangle* (Mansar, Reijers, 2007).

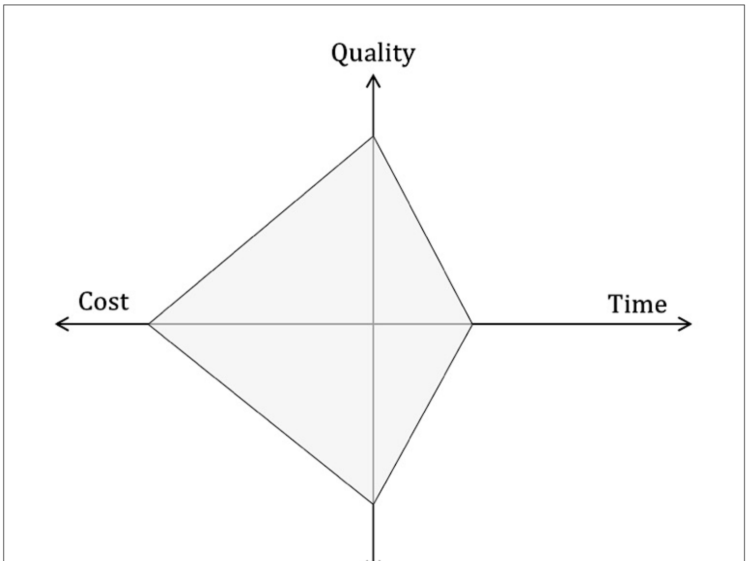


Figure 3.13: The Devil's Quadrangle

In general, a business process enhancement, by means of re-designing processes, is committed to decrease the time needed for handling a case, while it should lower costs related to their execution. Moreover, this should improve the quality of the product/service delivered to the costumer, make process ready to properly manage changes and to react according to the external scenario variations. Therefore, also taking into account the main KPIs dimensions, company should put their strength - both financial or human - in the direction of improving the desired dimension, without causing any loss in the results of another one.

As seen, the possibility of exploit information stored into event logs enables organization to reach a deepen level of understanding of their internal processes. This permits a stronger manipulation of their performance measures and guarantees a better control of the relevant dimensions. In this sense, process mining provides many techniques supporting companies to enhance business processes in each of the specific dimension identified. Accordingly, in the remaining part of the paragraph, each one will be described.

Dimension 1: Time

As evidenced in the beginning of this chapter (3.1.2: The required data for Process Mining), the presence of a timestamp is committed to be one of the three basic requirement to conduct process mining (Van der Aalst *et al*., 2003). Therefore, a trace directly addressed to the time of registration of the task is very frequent (Hornix, 2007). Timestamp indicates the precise moment of registration of the activity in the information system, which usually refers to temporal occurrence of the related event. Moreover, other information such as case ID and activity trace are necessarily contained into logs, while others like the resource name might be present. By exploiting these data, analysts could implement *dotted charts* which are diagram characterized by the time occurrence in the first axis, and another value (such as case ID, resource name, etc.) in the second axis. This very simple tool provides for a powerful visualization of processes' time occurrence, and may enable important evidences, cases grouping and enhancement possibilities (Process Mining Group, 2009).

Another time-related tool is the *timeline chart* of logs, which graphically depicts the duration of the activities that made up the process. Evidently, this is possible only when additional information about the end time of activity is present, or when is it feasible to reconstruct the sequence by using case ID. This technique helps analyst to understand which are the activities that require more time for their execution and to underline delays or time-related inefficiencies.

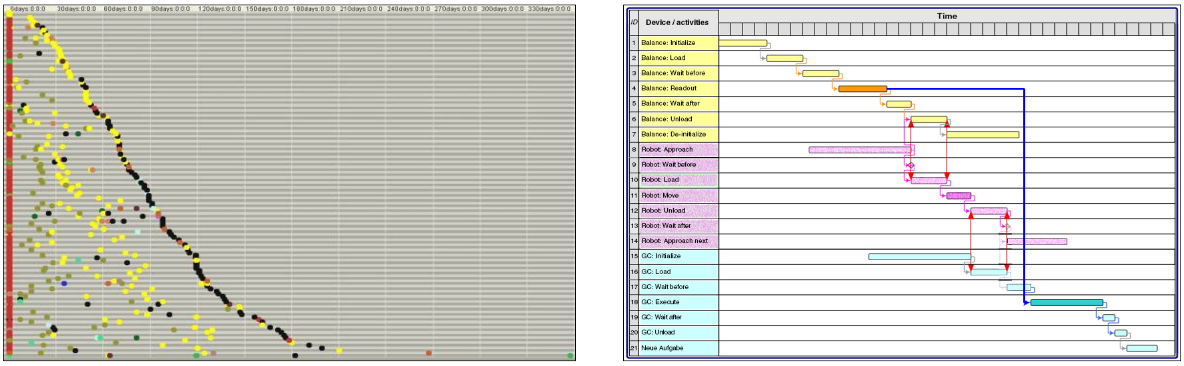


Figure 3.14: Examples of a dotted chart and a timeline chart of logs
(online available at processmining.org and labautopedia.com

Dimension 2: Cost

Even if cost-related information might be present within the event log (La Rosa, 2015), it is not common among companies to provide such a level of details in their information systems. Therefore, analyst have to use specific techniques in order to gain the needed consciousness of costs correlated to their internal activities. In addiction, the issue related to the correct assignment of indirect costs represent the main problem (Jansen-Vullers *et al.*, 2007). In fact, direct cost can more easily be addressed to the interested activity.

Due to that, the accounting philosophy of the *Activity Based Costing* (*ABC*) seems to help assigning these indirect cost properly. The basic idea under ABC accounting is to use the activities as starting point for allocating indirect costs. Then, this method just requires the presence of the start time of activities to be executed, which often is the precise time in which resources start handling the specific task. Anyway, it is necessary to consider the additional costs for achieving such a transparency. Nevertheless, even if it might be more onerous than other accounting techniques, it could be more precise. (Bartezzaghi, Spina, Verganti, 1999).

Dimension 3: Quality

Despite the fact that in general the quality of a product/service can be not easily determined from the information stored into information systems, event logs provide companies some detailed data that are relevant to quality measurement. Process mining can give analysts good indication about the presence of repetitions of tasks (called also reworks or re-loops). Indeed, such repetitions could be easily found within the sequences of activities. For instance, an approximation about the *repetition probability* r of a generic task can be calculated from a general series of event logs. Assumed CT as the complete cycle time of a task (the time that the resource spends on actually handling the case, comprehensive of possible rework times) and T the time required to execute the task only once, r is calculated through the following formula (Dumas *et al.*, 2013):

$$r = 1 - \frac{T}{CT}$$

For instance, giving the following execution times for task x:

1) 3 minutes, 7 minutes;
2) 7 minutes;
3) 4 minutes, 15 minutes, 6 minutes

CT is obtained from the average execution time of x per case CT = (3+7+7+4+15+6)/3 = 14 minutes, while T is calculated from the average execution time of x per occurrence T = (3+7+7+4+15+6)/6=7 minutes. According to this example repetition probability r would be $r = 1 - \frac{7}{14} = 0{,}5$.

Dimension 4: Flexibility

Finally, also the level of variation allowed by the process can be extracted by analysing event logs. This permits companies to put into relation the desired degree of flexibility with the real one and to improve their performances in a standardization/adaptability perspective.
Considering a generic event log, it is assumed that if two executions contain exactly the same sequence of activities, they will result in a single trace included in the workflow log. In this sense, workflow log contains traces having a unique sequence. By determining the quantity of *distinct executions* and the *degree of optionality* within a workflow log (calculating how many tasks not occur in at least one trace) analyst could assess the flexibility of the process taken into account (Dumas *et al.*, 2013).

Chapter 4 Case study

4.1 Introduction

The present chapter provides an application of process mining techniques within a practical business case. The case study involves a large Italian application infrastructure provider (with more than 700 employees in Italy) and HSPI Management Consulting who carried out the overall project. According to non disclosure agreement, the firm object of the case study will not be mentioned in the present dissertation. Therefore, from now on it will be named as *X-Provider*. The pilot project's prior aim is try to understand and to quantify which are the real benefits that could derive from process mining application.

4.1.1 About the Company

X-Provider business model is dual oriented: on the one hand it acts as a calculation centre hub providing costumers with strong electronic and automatic elaborations, while on the other hand it offers and manages its client's IT infrastructure and technological services.
Therefore, the organisation provides its customers (both the industrial sector and the public administration) with the computing resources and infrastructure for developing, deploying and managing enterprise-class applications. In addition, it offers services such as hardware maintenance, operating systems and management software, clustered servers, scalable storage, monitoring, physical or logical security. Generally, all these services are executed under a well defined *service level of agreement* (*SLA*) with clients. The underlying infrastructure can be accessed remotely over the Internet or within a secure virtual private network (VPN). In general, the physical resources are directly committed to a client who has complete control over the base operating system (OS) and the installed applications. Anyway, the infrastructure can be also shared in a multi-tenant architecture.

Nevertheless, X-Provided has recently implemented a new *issue tracking system* (ITS), which supports the firm in handling a portion of the ticketing workflow. It is important to immediately point out that tickets are not generated exclusively by customers who demand for X-Provider's assistance, but also by the company itself when some particular issue arises. In this situation, some tickets that have been created internally or externally are managed with the new issue tracking system, whereas other tickets are governed by the old one. Therefore, if the process mining pilot project proves the effectiveness of the new ITS, X-Provider would think about updating the overall management system of the tickets.

4.1.2 Goal and Scope

The goal of the pilot project is to demonstrate the useless of the process mining application within X-Provider's internal processes, in order to develop a tangible business case for then testing the utility of process mining in one or more clients' core processes. Indeed, X-Provider's long-term goal is the development of a new service for its customers able to measure, on a historical basis and in real time, the performances of those business processes which are supported by X-Provider's infrastructure and applications. Simultaneously, the objective is to discover the eventual advantages gained from using the new ITS for the selected portion of tickets, and to understand if there are eventual decline in the performance of those tickets still handled with old ITS.

The pilot project is carried out specifically on a ticketing process related to the technology support services that X-Provider delivers to its customers. It is appropriate that all the identified issues (for instance, ticket's higher execution time, or its higher variance) would be communicated in order to convey the analysts to identify the root-causes of such problems. The time horizon corresponds to nine months between 2015 and 2016 (from the beginning of 07/2015 to the end of 03/2016). The specific scope is defined on the basis of the available data, that are the information obtained from one or more applications used for the opening and the managing of tickets.

4.1.3 Activities and Deliverables

The pilot project is scheduled into four work packages which invoke the presence of various management levels and different degrees of commitment. In brief, the work packages' structure is organized as follows:

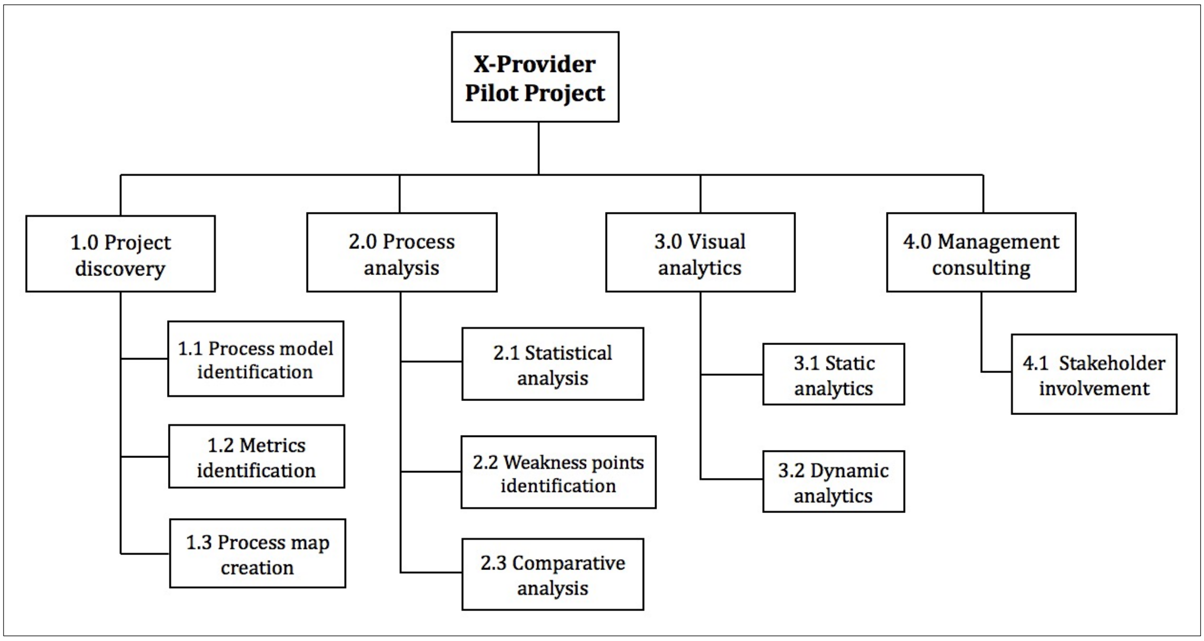


Figure 4.1: Work Breakdown structure of the pilot project

WP1: Process Discovery

- Automatic process discovery of the existing model, exploring different levels of abstraction (for instance on the basis of number/frequency of activities, or for specific variants such as slow/fast cases, compliant/non-compliant cases, etc.);
- Determine process metrics (e.g. time, frequency, and count metrics);
- Drawing a comprehensible diagram for the model identified with all the paths for presentation and comprehension purposes.

WP2: Process Analysis

- Providing statistics about the performance of the ticketing process (determine the most used paths);
- Identifying the weak points of the process (such as bottlenecks on states of the tickets or on specific resources, abnormal patterns in the workflow such as *ping-pong effect* of a ticket between resources, or the rollback of results previously achieved);
- Comparative analysis between the different process variants identified (for example fast variant vs. variant that suffers from slowdowns, compliant variant vs. variant that violates the SLA).

WP3: Visual analytics

- Static analytics: the provision of visual performance metrics (such as charts, time series, cumulative flow diagrams, etc.);
- Dynamic analytics: to obtain an iterative animation of data of the process models identified.

WP4: Management consulting

- Analysis of the results with the relevant project's stakeholders, supporting the preparation of a business case with the customers identified.

4.1.4 Considerations

If the pilot project works successfully, X-Provider will decide to extend the application of process mining techniques to other internal processes. In this situation, the economic value for carrying out the necessary activities are assumed to not to scale linearly, indeed:

- As the team involved become more familiar with the data and the analysis dimensions, the additional time required will be reduced significantly;
- Software applications might be developed in such a way that the exactly logging of data, required for process mining execution, is allowed;
- Progressively fewer support to internal and external communication will be required.

Therefore, unless subsequent activities do not face additional problems (such as the analysis of data gaps and noise), both effort and economic amounts will be reduced.

4.2 Pre-mining Phase

4.2.1 The Ticketing Process

The process taken into consideration is the ticket handling process, as registered into X-Provider information system software. The process started every time a ticket, which represents a generic request for managing an issue, is opened. The opening act of a ticket is the trigger of the process, which can be executed both externally by a customer and internally by a X-Provider's resource. In any case, an opened ticket triggers the process whose aim is to handle the request, solve the issue and close the ticket in accordance to the SLA projected with the customer (internal or external). The *issues* which represent the triggers of the process are four:

1. *Change*: an activity characterized by a change is requested;
2. *Help*: some kind of assistance is required;
3. *Fault*: an error has arisen;
4. *Task*: a particular internal issue has arisen.

Each issue originates a ticket with a unique case ID, that can be distinguished by different priorities chosen, activities or resources involved. Indeed, each case is characterized by a *priority* which can be defined as:

- *Low*: the lowest level of priority possible;
- *Medium*: the medium degree of priority;
- *High*: the high level of priority;
- *Very High*: the more than high grade of priority;
- *Red Code*: the maximum degree of priority, for special cases only.

Then, a ticketing handling process is characterized by different activities, which are actually the states enabled in accordance to the satisfaction of pre-defined events. The *activities* that made up the process are:

- *Created*: the state represents the creation of the ticket, related to a generic issue;
- *Waiting for Triage* (WFT): the state represents the creation of a ticket using the new ITS. Its value is totally equal to the "created" activity for a group tickets;

- *Waiting for Support* (WFS): the state indicates that an X-Provider's operator has started to handle the ticket;
- *Waiting for Customer* (WFC): the customer who had created the ticket is called for validate the support or to do something;
- *Waiting for Internal* (WFI): an X-Provider's specialist is called through an escalation because the operator involved in WFS in not able to resolve the issue;
- *Waiting for Approval* (WFA): after the escalation, a special approval is required;
- *Resolved*: the issue has been resolved by X-Provider;
- *Closed*: the ticket has been closed.

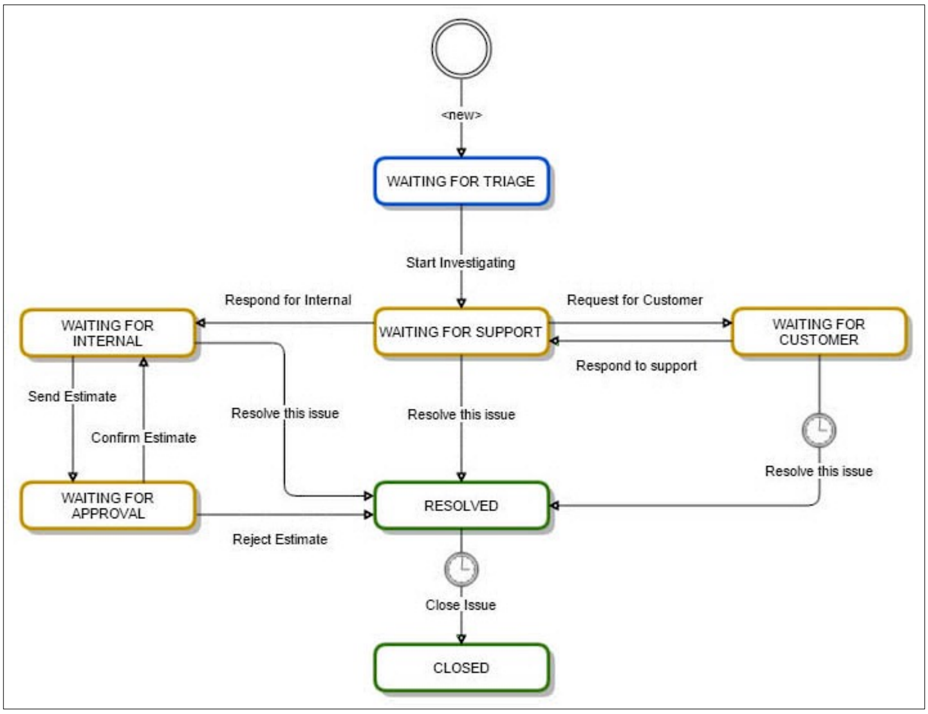


Figure 4.2: A framework for the ticketing process

Then, each log is related to a well defined *project*, which indicates the scope of application of the case. In particular, through the project's name is it possible to discover a portion of cases that were handled with the new ITS.
Finally, the cases are associated to a *customer*, that can be internal or external.

4.2.2 Data Cleaning

Firstly, X-Provider was asked to share some random data stored into its information system, in the form of CSV file or Extensible Event stream (an XLM-based standard for event logs), with the aim of understanding which ones were really pertinent to the intended analysis. The provided raw information's objective was to act as a *proof of concept* for the purpose of creating a mutual consciousness between the firm and the analysts involved in the project about which kind of data were needed to execute process mining techniques. In addiction, this initial proof of concept phase was necessary for X-Provider's specialists to progressively refine data before the mining stage. Furthermore, the acceptable degree of granularity was reached and the almost ready to use raw data were provided.

The raw event data processed derived entirely from X-Provider's central information system, which collected all the transitions related to the workflows of all the tickets opened. In particular, the data referred to a time-scope of nine months, between 07/2015 to the end of 03/2016 and counts an amount of over 75000 logs. The initial distribution in frequency of logs - both absolute and relative - among *issues*, *priorities, customers* and *projects* is summarized in the next table.

		ABSOLUTE FREQUENCY	RELATIVE FREQUENCY
ISSUE	Change	38071	50,52%
	Fault	26849	35,63%
	Help	10351	13,74%
	Task	82	0,11%
	Tot.	*75353*	*100,00%*
PRIORITY	Low	26230	34,81%
	Medium	25420	33,73%
	High	11968	15,88%
	Very high	6874	9,12%
	Red code	4861	6,45%
	Tot.	*75353*	*100,00%*
CUSTOMER	X-Provider	37659	49,98%
	External	30371	40,30%
	ND	7323	9,72%
	Tot.	*75353*	*100,00%*
PROJECT	New ITS	20067	26,63%
	Old ITS	5926	7,86%
	Others	49360	65,51%
	Tot.	*75353*	*100,00%*

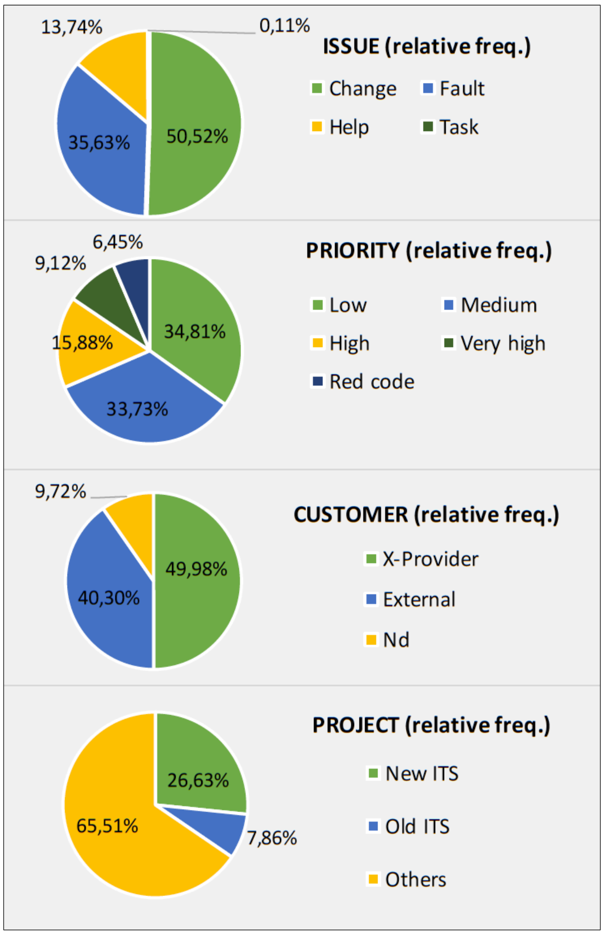


Table 4.1: Initial distribution of logs' frequency among issues, priorities and projects

However, event logs still provided too many information which were not pertinent for the expected study. For this reason, additional refining and retuning were needed. In particular, a consistent part of the logs captured within the observed time interval contained information about tickets which had been opened before the intended timeframe. On the other hand, some other tickets that had been correctly opened within the time interval, were not resolved and closed by the end of the window considered. In other words, original raw data provided a lot of information about incomplete tickets' lives. Therefore, the dataset was trimmed so that only the complete tickets (both opened and closed within the timeframe) were taken into account. All the other not relevant information was maintained, nevertheless the redundancy.

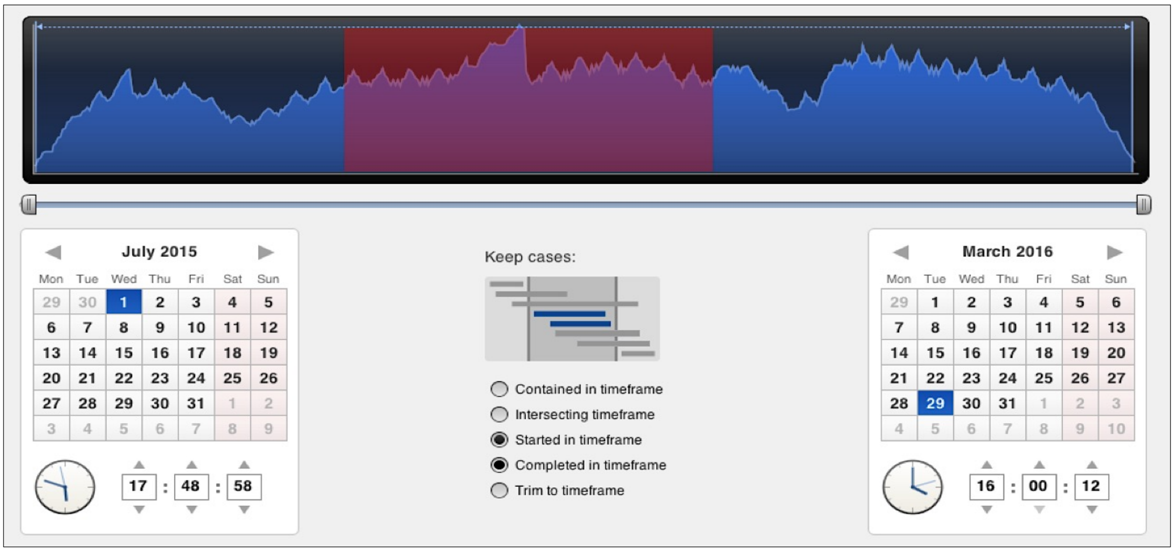
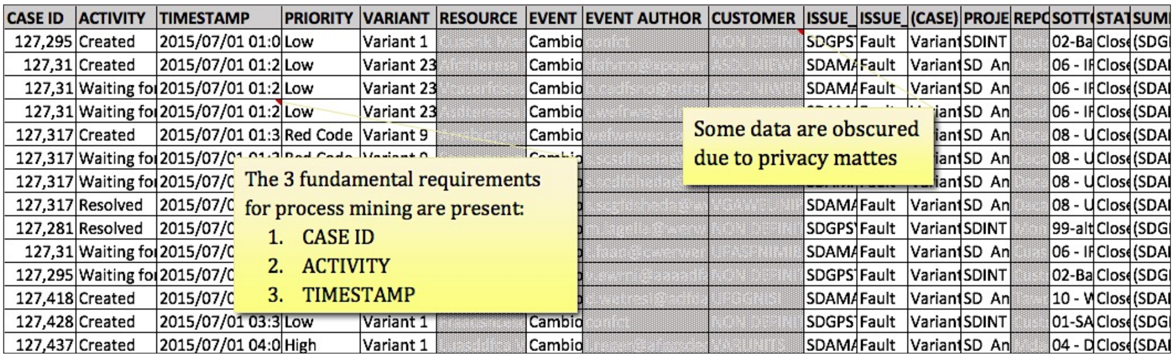


Figure 4.3: Data filtered by timeframe via Fluxicon Disco

The filter reduced the logs to approximately 16000 logs. Nevertheless, the obtained data "cleaned" contained all the three fundamental information required to perform process mining described in the previous chapter, namely reference to *case ID*, *activity* and *timestamp* (Chapter 3.1.2: The required data for Process Mining), so they were ready for process mining.



CASE ID	ACTIVITY	TIMESTAMP	PRIORITY	VARIANT	RESOURCE	EVENT	EVENT AUTHOR	CUSTOMER	ISSUE_	ISSUE_	(CASE)	PROJE	REPC	SOTT	STAT	SUMI
127,295	Created	2015/07/01 01:0	Low	Variant 1	[illegible]	Cambio	[illegible]	[illegible]	SDGPS	Fault	Variant	SDINT	[illegible]	02-Ba	Close	(SDGI
127,31	Created	2015/07/01 01:2	Low	Variant 23	[illegible]	Cambio	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	06 - IF	Close	(SDAI
127,31	Waiting fo	2015/07/01 01:2	Low	Variant 23	[illegible]	Cambio	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	06 - IF	Close	(SDAI
127,31	Waiting fo	2015/07/01 01:2	Low	Variant 23	[illegible]	Cambio	[illegible]	[illegible]	[illegible]	[illegible]	iant	SD An	[illegible]	06 - IF	Close	(SDAI
127,317	Created	2015/07/01 01:3	Red Code	Variant 9	[illegible]	Cambio	[illegible]	[illegible]	[illegible]	[illegible]	iant	SD An	[illegible]	08 - U	Close	(SDAI
127,317	Waiting fo	2015/07/0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	iant	SD An	[illegible]	08 - U	Close	(SDAI
127,317	Waiting fo	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	iant	SD An	[illegible]	08 - U	Close	(SDAI
127,317	Resolved	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	08 - U	Close	(SDAI
127,281	Resolved	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SDGPS	Fault	Variant	SDINT	[illegible]	99-alt	Close	(SDGI
127,31	Waiting fo	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	06 - IF	Close	(SDAI
127,295	Waiting fo	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SDGPS	Fault	Variant	SDINT	[illegible]	02-Ba	Close	(SDGI
127,418	Created	2015/07/C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	10 - W	Close	(SDAI
127,428	Created	2015/07/01 03:3	Low	Variant 1	[illegible]	Cambio	[illegible]	[illegible]	SDGPS	Fault	Variant	SDINT	[illegible]	01-SA	Close	(SDGI
127,437	Created	2015/07/01 04:0	High	Variant 1	[illegible]	Cambio	[illegible]	[illegible]	SDAMA	Fault	Variant	SD An	[illegible]	04 - D	Close	(SDAI

Figure 4.4: Cleaned data logs in Excel

4.3 Process Mining Phase

In order to conduct the process mining phase, the most used software was Disco by Fluxicon, while additional statistical analyses were executed via RapidMiner and Excel.
After having identified what column refers to the mandatory - and additional - information (that must be executed manually using Disco) finally all was set for the mining phase. During the project all the three techniques (Chapter 3.2: Process Mining Techniques) were applied in some way in order to gain the most relevant and precise information possible. Each of the techniques performed will be described in the following paragraphs.

4.3.1 Process Map Discovery

As described in Chapter 3, automated process discovery is the first process mining techniques that has to be applied. Indeed, it allows to gain a graphic overview about the process model. The software Disco is based on the *fuzzy miner algorithm* (see process mining algorithm section at Chapter 3.2.1: Automated Process Discovery), which provides users with a high-level representation of the processes through just two fundamental graphical objects: activity nodes and direct edges. The obtained process map shows clearly what are the most common routes followed by the totality of tickets.

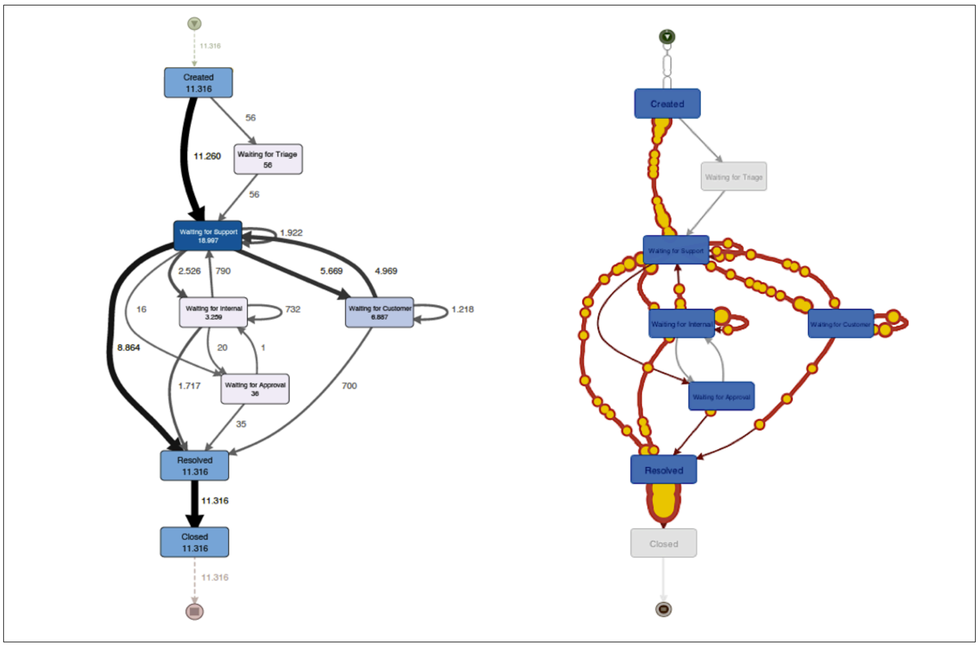
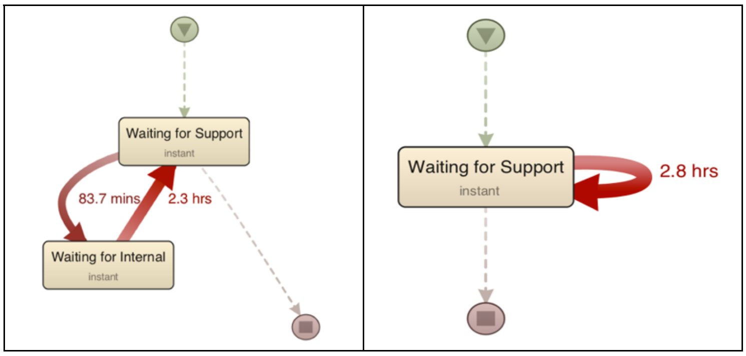


Figure 4.5: Process map and log animation of the ticket handling process
(raw data, activities = 100%; paths = 100%)

In particular, according to the model, the majority of the incidents are resolved after the WFS state only, while others need an additional passage through the WFC state. On the other hand, lesser cases ask for the support of an X-Provider specialist, through the escalation WFI and WFA routes. Finally, a particular portion of tickets handled with the new issue tracking system (ITS) are highlighted thanks to the WFT state.

Nevertheless, due to the fact that the initial process model takes into account all the logs representing different cases (with different customers, priorities, projects), the map inevitably leads to a graphical view characterized by a very low degree of representation. For this reason, it seems legit to explore the various form of handling ticket processes by filtering them by the main interesting variables: *issue*, *priority*, *customer* and *project*. This enables a better understanding of the model and gives analysts the possibility to face the problem in a structured way. For instance, by filtering event logs in this way, it is possible to find out that the state WFT compared only in a particular portion of cases handled with the new ITS. Filters enable to obtain specific process map views of the project and become the fundamental starting point for the next performance analysis stage.



Indeed, they let the first inefficiencies to emerge, such as *ping-pong effects* between two activities or apparently meaningless *self-loops* within the same activity, among different resources.

Figure 4.6: Instances of ping-pong effect and self-loop

Moreover, the software is capable of mapping the process, putting in evidence not only the frequencies, but also time-related metrics, such as median or maximum duration of routes (elapsed time). This provides a very first overview of the activities that act like *bottlenecks*, causing delays for all the other tasks.
All these problems and many others will be studied in the next phases of the project.

4.3.2 Process Analysis

According to the main dimensions of the project, the analysis phase is conducted in separate stages in order to face problems from different angles and to better understand the relation between root-causes and performance impacts.
Therefore, the perspectives to be considered are the following:

- *Issue/priority perspective;*
- *Customer perspective*;
- *Project/priority perspective*.

Each will be considered in the following paragraphs.

Issue/priority Perspective

The first view takes into account simultaneously the two process's dimensions of issue and priority and strives to find more insights by considering the inter-relations existing among them. Initially, looking at the occurrence's percentages of issues within the project (Table 4.1) it is clear that the majority of tickets are related to a *change* (approximately 50% of tickets), whereas the remaining ones are related to a *fault* and *help* situations. Instead, it was discovered that *task* tickets are limited to just a portion of very few cases (less than 1%). For this reason, tickets related to a task were neglected from the study.
The other dimension of priority is associated with ticket's urgency, therefore it is related to the service levels of agreement (SLA) projected.

During this phase it is possible to obtain evidences about what are the most frequent incident's cases and how the figures are sorted among urgency matters. So, for every issue, the principal six variants are considered. The choice of analysing just the main six behaviours is related to the fact that, generally, they are able alone to cover the majority of the cases available. In fact, by following a Pareto approach, the project pilot's aim is to focus firstly at those causes that influences process performances the most.
Anyway, the process maps achieved for each of the six behaviours show routes and median time between activities. Duration metrics are obtained without any pre-filtering stage, so the influence of *noise* must be considered. In this sense, the provided information about durations

are not to be considered to be exactly realistic, because just few tickets with abnormal performances (which are very often test-tickets) might influence greatly the overall performances results. The aim of this analysis stage is just to obtain a representative evidence of the routes characterizing major behaviours of each issue.

The first incident considered is *change*.

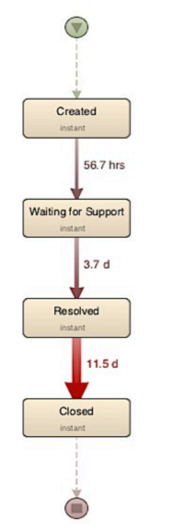
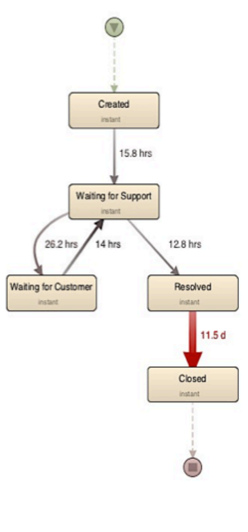
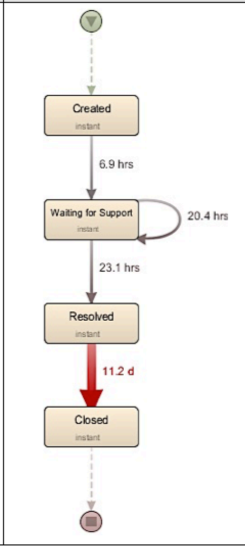
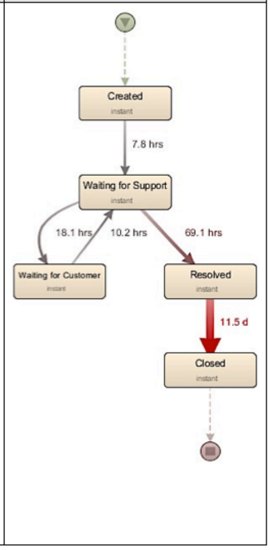
Change					
Main-stream	**2nd behaviour**	**3rd**	**4th**	**5th**	**6th**
					

Table 4.2: Process maps behaviour variations – change

	Change						
	mainstream (36%)	**2nd** (14%)	**3rd** (10%)	**4th** (5%)	**5th** (3%)	**6th** (2%)	
Red code	24%	17%	10%	2%	4%	7%	64%
Very high	32%	14%	7%	2%	7%	2%	64%
High	20%	19%	7%	3%	6%	5%	60%
Medium	31%	15%	10%	3%	4%	3%	66%
Low	45%	11%	12%	6%	2%	1%	77%

Table 4.3: Percentage of variations per priority – change

Firstly, it is clear why this kind of analysis takes into account just the six major variations: these behaviours alone are able to describe the 60-70% - per degree of priorities - of all the change tickets, so they could be seen as a quite robust depiction of the issue (see Table 4.3).

Thank to this analysis it is possible to notice that, even if the mainstream behaviour correctly represents the simplest route, it actually includes a notable part of cases characterized by remarkable priorities. Indeed, the fact that 24% of red code tickets and 32% of very high ones are included in the mainstream behaviour might reflect a distribution problem. The fact that too many tickets are characterized by great priority levels, could led to a situation in which a priority allocation does not exist at all. Moreover, a self-loop in the fourth most frequent behaviour, representing the 5% of all the change tickets, could be identified. Finally, the excessive median duration between the waiting for customer (WFC) and Resolved states (6.5 days), belonging to the 5th variant can be highlighted. Even if this bottleneck might just be influenced by noise, it is so much greater than other median durations and therefore would require additional analysis.

Change				
$n > 50$	$50 > n > 40$	$40 > n > 30$	$30 > n > 25$	$25 > n > 20$
◦ 162,699 ◦ 132,432 ◦ 231,124	◦ 312,24 ◦ 112,224 ◦ 179,917 ◦ 231,87	◦ 159,937 ◦ 345,234 ◦ 232,26 ◦ 148,485 ◦ 260,233	◦ 128,908 ◦ 135,074 ◦ 147,087 ◦ 149,94 ◦ 150,863 ◦ 155,80 ◦ …	◦ 147,393 ◦ 154,879 ◦ 155,324 ◦ 163,016 ◦ 165,116 ◦ 173,843 ◦ 179,252 ◦ 187,610 ◦ 192,040 ◦ …

Table 4.4: Case ID of tickets per number of events – change

Then, beside working on median values it may be interesting to approach to the maximal ones. According to this idea, change tickets characterized by the greatest number of events (n) are identified. Surely, these cases are not represented by the major behaviours seen before, so considering them more in depth might be substantial. In particular, a table showing which cases show the greatest number of events is derived (see Table 4.4). This allows to discover that some tickets are made up by even more than 50 events, and this may be the staring point for additional analysis on the the root-causes of such an alarming repetition.

Finally, the issue is filtered in a way that it is possible to highlight tickets with the poorest performances, in terms of handling time (see routes long more than 5 days in Table 4.5).

	Change				
	Low	**Medium**	**High**	**Very High**	**Red Code**
Created → WFS	248 (7,98%)	53 (2,94%)	30 (3,75%)	8 (2,23%)	0
WFS → WFI	59 (1,90%)	52 (2,89%)	20 (2,50%)	8 (2,23%)	0
WFS → WFC	85 (2,74%)	80 (4,44%)	31 (3,88%)	9 (2,51%)	3 (1,82%)
WFS → Resolved	251 (8,08%)	120 (6,66%)	49 (6,13%)	19 (5,31%)	2 (1,21%)

Table 4.5: Route's percentage of tickets with max duration > 5 days – change

For instance, even though service level of agreement (SLA) for change tickets characterized by very high priority states a response time within few hours, actual values shows that, during the timeframe considered, many cases have been handled after more than 5 days. The maximal waiting time point of view offers a new starting point for further investigation on those cases with the worst performances, accordingly to Table 4.5.

The, the second issue considered is *fault.*

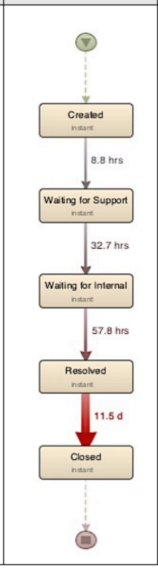
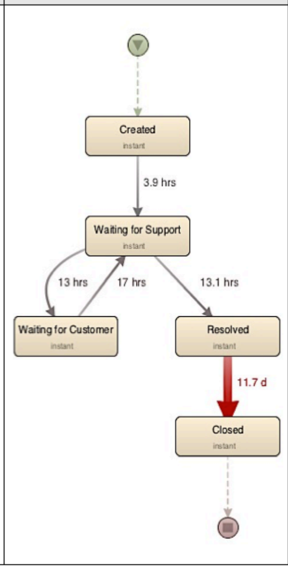
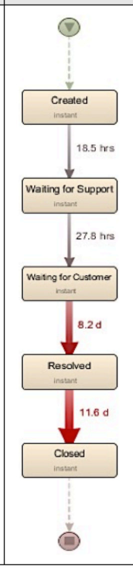
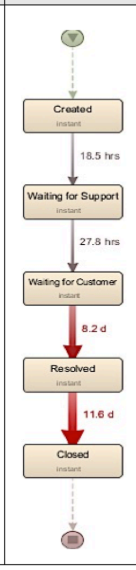
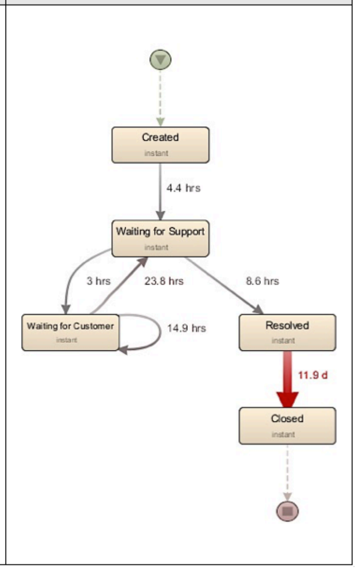
Fault					
Main.	**2nd**	**3rd**	**4th**	**5th**	**6th**
					

Table 4.6: Process maps behaviour variations – fault

Looking at the percentages of the variations (Table 4.7), it can be found out that here the allocation among priorities is better structured than the previous case. Indeed, the mainstream behaviour represents mostly tickets characterized by not high priorities (59% of medium and 59% of low ones). Then, considering process maps it is interesting to note the presence of the waiting for internal (WFI) state in the 2nd most frequent behaviour case (see Table 4.6). Since this state invokes another resource to handle the ticket, it is very onerous from both time and cost point of view, and it should be performed occasionally. For these reasons, the presence of WFI state will be analysed later separately.

	Fault						
	mainstream (49%)	**2nd** (11%)	**3rd** (7%)	**4th** (4%)	**5th** (2%)	**6th** (2%)	
Red code	19%	10%	15%	2%	2%	5%	53%
Very high	27%	7%	14%	1%	2%	4%	55%
High	30%	9%	10%	1%	4%	3%	57%
Medium	59%	13%	6%	0%	2%	3%	83%
Low	59%	9%	3%	10%	2%	1%	84%

Table 4.7:Percentage of variations per priority – fault

Moreover, considering process maps some other insights can be found. Firstly, median duration between WFC and Resolved states, within the the 4th and 5th behaviours, are again so much greater than others median values. These remarkable gaps suggest the need for supplementary checking. Secondly, the 6th maps highlighted the presence of a self-loop within the WFC state. Then, fault tickets made up by many events (*n>50; 50>n*>40; …etc.) are discovered again in such a way that additional examinations of the most problematic tickets would enabled.

Finally, tickets with exceptional mean durations (more than 5 days), compared to SLA values, are highlighted in Table 4.8. In particular, knowing that response time between a fault ticket creation and its handling by X-Provider have to be within 4 hours for very a high priority case (from SLA), the evidence that 27 tickets were managed after more than 5 days appears to be a serious problem. Evidently, a specific investigation of the issues should be carried by X-Provider.

	Fault				
	Low	**Medium**	**High**	**Very High**	**Red Code**
Created → WFS	40 (2,69%)	20 (1,02%)	16 (2,90%)	27 (6,01%)	3 (0,7%)
WFS → WFI	14 (0,94%)	30 (1,53%)	11 (2%)	8 (1,78%)	6 (1,41%)
WFS → WFC	14 (0,94%)	33 (1,69%)	28 (5,08%)	11 (2,45%)	7 (1,64%)
WFS → Resolved	136 (9,13%)	102 (5,21%)	57 (10,34%)	32 (7,13%)	18 (4,23%)

Table 4.8: Route's percentage of tickets with max duration > 5 days – fault

Finally, the last issue considered is *help*.

Again, for each priority process models and frequencies are captured.

Help					
Main.	**2nd**	**3rd**	**4th**	**5th**	**6th**

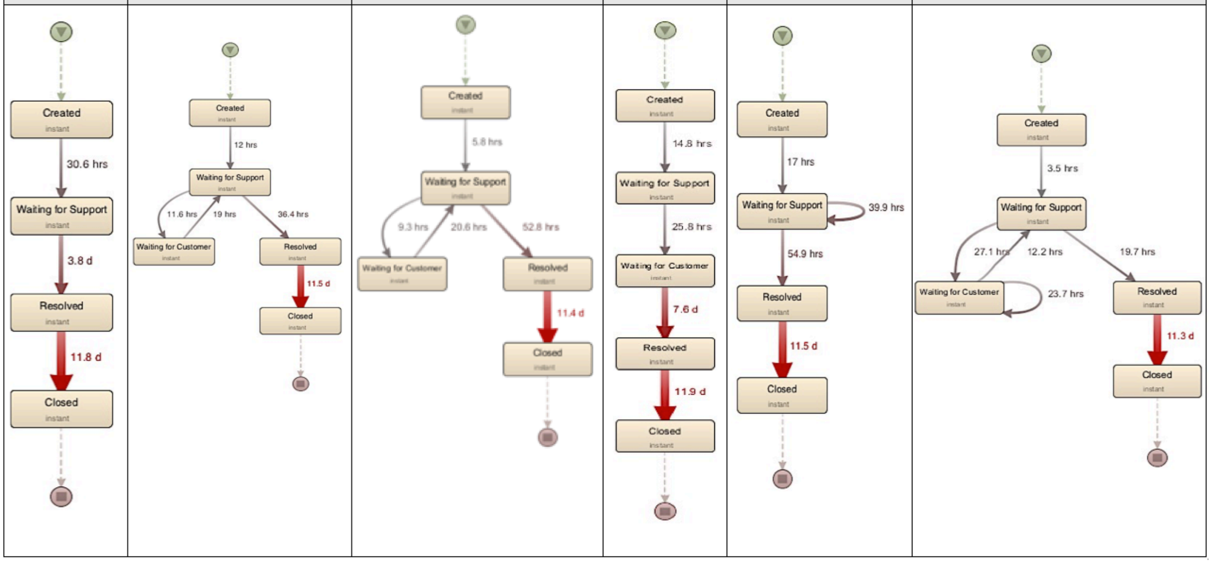


Table 4.9: Process maps behaviour variations – help

	Help						
	mainstream (30%)	**2nd** (18%)	**3rd** (5%)	**4th** (4%)	**5th** (4%)	**6th** (3%)	
Red code	18%	23%	3%	2%	4%	6%	56%
Very high	24%	21%	5%	3%	5%	3%	61%
High	26%	23%	6%	4%	2%	3%	64%
Medium	31%	18%	4%	5%	4%	2%	64%
Low	40%	11%	5%	4%	4%	3%	67%

Table 4.10:Percentage of variations per priority – help

Looking at the variation percentages, it turns out that this issue is handled through more routes than the others seen before. Indeed, the mainstream behaviour is shared by just the 30% of the tickets (against the 40% of faults and the 36% of changes) and this reveals that the way in which this incident is handled is less standardized.

Anyway, mainstream behaviour mostly reflects the routes of low priority tickets (40% low and 31% medium), and this shows a correct allocation of precedencies.

However, the second behaviour in terms of frequency depicts tickets with higher priority degrees (23% red code, 21% very high, 23% high). Then, the mainstream process map displays an alarming mean duration of 3.8 days between WFS and Resolved states (see Table 4.9). A great mean duration, in the same route, is evident in the 4th behaviour too. This would require specific analysis to understand if these trends are just influenced by noise (for instance a test ticket opened and then forgotten) of there is a real root-cause behind them. Moreover, the models highlight the presence of self-loops in the 5th and 6th behaviours, within the WFS state and the WFC state (median durations of 38,9 hours and 23,7 hours, respectively).
Finally, help tickets characterized by a huge number of events and those ones with duration time between tasks very greater than SLA values, are again traced for enabling singular inspection of the most problematic cases.

To sum up, by applying process mining within the issue/priority perspective it was possible to discover the following prior evidences, which would need additional investigation by X-Provider management or external consultants:

- E1. Change tickets mainstream behaviour reflects a suboptimal allocation of priorities, while other behaviours reveals a self-loop and a possible bottleneck;
- E2. Fault ticket 2nd most frequent behaviour indicates the presence of WFI state and the presence of a self-loop in the 6th;
- E3. Help tickets are handled in a less standardized way, and process maps highlight the presence of a possible bottlenecks in the mainstream and 4th behaviours, and two self-loops on the 5th and 6th most frequent ones;
- E4. For each issue, tickets characterized by the greatest number of events are identified (n>50, 50>n>40…etc.) as well as the distribution of slow tickets with a maximum duration of routes greater than 5 days.

Customer perspective

The second point of view of the study takes into account customers and puts them into relation with the other process dimensions of priority and issue. The aim here is to discover who are the customers whose tickets have the higher priority degrees and to find out evidences about how different incidents are managed by X-Provider.

For the ticketing management process considered, 125 different clients have been identified, besides the organization itself. As introduced, a considerable amount of tickets that X-Provider has to handle had been created by a resource working on the same company. In the specific, complete tickets distribution (in relative frequency) among customers is described in Table 4.11. Evidently, customers' name cannot be cited due to non disclosure agreement. For simplicity they are called "Customer 1, 2, 3…etc." in decreasing order of their relative frequencies.

According to the table, X-Provider itself is the major costumer of the ticketing management process (with approximately the half of tickets). Then, a considerable part of 17% of tickets is associated with an undefined customer. These tickets might be attributed to test tickets and for this reason are neglected from the analysis. Therefore, since X-Provider, Customer 1, Customer 2, 3 and 4 alone originate most of the tickets (60% together), a specific analysis will be carried out exclusively for them.

Value	Relative Frequency
X-Provider	**49,98%**
Customer 1	**6,03%**
Customer 2	**1,50%**
Customer 3	**1,30%**
Customer 4	**1,24%**
Customer 5	1,03%
Customer 6	0,86%
Customer 7	0,85%
Customer 8	0,84%
Customer 9	0,74%
Customer 10	0,71%
Customer 11	0,55%
Customer 12	0,54%
Customer 13	0,54%
Customer 14	0,52%
Customer 15	0,52%
Customer 16	0,52%

Value	Relative Frequency
Customer 17	0,43%
Customer 18	0,42%
Customer 19	0,38%
Customer 20	0,38%
Customer 21	0,35%
Customer 22	0,35%
Customer 23	0,33%
Customer 24	0,33%
Customer 25	0,32%
Customer 26	0,30%
Customer 27	0,29%
Customer 28	0,29%
Customer 29	0,29%
Customer 30	0,27%
Customer 31-125	10,41%
Undefined	17,04%
	100%

Table 4.11: Tickets relative frequency distribution per customer

Firstly, the frequency of the route WFC → WFS is taken into account. It is important to study this path because it could be seen as an indicator of X-Provider's *support quality*. Indeed, company answers to the costumer providing them a solution to the issue characterizing the ticket they have opened before. If customers are satisfied with this response, no more action are required and the ticket will be automatically closed after a premeditated period of time. On the contrary, if the issue is still unresolved they will ask again for X-Provider support and the cycle begin another time (see the process framework in Figure 4.2). In this sense, the frequency of the path WFC ➞ WFS indicates how many times first X-Provider's support is not able to provide the complete resolution of the issue.

Secondly, another parameter absolutely useful for the study is the percentage of WFC self-loops that reflects how many times data logs present a repetition of the event, which means that customer is demanded to answer. Evidently, this can happen for both internal and external customers. So, taking into account the way tickets are handled within the company, it is essential to understand that any distinction between customers are present. Therefore, Table 4.12 shows relative frequencies of paths and self-loops, per issue, for internal and external customer.

	Change					
	Tot.	%	**Internal**	Relative frequency	**External**	Relative frequency
WFC ➞ WFS	3.140	50,23%	1.827	58,18%	1.313	41,82%
WFC self-loop	772		497	64,38%	275	35,62%
	Fault					
	Tot.	%	**Internal**	Relative frequency	**External**	Relative frequency
WFC ➞ WFS	1.729	35,40%	662	38,29%	1.067	61,71%
WFC self-loop	466		180	38,63%	286	61,37%
	Help					
	Tot.	%	**Internal**	Relative frequency	**External**	Relative frequency
WFC ➞ WFS	1.343	85,34%	596	44,38%	747	55,62%
WFC self-loop	288		143	49,65%	145	50,35%

Table 4.12: Relative frequency of WFC ➞ WFS path and WFC self loop per issue

According to the data, WFC ➞ WFS route takes place half of the time a change ticket is handled. Taking into account customers, it can be observed that this trend is more evident for those tickets opened by an internal customer. The same tendency can be evaluated from the distribution of the WFC self-loops. These evidences suggest that X-Provider resources used to rework tasks and to be demanded for a response more than external customers. Although it can be supposed that the trend reflects the correct tendency of X-Provider to conduct tests or recovering, this evidence might require an additional investigation.

On the contrary, data shows that external customers own higher values in fault tickets. Again, this path seems to be quite reasonable if we consider that X-Provider's resources can ask for support in other ways than invoking WFS state another time. Anyway, figures show higher values in WFC self-loops for external customer too, so more examinations are recommended.
Finally, considering help tickets no particular trends emerge, since the distribution between internal and external customer is quite similar both for WFC ➞ WFS routes and for WFC self-loops. However, the important discovery here is that even more than 85% of the help tickets shows a WFC ➞ WFS path. Evidently, this requires further scrutiny.

Then, durations of routes between WFC and WFS (round trip), as well as for WFC self-loops, were studied for external customer specifically. Only the external customers are taken into account here because X-Provider has to respect with them pressing conditions defined by SLA projected. Indeed, these kind of durations can be seen as an indicator for X-Provider's *time related performances.*
Therefore, for each of the three issues, the facts in Table 4.10 are discovered.

	Change		
	Absolute frequency	**Mean duration (hrs)**	**Total duration (days)**
WFS ➞ WFC	1.470	25,2	1.544
WFC ➞ WFS	1.313	17,2	941
WFC self-loop	275	25,3	290
	Fault		
	Absolute frequency	**Mean duration** (hrs)	**Total duration** (days)
WFS ➞ WFC	1.198	19,3	963
WFC ➞ WFS	1.067	24	1.067
WFC self-loop	286	26,9	321
	Help		
	Absolute frequency	**Mean duration** (hrs)	**Total duration** (days)
WFS ➞ WFC	843	20,7	727
WFC ➞ WFS	747	19	591
WFC self-loop	145	39	236

Table 4.13:WFS and WFC routes duration among issues for external customers

According to the table, change's WFS ➞ WFC route shows a total duration (in days) very greater than other durations considered. Even if this huge handling time could be probably re-ferred to some tickets which have been opened and then lost or forgotten, more specific studies

are recommended. In fact, these tickets are associated to external customers only and therefore they should not include particular corruption caused by tests.
Similarly, considering fault tickets, the same route presents again a considerable duration, but not as significant as the opposite WFC ➞ WFS way. On the other hand, help ticket presents slightly greater time within the WFS ➞ WFC.

Next, total durations of routes and self-loops over time are traced, for each issue, for the purpose of discovering how the situation is evolving.
Before approaching to the graphs it is essential to state that last month values (march 2016) are to consider not representative because many states might not have been triggered within the end of the time period considered.

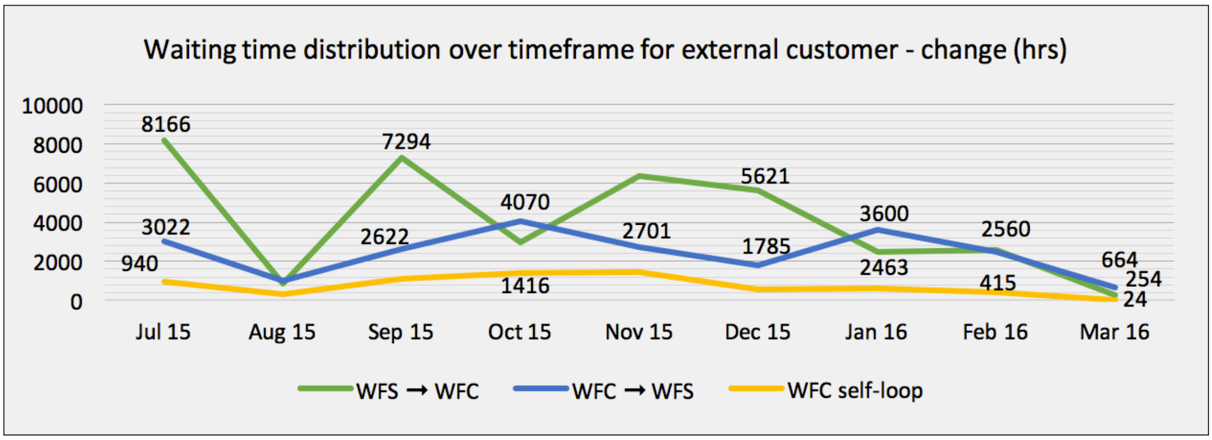


Figure 4.7:Waiting time distribution over timeframe for external customer – change

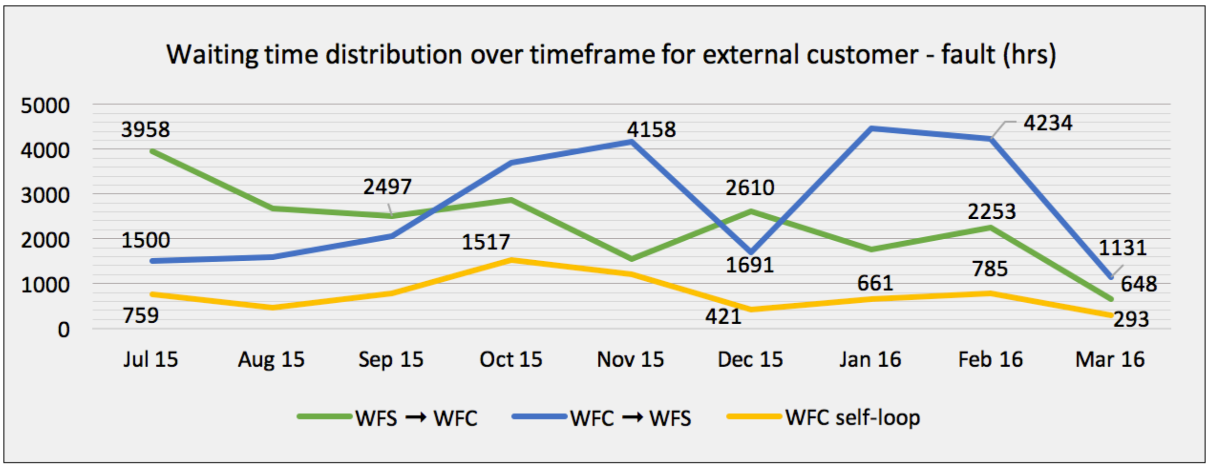


Figure 4.8: Waiting time distribution over timeframe for external customer – fault

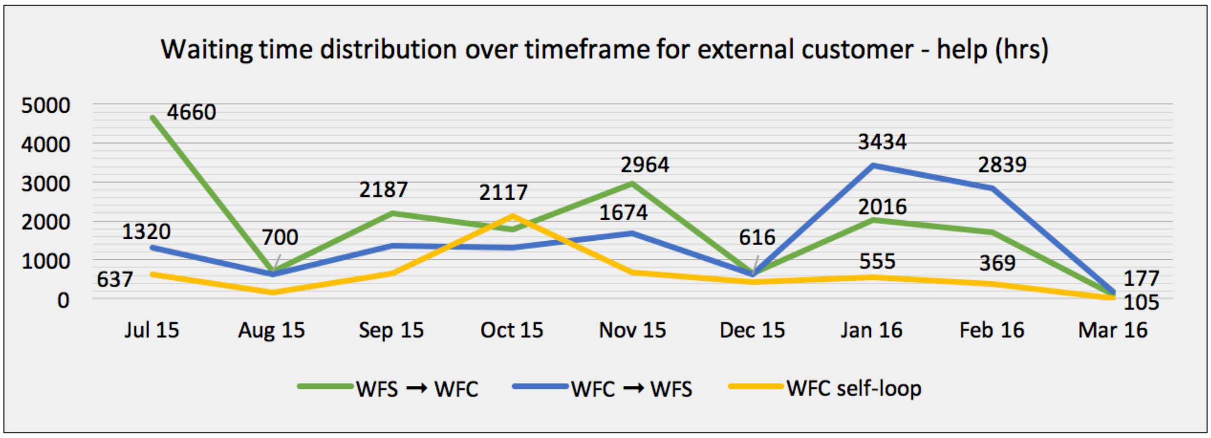


Figure 4.9: Waiting time distribution over timeframe for external customer – help

In general, line charts highlight too many variations to make any conclusion and therefore it seems not to offer any new meaningful insights. After all, waiting time values are very sensitive to exceptions and noises: just few tickets characterized by exceptional durations in these paths might neutralise the overall results.

Finally, rather than taking into account external costumer all together, just the most important ones, in terms of tickets ownership, are considered. In particular, for customers 1-4 and for each priority the percentage of WFS ➞ Resolved routes is calculated, as well as the percentage of tickets that require the call for an internal to be resolved. Indeed, the first variable indicates the X-Provider operator's ability to solve issues without the need of invoking an internal for his support (see the process framework in Figure 4.2).

	Customer 1				
	Low	Medium	High	Very High	Red code
WFS ➞ Resolved	57%	60%	35%	36%	34%
WFS ➞ WFC	12%	18%	40%	38%	37%
WFS ➞ WFI	13%	17%	15%	17%	17%
	Customer 2				
	Low	Medium	High	Very High	Red code
WFS ➞ Resolved	41%	38%	26%	25%	25%
WFS ➞ WFC	29%	37%	49%	47%	52%
WFS ➞ WFI	25%	19%	15%	18%	18%

	Customer 3					
	Low	Medium	High	Very High	Red code	
WFS ➞ Resolved	59%	55%	49%	44%	50%	
WFS ➞ WFC	30%	36%	43%	33%	38%	
WFS ➞ WFI	30%	36%	43%	33%	38%	
	Customer 4					
	Low	Medium	High	Very High	Red code	
WFS ➞ Resolved	53%	37%	23%	37%	46%	
WFS ➞ WFC	29%	59%	57%	49%	42%	
WFS ➞ WFI	18%	2%	13%	14%	13%	

Table 4.14: Customer 1-4 routes per priorities

According to the table, some common paths among major external customers can be discovered. For instance, as the ticket's priority increase, the percentage of issues resolved just within the WFS state decreases. As opposite, the need for a specialist's support (expressed through the WFI state) grows, as well as the need for the intervention to be validated by customers (WFC). Nevertheless, some exception can be pointed out. Customer 2's tickets with low priority present greater value in the WFS ➞ WFI route respect to other priorities, while Customer 4's WFS ➞ WFC path seems to not grow with the increase of priority. Again, also these evidences would require more investigation.

To sum up, within a customer perspective process mining techniques enable the following evidences to emerge:

- E5. Internal customers used to ask again for support more than external ones in the case of a change ticket. In addition, they are more often demanded for respond (WFC self-loop);
- E6. External customers used to respond to the support more than internal customers when a fault ticket is open. Moreover, they show higher relative percentage of WFC self-loops;
- E7. Customers, with no distinction between internal and external ones, ask again for support most of the times in the case of a help ticket;
- E8. External customer present very different waiting times among various issues;
- E9. Both 2nd and 4th customer (in order of tickets opened) present variation in the way tickets are handled from other major clients' tickets.

Project Perspective

Then, data logs are considered from a project point of view. This is the last perspective taken into account during the pilot project analysis.
As introduced in the beginning of the chapter, project names offer useful information to identify which cases are handled with the new *issue ticketing system* (ITS) and which are still managed with the old one. Even if the WFT state enables to discover a portion of these cases, by studying tickets from a project perspective it is possible to conduct a more detailed analysis about the expected benefits derived from the introduction of the new ITS in June 2015.

In particular, tickets handled with the new ITS are characterized by a specific project denomination, no matter the particular project. More in depth, the new ITS is applied for carrying some kinds of *Anomaly* and *Request (Technical Anomaly, Database Anomaly, Technical Request, Database Request...*etc.). For the sake of simplicity, all these variations are grouped in just two clusters, namely *Anomaly* and *Request*.
On the other hand, project names allow to identify which tickets of the same type are still managed with the old ITS. These cases are taken into consideration all together.

Therefore, all the principal paths durations are considered for both the Anomaly and Request clusters. In details, for each route the mean duration distribution over the 9 months is calculated. Firstly, durations are calculated through paths' frequencies and durations considering only the 95 percentile, in a way to neglect the outliers caused by noise. Then, total durations obtained are weighted for their corresponding route's frequency.

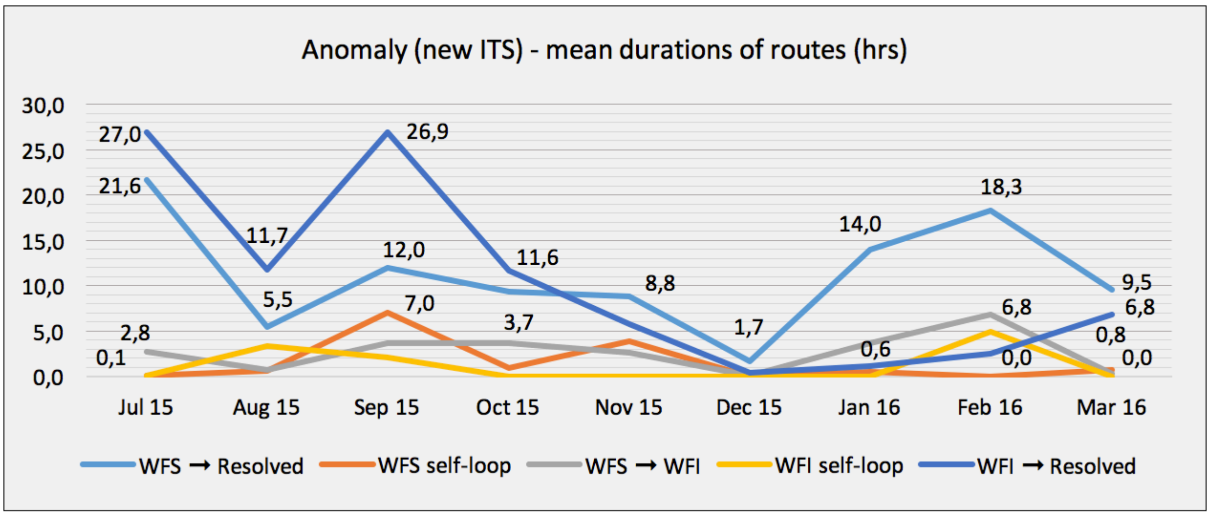


Figure 4.10: Anomaly mean durations of routes – new ITS

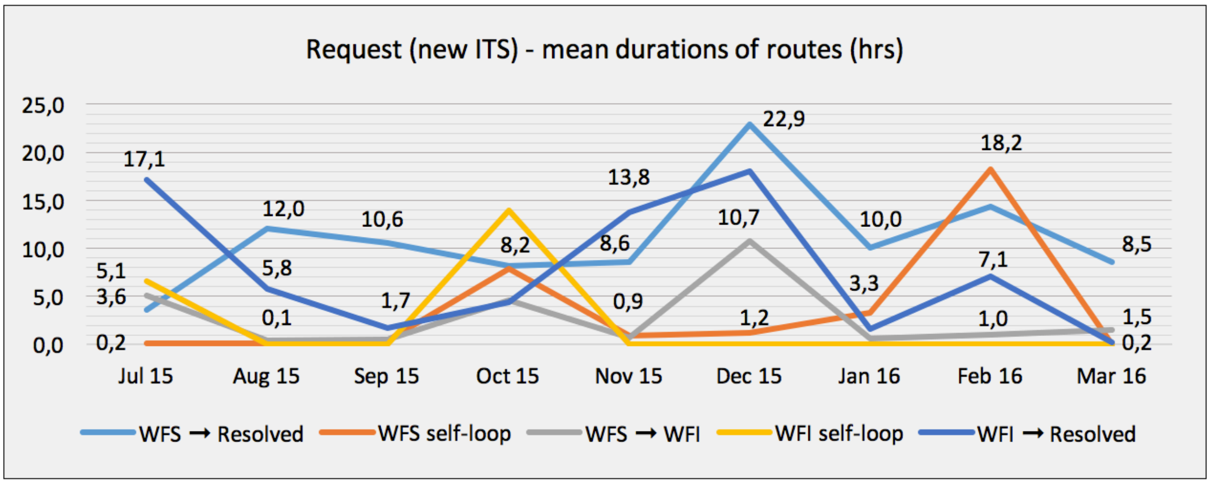


Figure 4.11: Request mean durations of routes – new ITS

Again, last month values should not to be considered trustworthy because some values might not be taken into account due to the timeframe of the project.

Then, a similar line chart is derived for tickets managed with the old ITS. Due to the fact that for these tickets there is no distinction between anomalies and request, the chart takes into account the mean duration of all values together.

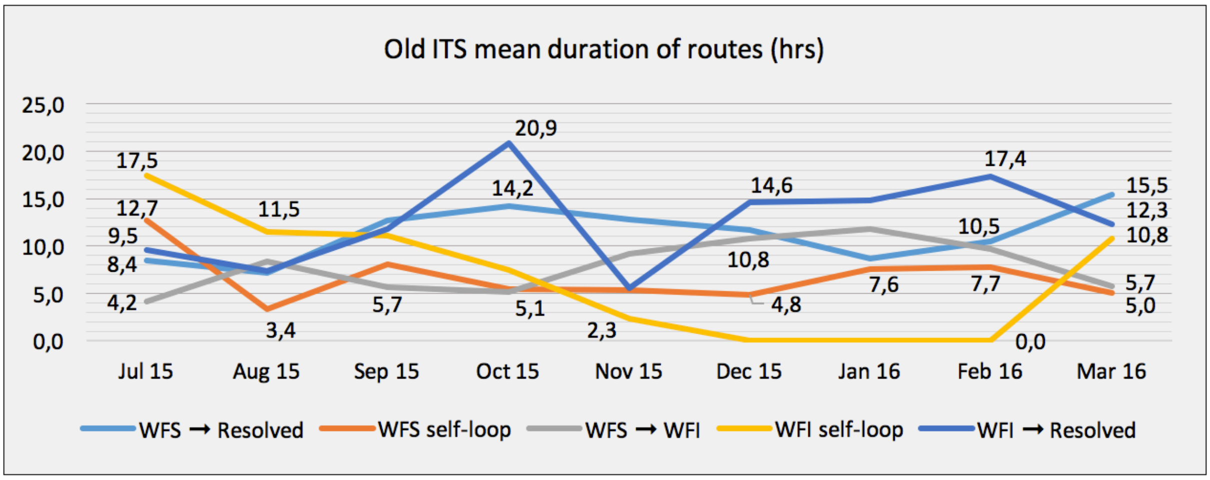


Figure 4.12: Mean duration of routes – old ITS

In general, by looking at the line charts no particular insights emerge. The only considerations that can be made are referred to the self-loops and to the WFI ➞ Resolved route durations. In fact, mean duration of self loops appear to be very different among situations considered. Moreover, anomalies handled with the new ITS show a slightly decrease in median time of the WFI ➞ Resolved route, from 27 hours in July 2015 to well below 7 hours during the last months studied (see Figure 4.10).

For this reason, just the durations of the self-loops are considered separately, as well as those of the WFI ➞ Resolved way. Therefore, both the self-loops mean durations obtained before are multiplied for their occurrence frequency, summed together and then weighted by the ticket creation frequency. The choice of weighting durations for such a value is aimed to gain a more realistic quantification on the impact that self-loops make to the handling of tickets. Nevertheless, also in this way any particular clues seem to emerge.

For this reason, in order to obtain a very high and easy to visualize graphic depiction of the distribution among months, just one value for New ITS (Anomalies), New ITS (Request) and Old ITS is derived. This evaluates the mean duration (at 95 percentile) of tickets, from the "Created" state to the final "Closed" one. Even though this led to a less structured answer, it would provide a good abstraction of the problem and allows X-Provider to finally discover if the launch of the new ITS has introduced some kind of advantages, otherwise it has caused eventual slowdowns in the performances of those tickets still handled with the old ITS.
Tickets' lead time distributions are associated to their demand distribution in such a way that it is feasible to pinpoint possible changes in performances due to demand's fluctuation.

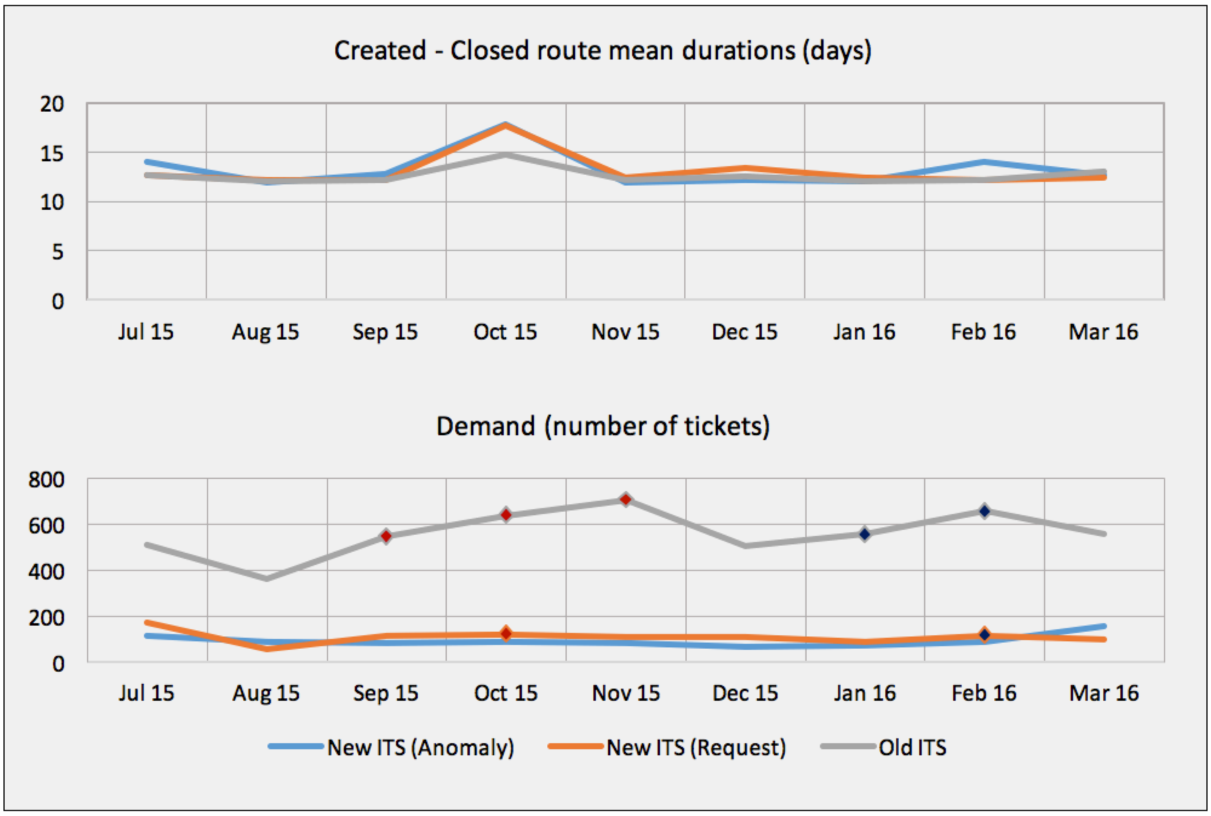


Figure 4.13: Comparison between new and old ITS mean lead time

Consequently, Figure 4.13 enables to derive some meaningful considerations. Indeed, according to the data, the seasonality of demand can be discovered. Between October and November, the first peak is reached, while the second one takes place in February. By looking at the mean lead times, it is clear that the introduction of the new ITS has not caused any decline in the performances of the other tickets. In particular, the old ITS seems to react better to demand fluctuation than the new one. Nevertheless, a general improvement can be highlighted during the second demand peak in February 2016. The marginal increasing of lead times in this period is less pronounced than the one observed during the first peak in November 2015. Anyway, it must be stated that the time period of logs taken into account inevitably limits the validity of the results. However, the better reaction against demand volatility observed cannot be contested.

To sum up this paragraph, process mining applied within the project perspective provides the following evidences:

- E10. The introduction of the new ITS seems to have not caused any deterioration in the handling performance of tickets still managed with the old ITS;
- E11. No evident improvements in the performances of Anomaly and Request with the new ITS can be highlighted;
- E10. After the new ITS implementation, X-Provider seems to better react to demand fluctuation.

Lastly, after having approach to the problem from issue, priority, customer and project point of views, a particular attention was dedicated just to the Waiting for Internal (WFI) stage. As seen, this phase seems to be problematic due to the fact that it implies the involvement of a specialist, who is asked to solve very challenging software incidents.
Clearly, the escalation is onerous especially from an economical dimension, because the hourly-wage of these specialists is significantly higher then that of the service desk resources. Moreover, the invoking of WFI interrupts SLA counters and will lead to the corruption of timers and indicators. Either way, the calling for an internal should be performed exclusively when the new ITS is used. By filtering the activity with Disco, the distribution of its cumulative frequency, among issues, can be discovered.

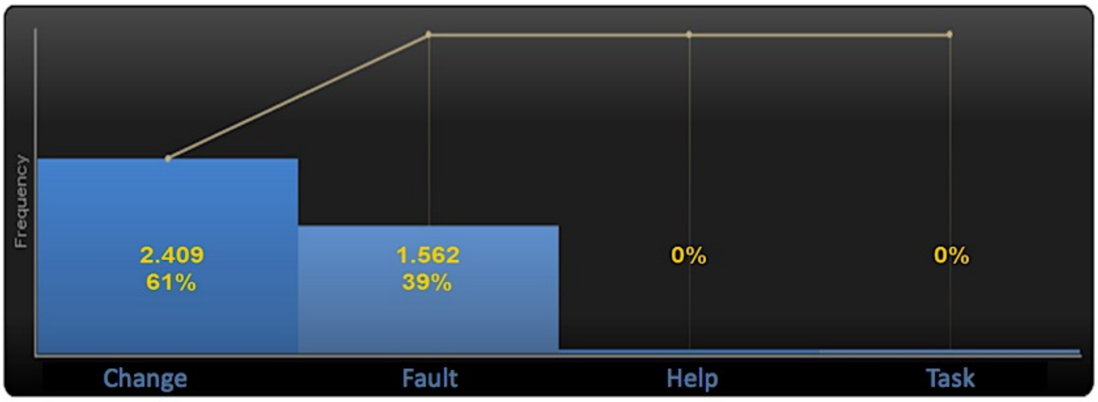


Figure 4.14: : WFI cumulative frequency among issues

According to the histogram, an internal participation is asked when a change or a fault ticket is being handled (respectively the 61% and 39% of cases). This was partially deducible from the major behaviours' process maps seen during the issue/priority perspective study. In particular, fault tickets have showed the presence of the undesired WFI state even in the 2nd most frequent case (see Table 4.6).

On the other hand, by filtering the same issue by priority, another problematic trend appears: an internal is called more frequently when a low or medium priority ticket is under investigation (nearly the 70% of total WFI frequency together).

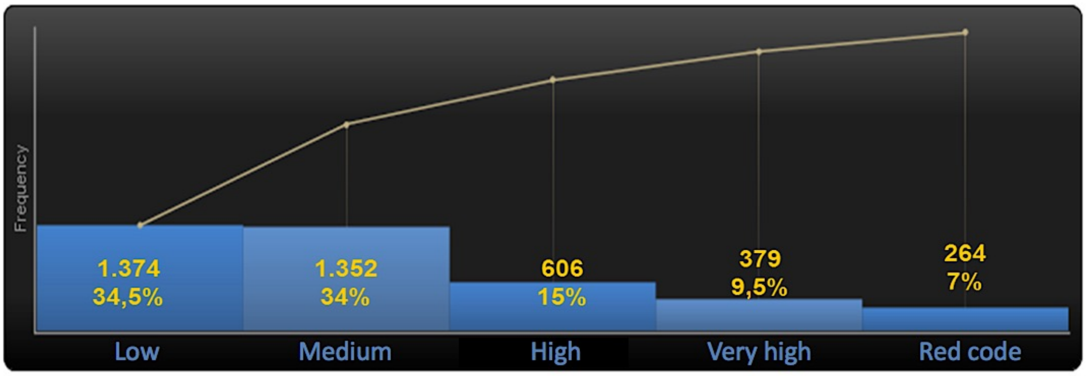


Figure 4.15: : WFI cumulative frequency among priorities

Despite using WFI exclusively for particular and complex necessities, the histogram reveals that actually the opposite occurs. Aware that WFI should be used for new ITS tickets only, it seems legit to analyse the state's frequency on the basis of the project perspective. The results obtained are indicative of the alarming tendency.

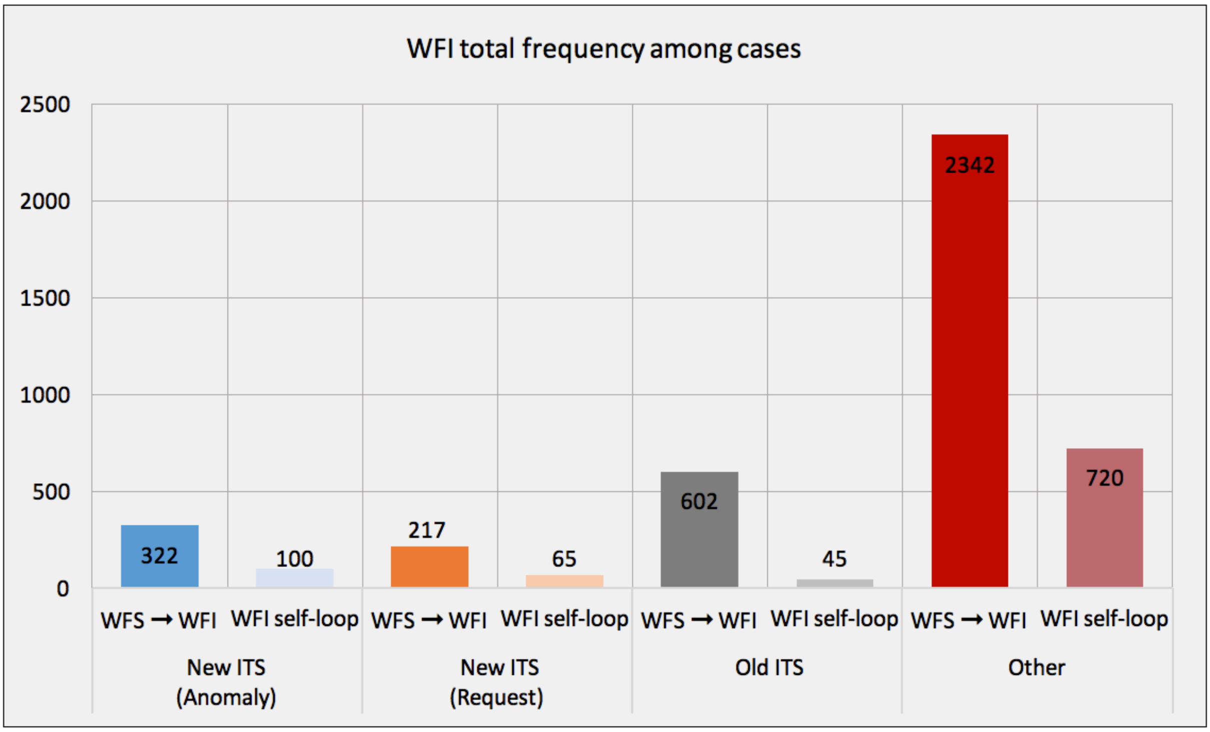


Figure 4.16: WFI total frequency among cases

Indeed, according to the histogram, the internals are not asked to participate only for the portion of tickets handled with the new ITS. Moreover, the WFI state is used not only for those tickets of the same type still managed with the old ITS, but also for other tickets for which the called for an internal is a complete nonsense. In particular, the values of route WFS→WFI are over-exploited especially for the "Other" cases. The WFI self-loops for this group are not negligible too. Thank to this evidences the phenomena was studied more in depth and this pointed out that the problem was related to the software that X-Provider service desk resources use for handling the incidents. In fact, the visual interface of this application always lets service desk resources to choose the "waiting for internal" option, no matter if the case is actually legitimate to be taken by a specialist. The lack of forbidden and mandatory options deteriorates X-Provider's performances in terms of SLA and costs the firm a lot of economic effort that should be avoided.

Finally, both the route distributions that involve the WFI state are plotted (Figure 4.17). Again, the time period considered is the one of the data logs taken into analysis.
The 9-month distribution plot highlights that the trend seems to be on the decrease. As seen for the mean lead time of the overall handling process (Figure 4.13), the line chart shows a better reaction during the second demand peak in February 2016.

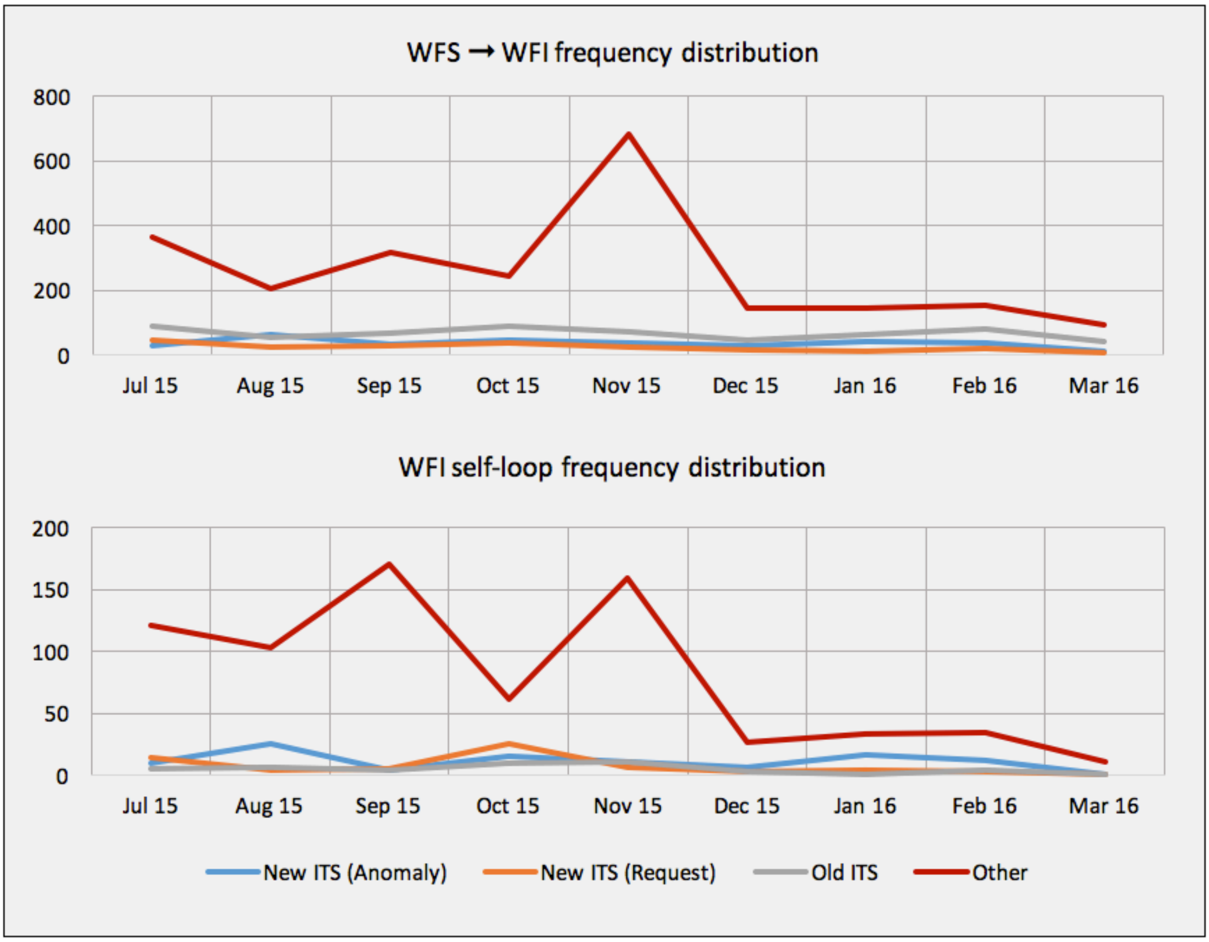


Figure 4.17: WFI frequency distribution

At the end, event logs are filtered by X-provider's service desk resources involved in the process. This enables to discover that the resources that use to ask for an internal intervention most of the time are ticket's queue administrators. These resources do their daily tasks, by means of the software, without any permission limitations due to their role. As the administrators see that the WFI command is enabled, they fell legitimate to use it, without thinking that this practise would be actually unfavourable in terms of process performance. Even though line charts indicate that the abuse of the command seems to be in decline, the phenomena can't be ignored. One solution might be changing the software user interface, so that it forbids - or limits - administrators' possibility to choose the command.

In order to completeness, the new most meaningful evidences are summarized. In particular:

E13. WFI state appears more frequently for change and fault issues, with low and medium priorities;

E14. Even though WFI should be present exclusively with tickets handled with the new

ITS, it actually compares at significantly rates in the other cases;

E15. Queue administrators use more frequently the WFI command because the software they use don’t apply any limitations;

E16. The trend of invoking internals’ support is slightly decreasing over the past 9 months.

Conclusions and Future Research

Management and business science have successfully theorized new organizational models in order to better fit the continuous growing dynamism and complexity of the real business situations. Even if this have surely enabled firm to pursuit economical advantages through economies of scales and to determine a well structured division of responsibilities, some disadvantages such as the searching for local efficiency instead of global improvement and the resistance to changes have not been reduced. Then, new organizational models, such as the *Business Process Reengineering* have spread, where managers and business scholars have firstly highlighted the central role that processes play within a company structure.

Finally, the theory of *Business Process Management* was introduced for partially solving some issues and limitations derived from the "one shot" philosophy at the base of the BPR.

Business Process management discipline is focused on the better understanding of event chains and the decision flows which characterized companies, rather than on the searching for an extraordinary, but just local and not replicable, improvement. It is a holistic management philosophy, instead of a tool for pursuing the enhancement in the business performances.

From BPR onwards, organizational models have continuously been finding their applicability and effectiveness on *information and communication technology* matters. With the specific aim of improving their internal efficiency, companies started to face to the fact that technology has become essential for a multitude of functions such as finance, accounting, human resources, purchasing, distribution, communication and so on. As a consequence, an ever-increasing relevance is earned by the role of data, created at incredible rates when whatever task is performed by any kind of computer based device. It is exactly from the outstanding growing of digital information in recent years that the expression of *Big Data* took shape. Anyway, today the object is not to create even more data, but to be able to collect them in such a way that it is possible to exploit their substantial information in an efficient way. Modern techniques strive to convert this source of information into tangible value and to take advantages from its administration. Hence, the methodologies and tools under the term of *business intelligence* try to extrapolate knowledge from such a volume of digital data.

Nevertheless, process related business organization models, such as the Business Process Management, are not data-driven and therefore are not able to manage the continuous workflow

of insights that data could provide in real time. Moreover, even though modern process modeling techniques like Petri, EPC, YALW and BPMN proper emphasize the centrality of processes, they are not capable of creating models by exploiting the amount of digital information now available. In other words, these process modeling methods cannot be defined as data-driven.
On the other hand, modern subfields of computer science born during the last decade, like data mining, are obviously data-driven but cannot manage the complexity and the multitude of drivers and stakeholders of today real business processes. Data alone, without the comprehension of such a countless number of variables, very often cannot provide an aware administration of resources and therefore the improvements desired.

In this context, *process mining* techniques can help executives and all the decision makers. Process mining can be seen as the natural bridge that links the process-driven discipline of Business Process Management with the data-driven methodology of business intelligence.
Since the starting point of this discipline is the real data stored into the ERP systems, it can be applied in almost every companies without any distinction of specific industries or business environments. To obtain a graphical model representation of the process, to support audit quality, ensuring the compliance with standards or regulations, and to allow performance enhancing of processes are the three main uses of process mining within a practical situation.
The *Database of Applications* I have developed, available in appendix, represents the most complete and comprehensive list of all the adoptions of process mining techniques worldwide over the past ten years. Despite the recent spread of the phenomena, it counts just less than 120 cases. Therefore, the question which arises spontaneously is why nowadays companies and consulting firm don't fully exploit process mining possibilities and remain still dependent on classical investigation methods such as expensive interviewing.
Process mining techniques has been applied within a real business case, in a big IT application and infrastructure provider, in order to strive to answer the question. The solutions that process mining application provided to the case study were surely relevant, especially the possibility of mapping processes in a transparent and effective way that Visio or other BPMN-based software could not provide. In addiction, the opportunity to visually discover mean waiting times or to isolate cases which performed worst, acted as a very useful starting point for a deeper analysis on the conformance with *service levels of agreements* stipulated with customers. Nevertheless, several not negligible issues related to process mining application came to light.
Firstly, the three mandatory requirements, without any kind of handling often cannot be fixed, and this causes process mining to not work at all. Secondly, the guidelines for logging provided

by Process Mining's Manifesto itself are not easy to respect in a practical corporate scenario. Quality of logs decides the success of the entire plan, and information system developers, database administrators and resources involved are still not aware of the improving that an effective and standardized way of logging could provide to their business. For this reason, process mining could be performed just in those companies where such a knowledge is inherent in firm's culture and it is shared among the different levels of the organization. Anyway, in the case study a significant data cleaning pre-analysis phase was required to convert logs in a format suitable for process mining techniques. Then, a trim stage was necessary in order to consider just the complete lifetime of tickets, by excluding those cases opened before or closed above the timeframe, in a way to avoid the so-called spaghetti process maps as result and to preserve performance metrics from corruption. This has significantly reduced data logs to be studied, from nearly 75000 to 16000 (a reduction of 78,7%). In addiction, the influence of noise could not be neglected, especially when mean or median values were taken into account. Surely, software like Disco, are committed to better manage this problem, thank to the possibility of excluding some data from the intended results. Finally, performance metrics often seem to lose their reliability, especially at the borders of the time period considered.

Even though some indisputable evidences have been identified, these often need a further specific examination. The process mining project evidently enables analysts to discover alarming situations or not receptiveness of standards and SLA, but these discoveries can be seen more as the starting point for future deeper investigation, rather than 100% trustworthy results. Without an effective logging of data and a proper database maintenance, process mining remains useful just for the very first stages of a change plan or a consulting project, serving more as an exploration phase of investigation. Even though this could reduce the initial need for intensive, time and cost consuming interviews and researches, process mining could not be seen as the only tool providing, with one wave of the wand, the solution of the projects.
Furthermore, an additional research on mining algorithms would be recommended to the scientific auditors or corporates who want to go deeper on process mining applicability in real business scenarios. Even if fuzzy miner has significantly improved the quality of insights obtainable from mining complex processes, some issues still come up in presence of noise (caused by tests, recovery or loading time of the software). Moreover, it would be required specific database management system knowledge to ensure data scalability, without losing any essential information.

Bibliography

Accorsi, R., Damiani, E., & van der Aalst, W. (2014). *Unleashing Operational Process Mining* (Dagstuhl Seminar 13481). Dagstuhl Reports, 3(11).

Adriansyah, A., van Dongen, B. F., & van der Aalst, W. M. (2011, August). *Conformance checking using cost-based fitness analysis. In Enterprise Distributed Object Computing Conference* (EDOC), 2011 15th IEEE International (pp. 55-64). IEEE.

Andersen, B., & Fagerhaug, T. (2006). *Root cause analysis: simplified tools and techniques*. ASQ Quality Press.

Atkinson, R. (1999). *Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria*. International journal of project management, 17(6), 337-342.

Bartezzaghi, E., Spina, G., & Verganti, R. (1999). Organizzare le PMI per la crescita. Come sviluppare i più avanzati modelli organizzativi: gestione per processi, lavoro per progetti, sviluppo delle competenze. *Milano: Il Sole24Ore*.

Burattin, A. (2013). *Applicability of process mining techniques in business environments*.

Carli, G., Grandi, A (2015), *Mappatura e Modellazione dei Processi Aziendali*, Corso di Laurea Magistrale in Ingegneria Gestionale, Cambiamento organizzativo e progettazione dei processi aziendali M (Università di Bologna).

Cook, J. E., & Wolf, A. L. (1998). *Discovering models of software processes from event-based data*. ACM Transactions on Software Engineering and Methodology (TOSEM), 7(3), 215-249.

Da Silva, L. F. N. (2014). *Process Mining: Application to a case study*.

Daniel, F., Barkaoui, K., & Dustdar, S. (Eds.). (2012). *Business Process Management Workshops: BPM 2011 International Workshops*, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers (Vol. 99). Springer.

Davenport, T. H. (2013). Process innovation: reengineering work through information technology. Harvard Business Press.

Dawson, M. (2003). *The consumer trap: Big business marketing in American life*. University of Illinois Press.

De Leoni, M., & van der Aalst, W. M. (2013). *Aligning event logs and process models for multi-perspective conformance checking: An approach based on integer linear programming*. In Business Process Management (pp. 113-129). Springer Berlin Heidelberg.

Dence, R. (1995). *Best practices benchmarking*. Holloway, JL, Performance Measurement and Evaluation, Sage Publishers, London, 124-52.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management*. Heidelberg: Springer.

Franz, P., Kirchmer, M., & Rosemann, M. (2011). *Value-driven business process management—impact and benefits*.

Freund, J., & Rücker, B. (2012). *Real-life BPMN: Using BPMN 2.0 to analyze, improve, and automate processes in your company*.

Gehrke, N., & Werner, M. (2013). *Process mining*. Das Wirtschaftswachstum, 42 (7), 934-943.

George, M. O. (2010). The lean six sigma guide to doing more with less: cut costs, reduce waste, and lower your overhead. 75-76. John Wiley & Sons.

Ghose, A., & Koliadis, G. (2007, September). Auditing business process compliance. In International Conference on Service-Oriented Computing. Springer Berlin Heidelberg.

Grandi, A. (2015). *Change Management*, Corso di Laurea Magistrale in Ingegneria Gestionale, Cambiamento organizzativo e progettazione dei processi aziendali M (Università di Bologna).

Grandi, A. (2015). *Progettazione del processo (I): Principi di progettazione per processi*, Corso di Laurea Magistrale in Ingegneria Gestionale, Cambiamento organizzativo e progettazione dei processi aziendali M (Università di Bologna), Materiale didattico in parte adattato da slide prodotte da Macrì D.M. (Università di Modena e Reggio Emilia) e Cagliano R. (Politecnico di Milano).

Grandi, A. (2015). *Organizzazione per processi e BPR Richiami teorici e metodologici*, Corso di Laurea Magistrale in Ingegneria Gestionale, Cambiamento organizzativo e progettazione dei processi aziendali M (Università di Bologna), Materiale didattico in parte adattato da slide prodotte da Macrì D.M. (Università di Modena e Reggio Emilia) e Cagliano R. (Politecnico di Milano).

Günther, C. W., & Van Der Aalst, W. M. (2007, September). *Fuzzy mining-adaptive process simplification based on multi-perspective metrics*. In International Conference on Business Process Management (pp. 328-343). Springer Berlin Heidelberg.

Hammer, M. (1990). *Reengineering work: don't automate, obliterate*. Harvard business review, 68(4).

Hammer, M. and Champy, J. A. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*, Harper Business Books, New York, 1993. ISBN 0-06-662112-7.

Head, S. (2005). *The new ruthless economy: work & power in the digital age*. Oxford University Press.

Hommes, L. J. (2004). *The evaluation of business process modeling techniques*. TU Delft, Delft University of Technology.

Hornix, P. T. (2007). *Performance analysis of business processes through process mining*. Master's Thesis, Eindhoven University of Technology.

Hounshell, D. (1985). *From the American system to mass production*, 1800-1932: The development of manufacturing technology in the United States (No. 4). JHU Press.

Hussein, B., Bazzi, H., Dayekh, A., & Hassan, W. (2013). *Critical analysis of existing business process reengineering models: towards the development of a comprehensive integrated model*. Journal of Project, Program & Portfolio Management, 4(1), 30-40.

Jadhav, S. (2011). *Business Process Discovery*. BP Trends (February).

Jansen-Vullers, M. H., Loosschilder, M. W. N. C., Kleingeld, P. A. M., & Reijers, H. A. (2007). Performance measures to evaluate the impact of best practices. In *Proceedings of Workshops and Doctoral Consortium of the 19th International Conference on Advanced Information Systems Engineering (BPMDS workshop)* (Vol. 1, pp. 359-368). Trondheim: Tapir Academic Press.

Joosten, S. (2000) *Why modellers wreck workflow innovations*. In: Intl. Conf. on BPM. Volume 1806 of LNCS. Springer 119-131.

Kanigel, R. (2005). *The one best way: Frederick Winslow Taylor and the enigma of efficiency*. MIT Press Books, 1.

Knuplesch, D., Reichert, M., Fdhila, W., & Rinderle-Ma, S. (2013). *On enabling compliance of cross-organizational business processes*. In Business Process Management (pp. 146-154). Springer Berlin Heidelberg.

La Rosa, M. (October 2015). *Our definition of a business process*, Queensland University of Technology, protected link.

La Rosa, M. (September 2015). *Roles in the BPM lifecycle*, Queensland University of Technology, protected link.

La Rosa, M. (September 2015). *The broader picture using Rummler's framework*, Queensland University of Technology, protected link.

La Rosa, M. (September 2015). *What else besides transparency?* Queensland University of Technology, protected link.

La Rosa, M. (September 2015). *What is the value of BPM*, Queensland University of Technology, protected link.

Limam Mansar, S., & Reijers, H. A. (2007). Best practices in business process redesign: use and impact. *Business Process Management Journal*, *13*(2), 193-213.
Murata, T. (1989). *Petri nets: Properties, analysis and applications*. Proceedings of the IEEE, 77(4), 541-580.

Panorama Consulting Solutions (2016), Clash Of The Titans 2016, *An Independent Comparison of SAP*, Oracle, Microsoft Dynamics and Infor.

Petri, C. A., & Reisig, W. (2008). *Petri net*. Scholarpedia, 3(4), 6477.

Process Mining Group, (2009) Math&CS department, Eindhoven University of Technology available at http://www.processmining.org/online/dottedchartanalysis.

R.S. Kaplan and D.P. Norton, *The Balanced Scorecard: Measures that Drive Performance*, Harvard Business Review (January-February 1992): 71-79.

Rashid, M. A., Hossain, L., & Patrick, J. D. (2002). *The evolution of ERP systems: A historical perspective*.

Reijers, H. A., Limam, S., & Van Der Aalst, W. M. (2003). *Product-based workflow design*. Journal of Management Information Systems, 20(1).

Rigby, D. K. (2015). *Management tools 2015: an executive's guide*. Boston, MA, Bain & Company.

Rozinat, A. A. (2010). *Process mining: conformance and extension* (Doctoral dissertation, Technische Universiteit Eindhoven).

Rozinat, A., De Medeiros, A. A., Günther, C. W., Weijters, A. J. M. M., & Van der Aalst, W. M. (2007). *Towards an evaluation framework for process mining algorithms.* BPM Center Report BPM-07-06, BPM center. org, 10.

Settis, B. (2016), *Fordismi. Storia politica della produzione di massa*, Il Mulino.

Sinur, J., Hill, J. B. (2010). *Magic Quadrant for Business Process Management Suites*, 2010. Gartner, 18 October 2010.
Van der Aalst, W. M. (1999). *Formalization and verification of event-driven process chains*. Information and Software technology, 41(10), 639-650.

Van der Aalst, W. M. (2015). *Extracting event data from databases to unleash process mining*. In BPM-Driving innovation in a digital world (pp. 105-128). Springer International Publishing.

Van der Aalst, W. M. (2015). *Extracting event data from databases to unleash process mining*. In BPM-Driving innovation in a digital world. Springer International Publishing.

Van der Aalst, W. M. P. (2008). *Process mining: discovery, conformance and enhancement of business processes*.

Van der Aalst, W. M., & Gunther, C. W. (2007, July). *Finding structure in unstructured processes: The case for process mining. In Application of Concurrency to System Design*, 2007. ACSD 2007. Seventh International Conference on (pp. 3-12). IEEE.

Van Der Aalst, W. M., & Ter Hofstede, A. H. (2005). *YAWL: yet another workflow language*. Information systems, 30(4), 245-275.

Van der Aalst, W. M., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. (2003). *Workflow mining: a survey of issues and approaches. Data & knowledge engineering*, 47(2), 237-267.

Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Burattin, A. (2011, August). *Process mining manifesto*. In International Conference on Business Process Management. Springer Berlin Heidelberg.

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

Van Dongen, B. F., & Van der Aalst, W. M. (2004, November). *Multi-phase process mining: Building instance graphs*. In International Conference on Conceptual Modeling (pp. 362-376). Springer Berlin Heidelberg.

Van Dongen, B. F., & Van der Aalst, W. M. (2005). *Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets*. In PNCWB 2005 workshop (pp. 35-58).

Van Hee, K. M., & Reijers, H. A. (2000). *Using formal analysis techniques in business process redesign.* In Business process management, pp. 142-160. Springer Berlin Heidelberg.

Verner, L. (2004). *The challenge of process discovery*. BPM Trends, May.

Vom Brocke, J., & Rosemann, M. (2010). *Handbook on business process management*. Heidelberg: Springer.

WBI Evaluation Group tool, first published in 2007. Available at http://sitere-sources.worldbank.org/WBI/Resources/213798-1194538727144/9Final Fishbone.pdf.

Weijters, A. J., & Van der Aalst, W. M. (2003). *Rediscovering workflow models from event-based data using little thumb*. Integrated Computer-Aided Engineering, 10(2), 151-162.

Weske, M., van der Aalst, W. M., & Verbeek, H. M. W. (2004). *Advances in business process management*. Data & Knowledge Engineering, 50(1), 1-8.

White, S. A., & Bock, C. (2011). *BPMN 2.0 Handbook Second Edition: Methods, Concepts, Case Studies and Standards in Business Process Management Notation*. Future Strategies Inc.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *Machine that changed the world*. Simon and Schuster.

Appendix


HSPI MANAGEMENT CONSULTING
Process mining:
A DATABASE OF APPLICATIONS
HSPI Management Consulting
BOLOGNA | ROME | MILAN

Index

Goal

The reason for carrying out this research is to create the most complete **list of all the adoptions of process mining techniques** and to collect, directly from who has been involved, basic information about the utilization of this new methodology.
The purpose of the study is to create a database of practical cases, no matter the specific industries and the final results, with the only aim of *completeness* and *validity*.

The intended audience includes all those **researchers, business analysts, managers and companies that want to implement process mining** or simply **explore business potentials of process mining** in order to improve internal processes or to develop performance management practices.

The final goal of this knowledge endeavour is to **build awareness and confidence about process mining techniques**.

N.B. HSPI Management Consultants is a **vendor-independent** company

Scope

The cases recorded come from companies which operate **in very different industries and markets**. Specifically, six industries have been identified, as follows:

1- Service	*4- Chemical*
2- Manufacturing	*5- Utility*
3- Healthcare	*6- Construction*

The firms analysed do their businesses in many countries, characterized by **significant differences in terms of competitive scenarios, political, legal and tax contexts**. So far, the most frequent countries are:

- Australia	*- Colombia*	*- France*	*- Italy*
- Austria	*- Denmark*	*- Germany*	*- Netherlands*
- Belgium	*- Finland*	*- Indonesia*	*- Portugal*

The projects can be executed under the supervision or technical support of some partners which can be **universities, vendor firms, consulting firms**, etc., etc.

Definition

The database consists of some mandatory features (Industry, Company, Internal Process, Description, Date, Partner) related to the projects. Each element has been validated through a web-based research, taking into account all the working papers, conference acts, relevant documents or direct contacts.

1- "**Industry**" refers to the specific industry in which the company operates.

2- Inside the industry category "**Service**", another sub classification in present (e.g. service-banking, service-logistic, etc.).

3- "**Company**" indicates the name of the firm where process mining techniques were adopted in a project.

4- " **Process**" is the specific process - or group of activities - in which process mining was used.

5- "**Description**" contains a brief overview of the project: the context, objectives and results (if present).

6- "**Date**" indicates the year (or months, weeks) in which the project was conducted. If not present, the date was assumed to be the same of the working paper or the conference in which the cases were described the first time.

7- "**Partner**" refers to the institution (university, firm) or the resource (researcher, scientist) that has supported the organization during the project.

Database

Industry	Company	Process	Description (goals, results)	Year (duration)	Partner
Service - IT infrastructure	IT Service Provider company (Italy)	Help Desk	This case study applied process mining techniques to help desk data collected from the company to expose performance issues. The study involved process discovery and comparison of execution traces associated with various cohorts of customers including (i) requiring assistance, (ii) presenting malfunctioning with their system, and (iii) requiring changes. Results: - Real process map was identified - Origins of bottlenecks and re-loops were detected. - More transparency about the ticket processes was obtained in order to improve customer orientation of its Service Desk.	2016 (4 months)	Queensland University of Technology (Australia); HSPI Management Consultants (Italy)
Service - Public	Copyright mediator company (Italy)	Event Licence Approval	This case study applied process mining techniques to event licence approval process to expose deviations and performance issues. Specifically, the study involved process discovery of the "as is" model and the conformance checking of the "as is" process to the expected process. Aim of the project: finding out the root of the problem that was affecting company's core processes. Results: - Anomalies and bottlenecks were clearly detected. - Was found that core processes didn't perform well because of the lack of quality data and transparent communications.	2016 (4 months)	Queensland University of Technology (Australia); HSPI Management Consultants (Italy)
Manufacturing	Production Company in a B2B environment	Production Process	Goal: to reduce throughput time from 3 months to 1 months (later further reduction was anticipated). Process mining was used for: - Analysing bottlenecks, which revealed also unexpected ones. - Measuring “as-is” situation (throughput/waiting time per resource). Improvements to be performed based on event data were identified.	2016 (ongoing)	Novo Consilium B.V (Netherlands)

Service - Logistic	Smart Coat Inc. (Belgium)	Logistic	- Real business processes discovery. - Removing unnecessary and divergent process activities. - Benchmarking various departments, plants, products or sales channels. - Identification of the bottlenecks, predicting and preventing process errors. - Visualizing the interactions among the employees. - Reporting the exact cost prices of activities.	2016	Horsum - Accelerating technology companies (Belgium)
Service - Telecom.	Telefónica (Spain)	Digital Operations	- Identification of the sources of delays, inefficient communication patterns, and bad practices such as work orders performed out of the scheduled window. As a result, improvements could be made with measurable effects on both the operation costs and the quality of the services.	2016	Fluxicon (Netherlands)
Service - Logistic	Deutsche Post DHL Group (Germany)	Audit	Integration of process mining into DHL's audit process in order to improve both the time spent for the analysis and the depth of the information audited. - They found that process mining helps to reduce the audit time by 25% in comparison to classical data analytics. In addition, they are now able to identify unknown risks in processes, which helps to add more value to the audits.	2016	Fluxicon (Netherlands)
Service - Housing	Zig Websoftware (Netherlands)	Housing allocation process	- Improving the housing allocation process. Every day that a rental property is vacant costs the housing association money. After process mining analysis, these vacancy costs could be reduced by 4,000 days within just the first six months.	2016	Fluxicon (Netherlands)
Utility	SPARQ Solutions (Australia)	Root case analysis	- Improving of the operations. - Discovering the actual problems and involving relevant resources in the root cause analysis. - Analysing the overall dispatching process as well as the maintenance process for a single machine.	2016	Fluxicon (Netherlands)
Manufacturing	Zimmer Biomet (Switzerland)	Value Stream Mapping	- Creating the value stream mapping with a process mining-based analysis of the manufacturing flow in a easier and effective way.	2016	Fluxicon (Netherlands)

Service - Credit	ALFAM Credit (Netherlands)	Sales Process	They analysed variation, re-processing, waiting times, and service levels. - By visualizing the processes and the process problems, improvement opportunities were designed in a powerful way.	2016	Fluxicon (Netherlands)
Service - IT infrastructure	Dimension Data (South Africa)	Compliance	Each region was responsible for running their own operations with very little enforced standards from a group perspective. The changing business landscape made it necessary for Dimension Data to standardize all their processes across all continent. Process mining was used in order to support the project.	2016	Fluxicon (Netherlands)
Service - Transport	Brisbane Airport Corporation (Australia)	NDA	Not started yet	2016	Queensland University of Technology (Australia)
Service - Telecom.	Vodafone (UK)	Process Improvement	Process Mining facilitated Vodafone's existing SAP infrastructure enabling continual real-time analytics and seamless transition to new process mining functions. Vodafone mentions that process mining also enabled faster GTM: they could resolve things faster and more proactively because they gained more visibility into their processes and operations	2016	Celonis (Germany)
Service - IT infrastructure	Xerox (India)	Process Optimization	Xerox is currently starting to use process mining in order to develop new technology projects. The focus of these projects will be on analysing complex business processes, designing cost and performance optimized policies for execution, monitoring, and identifying scope for process improvements.	2016	Xerox Algorithms & Optimization group
Service - Insurance	(Employee Insurance Agency (Netherlands)	Correlation analysis	Authors presented such a framework and its implementation in ProM by defining an analysis use-case composed of three elements (one dependent characteristic, multiple independent characteristics and a filter), in order to create a classification or regression problem.	2016	Eindhoven University of Technology (Netherlands)

Service - IT infrastructure	Basware (Finland)	Invoice Management	The analysis was conducted for the invoicing process and for a number of selected customers, all using Basware's invoicing system. They managed to analyse the number of open invoices in order to make comparisons between different invoice types or vendors.	2015	QPR (Finland)
Service - Retail	EDEKA (Germany)	IT Service Management	The Aim: gaining a scalable on-demand visualization of processes to fully exploit the hidden potential of the ticket data, for optimizing the efficiency and thereby costs of the process. Results: - Quick identification of sources of errors and deviations from the to-be process. - Better workforce planning based on the number of incidents in a given period. - Supporting the standardization of the process.	2015	Celonis (Germany)
Healthcare	AMC Hospital (Netherlands)	Conformance analysis from Billing system	- Mining the complex hospital processes giving insights into the process. - Deriving the understandable models for large groups of patients. - Comparing results with a flowchart for the diagnostic trajectory of the gynaecological oncology healthcare process.	2015	Eindhoven University of Technology (Netherlands)
Service - Transport	Dockwise (Netherlands)	Procure-to-Pay processes	By using Process Mining, Dockwise was able to: - Discover that 15% of the orders go through a different process. - Determine that are not always adhered to certain rules and procedural arrangements. - Optimize the quality and usefulness of the KPIs. - Create business cases for improvement based on facts. - To prepare the BI environment for the use of Process Mining.	2015	Zuiver ICT (Netherlands)
Healthcare	Atrium Hospital (Germany)	Conformance analysis	Process Mining was used in many ways, in order to obtain the following results: - Visualizing the pathway "Malignant Lymphoma". - The duration of the different patients was easily fixed and then analysed. - Finding the difference in fixed times for patients in whom a case manager was involved.	2015	Zuiver ICT (Netherlands)

Manufacturing	Veco (Netherlands)	Quality Management (Six Sigma Analysis)	Veco is a precision metal manufacturer. With more than 15 years of experience in supply chain management, Joris is the operations manager and Six Sigma expert at Veco. He used Minitab to statistically analyse the processes and to drive improvements. According to him, Process mining can leverage the human process knowledge in a powerful way that classical Six Sigma analyses can't.	2015	Fluxicon (Netherlands)
Healthcare	Radboudumc (Netherlands)	Process improvement	Radboud University Medical Centre is an academic hospital that was quite advanced in the adoption of electronic patient record systems, but process analysis and improvement remained as big a challenge as in all other hospitals as well. Process mining gave advantages to the improvement of healthcare processes based on the example of the Intensive care unit and the Head and Neck Care chain at Radboudumc.	2015	Fluxicon (Netherlands)
Service - Credit	DUO (Netherlands)	Process improvement of finance requenst	The new system was introduced with the goal to improve the speed of DUO's student finance request handling processes and to save 25% of the costs. Process mining was used to uncover technical errors in the pilot phase of a new system, as well as to gain transparency in the business KPIs for the new process.	2015	Fluxicon (Netherlands)
Service - Public	LatinAmerican University (Colombia)	Risk evaluation	- Quantifying the level of financial risk associated with each IT service supporting the business process, taking into account different scenarios. - Measuring the expected incomes of business processes, the probability for IT threats, and the changes on the performance of its quality attributes. - Analysing historic events to quantify the impact of IT failures in relation to different time horizons and desired confidence levels.	2015	Systems and Computing Engineering Department, School of Engineering, Universidad de los Andes, Bogota (Colombia)
Service - Public	Opéra de Lausanne (Switzerland)	Users accesses analysis	Analysis of the users accesses of the Opera's Storage Area Network (SAN) in order to refine the organisation of the SAN.	2015	Ville de Lausanne (Switzerland)

Service - Telecom.	Telecomm. Company (Indonesia)	Customer fulfilment analysis	The aim of process mining implementation was to discover the typical customer fulfilment business process. It was also aimed at assessing the current rate of completed customer fulfilment. - The company could use the findings as a foundation to improve their business process. First, the completion rate of the customer requests was found to be very low deserved further investigation. Then, findings regarding typical processes could be used to set standard sets of services which will be useful for prediction and planning of capacity.	2015	Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya, (Indonesia)
Healthcare	General Hospital of Valencia (Spain)	Health Process Tracking	Process mining techniques provided an easy to use way to achieve a view of the deployed process. The algorithm perfectly captured the features of the processes, showing them in an easy and understandable view that was accepted by the medical staff in a real environment. - With this information, the health professionals and managers could achieve a real view of the problems that are currently happening. This enabled the improvement of protocols with a better knowledge of the problems, increasing their efficiency and the probability of success.	2015	Instituto Universitario de Investigación de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politecnica de Valencia (Spain);
Healthcare	Toulouse Hospital (France)	Outpatient clinic redesign	Toulouse Hospital decided to redesign an outpatient clinic in order to mutualize the 11 consulting services of 6 medical specialties. - Process Mining clearly appeared as a good solution to support continuous improvement of complex and continuous (24/24) hospital processes. Furthermore, it became a relevant tool in diagnosis phase and also to monitor activities.	2015	Toulouse Hospital (France)
Healthcare	Integrating the Healthcare Enterprise (IHE)	Audit and node authentication	Integrating the Healthcare Enterprise (IHE) defines in its Audit Trail and Node Authentication (ATNA) profiles how real-world events must be recorded. Since IHE is used by many healthcare providers throughout the world, an extensive amount of log data is produced. In the research they investigate if audit trails, generated from an IHE	2015	Integrating the Healthcare Enterprise (IHE)

			test system, will carry enough content to successfully apply process mining techniques. Furthermore they assess the quality of the recorded events.		
Healthcare	Hospital in Mainkofen (Germany)	Care station for elderly people	The ward is an intensive care station for elderly people suffering from dementia and similar old-age diseases. Each of the patients needs care around-the-clock. - Discovering how real-world processes are executed. - Discovering that the process exhibits a relatively high repetition rate. The process could be documented directly and time-saving in comparison with the past.	2015	Institut für Parallele und Verteilte Systeme (IPVS) der Universität Stuttgart (Germany)
Healthcare	Isala Hospital (Netherlands)	Patients' records management	- Compliance analysis of the whole patients' records management. In average, 30 medical steps were avoided. - Reduction of the emergency management total duration.	2014	Eindhoven University of Technology (Netherlands)
Construction	Caverion (Finland)	Performance Management	- Measuring process performance based on ready defined indicators to ensure proactive actions to any discrepancies. - Quicker invoicing and improved cash flow from discovering and removing process bottlenecks. - Ability to continuously compare and value process performances and variations per country.	2014	QPR (Finland)
Manufacturing	Electronic Manufacturer (Netherlands)	Service Refund Process	- A critical bottleneck at a subcontracting forwarding company could be detected. - Discovering that additional documents were requested due to incomplete information at the beginning of the process if it was started through a particular channel. Understanding the problem could reduce this wasteful activity by more than 85% and significantly speed up the process for the customer and reducing customer complaints as well.	2014	Fluxicon (Netherlands)
Service - Insurance	Bridge Loans (South Africa)	Loan Processes	- Making right corrective actions and making the loan application process 40% faster. - Gaining a better understanding of why actual processes may differ, and measuring the performance of the system processes. - Monthly reports for comparing and analysing process performances. - Proactive process management via quick discovery of problem areas.	2014	QPR (Finland)

Service - Telecom.	Norddeutscher Rundfunk (Germany)	IT Service Management	- Analysing of Service Desk processes and building the foundation for an optimized Services Management.	2014	Celonis (Germany)
Service - Telecom.	Fiducia (Germany)	IT Service Management	Fiducia wished to implement automatic reporting. Results: - Reconstruction of the entire dataset based on HP Service. It is now possible to perform long-time evaluations and process reconstructions based on the data saved in archives of the last 2-10 years. - Using live process reconstruction, the identification and elimination of bottlenecks, long-running tickets and process inefficiencies became possible. To keep track of current trends, live monitoring dashboards were established.	2014	Celonis (Germany)
Service - Telecom.	Hessischer Rundfunk (Germany)	IT Service Management	- Analysing Service Desk processes and building the foundation for an optimized Services Management.	2014	Celonis (Germany)
Manufacturing	Siemens AG (Germany)	Service Process Management	With the continuously monitoring and analysis of new data from a multitude of SAP systems around the world, they obtained the following results: - Evidence of weak points. - Enabling constant improvement, harmonization and standardization of processes	2014	Celonis (Germany)
Healthcare	Berufsgenossenschaftliche Unfallkrankenhaus Hamburg (Germany)	Service Process Management	- Clearly assigned tasks, optimized flows of information as well as communication/collaboration across departments and occupation groups that resulted in smooth work flows, short decision making processes and individual solutions. This enabled the best possible treatment and rehabilitation of patients across all medical fields.	2014	Celonis (Germany)
Chemical	Bayer (Germany)	Process compliance	The goal of the project was to bring global transparency to the core processes (procurement, sales and logistics) in order to identity inefficiency potentials and ensure process compliance. Solution: process mining was used to reconstruct and to monitor global processes in relation to efficiency and risk beyond country, system and company	2014	Celonis (Germany)

			borders. Now, processes as well as performance and risk indicators can be dynamically analysed by users.		
Healthcare	Kliniken Südostbayern (Germany)	Performance management for medicinal treatment	The Kliniken Südostbayern decided to use the Process Mining solution for hospital management as a tool to obtain all the needed information. With Process Mining was possible to extract all necessary data from hospital information system (HIS) and to provide a detailed view of treatments.	2014	Celonis (Germany)
Service - Banking	DZ-Bank (Germany)	Process data analysis	Business processes require the highest possible level of transparency. Especially in the banking sector, the analysis of process data from source systems plays a very important role. Employees of the banking sector work with IT-systems every day, for example in relation to electronic files, creating process data continuously. Process mining enabled the improvement of the analysis required.	2014	Celonis (Germany)
Manufacturing	IG Metal (Germany)	IT Service Management	IG Metall placed high expectations on its customer service and internal IT Service Management. That's why the IG Metall had opted for the use of Process Mining. Process Mining made it possible to significantly improve efficiency and quality in the handling of customer requests by visualizing how inquiries are being processed in reality and thus uncovering process weaknesses.	2014	Celonis (Germany)
Service - Telecom.	SWR (Germany)	IT Service Management	SWR used the software "assyst" as its service desk solution and integrated the Process Mining for IT service management in order to analyse its service processes. This enabled the company to substantially improve its Service Management.	2014	Celonis (Germany)
Service - Logistic	Schukat Electronic (Germany)	Order Process management	As a catalogue distributor of electronic components, the company put high emphasis on the ability to deliver and processing orders. Since complete transparency of business processes is also an important component of constant optimization, Process Mining is now part of the IT landscape of the company. In conjunction with SAP HANA, Process Mining every day creates real-time transparency over the actual processes.	2014	Celonis (Germany)

Manufacturing	Essmann Automotive (Germany)	Production process system	By using process mining the, Eissmann created efficient production processes. Its integration into lean corporate processes and into the management system, represented a key pillar of long-term success. By involving suppliers at an early stage in the product creation process and promoting team-oriented partnerships, Eissmann achieved excellent and competitive products.	2014	Celonis (Germany)
Service - Public	Centraal Bureau voor de Statistiek (Netherlands)	Statistic analysis	Statistics Netherlands is responsible for collecting and processing data in order to publish statistics to be used in practice, by policymakers and for scientific research. With process mining they have improved their performances.	2014	Fluxicon (Netherlands)
Service - Credit	ING (Netherlands)	Website and call center improvements	Making sure a customer has the best possible experience when interacting with the company, is one the most important goals many companies strive for. ING DIRECT Australia asked for an in-depth analysis of the behaviour of their customers on their website before they called the call center. Using process mining they were able to get valuable business insights to make better decisions on how to further develop both their website and call center.	2014	Fluxicon (Netherlands)
Manufacturing	Fewzion	NA	NA	2014	Queensland University of Technology (Australia)
Service - Public	Ville de Lausanne (Switzerland)	Construction permit process	Analysis of the construction permit process, in order to find bottlenecks.	2014	Ville de Lausanne (Switzerland)
Manufacturing	Volvo (Germany)	Paths discovery	A first dataset provided data about factory orders for the construction of trucks. The second dataset contained customer orders of trucks. It was discovered that the attribute 'ORDERNUMBER' of any event in a trace of the first log was also displayed in the attribute 'Omnumber' of the event 'Accepted' in the second.	2014	Ghent University, (Belgium)

Healthcare	Scottish Rite Emergency Department of Children's Healthcare of Atlanta (US)	Paediatric asthma emergency department (ED) processes	Process mining's visual analytics has played an important role in healthcare process analysis. The interactive visual approach enabled users to gain insight into the complexity of paediatric asthma care pro-cesses. - It helped with care quality improvement programs, providing comparison, benchmarking and analysis of conformance to existing care protocols.	2014	School of Interactive Computing & Tennenbaum Institute, Georgia Institute of Technology (US)
Healthcare	EU project's MOSAIC	Datasets of Type 2 Diabetes analysis	Process mining methods were executed in order to derive healthcare pathways. The approach started by processing raw data, derived from heterogeneous data sources, and created event logs, which contained meaningful healthcare activities. Once event logs have been obtained and tasks and transitions defined, it was possible to explore how state-of-art process mining techniques could be used to gain insights into patients care.	2014	International Conference on Biomedical and Health Informatics (Spain)
Healthcare	Chicago Outpatient Clinic (US)	Analysis of Workflows in Clinical Care	Process Mining was used for workflows analysis for outpatient clinic center, admitting high-risk patients and low-risk patients. Based on the results from process mining, a discrete event simulation model was proposed to quantitatively analyse the clinical center. Sensitivity analyses have also been carried out to investigate the care activities with limited resources such as doctors and nurses. The results suggested that the methodology was a useful tool for healthcare process improvement.	2014	Department of Mechanical and Industrial Engineering, University of Illinois at Chicago (US)
Manufacturing	A steel manufacturer (UK)	Process improvement	Objectives: - To investigate the flows of material through the route. - To get insights and knowledge on the approach by using internal data only. Results: - Identification of some issues with the flows. - Discovering that the large number of processes actually undertaken. - Identification of the issues in relation to the interpretation of the process.	2014	A steel manufacturer (UK)
Service - Insurance	Suncorp (Australia)	Home Insurance Claim	This project aimed to apply process mining to Home Insurance Claims, processing records provided by Suncorp with the aim of finding insights into the reasons behind long processing times.	2013 (6 months)	Queensland University of Technology (Australia)

			Results: - Evidence of two major loops, which represented bottlenecks for the entire process. - Processing time reduction from 30-60 days to 5 days (within the SLA conditions).		
Service - Transport	Copenhagen Airports A/S (Denmark)	Bag-tag Analysis	- Identification of the reasons for KPI discrepancies. - Finding areas with potential process challenges more in depth. - The easy and fast way of looking at the process from different perspectives revealed many new insights. - The perspective could shift from KPIs and bottlenecks, to process performance related to locations.	2013	Fluxicon (Netherlands)
Construction	Ruukki (Finland)	Process Management	Process mining was used to respond to the needs of both system management and business. For System Owners: - Gaining transparency to system usage and enabling more focused guideline enforcement and modifications. For Business: - Supporting prioritization of process improvement activities. - Highlighting the importance of transparent process management over functional siloes.	2013 (3 weeks)	QPR (Finland)
Manufacturing	Samsung Electro-Mechanics (South Korea)	Conformance analysis and machine performance analysis	Results: - The derived process model showed real process flows in the factory and it was used to understand the manufacturing process. - The conformance checking showed how traces fit with the derived model. - The machine performance analysis showed the utilization of their resources. - The analysis results were presented to the managers of SEM, who were impressed by the obtained results.	2013	Ulsan National Institute of Science and Technology (South Korea)

Healthcare	Seoul National University Bundang Hospital (South Korea)	Performance Analysis per patient type	A performance analysis was conducted in order to make a simulation model and to analyse the process patterns according to patient types. The results: - According to the result of comparing the event log and their standard process model, the matching rate was as 89.01%. - Using the performance analysis result, they generated the simulation model. The simulation showed that the 10% increase of patients made the largest change in consultation waiting time. - Extraction of the process models and analysis of process patterns according to patient types. The most frequent pattern of each patient type was discovered.	2013	Ulsan National Institute of Science and Technology (South Korea)
Service - Transport	Ana Aeroports de Portugal (Portugal)	Service Process Management	- Finding a more effective method to balance the workforce. - Changing the process to be much more lean for particular technical categories. - Eliminating non-value-add tasks. - Identifying unambiguous performance metrics for the process. - Making changes in the way “Change Orders” are created and recorded in order for technical people to focus on what is really important and improve how they identify execution priorities. - Applying the same practices to other ITIL processes. - Making sure that no “Change Order” was implemented without being previously authorized.	2013	Process Sphere - End to end BPM
Manufacturing	IBM i (US)	Database management	For IBM i users, the event data were perhaps most prolific and most commonly available in database journals. This provided an event log of potentially thousands to millions of database events related to the files journaled. - Setting Journaling Parameters. - Extracting Journal Data. - Creating the animation.	2013	IBM
Service - Public	Auditdienst Rijk (Netherlands)	Assurance on the financial statements	The Dutch National Auditing Service monitors the annual reports of all Dutch ministries and provides assurance on the financial statements that are included. They used process mining in order to perform their audits in a more efficient way.	2013	Fluxicon (Netherlands)

Healthcare	St Andrew's War Memorial Hospital (Australia)	Emergency Department Patient Treatment	This project aimed to apply process mining to provide insights into St Andrew's War Memorial Hospital's (SAWMH) process for treating patients presenting at the Emergency Department (ED) with chest pain. The study involved process discovery and comparison of patient flows associated with patients whose stay in ED was less than 4 hours, with those whose stay was longer than 4 hours. The study also aimed to investigate potential delays introduced to the patient flows as a result of conducting routine clinical activities and the determination of factors that influence patients' length of stay in ED.	2013 (1 year)	Queensland University of Technology (Australia)
Healthcare	Seoul National University Bundang Hospital (South Korea)	Modeling of Outpatient Care Mega-Process	- From a total of 698,158 event logs, the most frequent pattern was found. - The matching rate between the expert-driven process model and the machine-driven model was found to be approximately 89.01%. - Was found that process mining techniques could be applied in the healthcare area, and through detailed and customized analysis in the future it can be expected to be used to improve actual outpatient care processes.	2013	Industrial Strategic Technology Development Program funded by the Ministry of Knowledge Economy, Korea (MKE).
Manufacturing	Boxes manufacturing unit, (India)	Process modelling and improving	- The generated process model reflected the actual process as observed through real process executions. - The heuristic mining algorithm gave clear information on how the process was executed, and analysis of the process could be evaluated to improve the performance of manufacturing.	2013	Department of Computer Science and Applications, Dayananda Sagar College of Arts (India); Science and Commerce, (India)
Healthcare	Dentistry (Netherlands)	Patient treatment	For a complex dental process, it turns out that the introduction of new digital technologies is largely beneficial for patients and dental lab owners, whereas for dentists there is hardly any benefit. - The effects of digital dentistry on the implant value chain was investigated using process mining and discrete event simulation. The implant value chain was concerned	2013	Eindhoven University of Technology (Netherlands); Perceptive Software, (Netherlands)

			with all steps that could be associated with dental implants, covering the stages from patient diagnosis until implant placement.		
Healthcare	Children's National Medical Center (Columbia)	Adherence to ATLS protocol analysis	- Through process mining they determined compliance with the ATLS protocol sequence, reviewed the most commonly occurring sequence and individual deviations, detected differences in clinical behaviour after the introduction of the checklist. Although the frequency of activations without notification was not reduced, the addition of the checklist to the trauma resuscitation routine helped to standardize the care provided specifically for these events.	2013	American College of Surgeon's 99th Clinical Congress, Surgical Forum (US)
Healthcare	IBM T. J. Watson Research Center, NY, (US)	Paths discovery	Identifying care pathways correlated with outcomes from patient event data were of vital importance for gaining the insights of which specific care pathway will lead to a good/bad outcome. Once identified, such care pathways were used by medical boards for refining care plan descriptions for treating particular diseases such as congestive heart failure etc.	2013	IBM
Healthcare	Gynecologic Oncology Department (Belgium)	Patient treatment deviation analysis	Using process mining techniques, that research has demonstrated that the patients' diagnosis-treatment cycles often significantly deviate from the standardized clinical pathways. - Analysing these deviations might result in the further enhancement of the quality of care, the promotion of patient safety, an increase in patient satisfaction. - Understanding pathway behaviour and deviations became possible.	2013	Department of Decision Sciences and Information Management, Faculty of Economics and Business (Belgium)
Healthcare	37 hospitals located in the Lombardia Region (Italy)	Patient treatment	The work showed that process mining and case retrieval techniques can be applied successfully to clinical data to gain a better understanding of different medical processes for different groups of patients). In this way, not only different practices used to treat similar patients may be discovered, but also unexpected behaviour may be highlighted.	2013	Computer Science Institute, Università del Piemonte Orientale, Alessandria (Italy); Dipartimento di Informatica e Sistemistica, Università di Pavia (Italy)

Healthcare	Maastricht University Medical Centre (Netherlads)	Patient routes in a medical Treatment process	The study used both the heuristic and the fuzzy miner for the process analysis. It was concluded that the heuristics miner is not able to show all low frequent behaviour which makes it difficult to use for extension/improvement research in the medical domain. The fuzzy miner is able to show this behaviour but must be accompanied by the Conformance Checker to make sure that all discrepancies are found between the original process and the acquired event log.	2013	Eindhoven University of Technology, University of Technology (Netherlands); Maastricht University Medical Centre (Netherlands)
Healthcare	Many European academic hospital	Diagnosis treatment cycle	Research demonstrated that the patients' diagnosis-treatment cycles often deviate from the approved and standardized clinical pathways. By Studying these differences may result in the further improvement of the quality of services, the promotion of patient safety, an increase in patient satisfaction and an optimization of the use of resources. Understanding pathways behaviour and deviations becomes possible because of an increased availability of reliable data logs, originated from every hospitals information systems.	2013	Department of Decision Sciences and Information Management, Faculty of Economics and Business, Leuven (Belgium)
Healthcare	Dutch Clinic (Netherlands)	Ambulant surgery process	It was used the log of a Dutch clinic for the ambulant surgery process. This is a sequential process that deals with both ambulant patients and ordered stationary patients.	2013	Eindhoven University of Technology, University of Technology (Netherlands)
Healthcare	Chania Hospital (Greece)	Clustering healthcare processes	The aim was to support decision making by providing comprehensible process models in the case of such flexible environments. Following a process mining approach, they proposed a methodology to cluster customers' flows and produce effective summarizations. Then, they proposed a novel method to create a similarity metric that was efficient in downgrading the effect of noise and outliers. It was used a spectral technique that emphasized the robustness of the estimated groups, therefore it provided process analysts with clearer process maps.	2013	Eastern Macedonia and Thrace Institute of Technology (Greece); Technical University of Crete, University Campus (Greece)

Service - Logistic	Package Delivery Company (Belgium)	Machine Configuration	- Evaluation of the correctness of the configuration of the state machine. - Investigation of a huge number of abnormal flows that have been identified by business users. - Linking the different states and events back to the business process.	2012 (2 months)	AE architects for business & ICT (Belgium)
Utility	Alliander (Netherlands)	Purchasing	The Challenges: - Managers thought they knew how processes worked, but it wasn't true. - To find a solution that can be applied to multiple processes and departments. - To discover areas of inefficiency. - To find why staff members had different ways of completing the same process. The Results: - Quick insight into how processes really worked. - Extensive list with potential areas for improvement. - Improved process insight delivers efficiency improvements. - A complete picture of eight business processes, that allowed for standardisation and staff re-training, was provided.	2012	Perceptive Process Mining, Lexmark (Kentucky)
Manufacturing	AkzoNobel (Netherlands)	Procure-to-Pay processes	- Management obtained insights into exceptions where the "First time right" principle was not realized. - Peer comparisons between countries helped to identify best practices that could be adopted on the corporate level. - The direct insights in process improvements enabled the desired "value extraction" from the P2P processes. - Compliance control was realized to execute on corporate guidelines that must be followed.	2012	Capgemini
Manufacturing	Vaisala (Finland)	Process Management	- Making effective operations and improved customer satisfaction through clear visual understanding of the real process and the deviations. - Reducing operational costs and time to corrective actions by having the means for effective change management, through fast verification and follow-up of process changes. - Sales process optimization through understanding of the process flows, and the ability to benchmark performances.	2012	QPR (Finland)

Service - Telecom.	WDR (Germany)	IT Service Management	NA	2012	Celonis (Germany)
Service - Banking	Bank of Queensland, (Australia)	NDA	NDA	2012 (3 months)	Queensland University of Technology (Australia)
Healthcare	Princess Alexandra Hospital (UK)	Emergency Department Patient Treatment	This project aimed to apply process mining to provide insights into Princess Alexandra Hospital's (PAH) "as is" processes for treating patients presenting at the Emergency Department with multiple traumatic injuries. The study involved process discovery and comparison of patient flows associated with various cohorts of patients including (I) patients presenting with minor and major trauma, and (II) patients presenting at different times of the day.	2012 (1 year)	Queensland University of Technology (Australia)
Healthcare	Metro North Hospital (Australia)	Outpatient Referral Process	Process mining was used to discover models from patient referral, appointment, to provision of a service with a specialist. Key insights obtained from this analysis include lack of implementation of the standards across hospitals, variation and impact of delays on the health and well being of patients.	2012	Queensland University of Technology (Australia)
Service - Public	Queensland University of Technology (Australia)	Student Services	The project aimed to apply process-oriented data mining (process mining) to analyse student behaviour (through the use of Blackboard data) in order to increase student retention.	2012 (6 months)	Queensland University of Technology (Australia)
Service - Retail	Woolworths (Australia)	Logistic Process	The project aimed to apply process-oriented data mining (process mining) to provide insights into Woolworths' Delivery Process. Specifically, the study involved process discovery of the delivery process, and the identification of optimal delivery routes.	2012 (3 months)	Queensland University of Technology (Australia)
Healthcare	Hospital of Sao Sebastiano, (Portugal)	Emergency Services	The process mining methodology was applied in the emergency service of a hospital that had its own electronic patient record system, developed in-house. Event data collected from this system was analysed. Using the radiology workflow as an example, they showed how the proposed methodology could provide insight into the flow of	2012	Hospital de Sao Sebastiao, Santa Maria da Feira, (Portugal); Technical University of Lisbon (Portugal)

			healthcare processes, their performance, and their adherence to institutional guidelines.		
Healthcare	Medical University of Vienna (Austria)	Compliance analysis for treatment processes	The goal of the project was to analyse skin cancer treatment processes regarding their compliance with relevant guidelines. Focus was put on the transformation and integration of the available data sources as well as billing data of the Main Association of Austrian Social Security Institutions. The challenge was to extract and integrate the data in a process-oriented way in order to apply process mining techniques in the sequel.	2012	University of Vienna (Austria); Medical University of Vienna (Austria)
Healthcare	Department of Ophthalmology at the University Hospital of Leipzig (Germany)	Workflow Management	The objective was the design and the implementation of a surgical workflow management system (SWFMS) that could provide a robust guidance for surgical activities. Result: - They demonstrated that a SWFMS with a workflow schema that was generated from a subset of 10 patient individual surgical process models (iSPMs) was sufficient to guide approximately 65% of all surgical processes in the total set, and that a subset of 50 iSPMs was sufficient to guide approx. 80% of all processes.	2012	University of Leipzig, Innovation Center for Computer Assisted Surgery (ICCAS) (Germany); University Hospital of Leipzig Department of Ophthalmology and Neurosurgery (Germany);
Healthcare	Dutch Academic Hospital (Healthcare)	Analysis of Patient Treatment Procedures	Given the heterogeneous nature of the cases, the research first demonstrated that it was possible to create more homogeneous subsets of cases (e.g., patients having a particular type of cancer that need to be treated urgently). Such pre-processing was crucial given the variation and variability found in the event log. The discovered homogeneous subsets were analysed using state-of-the-art process mining approaches. More specifically, they reported on the findings discovered using enhanced fuzzy mining and trace alignment. A dedicated pre-processing ProM plug-in was developed for the challenge.	2012	Eindhoven University of Technology; Philips Healthcare
Healthcare	Royal Prince Alfred Hospital (Australia)	Patient Treatment	- Gaining Insight from HIV/AIDS Patient Journey Data by using a Process-Oriented Analysis Approach with process mining.	2012	School of Information Technologies, University of Sydney, (Australia)

Healthcare	Mercy Health System St. Louis, MO (US)	Clinical workflow management	- Automating the method of documenting clinical workflows. - Identifying variations of clinical workflows and optimizing them. - Optimizing clinical workflows.	2011	The Healthcare Business Process Management Blog
Service - Insurance	Queensland Nominal Defendant (Austrialia)	NA	NA	2011	Queensland University of Technology (Australia)
Service	Multinational enterprise	Conformance analysis	From the analysis performed it was possible to highlight that the most striking of the variations was the difference between the processes executed in each of the order management teams around the world. - The process improvement teams took this information to one of the regular meetings of representatives from the regional teams so they could present findings in order to obtain the standardization level needed.	2011	Khalifa University, P.O., Abu Dhabi, (UAE)
Service - Insurance	United India Insurance Company LTD. (India)	Paths discovery	Process mining techniques were used to obtain meaningful knowledge about flows, in order to discover typical paths followed by particular groups of Insurance holders. - Obtaining understandable mined process models for large groups of services to identify the same and different insurance holder process. - The results were not derived by human thinking: the automated mined process model helped the insurance agent in their daily activities.	2011	Bharathiar University, Technical University, Avadi, Tamil Nadu, (India); Department of Computer Science, Rashtriya Sanskrit Vidyapeetha, Tirupati, Andhra Pradesh, (India)
Service - Insurance	Motor Accident Insurance Commission (Australia)	Compulsory Third Party (CTP) Claim	The project aimed to apply process mining to historical CTP claims processing records, provided by multiple CTP insurance providers with the aim of exposing impediments to efficient (time & cost) claims handling and to determine the impact of various "context" factors on the process execution.	2011 (2 years)	Queensland University of Technology (Australia)

Service - Insurance	Association of Certified Fraud Examiners (ACFE) (US)	Transactional logs analysis	Authors presented a case study in which they applied process mining in the context of transaction fraud. Given the procurement process of an organization using SAP as ERP system, they applied the process diagnostics approach to discover the real process and to analyse flaws. - This enabled the explicit possibility of checking internal controls and business rules in more general. This way, process mining enabled auditing by not only providing theory and algorithms to check compliance, but also by providing tooling that helps the auditor to detect frauds or other flaws in a much earlier stage.	2011	Faculty of Business Economics, Hasselt University, Agoralaan, (Belgium); Eindhoven University of Technology, (Netherlands)
Healthcare	Clinical Application Domain (Hospital for Children, Toronto & Women and Infants Hospital, Providence,) (Canada, US)	Patient treatment modeling	The paper presents a framework for process mining in critical care. The framework uses the CRISP-DM model, extended to incorporate temporal and multidimensional aspects (CRISP-TDMn), combined with the Patient Journey Modeling Architecture (PaJMa), to provide a structured approach to knowledge discovery of new condition onset pathophysiologies, in physiological data streams. The approach is based on temporal abstraction and mining of physiological data streams to develop process flow mappings.	2011	University of Ontario Institute of Technology, Oshawa (Canada); The Hospital for Sick Children, Toronto (Canada) Department of Paediatrics, University of Toronto, Toronto (Canada)
Healthcare	Belgium Hospital (Belgium)	Process improvement for breast cancer patients	They analysed a dataset consisting of the activities performed to 148 patients during hospitalization for breast cancer treatment in a hospital in Belgium. They exposed multiple quality of care issues process variations, best practices and issues with the data registration system. For example, 25 % of patients receiving breast-conserving therapy did not receive the key intervention "revalidation". They found out this was caused by lowering the length of stay in the hospital over the years, without modifying the care process.	2011	NA
Service - Telecom.	Bayerischer Rundfunk (Germany)	IT Service Management	Aim: To establish an improved Service Desk control station. Results: - Ticket data analysis and telephone routing systems at an hourly rate. - Pre-defined key indicators which provided a quick overview and showed existing trends.	2010 (2 years)	Celonis (Germany)

			- Ready-made analyses of business processes. Due to real time data assessment implemented, the effectiveness of correction measures could be reviewed immediately.		
Healthcare	Dutch Hospitals (Netherlands)	Process improvement for diabetes foot patients	The project was divided in three phases: process visualization, process analysis, and evaluation. During these phases, two approaches, process mining and visual analytics were used to visualize and analyse a business case. Based on the outcomes of this, the method was developed. The main finding was the fact that process mining and visual analytics did not provide sufficient process insights. Rather, a combination of both approaches was required.	2010	Eindhoven University of Technology, University of Technology (Netherlands)
Healthcare	Verbeeten Institute (Netherlands)	Achieve standardization in healthcare processes	Goal: to help healthcare organizations achieving a standardized and high quality care process by using historic information, gathered by registering the day-to-day operations with a healthcare information system. The research project successfully evaluated the applicability of process and data mining techniques in the context of the problem definition. However, it must be stated that the unavailability of exact activity and waiting time metrics significantly restricted simulation capabilities.	2010	Eindhoven University of Technology, University of Technology (Netherlands)
Healthcare	Maastricht University Medical Centre (Netherlands)	Conformance analysis on clinical pathways	The researchers developed and tested dynamic programming formulations for adherence measurement in clinical pathways, based on partially ordered data in medical records and pathway definitions. With these new methods at hand, they analysed clinical pathway adherence at the Cardiovascular Centre of Maastricht University Medical Centre.	2010	Institute of Health Policy & Management, Erasmus Medical Centre, (Netherlands); Maastricht University Medical Centre, (Netherlands);
Service - Public	Dutch governmental organization (Netherlands)	Process Diagnostics	Authors proposed a process diagnostics methodology, that gave a broad overview of the process supported by the information system. - In the process diagnostics methodology, several perspectives of the process were highlighted. The outcome covered the control flow perspective, i.e. "how the process model actually looks like", the performance perspective, i.e. "how well does the system	2009	Eindhoven University of Technology (Netherlands)

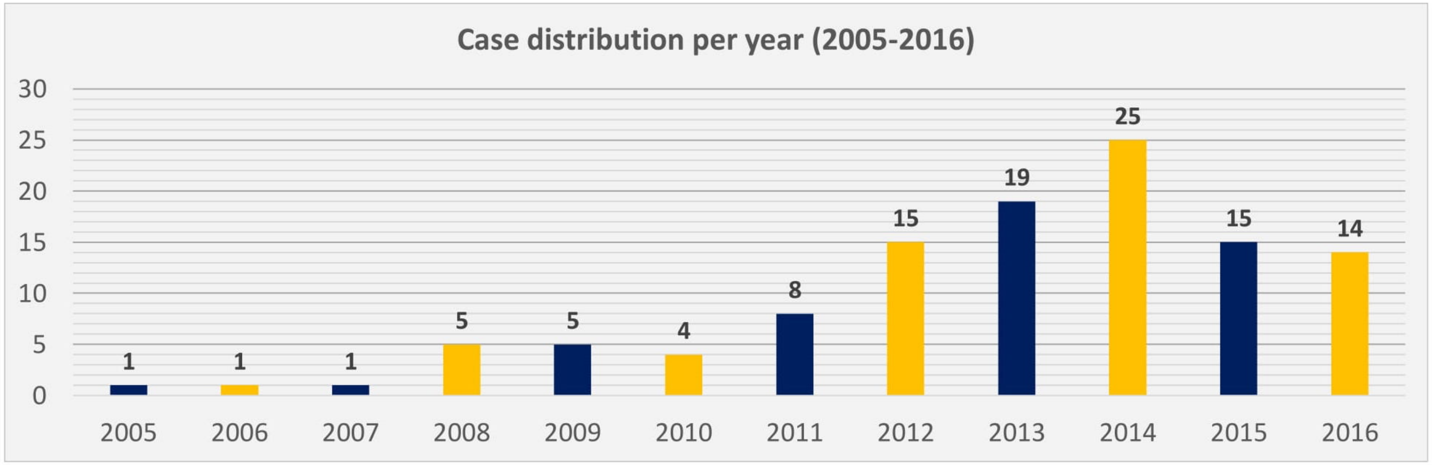
			perform" and the organizational perspective, i.e. "who is involved in the process and how".		
Manufacturing	ASML (Netherlands)	Test Processes	Authors demonstrated that current process mining techniques could answer many questions, even yield concrete suggestions for process improvement. However, due to the rapid technological advancements, the analysis results presented are likely to be outdated already for the next series of wafer scanners To enable a continuous improvement of the test process in ASML, process analysis should be best carried out in an iterative manner.	2009	IEEE transactions on systems, man, and cybernetics
Healthcare	Maxima Medical Centre (Netherlands)	Patient treatment analysis	Research objectives: to find the applicability of process mining on acquiring objective process information in the healthcare domain. Several process mining objectives were set: - Researching all the possibilities by discovering the care flow of the rheumatoid arthritis patients. - Checking if the process model discovered in the first objective corresponded to the predefined care paths and to reality.	2009	Eindhoven University of Technology, University of Technology (Netherlands)
Healthcare	University Hospital Leipzig (Germany)	Analysis of surgical intervention populations	According to differences in patient characteristics, surgical performance, or surgical technological resources used, surgical interventions had high variability. - Statistical differences between the gSPMs of ambulatory and inpatient procedures of performance times for surgical activities and activity sequences were identified.	2009	Universität Leipzig, Leipzig (Germany); Faculty of Medicine, INSERM, Rennes (France); VisAGeS Unit/Project, INRIA, Rennes, (France);
Service - Insurance	ING (Netherlands)	Internal Auditing	Process Mining empowered ING to make a difference in auditing quality by providing better focus on possible risk, control and efficiency issues. - Identification of the most complicated cases. - Finding out exceptions in processes. - Detecting policy violations and unusual transactions.	2008	Bitz Clarity LTD (UK)

			- Checking if processes, as designed, are also being executed the same way. -Verifying if internal controls, such as authorizations, are performed correctly.		
Healthcare	Erlangen University Clinic (Germany)	Business Process Analysis	In order to support the analysis of the radiology workflows at the clinic, the authors developed a data warehouse for process mining. Despite the limitations, the authors concluded that process mining has a great potential to facilitate the understanding of medical processes and their variants.	2008	International Congress of the European Federation for Medical Informatics
Healthcare	4 Italian Hospitals (Italy)	Patient treatment	Process mining was used to discover how stroke patients are treated in different hospitals. There was a need for intensive pre-processing of clinical events to build the event logs. It was concluded that process mining could be successfully applied to understand the different clinical pathways adopted by different hospitals and different groups of patients.	2008	Eindhoven University of Technology (Netherlands); University of Pavia (Italy)
Service - Public	Municipality in the Netherlands (Netherlands)	Modelling of social network and information flows	Authors addressed three issues: (I) Organizational model mining, (II) Social network analysis, and (III) Information flows between organizational entities. With a case study, they have shown how each of these issues can be supported. Moreover, they showed how organizational mining can benefit from creatively using approaches developed for the process perspective.	2008	Eindhoven University of Technology (Netherlands)
Healthcare	Zhejiang Huzhou Central Hospital (China)	Patient workflow	The study adopted process mining to analyse clinical pathways. The key contribution of the paper is to develop a new process mining approach to find a set of clinical pathway patterns given a specific clinical workflow log and minimum support threshold. The experimental results indicate the applicability of the proposed approach, based on which it is possible to discover clinical pathway patterns that can cover most frequent medical behaviours that are most regularly encountered in clinical practice.	2008	College of Biomedical Engineering and Instrument Science, Zhejiang University (China)
Healthcare	Catharina Hospital (Netherlands)	Analyse careflows of an Intensive Care Unit	The clustering approach of the DWS Algorithm was able to discover some patterns; however, the discriminants rules were hard to understand. To handle this problem, the author introduced the Association Rule Miner (ARM) plug-in, which aimed at discovering association rules and frequent item sets in the event log. The technique has proved to be useful to obtain patterns in the event log and to group similar patients.	2007	Eindhoven University of Technology (Netherlands)

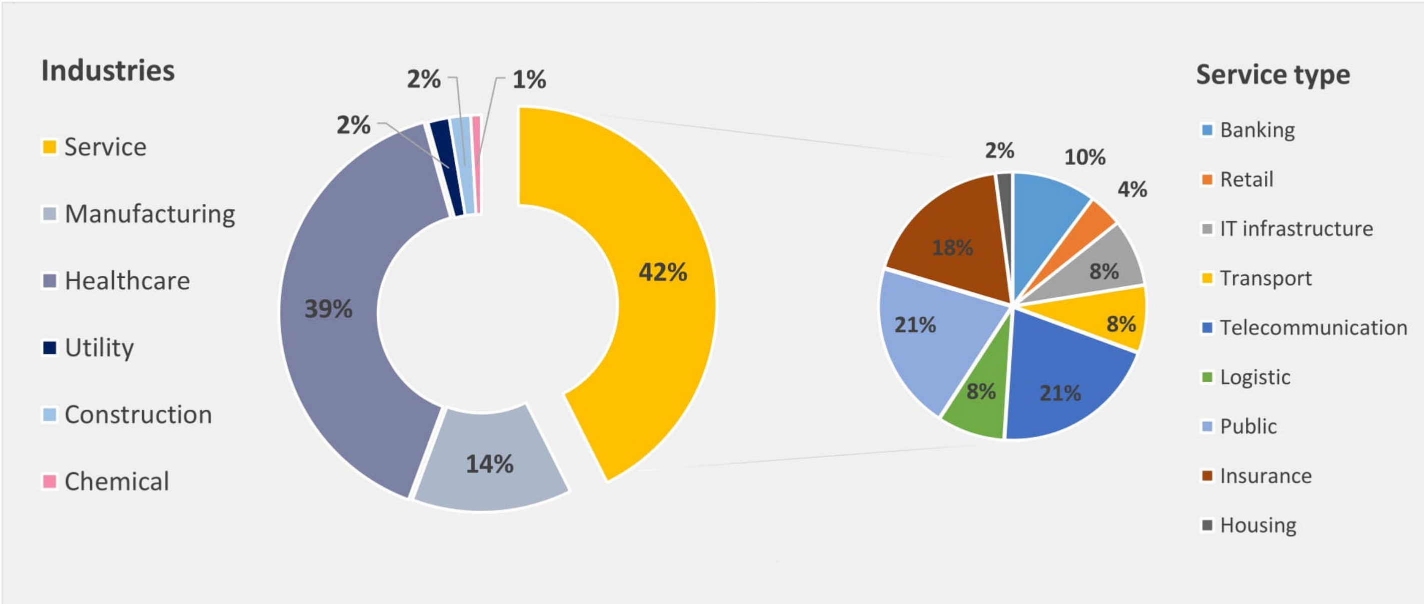
Service - Public	Dutch National Public Works Department (Belgium)	Invoice Management	Dutch National Public Works Department is responsible for the construction and maintenance of the road and water infrastructure. - Using a variety of process mining techniques, they analysed the processing of invoices sent by the various subcontractors and suppliers.	2006	Department of Technology Management Eindhoven; University of Technology (Netherlands)
Healthcare	Dutch hospital (Netherlands)	Logistic process of treating patients	Authors proposed a knowledge management perspective to provide a strategy for modelling and redesigning a business process. The specific group of patients required the involvement of different specialties for their medical treatment that lead to more efforts regarding the coordination of care for these patients.	2005	University of Groningen (Netherlands)
Service - Telecom-	NDA	NDA	NDA	NDA	Queensland University of Technology (Australia)
Service - Insurance	NDA	NDA	NDA	NDA	Queensland University of Technology (Australia)
Healthcare	4 South Australian Hospitals (Australia)	Emergency Department Patient Treatment	The case study applied process mining techniques to patient flow data, collected from patients with chest pain, at four South Australian hospitals. The study was a cross-organisational, comparative analysis that aimed to use routinely collected patient and treatment data to describe differences in the care processes, associated with management of Acute Coronary Syndrome (ACS) practiced in the four hospitals.	NA (2 years)	Queensland University of Technology (Australia)

Infographic

In general, the results of the database show that the number of projects based on process mining techniques are **getting higher every year**. This is probably related to the fact that this theme is becoming every day more actual and that many companies all around the world are becoming more conscious about the tangible advantages that this methodology may produce. More specifically, the histogram shows a peak during the 2014, and a slightly decrease during 2015. Anyway, according to the data, the trend is supposed to increase again: in just seven months of 2016, the cases already collected (14) are almost the same of those the entire 2015 (15).



In accordance to the data, the most frequent utilization of process mining methods are related to those companies operating in **services** (42%), followed by the ones working in **healthcare** (39%) and **manufacturing** (14%). In particular, the cases listed in the service group are analysed more in depth in the pie chart.



References

Author:

Philippe Gonella, Junior Consultant at HSPI

e-mail: philippe.gonella@hspi.it

Contact:

Serena Dal Molin, Junior Consultant at HSPI

e-mail: serena.dalmolin@hspi.it

The company:

HSPI SpA – Management Consulting

Viale Aldo Moro 16 – 40127 Bologna (Italy)
Tel. +39 051509861 | Fax. +39 051 509737
BOLOGNA | ROME | MILAN
www.hspi.it | e-mail: info@hspi.it