Process Analysis in Issue Tracking Environment

Iryna Oleksandrivna Shvydyuk

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Professor Doctor Miguel Leitão Bignolas Mira da Silva

Examination Committee

Chairperson: Professor Doctor José Manuel da Costa Alves Marques

Supervisor: Professor Doctor Miguel Leitão Bignolas Mira da Silva

Member of the Committee: Professor Doctor Fernando Manuel Pereira da Costa Brito e Abreu

November 2016

Agradecimentos

Gostaria de agradecer aos meus orientadores, Miguel Mira da Silva e Diogo R. Ferreira, pelo apoio e pela disponibilidade que me foram proporcionados ao longo deste percurso.

A todos os profissionais das empresas onde estive inserida. Sem a vossa participação, colaboração e compreensão, este trabalho não teria sido possível.

Aos meus colegas do IST por todos os momentos que passámos ao longo dos últimos anos. Independentemente de quão negras as coisas estivessem, vocês foram a fonte da boa disposição que, aparentemente, nos carateriza tanto.

Gostaria de agradecer a todos aqueles que sofreram e presenciaram o desespero associado a este trabalho e disponibilizara o seu tempo para o esclarecimento das minhas dúvidas existências, nomeadamente David, Diana, Gonçalo, Rafael e Rita.

Aos meus amigos do secundário e alguns ainda do básico, nomeadamente Alex, Ângela, Cátia, Daniela, Diana, Gonçalo e Zé. Obrigada por todos os momentos e todas as aventuras que vivemos. Podemos não nos ver durante anos, porém esse facto passa a um pequeno pormenor quando nos encontramos.

Gostaria de agradecer a pessoa que sempre será uma criança para mim, a minha irmã Kateryna Shvydyuk por me ter ouvido e distraído quando mais precisava.

Aos meus pais, Olena e Oleksandr Shvydyuk pelo seu apoio incondicional e preocupação, por terem mudado as suas vidas só para proporcionar melhores condições aos seus filhos. Por me terem ensinado que esforço, dedicação e, por vezes teimosia, são necessários para conseguir alcançar o que se pretende. Obrigada por serem um exemplo de que desistir não é a opção. Sem eles, eu não teria chegado onde cheguei.

Resumo

Os sistemas de gestão de incidentes são bem conhecidos pela sua capacidade de registo, monitorização, resolução e arquivamento dos problemas encontrados. Devido à sua natureza, estes desempenham a função de ligação entre diferentes *stakeholders* presentes no processo que apoiam. Tradicionalmente, a investigação efetuada sobre estes sistemas tem-se focado na componente operacional dos mesmos, ou seja, na sua utilização na prática em termos da sequência de passos. Os resultados destas pesquisas traduzem-se nas sugestões de implementação ou melhoria, deixando a componente da informação guardada na perspetiva dos processos esquecida. Em termos do processo de investigação utilizado, os investigadores têm-se focado no método qualitativo que, por sua natureza, é muito moroso e requer a presença física das pessoas. Esta tese propõe a deteção dos processos presentes do ambiente de tratamento de incidentes com as técnicas de mineração de processos, abordando deste modo o problema de falta de sensibilização existente sobre a informação guardada, porém na perspetiva dos processos. A metodologia utilizada para conduzir esta pesquisa foi *Design Science Research* e foi composta pelo *design*, desenvolvimento, demonstração e avaliação. Três aplicações da proposta apresentada na prática foram efetuadas num dos mais usados sistemas de tratamento de incidentes. Além disso, esta foi avaliada na perspetiva global dos critérios e métodos standard. O processo de avaliação foi composto por atividades de raciocínio lógico e análise e entrevistas não estruturadas.

Palavras-Chave: Sistemas de Gestão de Incidentes, Mineração de processos, Ágil, Metodologia, Sensibilização.

Abstract

Issue tracking systems are well-known tools for the recording, tracking, solving and archiving the issues. Due to its nature, they serve as a link among the stakeholder involved in the supported process. Traditionally, the research community focused its attention on the operational level of these tools, i.e. on the understanding of how they are being used in practice and on suggesting some improvements. As a result, the information that is heavily intrinsic in issue tracker's processes is completely forgotten. Additionally, the method used for reasoning is too time consuming and requires people's presence. This thesis proposes the uncovering of the existing processes in the issue tracking systems' environment with process mining techniques in order to address the lack of awareness of the stored information from process perspective. Design Science Research Methodology was used to conduct the present work which comprised proposal's design, development, demonstrations and evaluation. Three applications of the proposed solutions were performed on one of the most used issue tracker and these were evaluated according to a holistic view of criteria and standard methods. It was composed by logical reasoning and analysis and unstructured interviews throughout the applications.

Keywords: Issue Tracking Systems, Process Mining, Agile, Methodology, Awareness.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ARIS PPM	ARIS Process Performance Manager
CSV	Comma Separated Values
Disco	Discovery
DSRM	Design Science Research Methodology
EPC	Event Driven Process
HTML	HyperText Markup Language
IS	Information System
IT	Information Technology
ITIL	Information Technology Infrastructure Library
ITS	Issue Tracking Systems
JSON	JavaScript Object Notation
ProM	Process Mining
XES	eXtensible Event Stream
XML	eXtensible Markup Language
XSLT	eXtensible Stylesheet Language

1. Introduction

Among a wide range of tools, ITS are the ones that are well-known for its organizational ability. Traditionally, it is viewed as a tool for assisting the software development process. R.S. Pressman in his book stated that they enable the “development team to record and track the status of all outstanding issues associated to each configuration object” [1]. Reis et. al. stated that, in spite of recording and tracking the status of an issue, the tool also helps on its resolution and archiving [2]. Moreover, the authors discussed the role of stakeholders involved in the process, such as software development team, customer, quality assurance, etc.

As already perceived, these systems are built upon a concept of an issue. Habitually, it is related to a bug, a feature or an inquiry [3]. The main difference relies on the content and on the person who reported it. A bug represents something that does not allow the software program to go as it is supposed to and can be reported by anyone. A feature is a new functionality or improvement and usually is reported by project managers or team leaders. Finally, an inquiry is reported by customers and can include any information they think is relevant.

Through the time, ITS started to support the project tracking functionality and its role has evolved. In software projects, they are referred as bug tracking systems and in other context as project management tools. Consequently, the issue started to represent the main tasks performed.

Over the past few years, the research community has been performing studies on it. However, the major focus was done on the operational level, i.e. understanding of how these systems were used in practice [3]–[5] and on suggesting some improvements for already existing tools or the future ones [3], [4], [6]–[8].

Bertram et. al., conducted their research with the objective of understanding the design and use of issue tracking system from a social perspective. They wanted to explore both the tool and the process and provide some additional insights [3]. Thus, the decision made was to perform a qualitative study based on questionnaires and interviews. Since the focus of the problem was at the social level, all the participants were gathered according to different roles they performed in the software team.

Thus, ITS were justified to be a key knowledge repository, a communication and collaboration hub and a channel of communication. Additionally, the authors discovered that sometimes stakeholders hold different and/or conflicting perspectives about some aspects of the issue tracking process. A clearer picture was presented by people that are working with these types of systems. Consequently, the results obtained were used to present seven design configurations for future development. Those configurations aimed to cope with the development team’s needs.

Ortu et. al. explores the existence of communities in developer network and focused their attention on the developers’ productivity [5]. This aspect was measured through the time required to resolve an issue.

The techniques used, despite allowing to reason about those two aspects, required additional investigation.

Some researchers tried to appeal to completely different fields to explore ITS from other perspectives. Boyasal et. al. addressed the problem of situation awareness in development teams that rely on these tools [9]. The authors identified the most common tasks and to support the issues found, implemented a prototype. Although, the most frequent target is the issue, authors decided to focus their attention on individuals. Through a set of interviews, they managed to identify the main tasks performed by people. Note that method chosen to perform the reasoning allows a better understanding of the problem, nevertheless is time consuming and requires people's presence.

Despite trying to escape the cliché and approach the ITS from different perspectives, at the end, the researchers' focus was still on the operational level, i.e. its usage in practice and suggesting some improvements. Thus, there is a **lack of awareness of the stored information from process perspective**.

Process mining has appeared recently and offers the opportunity to understand actual behavior of business processes through the event log analysis, i.e. diagnose its weaknesses based on omnipresent information produced by the information systems. Moreover, it allows to improve the existing processes. For the last few years, this area is becoming more and more popular among organizations [10]–[14].

In order to address the identified problem, the main proposal of this thesis relies on exploring the process mining techniques and uncover the stored processes in the ITS environment.

The present research was conducted in compliance with DSRM's specifications and incorporated the application of the proposed solution in three different contexts of the one of the most used ITS. The evaluation process was based on the holistic view of criteria and standard methods. The results of this research were analyzed according to previously selected criteria, such as efficacy, validity and accuracy.

1.1. Research Methodology

The research methodology elected to conduct this work is DSRM. The methodology incorporates a set of principles, practices and procedures required to carry out the research in the context of IS [15].

The creation of this particular methodology was triggered by the need of overcoming all the existing and non-fitting paradigms, such as traditional descriptive and interpretive research paradigms.

In the context of IS, DSRM involves the activities of creation and evaluation that are performed upon an artifact intended to solve an identified problem. The artifacts are defined as *constructs* (vocabulary and symbols), *models* (abstractions and representation), *methods* (algorithms and practices) and/or *instantiations* (implemented and prototype systems).

DSRM is based on an iterative process and is composed by 6 phases [16].

Problem identification and motivation

Definition of the specific research problem and justification of the value of a solution. The problem definition will be used to develop an artifact that can provide a solution to the proposed problem. The knowledge of the state of the problem and the importance of its solution is imperative for this activity.

Define the objectives for a solution

Infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible. The objectives can be quantitative or qualitative and should infer rationally from the problem specification. The resources include knowledge of the state of problems specification. The resources required include knowledge of the state of problems and current solutions and, if so, their efficacy.

Design and development

Create an artifact. Such artifacts are constructs, models, methods and instantiations or even new properties of technical, social and informational resources. Theoretically speaking, an artifact can be any designed object in which a research contribution is embedded in the design. This phase combines the specification desired functionality and its architecture, and the creation of an actual artifact. To move from objectives phase to design and development, it is required to understand the knowledge about the theory that supports the offered solution.

Demonstration

Demonstrate the use of artifact to solve one or more instances of the problem. This could involve its use in experimentation, simulation, case study, proof or other appropriate activity. The essential resource for this phase is effective knowledge of how to use the artifact to solve the problem.

Evaluation

Observe and measure how well the artifact supports a solution to the problem. This phase involves comparing the objectives of a solution to the actual observed results from use of the artifact in the demonstration. The knowledge of relevant metrics and analysis techniques is required. It is important to denote that evaluation can take many forms depending on the nature of the problem and the artifact per se. Since DSRM is an iterative process, the evaluation phase allows to choose between iterating back or moving forward to the next stage.

Communication

Communicate the problem and its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences, such as practicing professionals, when appropriate. This phase requires knowledge of disciplinary culture.

DSRM process was adapted to fit the requirements of the present research and is presented in Figure 1.

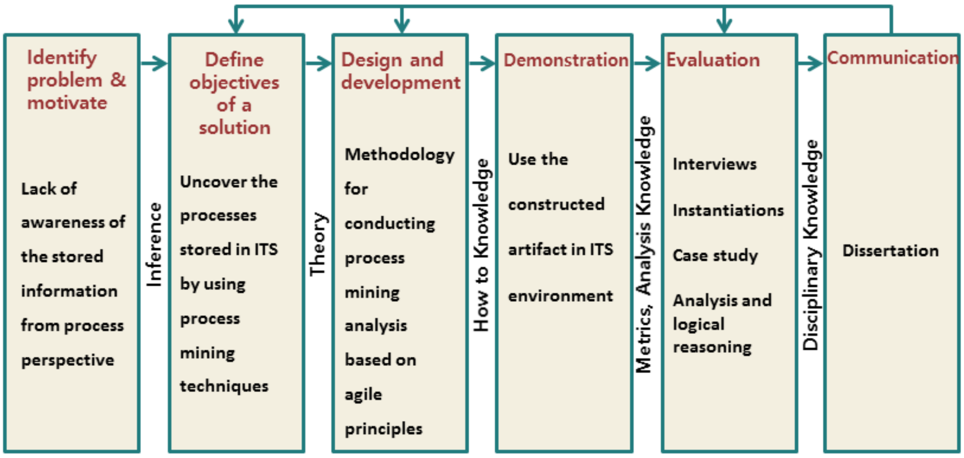

Figure 1 DSRM Adapted to the Present Research (adapted from [16])

1.2. Thesis outline

The organization of this document is strongly influenced by the methodology used to conduct the research.

Chapter 2 aims to raise the awareness about the problem that motivated the research.

Chapter 3 develops the theoretical background related to the present work.

Chapter 4 provides a detailed description of the approach proposed to solve the identified problem.

Chapter 5 aims to demonstrate the applicability of the proposed solution. Since the solution was applied in three contexts, their full description among with the obtained results can be found in the respective sub-chapters.

Chapter 6 comprises the evaluation of the results of each application. Note that the decision made was to leave the demonstration and the respective evaluation in different chapters. The correspondence is made by the sub-chapters.

Chapter 7 concludes the present report with the summary of the performed work, the indication of the overall findings and future work.

2. Related work

Throughout the following chapter, the theoretical background related to the present work will be developed. In **section 3.1**, the process mining concepts and techniques are introduced. Moreover, its application in real-life cases is briefly discussed along with the suggested methodology and tools. **Section 3.2** provides some insights about agile approach with a special focus on Scrum methodology. Finally, the chapter is concluded with an overview and discussion of the presented areas (**section 3.3**).

2.1. Process mining

Process mining is a recent discipline that aims to extract knowledge from event logs recorded by an information system in order to discover, monitor and improve business processes [17].

Usually an event log contains information about events that occurred in an information system and mention a *process instance*, also referred as case, and an *activity*, also known as task. A process instance represents what is being handled, e.g. a job application, student enrolment, etc. The activity is an operation performed by someone or something on the process instance. There may be some additional information stored.

To apply process mining, one assumes that beyond process instance and activity, it is possible to record the person that is performing an activity (*performer*) and all events have a *timestamp* associated to it and are ordered. In Table 1 one can see an example of a log regarding nineteen events, five activities and six originators adapted from [10](full table presented in **Appendix A**).

Table 1 An Event Log (partial)

Case ID	Activity ID	Performer	Timestamp
Case 1	Activity A	John	9-3-2004: 15.01
Case 2	Activity A	John	9-3-2004: 15.12
Case 3	Activity A	Sue	9-3-2004: 16.03
Case 3	Activity B	Carol	9-3-2004: 16.07
Case 1	Activity B	Mike	9-3-2004: 18.25
Case 1	Activity C	John	10-3-2004: 9.23
Case 2	Activity C	Mike	10-3-2004: 10.34
Case 4	Activity A	Sue	10-3-2004: 10.35

The analysis of information shown in the previous table suggests some aspects about the event log such as [18]:

- Every case is starting with activity A and ends with activity D;
- For the Case 1, the activities executed are A, B, C and D;
- The activities B and C are parallel to the activity E;
- The activities performed by "John" are A and C;
- "John" is always working with "Sue".

With the type of event log such as presented in Table 1 as a starting point, one can distinguish three types of process mining that will be addressed in the following subsection.

2.1.1. Types of process mining

As we mentioned before, there are three types of process mining, such as *discovery*, *conformance checking* and *enhancement* (Figure 2).

Discovery stands for a construction of a model based on the information contained in an event log, i.e. there is no a-priori model. It allows for the comprehension of the original process model, some execution properties and the organizational context. Traditionally, this has been the focus of process mining.

In contrast to discovery, conformance checking and enhancement require some a-priori models. The accordance among what was specified in the model and what happens through comparison between event data and the defined process is accomplished through the **conformance checking**.

The **enhancement** aims to enrich the existing model with data from event log. By using the decision mining, user profiling and performance analysis, the reasoning about the data dependencies and attributes that influence performance, routing, etc. can be done.

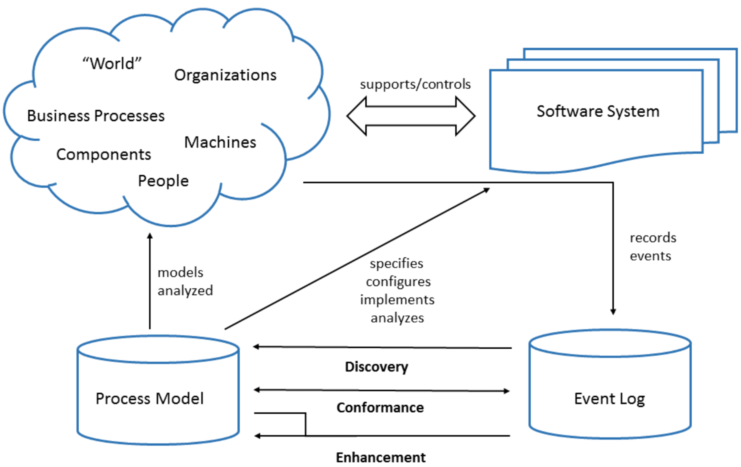

Figure 2 Types of Process Mining (adapted from [17])

Additionally, process mining can be applied to analyse three different perspectives, such as process perspective (*how?*), the case perspective (*what?*) and organizational perspective (*who?*) [19].

2.1.2. Perspectives of process mining

Every analysis that implies **control flow**, i.e. ordering of activities, is addressed by the process perspective. Here, the goal is to find a good characterization of all possible paths.

The **case perspective** handles each case as a separate one and spotlights its properties. Cases may be characterized according to several aspects, such as the paths they take, the performers working on it or simply according to the values of the corresponding data elements.

Finally, **the organizational perspective** focuses on the performer and tries to uncover the performers involved and their relations. Classification of people in terms of their roles and organizational units or showing the relation between individual performers is the final goal of this perspective.

Over the past few years, process mining area has gathered much interest from research community. Various papers were published with the presentation and description of its theoretical component, i.e. its algorithms and theory behind it, and the evolution of what has become the main framework for this kind of analysis. Likewise, a more practical component has been approached.

2.1.3. Process mining applicability: real-life cases

Process mining's application has been conducted in several real-life cases incorporated in different areas [10], [11], [20]–[22], [12].

Aast et. al. [10], aimed to show the applicability of process mining techniques. Their goal was to improve the main process responsible for efficiency loss in an organization. The analyzed process involves a large scale of stakeholders and information; thus, the accuracy was vital and the process of data collection took a long period of two years. The authors focused in all three perspectives known in process mining.

The process perspective was conducted in an iterative way over which several meetings with organization people were made to validate the obtained process and focus on the core process.

In case of an organizational perspective, two different metrics were used to reason about the social component intrinsic in the process. On one hand, with the first one, three different categories of personnel were discovered which were not involved in the process as expected. It pointed to a problem of assigning their work to someone else. On the other hand, with the second metric, the personnel's passivity was discovered.

Finally, by analyzing the case perspective, it was possible to discover another cause of efficiency problem related to the human component. Moreover, the results of previous perspective analysis were used in order to clearly understand its impact in the process (personnel's passivity).

Process mining techniques were also successfully applied in the health-care area. In Mans et. al., the subject of interest was finding a compromise between high-quality services and low costs [11], [12]. In order to address this issue, several approaches were made.

For example, the process of treating stroke patients in hospitals was attempted to be discovered [12]. Regarding the collection of information, two different data sets were considered in order to provide a better understanding and improve a process of interest. As a result, several performance decreasing components were found and comparison between treatments in different hospitals was possible.

Mans et. al. explores the process, case and organizational perspectives in favor of understanding gynecologic oncology process [11]. In terms of information gathering, when dealing with health-care area, one should take into account the large number of applications developed individually and the highly complex and ad-hoc nature of its processes. As a result, for performing any type of analysis, data manipulation must be done.

Similarly to Aalst et. al., during the analysis phase various meetings with experts were made in order to validate the information obtained. Here, the main output was expressed in terms of comprehensible and validated models [10].

Poncin et. al., applied process mining techniques to analyze data gathered from different software repositories [20]. In their work, mining poses the last phase and aims to explore the information obtained with their framework FRASR in order to show its applicability. The authors were interested in organizational analysis, i.e. social component, and process perspective, i.e. discovery of the bug's actual lifecycle.

As a result, it was possible to reason about different roles in development teams and their performance. Moreover, the authors concluded that there is a slight difference between the bug’s lifecycle expected and the real one. Note big data sets were chosen in order to perform process mining analysis and it has influenced the results. For example, when exploring the process perspective, the analysis was enriched with more process instances and, consequently, a wider number of scenarios was considered.

Process mining area was also applied in the IT service management area, to ITIL best practices. Ferreira et. al., addressed the subject of ITILs flexibility regarding its guidelines [22]. In their case study, the authors aimed to discover whether ITILs best practices are being followed during business processes’ implementation or not. For this purpose, a tool for Incident Management was chosen with a particularity of giving significant freedom to end-users, i.e. non-restrictive database population and end-users needs, practices and workarounds were considered.

The business processes regarding its run-time behavior had to be studied first. The authors advocated that process mining techniques would lead to less time-consuming extraction processes and increase the reliability of results. These were successfully obtained after focusing on the overall model. As a result, it was possible to conclude that the stages obtained matched the ones specified by ITIL.

Since the authors were dealing with a non-restrictive tool in terms of database population, it was possible to discover much more behavior than specified by ITIL guidelines. An example is when a behavior is an action simple to correct, i.e. the team skips some steps. These results and respective conclusions were communicated to the company and validated as expected.

Process mining was also proved to be valuable when incorporated in a software development project based on Scrum. Rubin et. al. experienced blocker performance issues and problems with system usage, i.e. it was not used according to the initial design specifications. The development team lacked the interactions with the end-user and, thus, did not have the whole clear picture to solve the problems found. The analysis of the produced data was attempted, however the team members found themselves lost in the details [23]. In order to solve the situation, process mining techniques were used. It turned out to be an asset that allowed to understand the overall behavior (system and users). At the technological level, the commercial tool Disco was used. Rubin et. al. appealed to its filtering capability in order to focus and analyze the cases that represented success, failure, suspicious behavior, etc. [23] All in all, they attempted to study each path taken and every combination done.

The performance issues were solved by analyzing major bottlenecks. In the case of system usage, the decision was made to store user's behavior after each sprint and analyze it later.

With these improvements, the Scrum team's planning and the functionalities of the final product were influenced. In the first case, the sprint tasks were based on the feedback provided by the users and, in the second one, new requirements were discovered and included in the product backlog.

As a final point, it became obvious that the applicability of process mining in real-life cases has taken a very successful path. The offered mining perspectives allow us to get inside the organizational processes and understand which occurs and is stored in information systems. Moreover, by streamlining their processes, the companies can assure its competitive advantage.

2.1.4. Process mining methodology

In the previous subsection, an overview through the concepts of process mining among with the introduction of some case studies that show its applicability was performed. Although the motive that triggers this type of study vary, usually it is well-known by organization or researchers. As well, one may notice a pattern in the approaches through the whole research process.

Bozkaya et. al. noticed that there was no methodology for conducting process mining research and, after doing some analysis, proposed one [13]. The authors stated that process mining can be a "repeatable service" and does not require any previous knowledge about the context of the problem, i.e. only the organization should analyze the outcome. As well, the methodology should provide whole information needed in "short time-period". Following these objectives, a six steps methodology was defined (Figure 3).

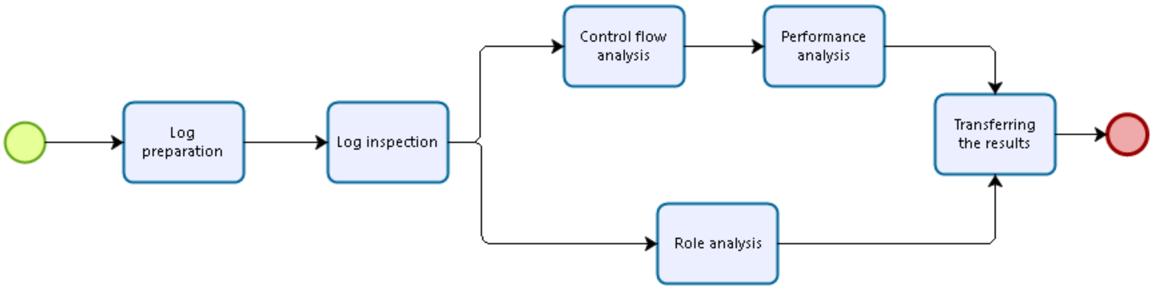

Figure 3 Methodology for Process Mining Analysis [13]

Log preparation

The data is pre-processed and the event log is built. The identification of problems is made. Afterwards the most suitable notion of case is chosen from the available ones. Similarly, the activities and their events are identified.

Log inspection

Helps with the familiarization with event log. It includes statistic collection about the log, such as number of cases, number of cases per event, number of roles per case, number of performers, total number of different sequences, etc.

The information gathered contributes to the overall knowledge of the event log and process in terms of size, choice of the mining algorithms and to the evaluation process. Throughout this phase is possible to perform data filtering to remove cases that do not belong to the time-period of interest.

Control flow analysis

The core process stored in the log is obtained by running some discovery algorithms, e.g. heuristic miner [24]. However, in the case of the having the specification of the process, the conformance checking can be performed instead.

Performance analysis

The fourth stage of presented methodology aims to check the business process discovered regarding the performance component. The authors suggest applying a specific type of analysis offered by ProM tool. This tool will be explained in more detail in **section 3.1.5**. Moreover, another suggested technique consists in replaying the filtered event log on the process obtained.

Role analysis

In parallel to control flow and performance analysis, the authors define role analysis phase. This phase allows to understand the interactions inside the organization and define roles in terms of the work that personnel (specialists and generalists) perform [25].

Transferring the results

Finally, all the results are presented and validated with the client. The authors advocate that with the information presented, the organization could make meaningful changes to solve the problem that lead it to process mining analysis at first place.

2.1.5. Software tools

As a field of study, process mining involves the construction of models through the analysis of event logs. Nowadays, there is a large number of tools for this purpose. Some of them are more suitable for academic research, others were developed for commercial purposes. In this work, the major focus will be made on two process mining tools: ProM and Disco.

ProM: a process mining framework

In the past few years, many tools have been developed in process mining. Each one of them focuses on a perspective that is more suitable for its context of interest and uses different process mining techniques. However, they manage to work in the same event log and produce similar type of models. With this in mind, ProM framework was created.

ProM integrated the functionality of several existing process mining tools and provides additional plug-ins [19]. It was developed in such a way that every plug-in set can be extended by simply adding a new one. For example, an alpha-algorithm, one of the first algorithms developed, is implemented by a plug-in that automatically generated Petri nets from event log [14]. Up until now, one can name five types of plug-ins which are represented in Figure 4.

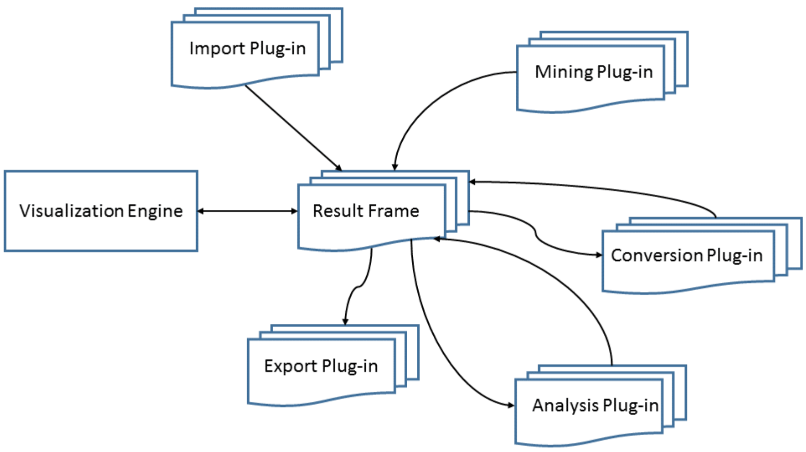

Figure 4 Modules of the Framework ProM (adapted from [19])

Mining plug-ins implement some mining algorithms, e.g., mining algorithms that construct a Petri net on some event log.

Export plug-ins implement some "save as" functionality for some objects, such as graphs. For example, there are plug-ins to save EPCs, Petri nets, spreadsheets, etc.

Import plug-ins implement an "open" functionality for exported object, e.g., load instance EPC from ARIS PPM.

Analysis plug-ins typically implement some property analysis on some mining result. For example, for Petri nets there is a plug-in that constructs place invariant, transition invariant and a coverability graph. Moreover, there can be found several plug-ins that compare the event log to the pre-defined process.

for event logs one can find several plug-ins that compare them to models.

Conversion plug-ins allow conversation between different data formats, e.g., from EPC to Petri nets.

To use the plug-ins described below, ProM requires the data imported to be in XES format [26].

Finally, it is important to note that despite of ProM's range of functionalities and advantages, it is mostly directed to academic research and, consequently, it lacks significant features for process exploration. Moreover, the analysis of the obtained processes is monotonous and non-interactive due to the number of steps needed for its accomplishment [27].

Disco: a process mining tool

The functionality of ProM is extraordinary in the way that there is no other tool with the same range of process mining algorithms and possibilities for analysis. Although, as stated below, as academic tool, it presents some disadvantages. To address them and explore other possibilities, the commercial tools were considered. Fortunately, nowadays the number of commercially available software products is growing. Some examples are Disco, ARIS Process Performance Manager, Reflec|one, etc. In this work, we choose to focus on Disco.

Disco stands for discovery and it is commercial tool for process mining analysis developed by Fluxicon. Its major advantage is good handling of large and complex data sets, i.e. high performance and ease to use [17]. Disco requires to all data being on CSV format [28].

Although Disco lacks on the number of algorithms, as a matter of fact, the process mining analysis is based on a single algorithm, it provides tools needed for quick and effective analysis. For example, one can choose a level of abstraction, inspect statistics about process of interest, understand what is happening and spot bottlenecks by using animations, analyze just the case of interest and, finally, can filter the data considered not necessary on that point of time [29].

2.2. Agile

The business environment continues to change at a dramatically increasing pace [30]–[33]. Throughout time, customers' satisfaction at the time of delivery managed to gain extra importance. Highsmith et. al. stressed the question of handling these inevitable changes throughout project's life-cycle [30]. As the changes, cannot be ignored because it results in a bad customer evaluation and ultimately in business failure, the new strategy is to reduce the costs of responding to arising changes.

The traditional methodology had as core principle following the plan in order to make the process more predictable, structured and efficient [33]. However, by strictly following the initial plan, without any deviation from it, the nature and unpredictability present in the process are not considered.

Agile approach differs from the traditional way of conducting software projects. Apart from reducing costs, it aims to maximize the value of software development processes, finding a compromise between no process and too much process. Moreover, it also tries to encourage the quality in design process [30]. Scrum, one of the most used agile methodologies, uses management practices, such as daily meetings and iteration reviews after each sprint (**section 3.2.3**) to maximize designs quality.

2.2.1. Agile proposal

Agile approach focuses on two main concepts: *working software* and *people working together* [30]. On one hand, the working code expresses what has been perceived by developers and is a real proof of performed work. Moreover, it can be validated and, in case of having some mismatches, can be modified. On the other hand, people are more effective when talking face to face. Thus, by encouraging interactions, the development process will achieve maneuverability, spend and costs savings.

The agile manifesto appeared as a combination of already existing practices and relies on four core values and twelve principles [34]. In the present dissertation, the focus will be done on the main pillars.

Individuals and interactions over processes and tools

At first sight, it might appear that agile approach aims to discriminate the processes and tools that are used to develop software and put social interactions and individuals first. In actual fact, it recognizes the contribution of structured and technological component and outlines the interactions between skilled individuals [34]. That is, the agile approach aims to highlight the relationships and communality between software developers and the human role reflected in the contract [35].

Working software over comprehensive documentation

There is nothing wrong with having complete documentation, although the primary focus must remain on final product. In agile, the development team decides which documentation is essential [34]. A balance must be found between too little (or partial) and too much (or exhaustive) documentation. Regarding code, agile movement advocates that it must remain simple, straightforward and technically advanced as possible [35]. With this principle, the customer will have the opportunity to see the working and tested software in small and frequent releases and witness its evolution, i.e. witness how fast the results are obtained and deliver quick feedback [30].

Customer collaboration over contract negotiation

Customer collaboration provides the opportunity to make sure that every actor involved in the project shares the same vision and same understanding of current situation. As a result, the need for change will be detected earlier and more suitable results will be produced. In terms of business, it will benefit from immediate value delivery from the very beginning of the project. Therefore, the risk present on traditional methodology regarding contract non-fulfillment is reduced.

Additionally, the final product will meet clients' expectations since through continuing collaboration, the development team will fully understand its needs [34].

Responding to change over following the plan

From the overall proposal, this part is the one that differs the most from the traditional methodology. As already mentioned, the business and technology world is turbulent and the inability of responding to changes can impact the organizational survival. Thus, when developing a project that aims to meet the environmental needs, every stakeholder involved should be able to recognize and adapt to every external change and emergencies that would appear during the development process life-cycle [34], [35].

2.2.2. Agile practices

Agile does not provide a set of rigid practices for an organization to follow. However, it requires certain management practices and tools to be implemented.

Agile approach suggests to perform **frequent releases**, i.e. between two and six weeks, in order obtain feedback from customers as early as possible and perform all required modifications. The concept of **dynamic prioritization** means that the requirements or features defined at the beginning of the project can be reordered or even excluded but only at the end of an iteration to better fit the needs of the customer.

This approach aims to address the problem of rapid change of business environment. Thus, it strongly emphasizes the importance of **constant feedback** on technical decisions, customer requirements and management constraints [30].

Finally, an organization using one of agile methods for system development process should understand the importance of **interactions among people** and the **teamwork** concept. If a client does not have a clear vision of the main goal, even the best and most efficient team will also be lost.

2.2.3. Agile Methodology: Scrum

Originally, the term "scrum" comes from the scrum formation used in rugby where it means "getting an out-off play ball back into the game" with teamwork [35]. In this formation, each player fulfils a very specific role, however they all collaborate to the end goal of winning the ball to their side. Its idea was adopted in development process to cope the problem regarding systems flexibility in changing environment, i.e. manage the changing environmental and technical variables that make the development process unpredictable and complex.

Scrum methodology aims to improve the existing engineering practices by focusing on the way software development team members should function [36]. It is based on continuous evaluation of performed work, detecting any problems that the team is experiencing in the development process and improving it.

Roles and responsibilities

Through the literature, one can find some complementary definitions of roles in Scrum [35], [37]–[39]. Some authors involve customer, users and management as performers in the Scrum process [35], and others focus only on the team, i.e. emphasize the product owner, Scrum master and development team [37], [38]. Since the goal is to cover both business and technological sides, in this work every actor will be considered.

- **Product owner**

The “sole person” that is responsible for assuring the overall success of the product being developed and guarantee that the development team is working at its full potential. This entity manages the product backlog and decides which item will be built and their order. Moreover, product owner acts as communication link between the Scrum team and all other participants, i.e. it is his/her work to communicate what the scrum team aims to achieve.

Moreover, to maximize the pace of development, product owner should continuously cooperate with other members of a Scrum team and be always available.

- **Scrum master**

Unlike traditional role of project manager that controls its team, the scrum master is a leader, i.e. the scrum master is responsible for guiding the team towards its own working process based on Scrums practices, theory and rules [37], [38]. Similarly to the product owner, Scrum master aims to maximize the value created by the Scrum team. In order to achieve it, he/she aims to remove all the impediments that affect the teams’ productivity [37].

- **Development team**

A group of self-organized professionals which main purpose it to deliver what product owner has requested. Scrum suggests that the team has between five to nine people and each of its members has requested skills for producing good quality software [37]. In a project, there can be more than one Scrum team though always with advisable number of individuals.

- **Customer**

In Scrum, customer is directly connected to product backlog and intervenes in the tasks that involve this aspect.

- **Management**

In Scrum's context, management was assigned to participate in goals and requirements decision making and oversees the final decision making.

Practices

Scrum is a non-restrictive process in terms of methods or practices. It can be used without additional worrying about following specific set of methods or practices. Nevertheless, in order to deal with unpredictability and complexity, it requires certain management practices and tools in the various phase [35].

- **Product Backlog**

Item that defines the work to be completed in the project, i.e. contains everything that is required in the final product based on current knowledge. It is expressed in terms of prioritized requirements that are repetitively being updated. In order to get an extensive list, actors such as customer, marketing and sales, project team, management and customer support can participate in the specification process [35].

The maintenance of product backlogs list is assigned to product owner.

- **Effort estimation**

An iterative activity that is performed by Scrum team and product owner. As the project goes through, some new information can be available about each item of product backlog. Thus, effort estimation practice aims to re-evaluate those elements at more accurate level [35].

- **Sprint**

An iterative cycle that aims to develop new functionalities to produce new increment. Through this practice, the adapting to the changing environmental variables such as requirements, time, resource, etc. is performed. Each sprint lasts between one week to one month. During this time frame, the Scrum team commits to develop a part of the main product [35].

- **Sprint planning meeting**

During this practice, the functionalities and the goals for the next sprint and the way in which they are going to be implemented are defined. When the discussion is at business level, actors such as customers, users, management, product owner and Scrum team are present [35]. When the subject is more technological, the development team, product owner and Scrum master are present. Thus, one can assert that it is a two-phase meeting.

- **Sprint backlog**

This notion is similar to the product backlog. Sprint backlog contains items that will be implemented during the sprint and, unlike product backlog, cannot change. New iterations cannot start until every item present in sprint backlog is implemented. The items are chosen from product backlog by Scrum master, development team and product owner during sprint planning meeting.

- **Daily Scrum meeting**

Throughout daily scrum meeting, Scrum team led by Scrum master discuss their progress and plan what needs to be done. Moreover, any problems experienced are also discussed in order to improve the process. Business actors, such as management, can also be present and intervene in the meeting [35].

- **Sprint review meeting**

Sprint review meeting joins the Scrum team, management, customer and users in order to present and discuss the results of the sprint, i.e. review all the features that were completed from the angle of general development effort [35], [37]. It takes place on the last day of the sprint and can set new directions for the system that is being developed or add new requirements for the product backlog list.

Process

Scrum process is composed by *pre*-game, *development* and *post-game* phases that are presented in Figure 5.

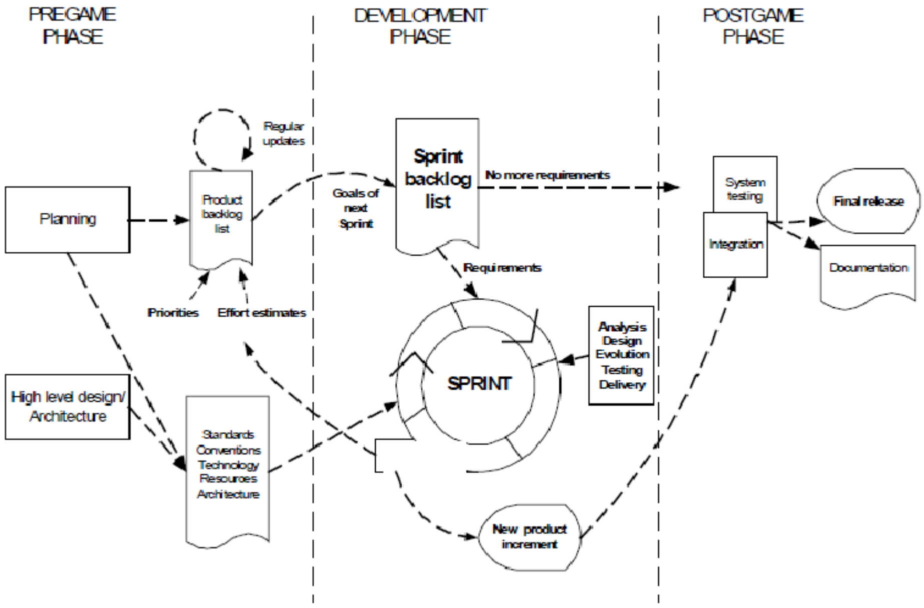

Figure 5 Scrum Process [35]

The **pre-game phase** aims to perform initial planning based on the current knowledge. In one hand, all presently known requirements are added to the product backlog. These are prioritized and the effort required for their implementation is estimated [35]. Additionally, to promote the teams' engagement from the start, it is up to the team to perform the review of updated product backlog, after each iteration. On the other hand, high-level design of the system based on product backlog is developed. Note that, as the project goes through, the design is changing along with the requirements.

The **development phase**, as the name implies, covers the whole system development process. It is performed in sprints and aims to observe and constantly control different environmental and technical variables which may change during the process. As a result, the ability to flexibly adapt to the changes is assured.

Finally, the **post-game phase** represents the preparation for system's release. Includes such activities as integration, system testing and documentation [35].

2.3. Discussion

In this chapter the theoretical background related to the present work was introduced. Process mining concepts and techniques were presented. Throughout the analysis of its application in real-life cases, it became evident that process mining has taken a very successful path. The organizations gained the possibility of getting inside its processes and understanding what was happening. Thus, they gained the opportunity to analyze the problems faced from different perspectives and increase its efficiency and productivity.

Among lessons learnt by the performers of the analysis, staying in touch with organization was considered crucial. The communication and feedback provided during the meetings allowed to validate all the obtained results, i.e. with the help of experts it was possible to understand what makes sense in the context of analysis and what does not. Moreover, the possibility to explore a specific aspect, e.g. a suspicious behavior or path, emerged.

The context was proved to play decisive role in the data gathering process since the nature of the processes changes according to the business area. For example, in case of the health-care area its processes are extremely complex and ad-hoc. Additionally, it is up to the organization to decide the way in which the information is stored. Thus, the extraction process should be customized.

At technological level, as was shown, both commercial and academic tools can support the organizational requests. The wide range of mining algorithms present in the academic tool allowed to choose the more suitable one. However, ProM lacks features for process exploration and is difficult to use.

Disco, besides being easy to use, was proved to deliver high performance and provide quick and effective analysis. For example, this commercial tool was used in analysis of the data produced after each sprint in a software development process. Also, a generic methodology for conducting process mining analysis was presented.

In addition to process mining area, the agile approach and one of its most used methodologies were presented. The agile manifesto aims to address the problem of rapid change intrinsic in the business environment and is supported by four core values. It is a people oriented approach and emphasizes the importance of working software. Thus, it heavily supports the interactions among people, teamwork, constant feedback, frequent releases and dynamic prioritization of the requirements.

Scrum, the one of the most used methodologies, focuses on the way that software development team members should function. It starts with initial planning and functions with iterations, i.e. sprints. After each sprint, the results are presented and discussed with the business individuals. The feedback obtained is integrated in form of requirements in the product backlog. Moreover, the team self-evaluates its own performance after each sprint. With this iterative process Scrum intends to implement the agile philosophy in the organizations.

When comparing process mining and agile areas, several crossing points can be found. As already mentioned, staying in touch with the organizations allows to clarify all doubts about processes that are being analyzed and validate the results. Agile approach recognizes the importance of customer collaboration and experts' feedback. As an example, Scrum provides a set of practices, such as sprints reviews. Likewise, the collaboration and interactions among individuals help to understand the context in which the analysis will be performed and the nature of the processes.

Over the time, the customer may want to perform some additional analysis or explore some specific aspect. This results in changing the current knowledge and, consequently, in development of new requirements. Moreover, to perform these changes quickly, relying on interactions between team members eases the process.

3. Research proposal

Process mining discipline is a relatively new area that offers the opportunity to understand the actual behavior of business processes through event log analysis, i.e. diagnose its weaknesses based on omnipresent information produced by information systems. Furthermore, with the application of its techniques, the organizations obtain additional insights about its processes and improve the existing processes. It is not surprising, as was already perceived, that this area is becoming more and more popular among organizations.

In order to make the analysis more systematic, a general methodology was provided as a repeatable service and focused on specific perspectives and features, e.g. algorithms to use. The aspects that were considered important in each application of process mining in business environment were not considered.

With these considerations in mind and, in order to solve the presented problem, we propose to **uncover the processes stored in ITS by using process mining techniques** with a development of a method. It will combine process mining analysis with the agile approach to **impose continuous feedback** and **involve the customer in the analysis process**. Despite the agile approach being more software development oriented, it is believed that the main ideas can be applied to process mining. According to DSRM, it will be the developed artifact.

3.1. Methodology

The proposed methodology is based on agile and Scrum development processes, heavily incorporating their iterative nature. During the whole process, it is important that everyone is on the same page and understands which course the analysis will take. Thus, the full potential of process mining can be explored.

3.1.1. Preparation phase

In order to address the problem faced, every stakeholder involved in the process has to be present, i.e. business individuals and Scrum team. The preliminary step consists on the definition of the aspects of the process that the customer wants to explore. Based on the current knowledge, requirements are gathered and added to product backlog. It should be noted that the items present in the backlog are process mining oriented. These also will be prioritized according to information available and will refer the process chosen.

Furthermore, this phase is used to gain some familiarization with the business process and technology. It will allow to have some previous knowledge and understanding of the nature of the process and the way the data is stored.

After these tasks are concluded, the sprints are planned by Scrum team along with the customer and management. To prove the value of process mining from the start, the will consider one of key but simple aspect to gain credibility along the customer.

3.1.2. Development phase

The development phase can be divided in six different components, such as data extraction, log inspection, control flow analysis, performance analysis, role analysis and validating the results. The Scrum team, represented by Scrum master, development team and product owner, is heavily present throughout the process. The customer and management are required when the results are transferred.

Data extraction

During the first sub-phase, the data is extracted from ITS and pre-processed in order to satisfy the assumption than an event log must have at least events, activities, performers and timestamps. Nevertheless, some additional information can also be stored. The team must define rapid and simple way of extracting and storing information. The decision of technology to use is also up to the team.

Log inspection

Familiarization with the information stored in the event log will be performed. For this purpose, Disco will be used. Its ability for performing quick and effective analysis will play the central role in the adaptation process. Information regarding processes obtained by combining different attributes extracted in the previous stage, the statistical information, such as number of cases, number of performers, total number of sequences, etc. will be obtained and analyzed.

Control flow analysis

The focus of control flow analysis is obtaining the central process stored in the log by applying different discovery algorithms. The combination of ProM's and Disco's functionalities will be done to exploit its full potential and get a better understanding of the process. Moreover, the comparison between the results obtained will be possible.

Performance analysis

During this phase, as the name implies, the performance checking on the discovered process will be done. Likewise control flow analysis, ProM and Disco's functionalities are used.

Role analysis

The main objective of role analysis is analyzing the event log regarding to the social perspective stored on it. The interactions, social networks and performers roles stored in the ITS's log are uncovered with the usage of Prom's plugins and Disco tool.

Validating the results

When the sprint is over, the results are transferred to the client and the product backlog is updated. The feedback provided by experts will serve as input for the next sprint. The process mining analysis can benefit from this approach since the customer will have the results earlier and there is the possibility to focus on what matters and clarify all the doubts found in the log.

3.1.3. Transferring phase

During this phase, the conclusions about the overall process are drawn. The results obtained from each perspective and, if applicable, some suggestions of improvements are provided. Moreover, the documentation regarding the software developed and analysis performed is produced. In the case of in-person meeting, the presence of the management and the experts is required.

4. Demonstration

The following chapter corresponds to the demonstration step of DSRM and, as the name implies, will demonstrate the applicability of the proposed solution. In **section 5.1,** an overview of the issue tracker used is performed. **Section 5.2** provides the description of the applicability of the methodology in the issue management context. The second applications regarding the idea management is demonstrated in **section 5.3**. Finally, the chapter is concluded by the third application of the proposed solution (**section 5.4**).

4.1. JIRA

Nowadays, JIRA is one of the most popular ITS. It is a web-based application provided by Atlassian Software Systems that challenges the definition of an issue tracker. Beyond dealing with software bugs, JIRA handles problems and helps in administration of projects [40]. It is a highly customizable tool and is built upon the concept of an issue. For a better understanding, a screenshot of JIRA's issue is presented in Figure 6.

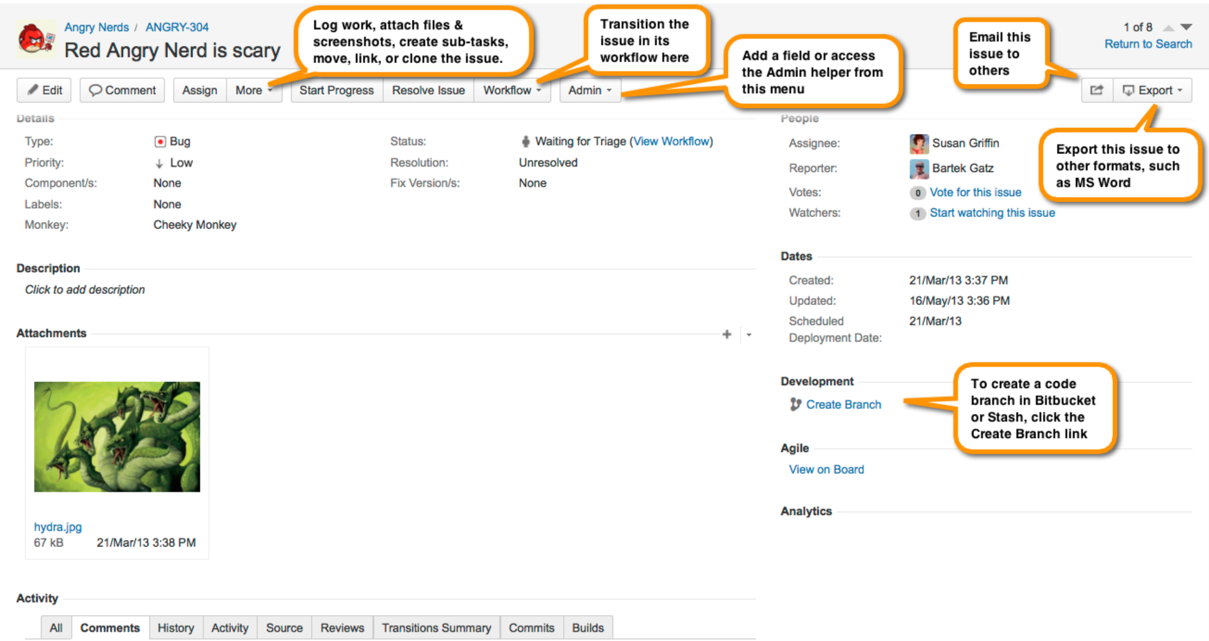

Figure 6 JIRA's Issue Screenshot[1]

In present context, an issue can be a bug, an improvement, a new feature, a task and a custom issue. The last one can be used in the customization process by the organization that decided to adopt this tool.

[1] https://confluence.atlassian.com/jira064/what-is-an-issue-720416138.html

After being created, an issue follows a predefined life-cycle, i.e. at some point of time is stays at one of the predefined statuses. The definition of the statuses and transitions among them is up to the organization and is performed at the moment of installation.

Each issue stores information regarding its reporter and assignee. The main difference relies on the fact that the reporter is the person that detects an issue and the assignee is the person who aims to solve it. Another important part of an issue is the time information. In JIRA’s context, the moment an issue enters in the systems, its update and the date of its resolution are considered significant and are recorded. This information may be used to perform some statistical information.

The last component of an issue to be mentioned in the scope of this work is the History section. It aims to record every change made to a issue and can provide information about the creator, the changes made on issue field, deletion of a comment, etc. Also, for every change, one can look not only to a person that made the change, but also to the time at which it was made and, in case of changes to the issue field, to the new and old values of the field. To provide a better understanding, a screenshot of History section is presented in Figure 7.

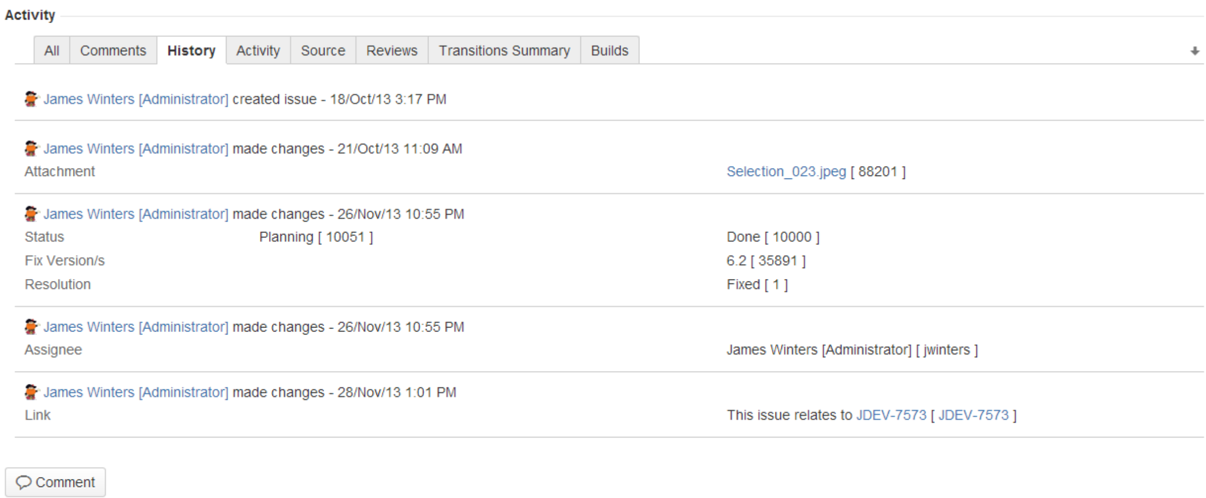

Figure 7 History Section of an Issue[2]

As a final point, it is important to note that in JIRA’s scope, a project is represented by a collection of issues and its definition depends on the organization’s needs.

[2] https://confluence.atlassian.com/jira064/viewing-an-issue-s-change-history-720416460.html

4.2. Issue management

In order to provide a better understanding of the organizations processes, an overall analysis took place regarding four projects, namely Xray, Xray-Support, XPORTER and XPORTER-Support.

4.2.1. Preparation phase

Since all projects in JIRA's context are represented as collections of issues, they became the main focus of the analysis. The organization A was interested in exploring the issue's behavior in different contexts and understand where the main loses were. In terms of roles, Scrum team, management and the experts were present.

In terms of the attributes of an issue, the focus was made on reporter, status, assignee, priority, type and resolution fields. Thus, the issue will represent the case identifier and the moment it suffered a change will be the timestamp. The performer will be chosen between the reporter and assignee. In order to not restrict the analysis, the choice of the activity was left for the moment when the overall knowledge is more elaborated.

4.2.2. Development phase

First iteration

Throughout the first week, the familiarization with the process and technology was performed. Since the work would be performed on web-based application, the way JIRA was working had to be learnt, i.e. the tool and its documentation were studied.

The following four weeks were dedicated to data extraction and its storage according to the format assumed by the process mining analysis.

- **Data extraction**

The approach taken can be divided in three steps, such as obtaining the list of projects, extracting the issues per project and transforming the XML to CSV document.

I. **Obtain the list of projects**

During the first component, HTML connection is established with an explicit link. The username and password are specified. The link is concatenated with "rest/api/2/project" to obtain JSON object with required content. For example, in case of specifying "http://jira.- confidential +.com", the result will be "http://jira.- confidential +.com/rest/api/2/project".

After having the JSON object, an XML document is created with the corresponding list of projects and its attributes (Figure 8).

```xml
<root>
    <projects>
        <project>
            <name>XRAY FOR JIRA</name>
            <key>XRAY</key>
            <id>10001</id>
            <self>http://jira. - confidential + .com/rest/api/2/project/10001</self>
        </project>
        <project>
            <name>XPORTER for JIRA</name>
            <key>XPORTER</key>
            <id>10000</id>
            <self>http://jira. - confidential + .com/rest/api/2/project/10000</self>
        </project>
        <project>
            <name>XPORTER for JIRA Support</name>
            <key>XPORTERSUPPORT</key>
            <id>10201</id>
            <self>http://jira. - confidential + .com/rest/api/2/project/10201</self>
        </project>
            <name>XRAY for JIRA Support</name>
            <key>XRAYSUPPORT</key>
            <id>10202</id>
            <self>http://jira. - confidential + .com/rest/api/2/project/10202</self>
        </project>
    </projects>
</root>
```


Figure 8 XML Document with a List of Projects

II. Extracting the issues per project

In concrete terms, each project contains information about its name, key, id and link. The last attribute allows to access to the current content of a project. Moreover, the total number of issues is added. For this purpose, the link given at the beginning will be concatenated with "/src/jira.issueviews: searchrequest-xml/temp/+SearchRequest.xml?jqlQuery=project+%3D+ + PROJECT_NAME + &tempMax=0". Thus, another XML file that contains a list of projects with respective names, keys, ids and number of maximum issues is produced. Due to space limits, Figure 9 represents a partial information.


```xml
<project>
   <name>Xporter for JIRA Support</name>
   <key>XPORTERSUPPORT</key>
   <id>10201</id>
   <self>http://jira - confidential + .com/rest/api/2/project/10201</self>
   <maxIssues>124</maxIssues>
</project>
<project>
   <name>Xray for JIRA Support</name>
   <key>XRAYSUPPORT</key>
   <id>10202</id>
   <self>http://jira - confidential + .com/rest/api/2/project/10202</self>
   <maxIssues>234</maxIssues>
</project>
```


Figure 9 XML Document with Total Number of Issues Per Project (partial)

With the collected information, the issues presented in each project are extracted. Moreover, the issues that do not exist or are not available are filtered.

In terms of content, the focus was made on such fields as reporter, status, assignee, priority, type and resolution. Throughout the lifecycle of an issue, the fields status, assignee and priority can experience some changes. Thus, the History section was analyzed in more detail.

All the stored information was accessible through the concatenation of the link specified at the beginning and "/rest/api/2/issue/ + PROJECT_NAME +? expand=changelog". Here, the previously extracted number of issues was used in order to bound the extraction.

With all the data gathered, the final XML file with all the issues per project and its content is produced. Due to space limits, only a partial representation is presented in Figure 10.


```
<project>
    <name>Xray for JIRA</name>
    <key>XRAY</key>
    <id>10001</id>
    <self>http://jira. - confidential + .com/rest/api/2/project/10001</self>
    <maxIssues>476</maxIssues>
        <item>
           <key>XRAY-68</key>
           <Summary>Actions gear is missing in the Test Execution page</Summary>
           <type>Bug</type>
           <priority>Major</priority>
           <initialPriority>Major</initialPriority>
           <status>Resolved</status>
           <initialStatus>Open</initialStatus>
           <assignee>Bruno Conde</assignee>
           <initialAssignee>Bruno Conde</initialAssignee>
           <resolution>Fixed</resolution>
           <reporter>Bruno Conde</reporter>
               <histories>
                <history>
                    <firstAuthor>Bruno Conde</firstAuthor>
                    <firstCreated>2014-05-06T18:20:25.162+0100</firstCreated>
                    <currentStatus>Open</currentStatus>
                    <currentAssignee>Bruno Conde</currentAssignee>
                    <currentPriority>Major</currentPriority>
                    <currentField>Created issue</currentField>
                </history>
                <history>
                    <authorName>Bruno Conde</authorName>
                    <timestamp>2014-05-06T18:20:35.416+0100</timestamp>
                    <currentStatus>Open</currentStatus>
                    <currentAssignee>Bruno Conde</currentAssignee>
                    <currentField>Component</currentField>
                    <originalValue>null</originalValue>
                    <newValue>User Interface</newValue>
                </history>
```


Figure 10 XML Document with Issue's Data (partial)

III. Transforming the XML to CSV

In order to perform a process mining analysis, the XML file previously produced had to be converted to CSV format. To achieve this, XSLT which stands for transforming an XML document to another one was used. Mostly, the information regarding each issue was recorded in a new document and delimited by a comma.

The final log has seven original fields extracted directly from JIRA and four manipulated ones. The last ones aim to capture the actions performed on each issue represented in History section.

Each entry of the CSV file contains information regarding the case identifier, activity performed, performer, timestamp and, finally, the change experienced by activities throughout the time. More specifically, issue's name was recorded as case identifier and the rest of the fields (type, status, priority, resolution, reporter, assignee, "changedField", "statusHistory", "priorityHistory" and "assigneeHistory") were used as activities.

As previously presented in this report, an issue can experience changes throughout the time and these are stored in History section. In order to capture the issue's behavior, the decision was made to repeat each issue and just update those fields that have been changed. In practice, three new fields were created such as "statusHistory", "priorityHistory" and "assigneeHistory". Likewise, to explore the customization options, the "changedField" was created.

The following phases such as log inspection, control flow analysis, performance analysis and role analysis took one week.

- **Log inspection**

The provided projects were treated in separate ways, thus the constructed logs presented different number of cases and events. The data gathered was from one year and half period.

I. Xray

The resulted log contained 476 issues, a total number of 7476 events and 10 types of different activities. The dominant activity was "Open" (60,47%), followed by "In Progress" (11,84%), "Waiting for Testing" (8.86%), "Closed" (4,35%), "Testing" (4,17%), "Waiting for support" (0.8%), "Reopened" (0.68%), "Blocked" (0,33%) and, finally, "Waiting Approval" (0.24%). Regarding to the number of different sequences of activities, 299 variants were found. The constructed log presented 42 performers.

II. Xray-Support

The log comprised 234 issues, 1,987 events and 121 different sequences of activities. The number of activities was five, namely "Waiting for support" (49,32%), "Waiting for customer" (33,87%), "Resolved" (15,95%), "Frozen" (0.55%) and "Open" (0.3%). Regarding to the performers, their number was 113.

III. XPORTER

The constructed log contained 249 issues, 4564 events and 10 types of activities. The dominant activity corresponded to "Open" (45,42%), followed by "Resolved" (14,55%), "In Progress" (14,22%), "Waiting for Testing" (14,13%), "Testing" (7,1%), "Closed" (2,04%), "Reopened" (1,64%), "Waiting for support" (0,53%), "Waiting for Approval" (0,26%) and, finally, "Blocked" (0,11%). The number of performers associated to the activities was 25. In terms of variants, 202 sequences were discovered.

IV. XPORTER-Support

The resulted log contained 124 issues and 1512 events. It was possible to discover four different types of activities, namely "Waiting for support" (46,23%), "Waiting for customer" (36,18%), Resolved (16,73%)

and Frozen (0.86%). Regarding to the number of different sequences of activities, 97 variants were found. The constructed log presented 81 performers.

Throughout the inspection it became obvious that in terms of the number of events, the projects of support were considerably smaller. Consequently, the variability if the information and the ability for covering the higher number of paths possible was questioned. Despite this observation, the analysis was performed in all four projects.

- **Control flow analysis**

Throughout this phase, apart from project's scope analysis, analysis in an academic context was performed. However, due to space limitations, only more relevant and interesting processes will be presented.

To discover the workflow stored, Disco and ProM's functionalities were used. The results with Disco are presented in Figure 11 where is shown the information in terms of frequency. For better understanding, the activities and paths filters were set at 100% and 35%, respectively. In concrete terms, it means that all the activities and only 35% of the total number of paths will be displayed. Appendix B represents the process with all the transitions among its activities.

Before performing a deeper analysis, it is important to note that the frequency is shown by the thickness of arrows and the color of activities. The number that appears inside the activity box indicates the frequency of that activity and the number near the arrow describes the number of transitions made. For example, in Figure 11, "Open" activity occurred 476 times and the transition "Open -> In Progress" took place 241 times.

Among ProM's discovering algorithms, we chose to apply the Heuristic Miner algorithm due to its ability to focus on the main process. The obtained process is presented in Appendix B.

After a careful analysis, it can be argued that the model discovered are identical. However, the one discovered by Disco tool allowed to identify the main path taken by the issues ("Open -> In Progress -> Waiting for Testing -> Testing -> Resolved -> Closed"). The considerable number of transitions in both cases supports the presence of variants of the process.

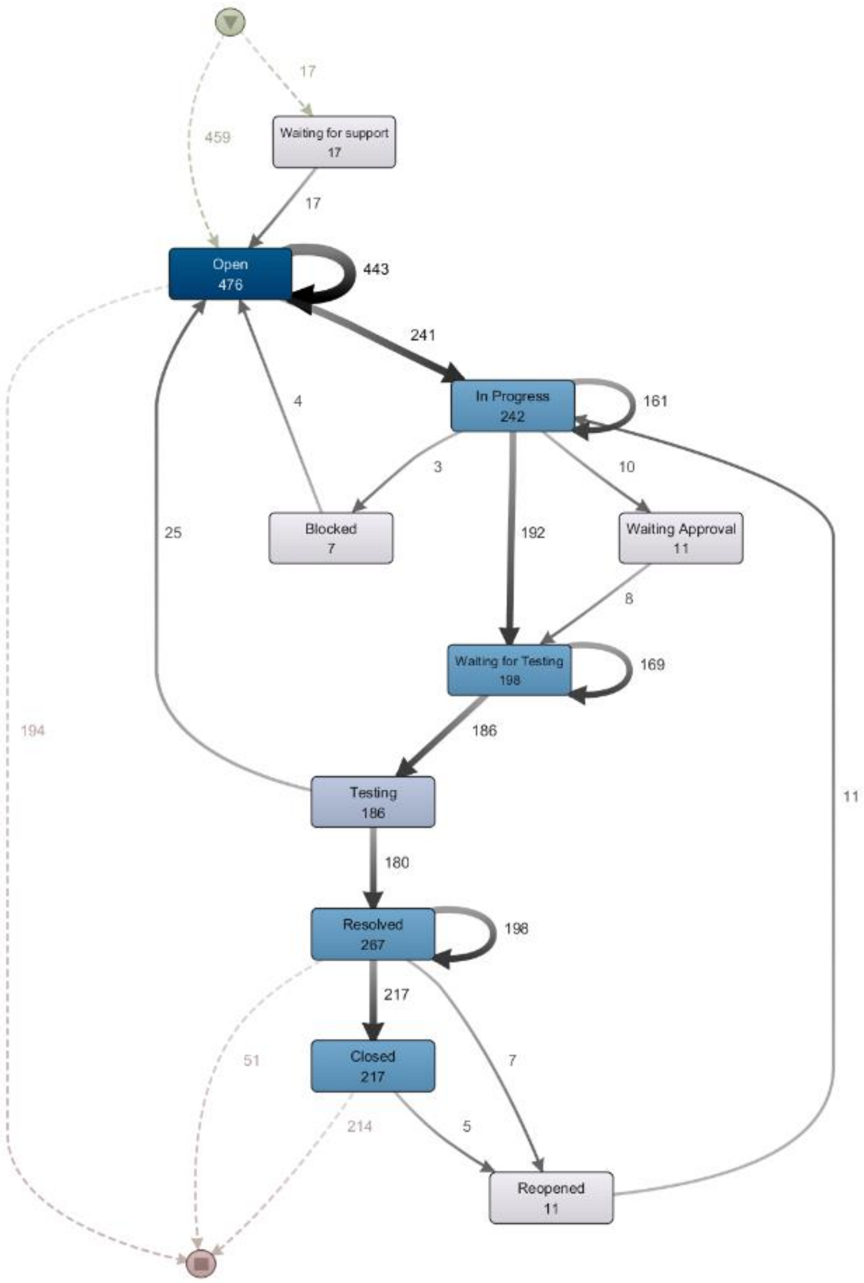

Figure 11 Issue Management Workflow – frequency (Xray project; 100% of activities and 35% of paths)

- **Performance analysis**

The results with Disco tool are presented in Figure 12 where is shown information regarding mean performance. Likewise control flow analysis, the activities' and paths' filters were set at 100% and 35%, respectively. Appendix B represents the process with all the transitions among its activities.

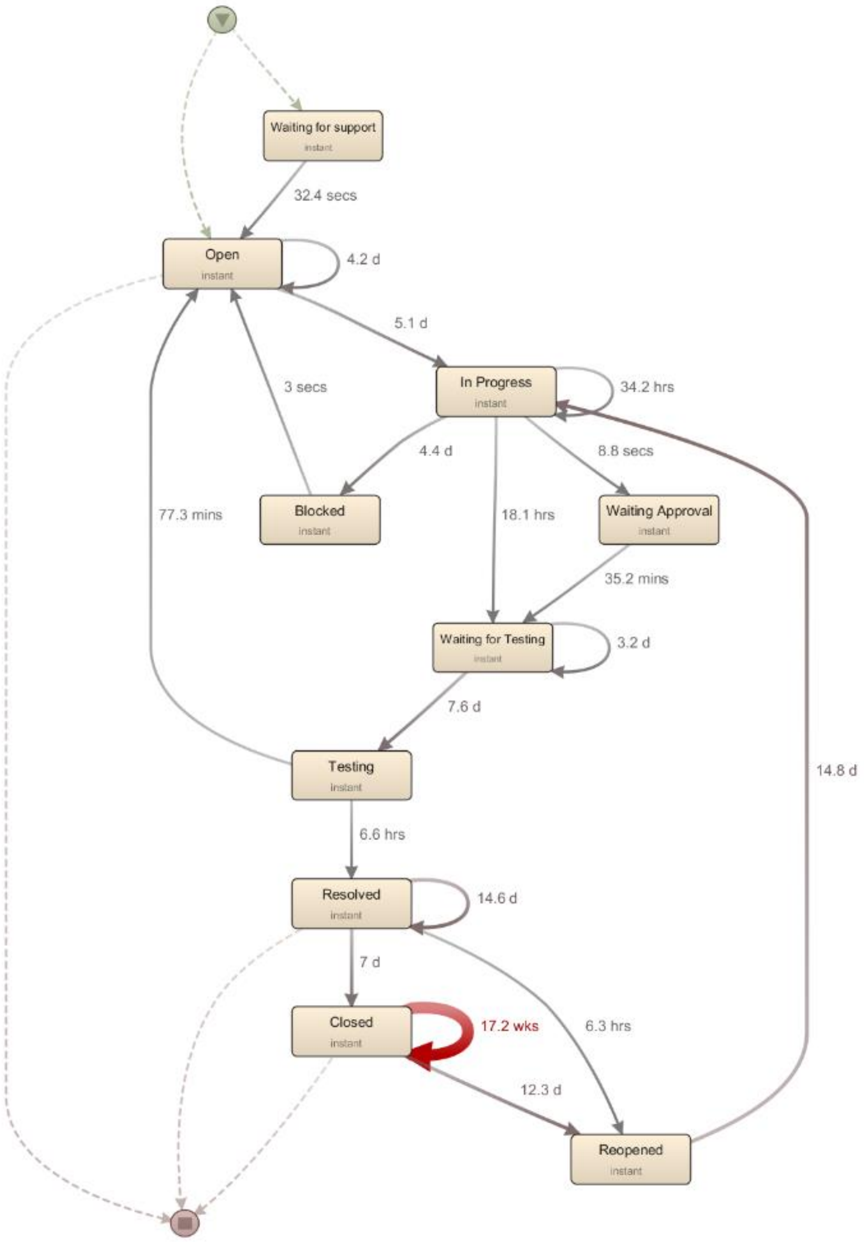

Figure 12 Issue Management Workflow – performance (Xray project; 100% of activities and 35% of paths)

According to mean metric, it takes 89.6 days from the moment that an issue is opened until its final resolution, i.e. in a real life the organization takes approximately 90 days to handle an issue.

In terms of bottlenecks, only one was detected and it concerns the final phase of the process. The transition “Closed -> Closed” takes, in average, 17,2 weeks. When applying in the organizational context, it means that the time required for an issue to be actually closed is approximately 4 months.

Although the followed methodology suggested using ProM’s functionality, in this case the results obtained did not allowed to extend the analysis performed with Disco.

- **Role analysis**

For this phase, the outcomes from the two projects will be provided. Since the goal is to discover the interactions between individuals of the organization, the results to be presented were based on the assignee field. At technological level, Disco and ProM's functionalities were used.

The results obtained with Disco regarding the projects are present in Figure 13 and Figure 14 where the information based on frequency is presented. For better understanding, the activities' and paths' filters were set at 100 % and 50% in the first case and, at 100% of activities and 0% of paths in the second one.

In Figure 13, seven different assignees can be found, however "Bruno Conde", "Pedro Rodrigues" and "Diamantino Campos" were identified as main performers. Closer interaction was observed between "Bruno Conde" and "Pedro Rodrigues", corresponding to 146 transitions. As well, "Bruno Conde" plays central role since in the event log it occurred 461 times, corresponding to 69.46% of all cases.

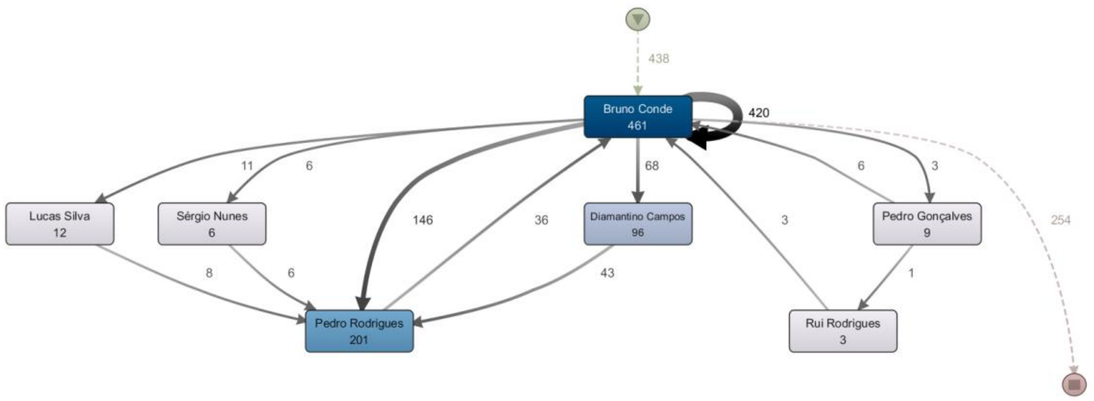

Figure 13 Social Network (Xray project; 100% of activities and 50% of paths)

In Figure 14, two main performers were found, such as "Bruno Conde" and "Diamantino Campos", and their strong interaction is represented by 103 transitions. The total number of assignees involved in the project is four. Similarly, in this case "Bruno Conde" plays a central role since it appeared 213 times, corresponding to 58.53% of all cases.

As a final point, one can argue that regardless of the project, "Bruno Conde" is the main performer, i.e. he is the leading person to whom the issues are assigned in both types of project. This observation was considered and explored with ProM's organizational mining algorithm. The obtained results did not allow to extend the analysis performed with Disco, i.e. there was no interactions found between performers. In Appendix B, it is represented the results regarding the Xray project.

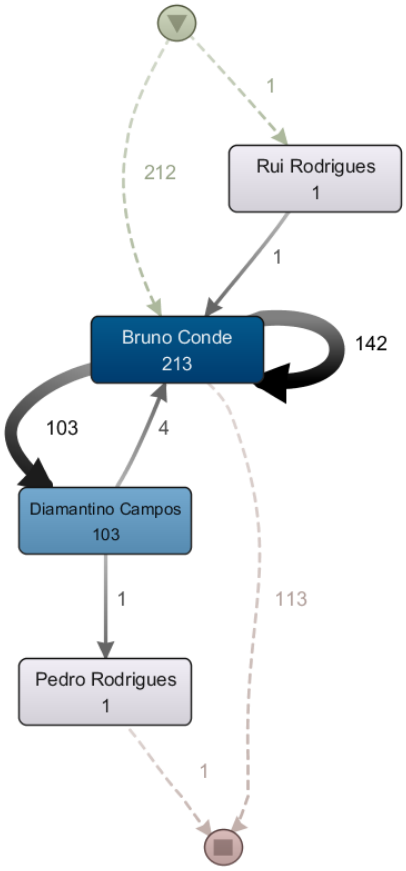

Figure 14 Social Network (Xray-Support project; 100% of activities and 0% of paths)

- **Validating the results**

Throughout the **data extraction** month, the experts from the organization A were contacted to present the developed software and clarify all technical issues. In order to ease the analysis and focus on one project at the time, the decision was made to extract the information per project. Since the URI paths were used, the total number of issues was gathered to bound the extraction. Moreover, during the data gathering process, the organization expressed interest in understanding which were the most customized fields. Thus, the “changeField” was created.

After the **log inspection**, the assignee and status field provided the most interesting results. Here, the opportunity to explore them and their relationship more closely was found.

In the case of type and priority fields, the analysis showed no interactions between them or a single predominant one. Thus, the decision that was made was not to perform further analysis of these fields. It would not bring any additional information about the subject of study.

Regarding to the reporter field, the situation in the four projects was similar. There was no interaction between the activities and its number was too high. Finally, the results presented in the previously created “changedFiled” were too scattered for academic purpose.

In order to present and validate the control flow, performance and role analysis, the in-person meeting took place. The Scrum team and business individuals, such as experts in JIRA and management were present. The client organization recognized the process models obtained.

In terms of **control flow** and **performance analysis**, the customer was surprised with the additional five activities found. The situation was explained by the way in which JIRA deals with the projects, i.e. some issues may be cloned from one project to another. The decision was made to filter those activities, namely “Waiting for support”, “Waiting Approval”, “Waiting for Testing”, “Testing” and “Blocked” from the Xray project.

In the case of **role analysis**, the centrality of “Bruno Conde” presented in both Xray and Xray-Support projects was surprising, i.e. the organization was not aware of its contribution.

Finally, the earlier observation about smaller projects was considered and XPORTER and XPORTER-Support were discarded to perform richer analysis (mentioned at the beginning of the **section 5.2**). At technological level, the organization was interested in continuing with Disco.

Second iteration

Throughout the following three weeks, the XML files regarding to Xray and Xray-Support were filtered according to the specifications, i.e. the activities were removed, and the CSV document was produced.

- **Log inspection**

The resulted log contained 6399 events and five types of different activities. The dominant activity continued to be “Open” (70,65%), followed by “In Progress” (13,86%), “Resolved” (9.61%), Closed (5,08%) and, finally, “Reopened” (0.8%). Regarding the number of different sequences of activities, 259 variants were found. The number of performers remained the same (42 performers).

- **Control flow**

In Figure 15 is presented the process obtained with 100% of activities and 50% of its paths.

The “Open -> In Progress -> Resolved -> Closed” was discovered to be the main behavioral pattern. In the case of “Reopen” activity, few issues were conducted to it. Thus, it represents the infrequent behavior of the process.

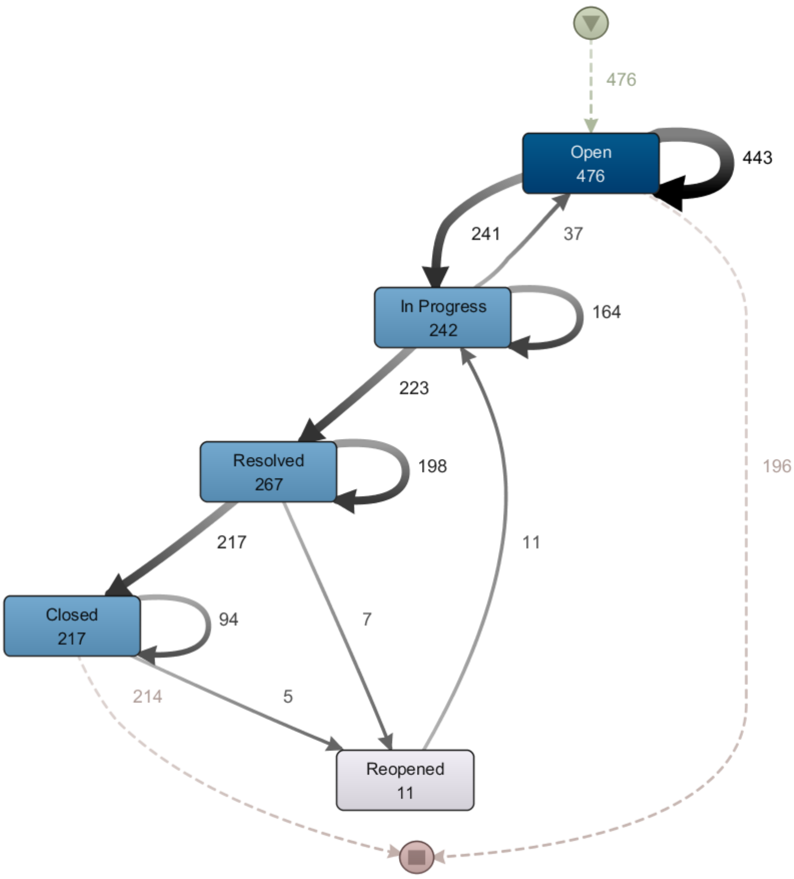

Figure 15 Issue Management Workflow – frequency filtered (Xray project 100% of activities and 50% of paths)

- **Performance**

After filtering the activities that did not belong to the predefined workflow, the results in terms of performance are represented in Figure 16. For better understanding, the activities' and path's filters were set at 100% and 50%.

According to mean metric, an issue takes approximately 89,4 days for being handled. When comparing with the previous iteration, the time remained approximately equivalent (decrease of 0.2 days).

The major bottleneck that was found in the previous iteration has remained. However, other transitions taking the longest time can be spotted in different points of the process. For example, an issue that has been reopened took 14.8 days to be handled ("Reopened -> In Progress "). A similar case can be verified in the transitions "In Progress -> Open" and "Resolved -> Resolved". This observation was confirmed after replaying the constructed event log on the process.

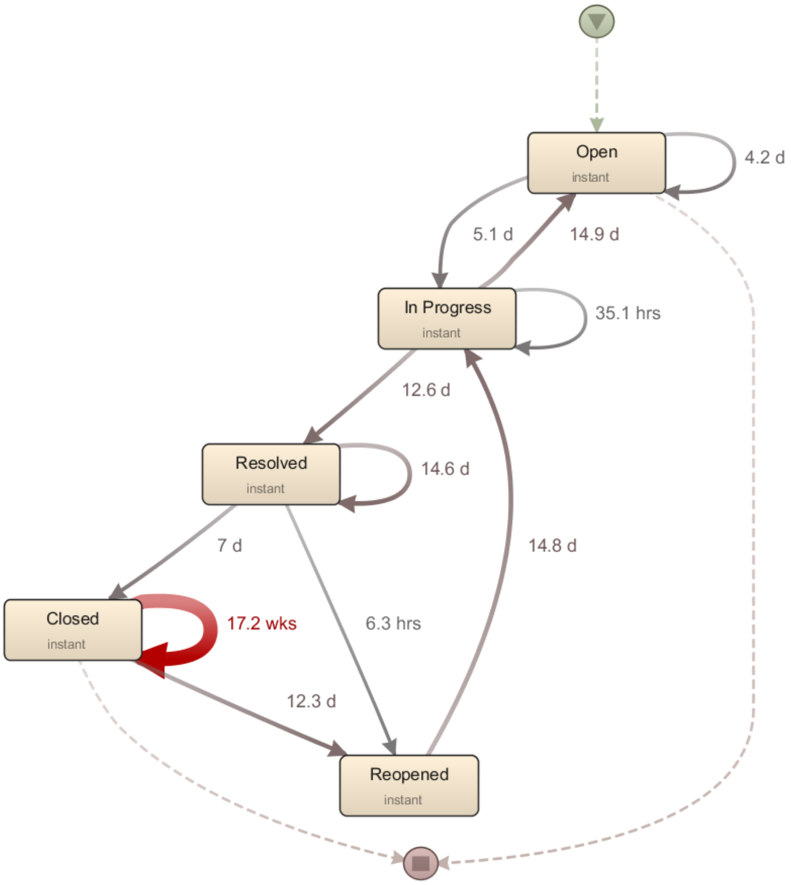

Figure 16 Issue Management Workflow – performance filtered (100% of activities and 50% of paths)

- **Validating the results**

In order to validate the obtained results, a meeting took place. The client organization (experts and management) recognized the process model discovered during the control flow analysis as the final one. Moreover, the path “Open -> In Progress -> Resolved -> Closed” was validated as the most used in those projects.

A special attention was given to the performance analysis. The second longest time presented in the process can be explained by the low number of cases present between “Reopened -> In Progress” and “In Progress -> Open”. When comparing to control flow analysis, these are represented by 11 and 37 cases, respectively. Moreover, these transitions represent the moments in which the issues return to the beginning of the process. Thus, this situation reflects the fact that the organization gives the higher priority to new issues.

4.2.3. Transferring phase

During one week, the documentation was produced to explain the performed work. The software developed among with process mining analysis from each iteration were documented and transferred to the organization.

In a summarized manner, the main findings after two iterations were:

- The discovery of a business process regarding to a workflow of an issue;
- The particularity of JIRA when dealing with issues that belong to more than one projects and its impact in the overall workflow;
- The uncovered average time for handling an issue is approximately 90 days;
- The observed major bottleneck present at the final phase of a process;
- The performance results regarding to possible bottlenecks present in different phases of the process ("In Progress -> Open" and "Resolved -> Resolved") representing the organizations tendency to give higher priority to the new issues;
- The centrality of one performer (assignee) among different projects.

At technological level, a Disco tool supported the needs of the organization. During the first iteration, its ability for a quick analysis allowed to get familiarized with the event log and explore the additional information that the extracted fields could bring. Furthermore, through the combination of these fields, the knowledge about the event log was improved and the more promising analysis were spotted. This fact allowed to clarify the doubts from the organization during the results transferring. Moreover, it allowed to make some suggestions. ProM, on the other hand, was difficult to understand and the analysis performed did not strength the results.

4.3. Idea management

Organization B explores JIRA's ability of administration of projects in the idea management context. Here, the goal is to support employee's creativity and stimulate ideas' production. The most valuable ideas are eventually implemented.

Additionally, during the second application of the methodology, the project involving process mining analysis was developed.

4.3.1. Preparation phase

The organization was not familiar with the process mining concept. During the first meeting, an overview of process mining area and its techniques was performed to provide some insights and do not create false hopes or mislead the client. The management, experts and Scrum team were present.

JIRA was used to manage the innovation process; thus, the focus was made on the business process regarding the idea and initiatives. The organization was interested in the possibility of uncovering the stored process and compare it with the predefined one. Moreover, the management felt especially enthusiastic for obtaining the results from performance analysis.

The second meeting took place with one of the experts and aimed to provide an overview of the process chosen for the analysis and some insights of the way JIRA was used.

Idea management process

Whenever an employee has an idea, he/she register it in the system (“Registada”). The idea is attached to its creator and remains private until is confirmed (“Confirmada”). The confirmation process aims to guarantee that the required information it is clear and comprehensive.

Afterwards, the idea is published without an indication of its creator. Here, the goal is to get people to vote and comment on each other’s ideas and do not be influenced by personal relationships. At this point, an idea can also be reviewed in the case of it is redundant or out of scope, i.e. avoid confusion between an idea in innovation area and a simple bug or some technological issues (“Em revisão”). In these cases, an idea is returned to its creator and he/she must perform the required modifications or include the missing information (“Registada”).

When the description of an idea is considered complete, it is analyzed and classified per four parameters (“Avaliação em Comité”). Even after the classification process, some additional clarification may be needed. Thus, an idea can transit to “Em Revisão para Comité”.

Then, the ideas are presented to a steering committee (“Avaliada em steering”) and, according to its classification, are approved (“Aprovada em steering”) or not (“Cancelada”). In the case thay still raise some doubts, the idea can be reviewed by a steering (“Em revisão para steering”).

Finally, when the idea is approved, it is promoted to an initiative (“Promovida a Iniciativa”). Since the ideas and initiatives are related, “Implementada” and “Implementação Avaliada” stages aim to express its evolution. Note that to an initiative is assigned an owner that is responsible for its implementation. He/she has to accept it in order to start working. Moreover, it is up to him/her to update the initiatives status.

4.3.2. Development phase

Throughout this application of the proposed methodology, the iterations were considerably smaller. The data extraction process was already implemented from the previous application in the issue tracking environment. Thus, the first iteration was one-week long and the second one took three weeks.

First iteration

- **Data extraction**

The data was extracted according to the approach described in **section 5.2.2**. In order to identify the specified project, the custom field was used.

The focus was made on such fields as status and assignee, and on the History section. Both, time information and the changes performed on an idea, were extracted from the History section. During the transformation of XML to the CSV document, the “statusHistory” and “assigneeHistory” were updated with the required information. However, the fields regarding to the assignee was marked as “Unassigned”. The decision made was to continue with the analysis and clarify this aspect with the experts.

- **Log inspection**

The data gathered was from two-year period. The resulted log contained 690 ideas, a total number of 13571 events and 12 different activities. The dominant activity was “Avaliada em Comité” (27,49%), followed by “Confirmada” (16,73%), “Registada” (15,16%), “Aprovada em Steering” (9,02%), “Avaliada em Steering” (7,93%), “Cancelada” (7,64%), “Promovida a Iniciativa” (6,68%%), “Em Revisão” (5,34%), “Implementada” (3,12%), “Em Revisão para Steering” (0,42%), “Impelementação Avaliada” (0,38%) and, finally, “Em revisão para Comité” (0,1%).

Regarding the number of different sequences of activities, 469 variants were found. The constructed log presented 166 performers.

- **Control flow**

Appendix C represents the discovered business process for idea management from the moment of its registration until its implementation.

Since the organization had a predefined process, the models were compared. The discovered model was very similar to the predefined one. Its main difference relied on several additional transitions, such as “Confirmada -> Registada”, “Confirmada -> Cancelada”, “Registada -> Cancelada”, etc. These transitions, apart from being a proof of deviations in process behavior, support the idea of its variability.

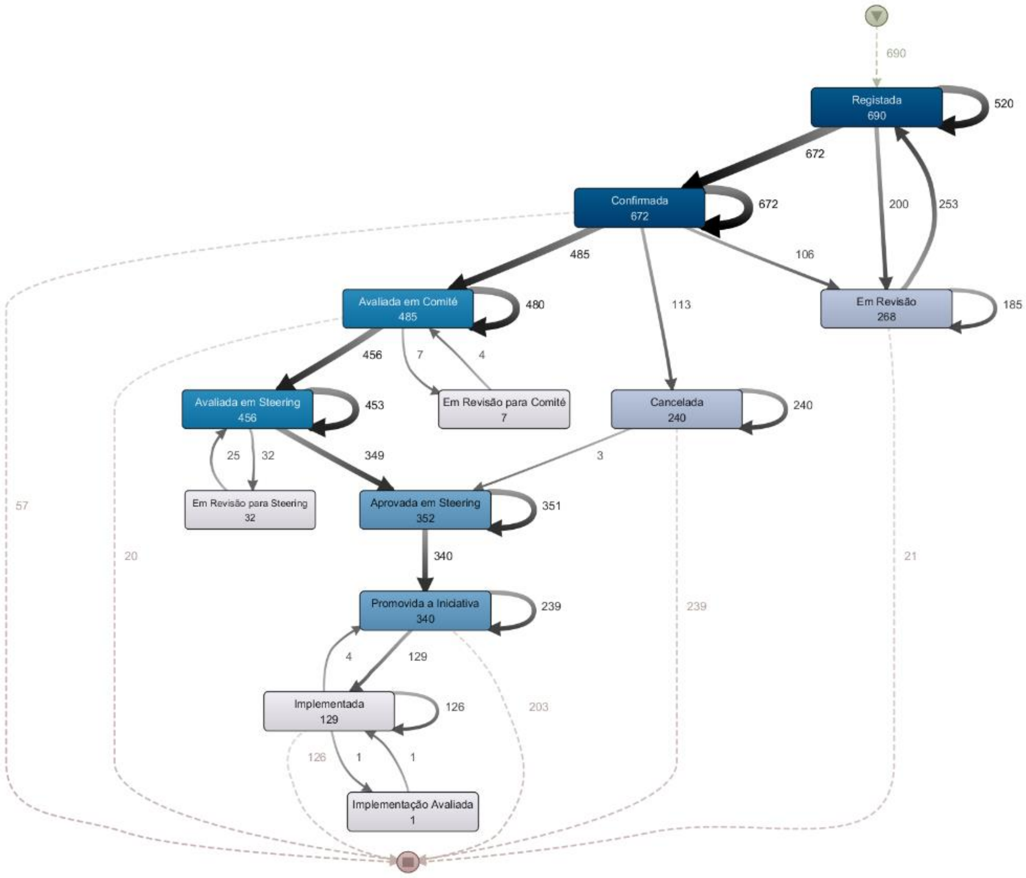

Figure 17 Idea Management Workflow – frequency (100% of activities and 50% of paths)

The behavioral pattern “Registada -> Confirmada -> Avaliada em Comité -> Avaliada em Steering -> Aprovada em Steering -> Promovida a Iniciativa” is stated as a dominant one for an idea. It is represented in Figure 17. For a better understanding, the activities and paths filters were set at 100% and 50%, respectively.

In terms of infrequent behavior, the patterns that involve “Em revisão para Comité”, “Em Revisão para Steering”, “Implementada” and “Implementação Avaliada” activities were considered as so. The first two activities express the situation when some additional information about an idea is required. The goal of the “Implementada” and “Implementação Avaliada” activities is to follow up the implementation developments. Thus, a possible explanation can rely on this fact.

The most interesting situation was found in the transition “Cancelada -> Aprovação em Steering”. In real life, it means that after the cancellation of an idea, the steering can still approve it. This aspect was analyzed and three ideas that make this transition possible were discovered. Due to confidentiality issues, their main workflow is represented in Figure 18.

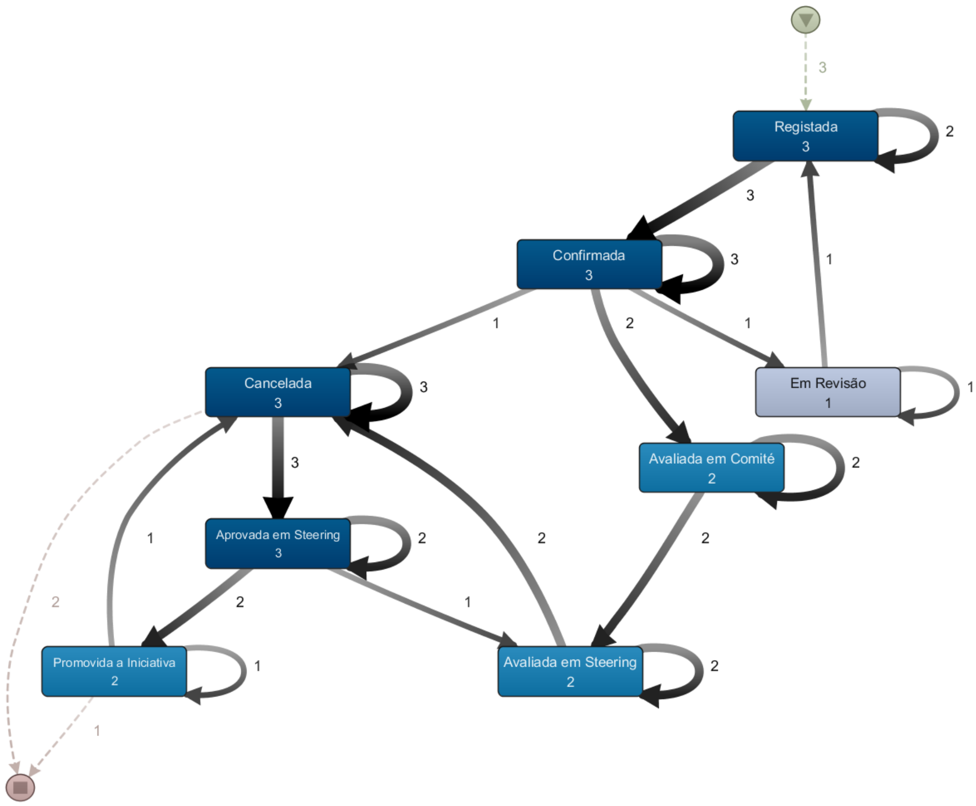

Figure 18 Idea Management Workflow – unusual behavior (100% of activities and 100% of paths)

Each one of the ideas expresses different variant of a process. However, two ideas were cancelled twice. In the first case, the idea was cancelled after being confirmed and then after being promoted to an initiative. In the second case, the cancellation was made after being evaluated in steering and lately after being approved in steering. The third idea was canceled after being evaluated in steering.

This observation was considered and were analyzed in more detail during performance analysis.

- **Performance analysis**

Appendix C presents the idea management process in terms of its performance. For a better understanding, the total number of activities and 50% of the paths are presented in Figure 19.

According to mean metric, an idea takes 81,75 days from the moment it is registered until it is implemented, i.e. in a real-life, the organization takes approximately 82 days to handle an idea and its implementation.

In terms of bottlenecks, these were detected during the whole business process, however the most worrying ones were only presented at the end of the process. For example, an idea that has been promoted to the initiative stays at “Promovida a Iniciativa” for 16 weeks and afterwards it takes another 21,6 weeks to transit to “Implementada”. As was already mentioned, the initiative is associated to an owner that will implement it and it is up to him/her to accept the initiative and update its status. Thus, the discovered time-period means that the acceptance process takes almost four months.

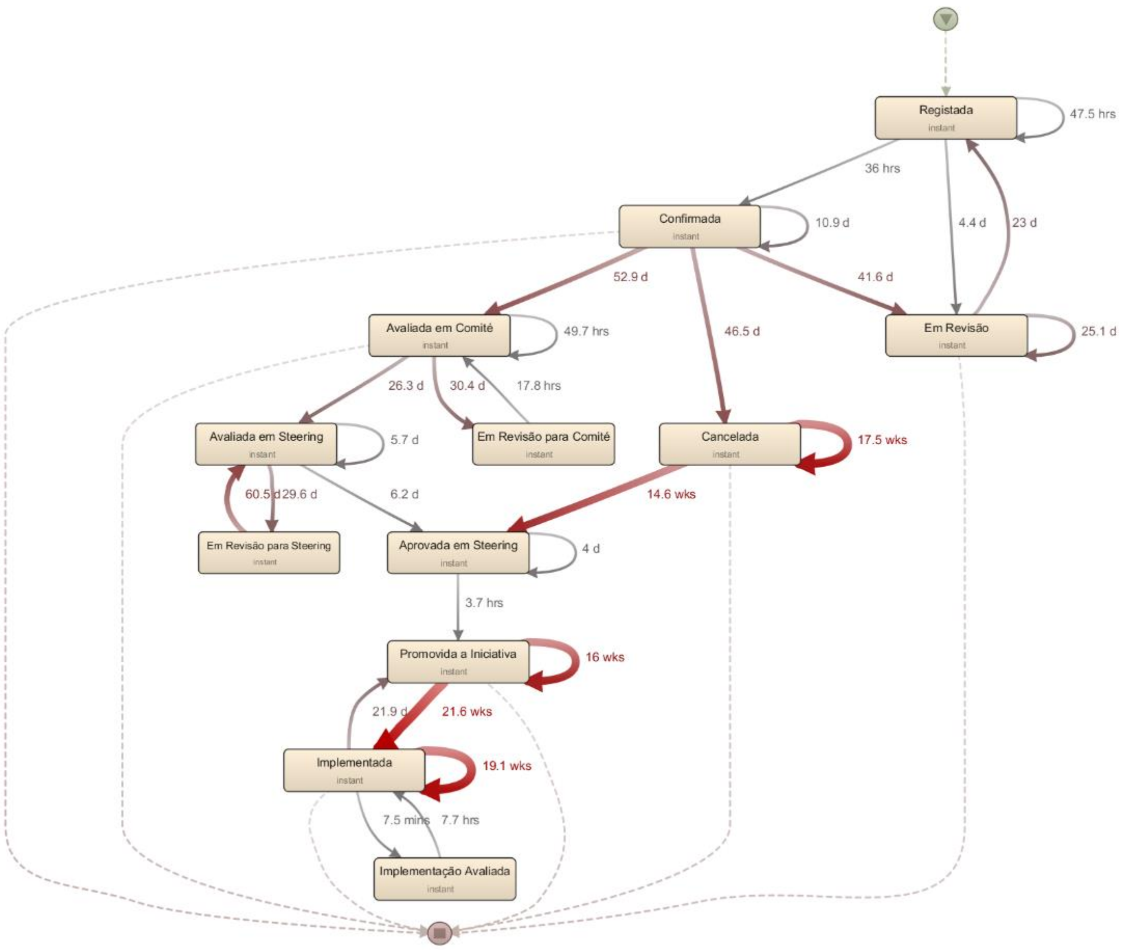

Figure 19 Idea Management Workflow – performance (100% of activities and 50% of paths)

Other transitions taking the longest time are “Cancelada -> Cancelada” (17,5 weeks) and “Cancelada -> Aprovada em Steering” (14,6 weeks). The second one corresponds to the deviating case spotted during the control flow analysis. Its time may be influenced by the number of ideas and the fact that some of them are cancelled more than once.

With an average transition time varying between approximately 42 and 53 days, the ideas that go through “Confirmada -> Avaliada em Comité”, “Confirmada -> Cancelada” and “Confirmada -> Em revisão” were considered as minor bottlenecks of the ideation process. This observation can be explained by the filtering performed in the early stages of the process.

Finally, the last case corresponds to another confirmation required by steering. The transition time takes 60,5 days from “Em Revisão para Steering” to “Avaliada em Steering”.

As was already stated, the most worrying bottlenecks were present at the end of the process. This observation was considered and Appendix C represents the issue’s workflow until the moment it is promoted to an initiative. Unlike the previous case, the process shows the overall number of activities and paths.

When comparing to the previous process' mean duration, the decrease of approximately 28 days was observed, i.e. from the moment an idea enters the system, it takes 53,8 days to go through the approvals and be promoted to an initiative. However, the bottlenecks at the end of the process remained. It became obvious that the path "Implementada -> Promovida a Iniciativa" influences a lot the results. Additionally, the bottlenecks regarding to the ideas approval process remained.

- **Validating the results**

To validate the obtained results from the first iteration, an in-person meeting took place. The client organization (experts and management) recognized the process model discovered during the control flow analysis as the expected one. Moreover, the pattern "Registada -> Confirmada -> Avaliada em Comité -> Avaliada em Steering -> Aprovada em Steering -> Promovida a Iniciativa" was acknowledged as the most used.

The deviating behavior discovered through the idea management process was not new for the organization. One of the experts stated that, at the beginning of the process, some additional filtering may be required. Since the employees are free to create an idea, there are cases with some ethical issues associated to. Moreover, as the process goes through, it is possible for an idea to be re-assessed. The organization is aware of the changing business environment and, in order to cope with this changes, does not completely remove the ideas.

Regarding the transition "Cancelada -> Aprovação em Steering", the experts stated this situation is very rare, however there were cases when an idea was approved and then the organization realized that would not bring any value.

With an average transition time of approximately 22 days, the main bottleneck observed at the end of the process ("Promovida a Iniciativa -> Implementada") was justified by estimating the time, resource and complexity of the idea's implementation. Moreover, the reason identified during the performance's analysis was complementary.

The issues found with the extraction of the creators of the ideas were also communicated. The organization explained that the creator remains private for the employees. The only individuals who have access to this information are the person who has created it and the system's administrator. Thus, the employee's would vote and comment on each other's ideas without personal influence. Despite the privacy associated to this field, it can be extracted from a customized one.

Inspired by the results, the experts and management were interest in exploring the relation between the ideas and the employees and in uncovering the most participative units of the organization.

Second iteration

Throughout the following two weeks, the information stored in the customized field was extracted and role analysis was performed.

- **Data extraction**

The approach that was used for the field extraction had to be changed. In this case, the work was performed on a nested JSON object. Thus, the creator's identification had to be extracted and, afterwards, matched to a corresponding one in the overall list. The XML file regarding the project was updated and a CSV document was produced.

- **Role analysis**

Appendix C represents the social network discovered with 10% of activities and paths. The high number of performers associated (116 performers) and the inexistence of interactions among them did not allow to perform social network analysis. Regarding to the areas that were more participatory/ active in the ideation process, for reasons of confidentiality was not possible.

- **Validating the results**

Despite not being able to perform the requested analysis, the information gathered about the creators was communicated to the experts.

4.3.3. Transferring phase

The process mining analysis from each iteration was documented and transferred to the organization. In a summarized manner, the main findings after two iterations were:

- The discovery of a business process regarding to a workflow of an idea with a well-defined beginning and end;
- The discovery of additional behavior corresponding to the supplementary filtering performed in order to optimize the process (e.g. remove the duplicate ideas), consider the ethical issues or respond to the changing business conditions (ideas re-assessment);
- The discovery of a dominant behavioral pattern represented by the pattern "Registada -> Confirmada -> Avaliada em Comité -> Avaliada em Steering -> Aprovada em Steering -> Promovida a Iniciativa";
- The discovery of the rare cases represented by a deviant transition "Cancelada -> Aprovada em Steering";
- The uncovered average time for handling an idea being approximately 82 days;
- The major bottleneck observed corresponding to the final phase of the process and representing the idea's preparative for (e.g. time, resource and complexity estimative) and its actual implementation;
- The influence of the idea's preparative for its implementation and the actual implementation on the overall average time being approximately 28 days;
- The possibility for idea's review being more significant at the beginning of the ideation process represented by minor bottlenecks.

At technological level, the lesson learned from the previous application of the methodology was considered. Disco tool was used to perform the requested analysis. Moreover, as the developed project was evolving, the possibility for performing workflow analysis arose. The results regarding the process are presented in Appendix C.

4.4. Request management

The third application of the proposed solution was performed on the same organization. Here, JIRA was used to assist the request management system, i.e. the request management workflow was supported. In terms of time-period, the first iteration took two weeks and the second one took one month.

4.4.1. Preparation phase

Despite the performed application regarding the idea management process, the expert was not familiar with the process mining concept. Thus, throughout the first meeting, an overview of process mining area and its techniques was performed in order to provide some insights and do not create false hopes or mislead the client. Moreover, a brief explanation to what was done in the idea management area was made.

Since JIRA was used to support the request management workflow, the decision made was to focus on the business process regarding the request. The organization was interested in the possibility of uncovering the stored process and compare it with the predefined one. Additionally, an overview of the process chosen for the analysis was provided.

Request management process

A request expresses the need to acquire a specific service before contacting the supplier. In order to perform it, several authorizations must be granted.

At the beginning of the process, the employee creates a request and provides all the needed details for its analysis (“Registada”). Once the information is obtained, the request is sent to the “Unidade de Âmbito” (“A Aguardar Validação da Uni. Âmbito”). This unity will settle the transaction with the supplier.

The “Unidade de Âmbito” redirects to the unity of procurement (“A Aguardar Validação do Procurement”). Every request made should pass through the procurement. In the case of an order being valid, it is sent to the management (“A Aguardar Validação da Chefia Directa (1)”, (“A Aguardar Validação da Chefia Directa (2)”).

The organization has two levels of management. The first one represents the level of the creator of the request and the second the level above.

To approve a request, there are a set of financial rules that should be verified. According to the request’s value, these can be automatically approved (“Requisição Aprovada”) or have to be accepted by DF (“A Aguardar pela Validação do DF”) or DG “A Aguardar pela Validação do DG). Once again, the value of the request will define the entity that will approve it.

At any point of the financial approval, the request can be rejected (“Rejeitada”). Thus, it returns to the beginning of the process (“Registada”).

The request may have several tranches associated to it. Thus, in order to be fulfilled (“Totalmente Satisfeita”), these have to be fulfilled first. Meanwhile, the request stays at “Encomenda Colocada”.

When the payments are finished and the request is fulfilled, the process ends.

4.4.2. Development phase

First iteration

The developed project allowed to perform the data extraction and obtain the workflow of the required entity. Thus, it was used in order to get some insights about the request management process.

Control flow and Performance analysis

The model was extracted and four additional activities were found (Appendix D). When comparing with the information gathered during the first meeting, there was no evidence of such. Thus, the experts were contacted in order to clarify the situation.

Transferring the results

Considering the expert’s availability, an in-person meeting took place. He/she was surprised to find those activities and stated that they belong to the tranche component that is intrinsic in a request. Therefore, the decision was made to filter the data regarding to “Criada”, “Satisfeita”, “Facturada” and “Parcela Cancelada” activities.

Second iteration

- **Data extraction**

After a careful analysis, it was realized that the tranche component was a separate entity. Thus, in order to filter the data and focus only on the requests, the tranche component was excluded from the XML file. For this purpose, the type field was used.

- **Log inspection**

The data gathered was from four-month period. The resulted log contained 489 requests, the total number of 13618 events and 13 different activities. The dominant activity was “A Aguardar Validação Uni. Âmbito” (21,57%), followed by “A Aguardar Validação de Procurement” (20,83%), “Requisição Aprovada” (12,51%), “Totalmente Satisfeita” (11,21%), “A Aguardar Validação da Chefia Directa (1)” (10,12%), “Registada” (7,98%), “A Aguardar Validação da Chefia Directa (2)” (6,52%), “Encomenda Colocada” (3,09%), “A Aguardar pela Validação do DF” (2,97%), “Despesa Encerrada” (2,67%), “A Aguardar Validação do DG” (0,36%), “Rejeitada” (0,11%) and, finally, “Cancelada” (0,05%).

Regarding the number of different sequences of activities, 104 variants were found. The constructed log presented 41 performers.

- **Control flow analysis**

Appendix D represents the discovered business process for request management with its total number of activities and paths. Since the organization had a predefined process, these two were compared.

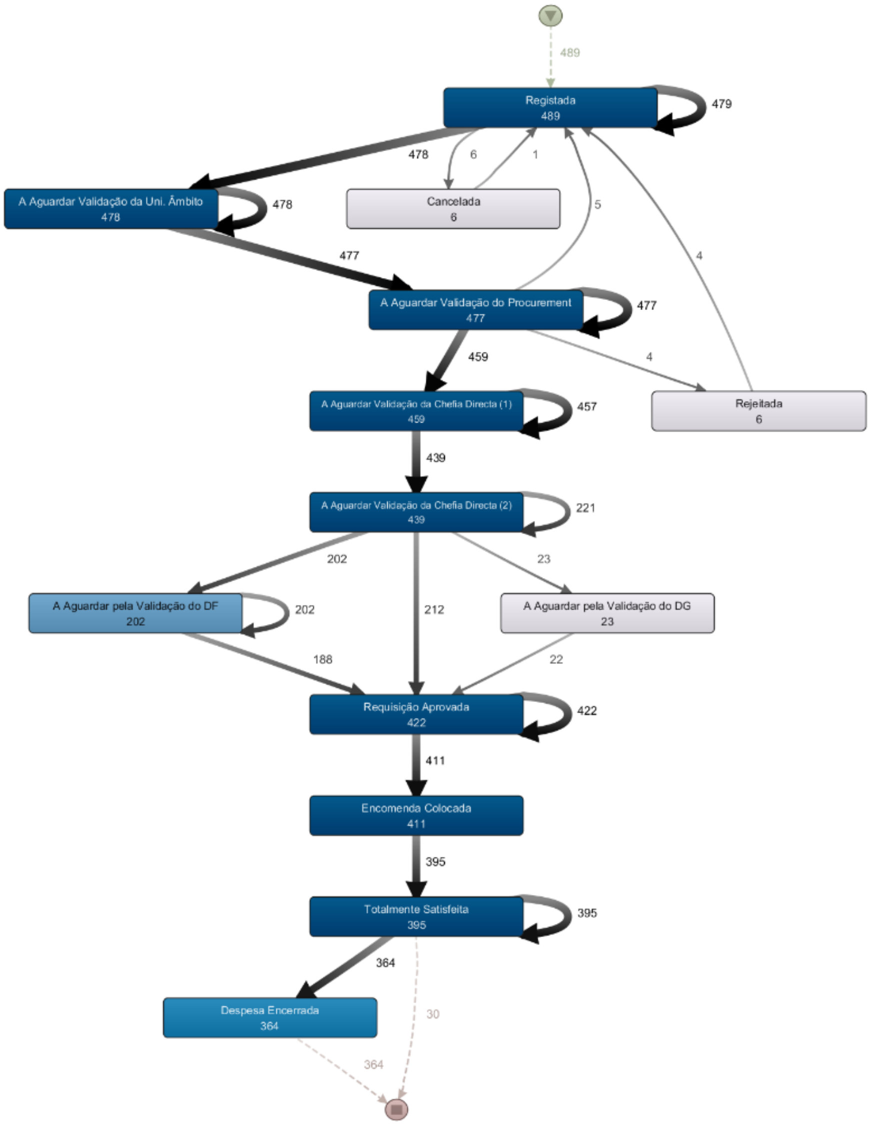

Figure 20 Request Management Workflow – frequency (100% of activities and 0% of paths)

The discovered model was very similar to the predefined one. Its main difference relied on several transitions to “Registada” activity, e.g. “A Aguardar Validação do Procurement -> Registada”, “A Aguardar Validação da Chefia Directa (1) -> Registada”, etc. The presented transitions are a proof of deviations in the process’ behavior. Moreover, the presence of different transitions supports the idea of its variability.

The behavioral pattern “Registada -> A Aguardar Validação Uni. Ambito -> A Aguardar Validação do Procurement -> A Aguardar Validação da Chefia Directa (1) -> A Aguardar Validação da Chefia Directa

(2) -> Requisição Aprovada -> Encomenda Colocada -> Totalmente Satisfeita -> Despesa Encerrada". is stated as a dominant one for a request. It is represented in Figure 20.

In terms of infrequent behavior, the patterns that involve "Cancelada", "Rejeitada" and "A Aguardar Pela Validação do DG" activities were considered as so. As was already mentioned, the activity "A Aguardar Pela Validação do DG" relates to the requests which financial value is above a predefined one. Thus, its low number was already expectable. ´

The constraint specified regarding the procurement unity was verified, i.e. each request that entered the workflow and was not rejected, passed through it.

- **Performance analysis**

Appendix D presents the request management process in terms of its performance. For a better understanding, the filtered one is presented in Figure 21.

According to mean metric, a request is processed 13,6 days from the date of its registration.

The transitions "Registada -> Cancelada", "A Aguardar Validação da Chefia Directa (1) -> A Aguardar Validação da Chefia Directa (2)", "A Aguardar pela Validação do DF -> Requisição Aprovada", "A Aguardar pela Validação do DG -> Requisição Aprovada" represents the main bottlenecks of the process.

The first case represents the time required to cancel a registered request (6,4 days). When compared with the results from the control flow analysis, it was noted that it belongs to the infrequent behavior.

The second bottleneck represents the need for approval from the higher level of management. This bottleneck is the most worrying one, since it means that, in a real-life, the organization takes approximately 10 weeks to perform the transition (71 days).

Finally, the last two bottlenecks correspond to the approval by specialized entities. Whereas the four-day period associated to the transition "Aguardar pela Validação do DG -> Requisição Aprovada" could be justified by the number of request associated to it, the transition "A Aguardar pela Validação do DF -> Requisição Aprovada" raised some questions.

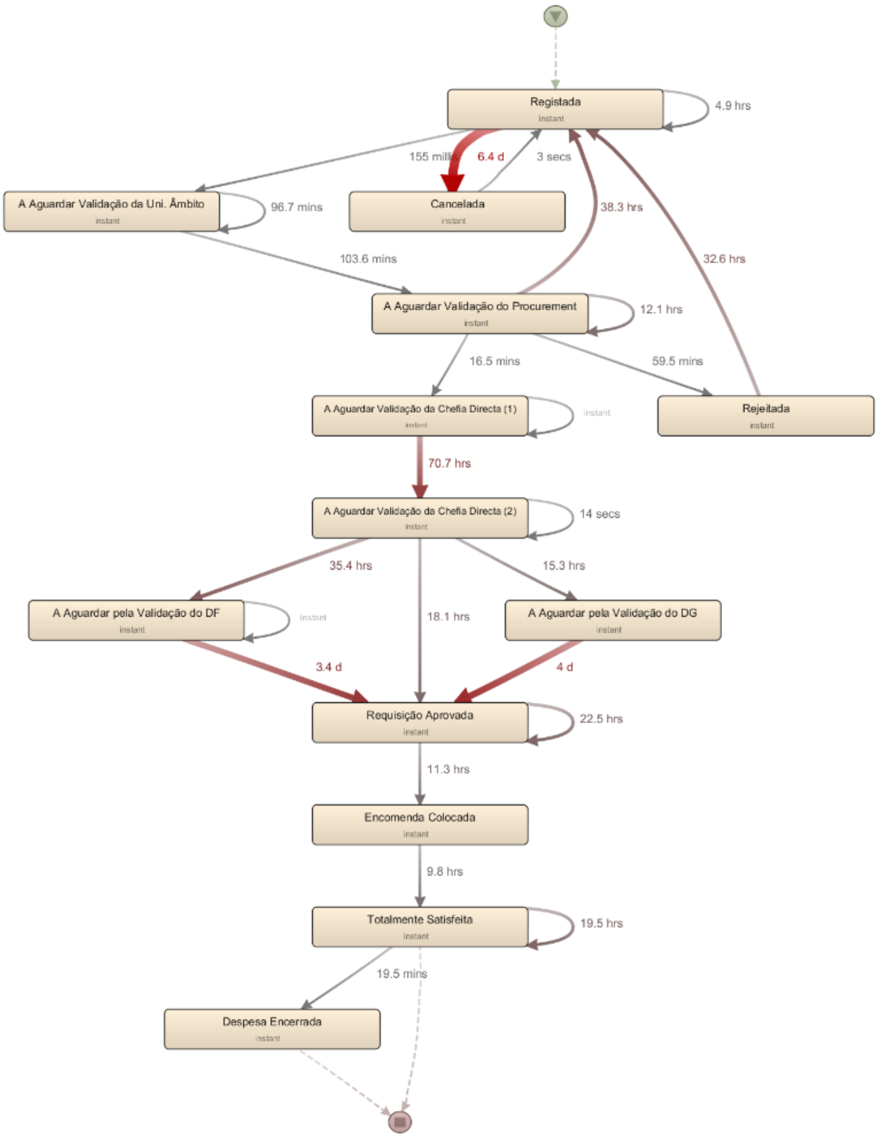

Figure 21 Request Management Workflow – performance (100% of activities and 0% of paths)

- **Validating the results**

To validate the obtained results from the first iteration, an in-person meeting took place. From business individuals, the expert was present. The process model discovered during the control flow analysis was recognized as the expected one. Moreover, the main behavioral pattern was recognized to be “Registada -> A Aguardar Validação Uni. Ambito -> A Aguardar Validação do Procurement -> A Aguardar Validação da Chefia Directa (1) -> A Aguardar Validação da Chefia Directa (2) -> Requisição Aprovada -> Encomenda Colocada -> Totalmente Satisfeita -> Despesa Encerrada”.

The deviating behavior discovered through the request management process was not new for the organization. The expert stated that, at any point of time, some additional information may be required. Thus, the request’s status can be updated to “Registada”.

With an average transition time of approximately 10 weeks, the main bottleneck observed ("A Aguardar Validação da Chefia Directa (1) -> A Aguardar Validação da Chefia Directa (2") was justified by the number of employees involved in the process.

The issues found with the transition time of the "A Aguardar pela Validação do DF -> Requisição Aprovada" were also communicated. The expert became curious and pointed out the transition time associated to the final phase of the process was too fast. Thus, a doubt arose about the influence of the filtered tranches on the overall time.

4.4.3. Transferring phase

The process mining analysis from each iteration was documented and transferred to the organization. In a summarized manner, the main findings after two iterations were:

- The discovery of a business process regarding to a workflow of a request with a well-defined beginning and end;
- The discovery of a dominant behavioral pattern represented by the pattern be "Registada -> A Aguardar Validação Uni. Ambito -> A Aguardar Validação do Procurement -> A Aguardar Validação da Chefia Directa (1) -> A Aguardar Validação da Chefia Directa (2) -> Requisição Aprovada -> Encomenda Colocada -> Totalmente Satisfeita -> Despesa Encerrada";
- The uncovered average time for handling a request being approximately 14 days;
- The major bottleneck observed being approximately 10 weeks and corresponding to the transition of the management level. It was justified by the amount of individuals involved in the process;
- The influence of the tranche component on the request's overall performance time.

At technological level, the lesson learned from the previous application was considered. The combination of the developed project's functionalities and Disco tool was used to perform the requested analysis.

Additionally, it should be noted that the third application was made during the vacation period, thus its durability was longer than expected.

5. Evaluation

Throughout this chapter, the evaluation of the outcomes and the validation of the proposal will be performed. The importance of evaluation process in the DSRM's context had been recognized among the research community. However, a predefined criteria and methods that could turn the evaluation process more systematic and repeatable was absent. Thus, many researchers try to address this matter in different ways and a wide range of solutions can be found.

Table 2 Criteria Selected for the Artifact Assessment

Dimension	Evaluation criteria	Description
Goal	Efficacy	Degree in which artifact produces its desired effect, i.e. achieves its goal
Goal	Validity	Degree in which artifact works according to expected, i.e. correctly achieves its goal [41]
Activity	Accuracy	Agreement between the expected and obtained results

In order to evaluate the presented artifact, the holistic view of criteria and standard methods proposed by Prat et. al. was considered [41].

Table 3 Method of the Artifact's Analysis

Evaluation criteria	Form of evaluation	Secondary participants	Level of evaluation	Relativeness of evaluation
Goal/Efficacy	Qualitative	None	Instantiation/Real examples	Absolute
Goal/Validity	Analysis and logical reasoning	None	Instantiation/Real examples	Absolute
Activity/Accuracy	Analysis and logical reasoning	None	Instantiation/Real examples	Absolute

The authors considered an artifact as a system and the criteria defined were based on five systems dimensions, such as *goal*, *environment*, *structure*, *activity* and *evolution*. Due to the nature of the work, the criteria found as the more suitable ones belong to goal, environment and activity dimensions. These are present in Table 2.

Regarding to evaluation method, the authors defined it according to a set of characteristics, such as *form of evaluation*, *secondary participants*, *level of abstraction and relativeness of evaluation*. To obtain a method suitable for the artifact, different values of these characteristics will be combined. These are present in Table 3.

5.1. Issue management

Efficacy

From the feedback obtained during the analysis process and the final meetings, it is believed that the criterion was achieved. The experts and management from issue management application stated that they were interested in continuing with the analysis in the future.

Validity

The proposed artifact was applied in the isolated environment, i.e. there have been no changes or modifications made during the process. The analyzed data was gathered from one year and half time-period. The performed manipulations that were used as input in each iteration were based on the experts' feedback from process perspective.

Accuracy

Throughout the iterations, the required process was discovered and analyzed according to process, organizational and case perspectives defined by process mining. The analysis allowed to provide additional knowledge expressed by the number of cases between the activities, the time required to perform the transition between each activity, the performers associated to the issue management process and its variants. However, in the case of organizational perspective, its potential was not fully explored with the academic tool. A possible explanation can rely on the small number of performers (assignees) and, consequently, the impossibility for more interactions.

Concerning the questions about feedback and involvement the customer into the process, its impact was heavily noticed during the whole process. With the frequent releases and continuous collaboration, it was possible to solve the issues found earlier.

5.2. Idea management

Efficacy

The experts said that the analytic information could be used as input in process optimization or definition of specific metrics. In terms of collaboration, they were interested in scheduling additional in-person meeting to perform a deeper analysis of the provided information.

Validity

Similarly to the previous application, the proposed artifact was operating in the isolated environment. The analyzed data was gathered from two-year period and all the performed manipulations were based on the feedback obtained from the experts.

Accuracy

In this application, the process was correctly discovered through the iterations and the compliance with a predefined workflow was verified. Likewise, the issue management, the process, organizational and case perspectives were used for its analysis. The results allowed to provide additional information from idea management process perspective. Moreover, the deviant behavior expressed by three cases was discovered and communicated.

Concerning the question of the customer involvement, with a frequent release, i.e. by having the results early and seeing the process mining in practice, it was possible to stimulate ideas for the process mining analysis.

5.3. Request management

Efficacy

The proposed artifact was applied for the third time in the request management context, i.e. it was its third instantiation. Due to the time limits and the fact that it comprised the vacation period, there was no possibility to fully evaluate this criterion.

Validity

The proposed artifact was operating in the isolated environment, i.e. there have been no changes or modifications made during the process. The extracted data was from four-month period. Additionally, the performed manipulations that were used as input in each iteration were based on the experts' feedback from process perspective.

Accuracy

Due to the impossibility to conclude the third application of the artifact, the performance analysis was left open. However, the process discovery and its compliance with the predefined one was possible. The variants of the process were discovered along with the performers associated to it.

Additionally, the vacation period heavily influenced the iteration duration and, consequently, the transferring process. The communication with the experts was too scattered.

6. Conclusion

Throughout this chapter, an overview of the performed work is provided along with the list of lessons learnt. The suggestions for the future work are presented and the issues with the communication phase are addressed.

Among a wide range of tools, ITS are the ones that are well-known for its organizational ability. These systems were proved to be a main repository for recording, tracking, solving and archiving of issues. Due to its nature, it also serves as a link between different stakeholders of the process that it supports.

Throughout the time, ITS stared to support the project tracking functionality. Thus, in the software projects, they are referred as bug tracking systems and in other context as project management tools.

From the research perspective, the focus was made on its operational level and the information that was intrinsic in its processes was completely forgotten. Thus, the main driver for this work was to address the problem of lack of awareness of the stored information from process perspective.

With this consideration in mind, a development of a method was proposed that would uncover the processes stored in the ITS environment with process mining techniques. It was enriched with the notions of agile approach and incorporated the iterative nature of Scrum. To conduct the research in a structured way, DSRM was followed.

The proposed methodology was applied in three projects available in JIRA, one of the most used ITS. The first one was incorporated in the software development environment, the second on the idea management and the third on the request management.

The data regarding each project was gathered and analyzed with academic and commercial tools for process mining analysis, such as ProM and Disco. The analysis allowed to discover the stored processes and, in the idea and request managements, to verify whether those processes were similar to the predefined ones.

to verify the compliance with the predefined ones. Moreover, additional knowledge of the processes was provided in terms of frequencies among activities, number of variants, performers and its centrality, performances and bottlenecks. Moreover, JIRA's particularity for handling the projects was discovered.

By comparing the applications, it can be advocated that each one of them allowed to explore a different component of the methodology.

In the first case, the most interesting component was represented by data extraction phase. The analyzed projects contained a low number of events and, thus, the variability of the information and the ability to cover a high number of paths was questioned.

With the application of the methodology in the idea management environment, it was possible to explore the requirements variability and prove that involving the customer in the process is a great asset.

Finally, the third case is an example of the impact of the external factors in a business environment.

After each application, the assessment of the obtained results took place. The evaluation process was based on the holistic view of criteria and standard methods. It was composed by logical reasoning and unstructured interviews (feedback) throughout the applications.

6.1. Lessons learned

Throughout the three applications of the proposed methodology, the most valuable lessons learned were:

- The importance of the experts' involvement from the beginning of the process mining analysis;
- The organizational interest in further use of process mining;
- The impact that the recurrent iterations have on the result validation and, consequently, on the formulation of requirements;
- The bureaucracy associated to the business environment should be considered;
- Regardless of the degree of interest and usefulness, the terms of confidentiality establish the boundaries for process mining analysis.

6.2. Future work

Considering the expressed interest by the experts in the **issue management** environment and the doubts that have arose about the variability of information, the application of process mining in a bigger project will be performed. This opportunity will be seized to explore the organizational component intrinsic in the process.

In the case of the **request management**, the process mining analysis was not concluded. Thus, the influence of the intrinsic tranches in the overall process will be investigated.

6.3. Communication

The business processes are well-known for their ability to be very bureaucratic and involving a high number of individuals. Thus, throughout the presented research it was not possible to submit any paper to get appraisal from the scientific community.

Bibliography

[1] R. S. Pressman, *Software engeneering: a practitioner's approach*, 7th ed. Palgrave Macmillan, 2005.

[2] C. R. Reis and R. P. de Matos Fortes, "An overview of the software engineering process and tools in the mozilla project," in *roceedings of the Open Source Software Development Workshop*, 2002, pp. 155–175.

[3] D. Bertram, A. Voida, S. Greenberg, and R. Walker, "Communication, collaboration and bugs: the social nature of issue tracking in small, collocated teams," in *ACM conference on Computer supported cooperative work (CSCW '10)*, 2010, pp. 291–300.

[4] N. Bettenburg and S. Just, "What makes a good bug report?," in *16th ACM SIGSOFT International Symposium on Foundations of software engineering (SIGSOFT '08/FSE-16)*, 2008, pp. 308–318.

[5] M. Ortu, G. Destefanis, M. Kassab, and M. Marchesi, "Measuring and understanding the effectiveness of JIRA developers communities," in *Sixth International Workshop on Emerging Trends in Software Metrics (WETSoM '15)*, 2015, pp. 3–10.

[6] S. Breu, R. Premraj, J. Sillito, and T. Zimmermann, "Information needs in bug reports: improving cooperation between developers and users," in *ACM conference on Computer supported cooperative work (CSCW' 10)*, 2010, pp. 301–310.

[7] S. Just and T. Zimmermann, "Towards the next generation of bu tracking systems," in *Visual Languages and Human-Centric Computing*, 2008.

[8] T. Zimmermann, R. Premraj, J. Sillito, and S. Breu, "Improving bug tracking systems," in *ICSE Companion*, 2009, pp. 247–250.

[9] O. Baysal, R. Holmes, and M. W. Godfrey, "Situation awareness: personalizing issue tracking systems," in *International Conference on Software Engineering (ICSE 2013)*, 2013, pp. 1185–1188.

[10] W. M. P. Van Der Aalst, H. A. Reijers, A. J. M. M. Weijters, and B. F. Van Dongen, "Business process mining : An industrial application," *Inf. Syst.*, vol. 32, pp. 713–732, 2007.

[11] R. S. Mans, M. H. Schonenberg, M. Song, and W. M. P. Van Der Aalst, "Application of process mining in healthcare - a case study in a dutch hospital," in *Communications in Computer and Information Science*, vol. 25, A. Fred, J. Filipe, and H. Gamboa, Eds. Springer Berlin Heidelberg, 2009, pp. 425–438.

[12] R. Mans, H. Schonenberg, and G. Leonardi, "Process mining techniques: an application to stroke case," *Stud Heal. Technol Inf.*, vol. 138, pp. 573–578, 2008.

[13] M. Bozkaya, J. Gabriels, and J. M. Van Der Werf, "Process diagnostics : a method based on process mining," in *International Conference on Information, Process, and Knowledge Management (eKNOW '09)*, 2009, no. 1, pp. 22–27.

[14] W. M. . van der Aalst, K. M. van Hee, J. M. van Werf, and M. & Vendronk, "Auditing 2.0: using process mining to support tomorrow's auditor," in *Computer (43)*, 2010, pp. 90–93.

[15] R. A. Hevner, S. Ram, S. T. March, and J. Park, "Design science in information systems research," in *MIS quarterly*, vol. 28, no. 1, 2004, pp. 75–105.

[16] K. E. N. Peffers, T. Tuunanen, and M. A. Rothenberger, "A design science research methodology for information systems research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2008.

[17] W. M. P. van der Aalst, *Process mining: discovery, conformance and enhacement of business processes*, 1st ed. Springer-Verlag Berlin Heidelberg, 2011.

[18] W. M. P. Van Der Aalst and M. Song, “Mining social networks: ucovering interaction patterns in business processes,” *Lect. Notes Comput. Sci.*, vol. 3080, pp. 244–260, 2004.

[19] B. F. Van Dongen, A. K. A. De Medeiros, H. M. W. Verbeek, and A. J. M. M. Weijters, “The ProM framework: a new era in process mining tool support,” *Lect. Notes Comput. Sci.*, vol. 3536, pp. 444–454, 2005.

[20] W. Poncin, A. Serebrenik, and M. Van Den Brand, “Process mining software repositories,” in *Software Maintenance and Reengineering (CSMR)*, 2011.

[21] V. Rubin, C. W. Gunther, W. M. . van der Aalst, E. Kindler, B. F. van Dongen, and W. Schafer, “Process mining framework for software processes,” in *Software Process Dynamics and Agility*, Springer Berlin Heidelberg, 2007, pp. 169–181.

[22] D. R. Ferreira and M. Mira da Silva, “Using process mining for ITIL assessment: a case study with incident management,” in *Proceedings of the 13th Annual UKAIS Conference*, 2008.

[23] V. Rubin, I. Lomazova, and W. M. P. van der Aalst, “Agile development with software process mining,” in *International Conference on Software and System Process (ICSSP 2014)*, 2014, pp. 70–74.

[24] A. J. M. M. Weijters, W. M. P. Van Der Aalst, and A. K. A. De Medeiros, “Process mining with the heuristics miner algorithm,” in *BETA Working Paper Series WP 166*, Edindhoven University of Technology, 2006.

[25] M. Song and W. M. P. Van Der Aalst, “Towards a comprehensive support for organizational mining,” in *Decision Support Systems*, vol. 46, no. 1, 2008, pp. 300–317.

[26] H. M. W. Verbeek, J. C. A. . Buijs, B. F. van Dongen, and W. M. P. van der Aalst, “ProM 6: the process mining toolkit,” in *Proc. of BPM Demonstration Track 615*, 2010, pp. 34–39.

[27] S. J. J. Leemans, D. Fahland, and W. M. P. Van Der Aalst, “Exploring processes and deviations,” in *Business Process Management Workshops*, vol. 202, F. Fournier and J. Mendling, Eds. 2015, pp. 304–316.

[28] A. Rozinat, “Disco user’s guide.” .

[29] A. Rozinat, “Disco : discover your processes,” in *Business Process Management (BPM 2012)*, 2012, pp. 40–44.

[30] J. Highsmith and A. Cockburn, “Agile Software Development: The Business of Innovation,” *Computer (Long. Beach. Calif).*, vol. 34, no. 9, pp. 120–127, 2001.

[31] B. Boehm, “Get ready for agile methods, with care,” *Computer (Long. Beach. Calif).*, vol. 35, no. 1, pp. 64–69, 2002.

[32] A. Cockburn, *Agile software development the cooperative game*, 2nd ed. 2001.

[33] M. Fowler, “The New Methodology,” *Wuhan Univ. J. Nat. Sci.*, vol. 6, no. 1, pp. 12–24, 2001.

[34] M. Fowler and J. Highsmith, “The Agile Manifesto,” *Softw. Dev.*, pp. 28–35, 2001.

[35] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta, “Agile software development methods: review and analysis,” 2002.

[36] L. Rising, N. S. Janoff, and A. G. C. Systems, “The Scrum software development process for small teams,” in *IEEE software*, 2000, no. 26.

[37] K. S. Rubin, *Essential Scrum: A practical guide to the most popular agile process*. Addison-Wesley, 2012.

[38] J. Sutherland and K. Schwaber, “The Scrum guide™,” in *The definitive guide to Scrum: the rules of the game*, 2011.

[39] K. Schwaber, *Agile Project Management with Scrum*. Microsoft Press, 2004.

[40] C. R. Prause, M. Scholten, A. Zimmermann, M. Eisenhauer, S. Birlinghoven, and S. Augustin, “Managing the iterative reqirements process in a multi-national project using an issue tracker,” in *2008 IEEE International Conference on Global Software Engineering*, 2008, pp. 151–159.

[41] N. Prat, I. Comyn-Wattiau, and J. Akoka, “Artifact evaluation in information systems design-science research - a holistic view,” in *PACIS.*, 2014, p. 23.

Appendixes

Appendix A

Table 3 An Event Log

Case ID	Activity ID	Performer	Timestamp
Case 1	Activity A	John	9-3-2004: 15.01
Case 2	Activity A	John	9-3-2004: 15.12
Case 3	Activity A	Sue	9-3-2004: 16.03
Case 3	Activity B	Carol	9-3-2004: 16.07
Case 1	Activity B	Mike	9-3-2004: 18.25
Case 1	Activity C	John	10-3-2004: 9.23
Case 2	Activity C	Mike	10-3-2004: 10.34
Case 4	Activity A	Sue	10-3-2004: 10.35
Case 2	Activity B	John	10-3-2004: 12.34
Case 2	Activity D	Pete	10-3-2004: 12.50
Case 5	Activity A	Sue	10-3-2004: 13.05
Case 4	Activity C	Carol	11-3-2004: 10.12
Case 1	Activity D	Pete	11-3-2004: 10.14
Case 3	Activity C	Sue	11-3-2004: 10.44
Case 3	Activity D	Pete	11-3-2004: 11.03
Case 4	Activity B	Sue	14-3-2004: 11.18
Case 4	Activity D	Pete	14-3-2004: 11.48
Case 5	Activity E	Clare	17-3-2004: 12.22
Case 5	Activity D	Clare	18-3-2004: 14.34

Appendix B

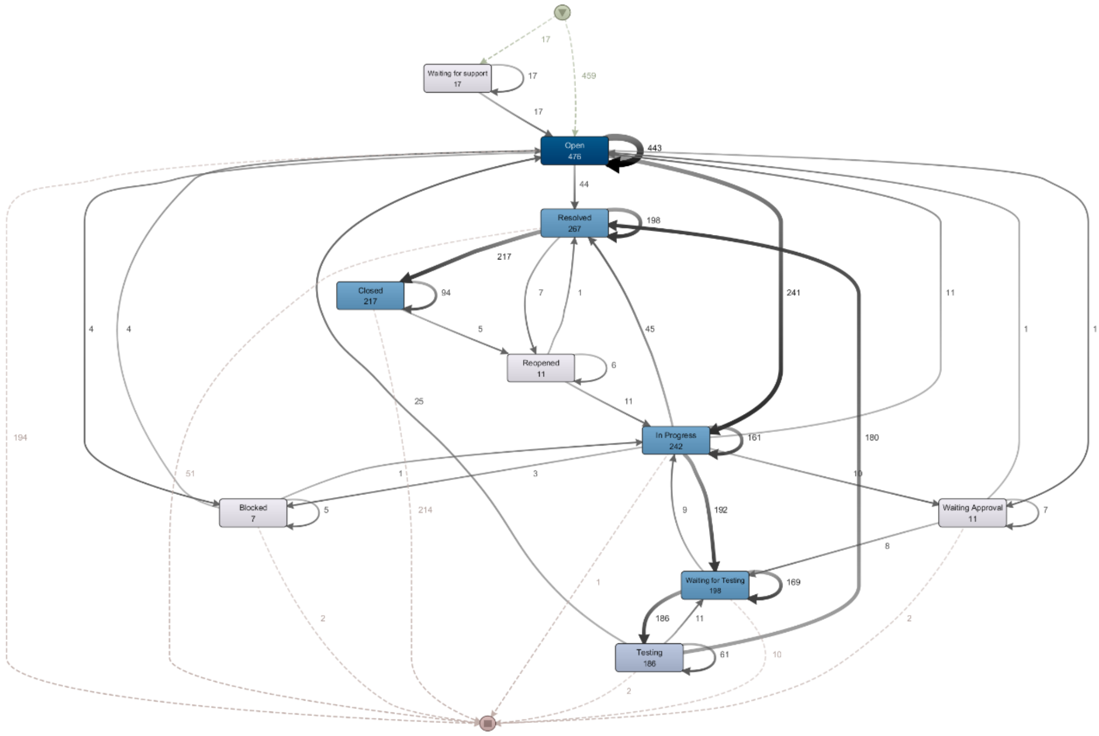

Figure 22 Issue Management Workflow – frequency (Xray project:100% of activities and 100% of paths)

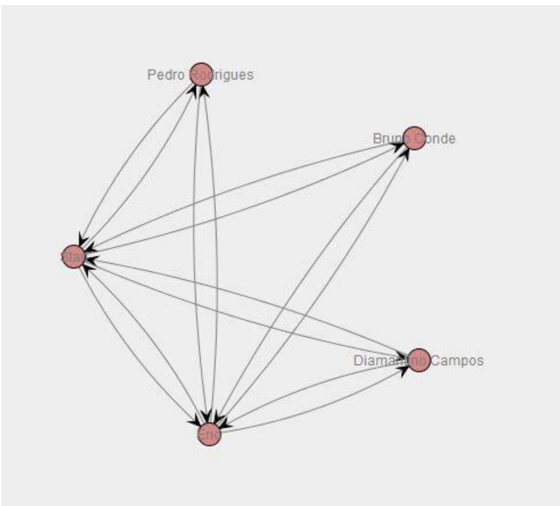

Figure 23 Social Component - handover of work metric (Xray project)

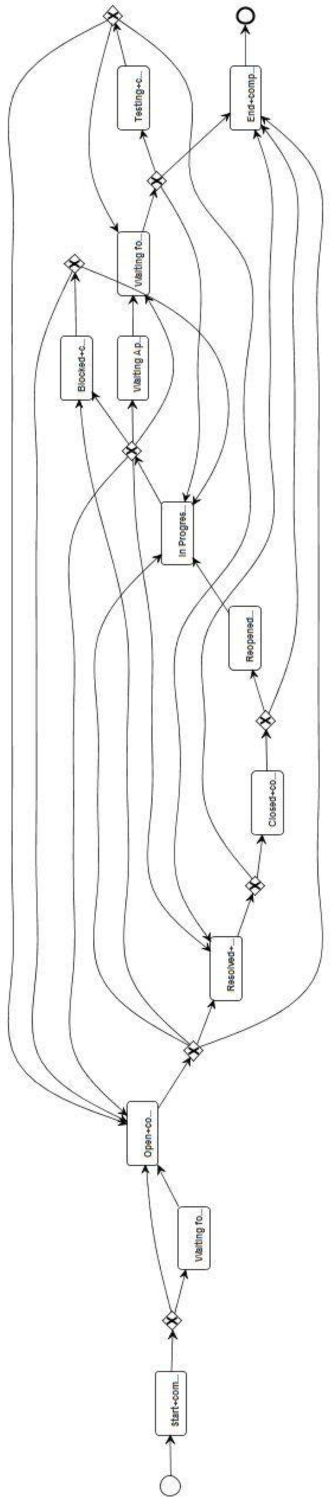

Figure 24 Issue Management Workflow (Xray project)

Appendix C

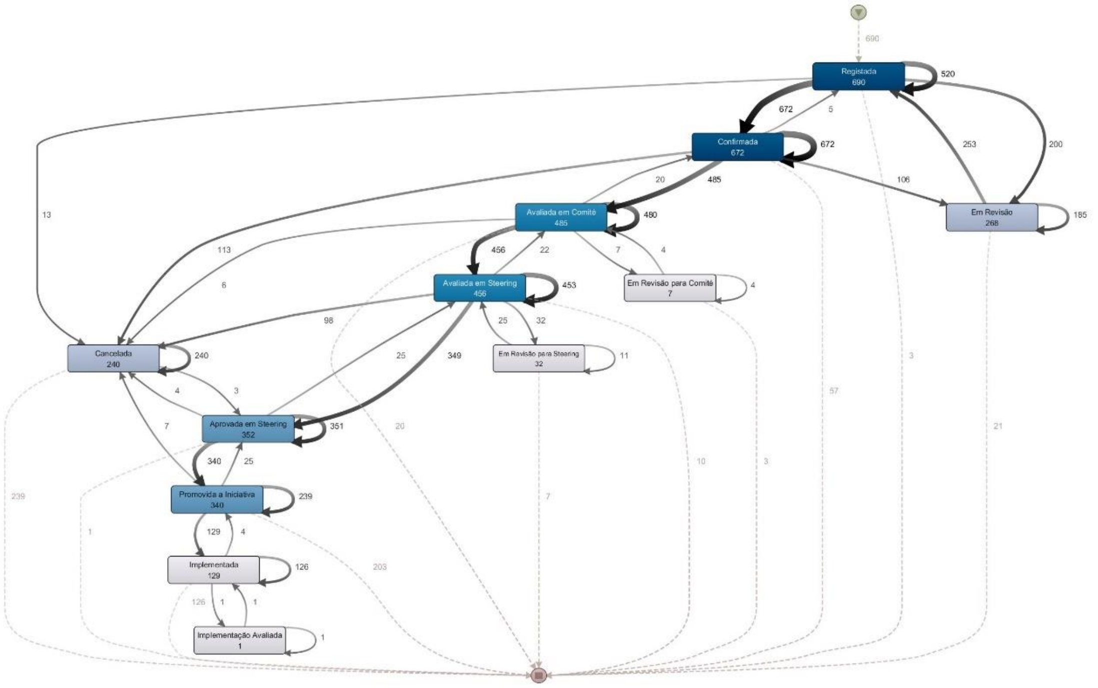

Figure 25 Idea Management Workflow – frequency (100% of activities and 100% of paths)

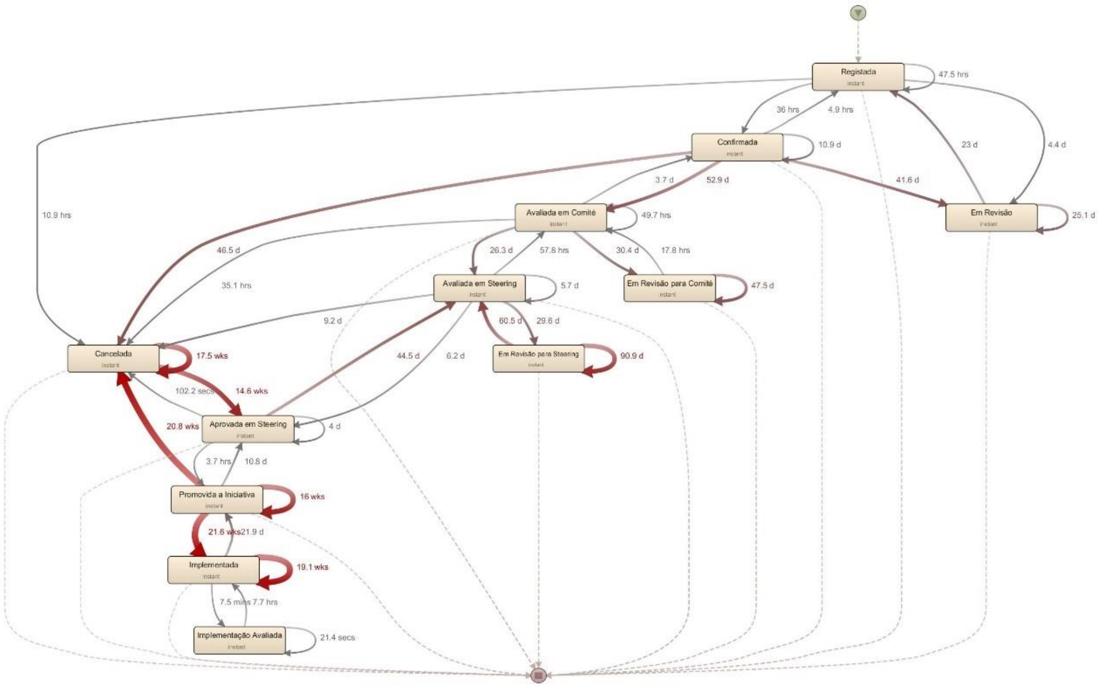

Figure 26 Idea Management Workflow – performance (100% of activities and 100% of paths)

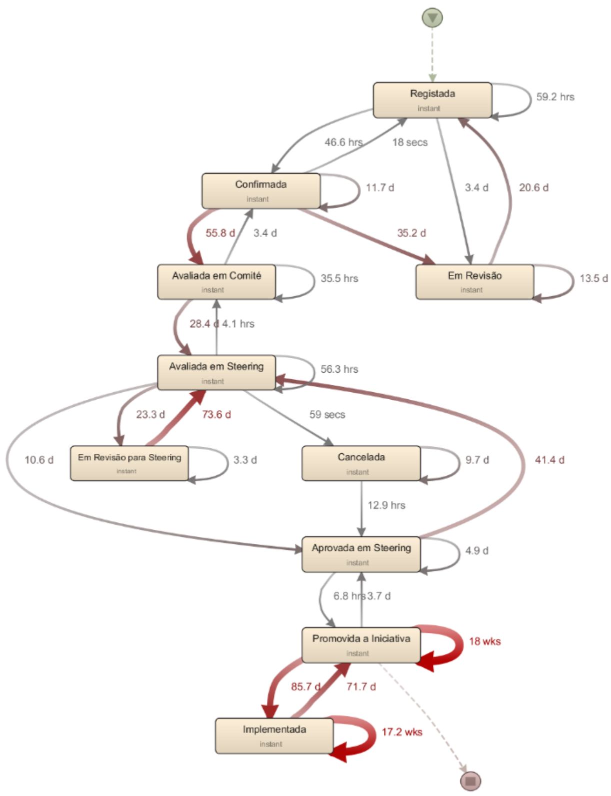

Figure 27 Idea Management Workflow – performance (filtered; 100% of activities and 100% of paths)

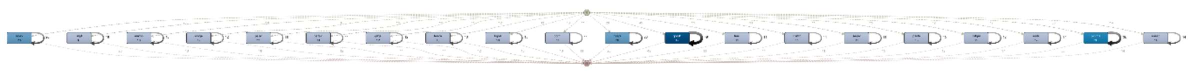

Figure 28 Social Network (10% of activities and 10% of paths)

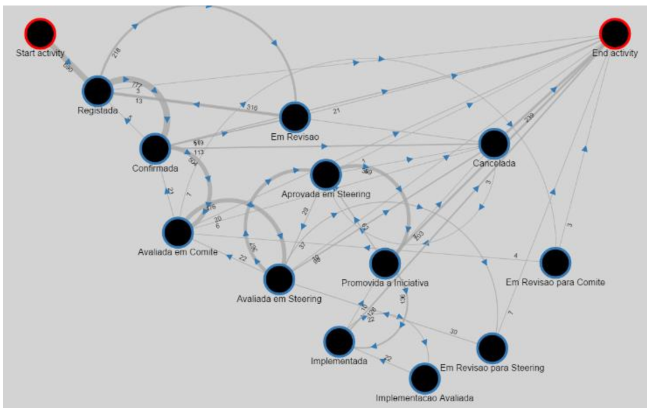

Figure 29 Idea Management Workflow – frequency

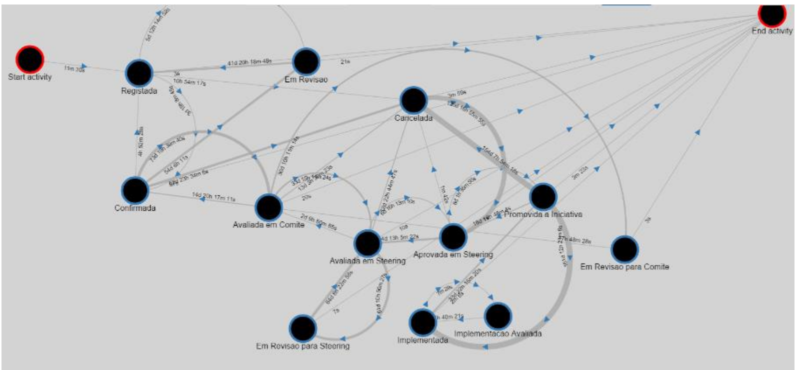

Figure 30 Idea Management Process – performance

Appendix D

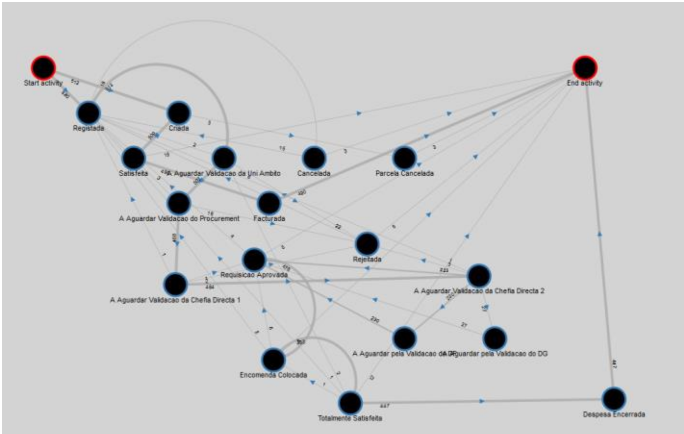

Figure 31 Request Management Workflow

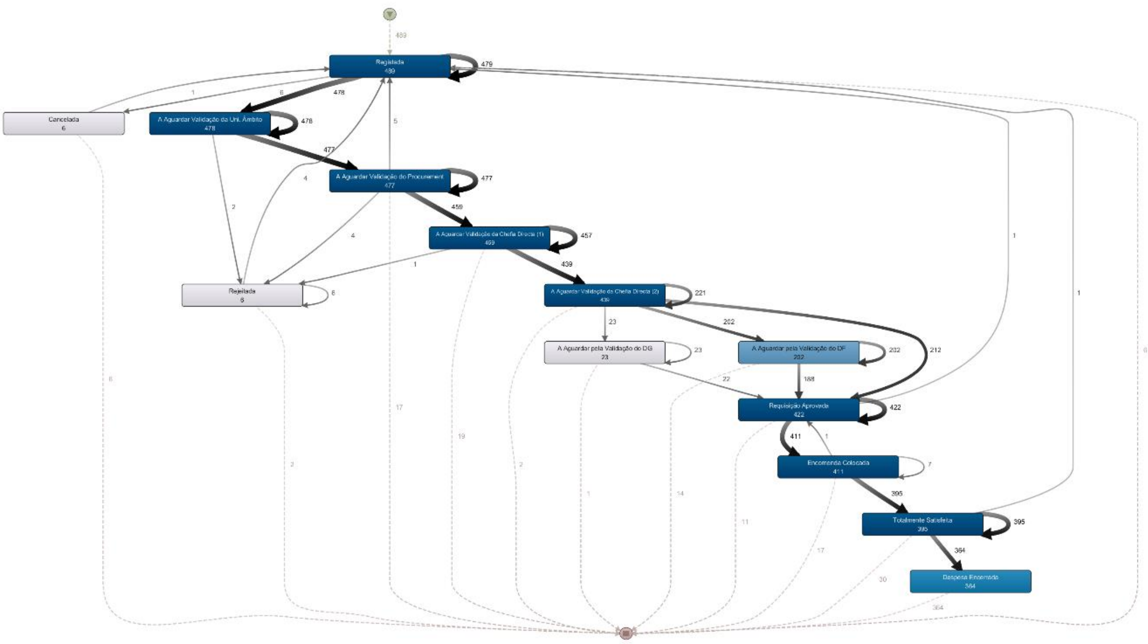

Figure 32 Request Management Workflow – frequency (100% of activities and 100% of paths)

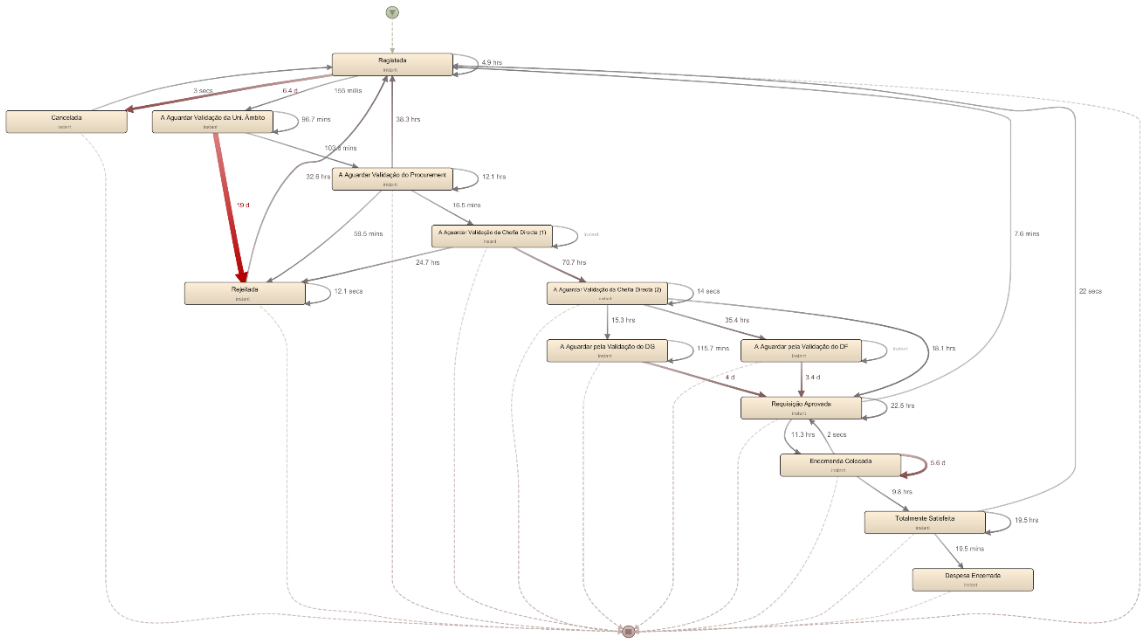

Figure 33 Request Management Workflow - performance (100% of activities and 100% of paths)