UNIVERSITY OF WESTERN ATTICA
DEPARTMENT OF ENGINEERING INFORMATION AND COMPUTER

bachelor's thesis

Analysis and Evaluation of Business Processes in Local Government using Process Mining Techniques Vassilios Christodoulou 711161028

Supervising professor: Alexandros Bousdekis
Co-supervising professor: Georgios Miaoulis
Diploma thesis submitted to the Department of Informatics and Computer Engineering of the University of Western Attica,
Athens, Winter semester 2021-2022

UNIVERSITY OF WESTERN ATTICA
DEPARTMENT OF INFORMATION AND COMPUTER ENGINEERING

Analysis and Evaluation of Business Processes in Local Government using Mining Process Techniques

Members of the Examining Committee

1. George Miaoulis
2. Dimitrios Magos
3. George Bardis

The degree/diploma thesis was successfully examined by the following Examination Committee:

a/a	Name	Capacity	Digital signature
1	George Miaoulis	faculty	
2	Dimitrios Magos	faculty	
3	Georgios Bardis	faculty	

DIPLOMATIC THESIS AUTHOR'S STATEMENT

The undersigned VASILIOS CHRISTODOULOU of CHRISTOU, with registration number 711161028 student of the Department of Informatics and Computer Engineering of the Faculty of Engineering of the University of Western Attica, declare that: "I am the author of this thesis and that any help I had in the preparation of , is fully acknowledged and referenced in the work. Also, any sources from which I have used data, ideas, or words, whether exact or paraphrased, are cited in their entirety, with full credit to the authors, publisher, or journal, including sources that may have been used from the Internet. I also certify that this work has been written by me exclusively and is a product of intellectual property of both me and the Foundation.

Violation of my above academic responsibility is an essential reason for the revocation of my degree".

Dilon Vassilios
Christodoulou

THANKS

For the preparation of this thesis, I would like to thank
my professor Mr. Alexandros Bousdekis, for his advice and support throughout
duration. Also, I would like to thank my family for their support and
their patience throughout my studies.

SUMMARY

In today's era a business/organization needs to manage large volumes of data, perform daily unlimited tasks and deal with issues that arise non-stop. The ability of a business to cope with these issues is of major importance for its viability. If in some cases the performance and its productivity lags, then a problem arises in the business or organization. The business aims, through tasks, to increase revenue, while the organization aims to solve the requests that arise. Process mining and its tools can help us monitor tasks and make inferences about them and how they are executed. Mining techniques and platforms enable us to analyze tasks and their behavior, identify problems and vulnerabilities in a system, and limit or eliminate problems.

Applying process mining to activities and tasks is a tool to study behavior and patterns that develop, discover anomalies and improve the way tasks are performed. In this way a business will be able to increase its performance and improve the way it operates, to support its sustainability. With a direct impact on the workers themselves, where they will be able to improve their standard of living.

In the thesis we use a data set from the 11th International Workshop on Business Process Intelligence 2015, Business Processing Intelligence Challenge (BPIC), which includes data from 5 Dutch municipalities. The data contains work carried out by the local government, over a period of 4 years .Many different tasks are included, denoted by both codes (attribute:name) and labels, both in Dutch (attribute taskNameNL) and in English (attribute taskNameEN).The cases contain tasks at many different stages of their development, as and the person in charge who had undertaken them and the estimated cost. The procedures in the 5 municipalities must be identical, but they can differ slightly from each other.

By using the open source software Prom Tools for process mining, which is widely used in the mining sector, we will be able for each municipality to analyze the processes and study them. Then we will be able to identify points of interest, such as inhomogeneities, pathogens and errors in each municipality, and further process them with filtering and mining algorithms. With this technique we will be able to discover vulnerabilities in each municipality that reduce performance and eliminate them. Ending up in a situation where we can perform tasks with the most efficient way, greatly increasing the desired productivity.

Keywords: Process mining, Local government, Event log, Analysis and evaluation, Business processes

TABLE OF CONTENTS

LIST OF IMAGES

CHAPTER 1

1.1 INTRODUCTION

BUSINESSES AND ORGANIZATIONS

Businesses and organizations have as their main purpose of operation, profit offering consumers goods or services. By the term business we also mean the purist economic unit which is an independent and responsible organization of producers rates and management of transactions with which he seeks the maximum possible profit.

As a rule, the profit should exceed the corresponding usual remuneration (as remuneration) of the administrative or executive work performed in it. Thus with the above definition profit is clearly given as the main criterion, which distinguishes it from any other economic one unit, since all the aforementioned characteristics are common and abstract, consequently to not be taken as criteria for distinguishing between the two concepts. Businesses and organizations they are generally productive units with various legal forms. In them, one or more people decide on production. The pursuit of maximum profit is key characteristic of them, because the greater the profit, the more certain is the survival and their long-term development. [1]

As far as the organizations are concerned, they are not part of the for-profit part businesses operate, but they aim to use legal means and actions that are intended to achieve a specific result for the satisfaction of the general interest of the citizens of a state.

CHARACTERISTICS AND TYPES OF BUSINESSES

Main characteristics of a business, which are also the necessary ones conditions are the following:

1. An economic unit that presupposes a permanent combination of factors of production. 2. Independent financial unit. 3. Its production should be addressed to an unknown consumer audience and finally 4. It should not be primary, but should be formed in a certain socio-economic an environment characterized by both monetary valuation and the pursuit of profit

In general, businesses are distinguished according to the object of their activity, the proprietary venture capital, by their legal form and by size:

Enterprises by object of action

1. Primary production enterprises. 2.
Production or processing or transformation enterprises. In this category
all industries and crafts are included.
3. General trade, or commercial enterprises. 4.
Insurance companies. 5.
Businesses providing services
6. Banking operations. 7.
Transport operations, which are distinguished into land, sea and air.

Businesses from their agency

1. Private enterprises 2.
Public enterprises 3.
Mixed enterprises

The organizations we are referring to include mainly state, self-governing ones more often, units where they aim to serve the needs of the natural and legal entities and in general the exercise of public administration. We point out here that the public administration of a state is part of the state apparatus and is absolutely intertwined with the government. In the specific example of the work we deal with the municipalities, an entity of the public administration, where they are part of the local self-government of a state. They are units independent to a degree, which include one city or town and its surrounding villages and sometimes includes only some district or suburb of a big city, or on the contrary only villages of an isolated mountainous region or an island. We clarify here that in Greece the municipality is *a primary* Organisation of Local Government (OTA). In other countries, primary local authorities are referred to indiscriminately as communities. [2]

In the above cases we have mentioned whether it is a business or an organization units are aimed at high productivity and efficiency, either for profit or serving the needs of individuals respectively. In the cases of organizations and especially the objective of the municipalities is generally the orderly functioning of the community of residents and the promotion of these interests. In each state, the respective governments draw up the guidelines and administrations implement them.

With the development of science and technology that we have at our disposal today era, we are able to address these issues and every business can develop its production and the availability of targeted products and services, through It is worth noting that in recent years, many companies have invested in rapid development of the internet to offer their services. Even with its use

process mining (process mining) we can analyze the daily tasks that take place in any business or organization. In this way by deeply analyzing the profile of these processes and the people involved in them, we are able to draw conclusions about them and to improve them. Below we will study the main issue of process mining, where through a process model it analyzes processes from the log files, in order to identify pathogens and deal with them.

1.2 OBJECT OF DIPLOMACY

The subject of the diploma is the analysis and evaluation of business processes in Local Government Organizations using process mining techniques, so we will be able to identify problems and improve the processes. The data of the work comes from the Technical Service of five Dutch municipalities and is presented in the form event logs, with the main goal to analyze, extract and evaluate process models. Process and process model mining techniques were done using the open source tool Prom Tools.

With diplomacy, we aim to provide answers to questions concerning these municipalities and other organizations, aiming to optimize their performance.

Some of the questions concern: ÿ
What are the most common jobs in each municipality, but also the rarest? ÿ Which organizations/employees handle the most tasks and the fewest respectively? ÿ Which municipalities have the highest productivity and performance? ÿ How do the tasks and employees in the 5 municipalities interact with each other? ÿ How well do the process models agree with the logs and measurements we have?

1.3 STRUCTURE OF DIPLOMACY

To better organize and structure the framework of diplomacy, in chapter 2 we will start with the theoretical background of process mining. There we will report the path from the early business acumen developed to increase business performance, to the need for business process management. processes in an agile way. With techniques such as data and process mining with their tools, such as event logs and process models, we will be able to study and manage the processes that are performed and improve them.

Then the 3rd chapter aims to introduce us from theory to practice, where at the beginning it presents a general guide for process mining that can be applied in any situation or business/organization we find ourselves in. In the continuation of the 3rd chapter, the aim is to specify the steps and the techniques we follow in the Prom Tools platform. Starting with the simple step-by-step steps from importing and analyzing the event logs, to the techniques for processing and checking the data. While with the tools we have we can find whether the theory with practice, checking if from the event logs, we get the expected results, in the representation of the elements with the process models. In this way, the theoretical element of diplomacy is completed.

With the aim in the 4th chapter to enter the experimental part with the Dutch municipalities and process mining in them. Applying the techniques mentioned in the previous chapters we go step by step from words to projects to study the data, process it appropriately and extract conclusions. In the 5th and last chapter, we comment on the conclusions and based on them, we answer the key diplomatic questions for municipalities, showing the contribution and the need of process mining in the business sector.

CHAPTER 2
2.1 BUSINESS INTELLIGENCE

As we mentioned before, every organization and every business aims to optimize the processes carried out in it. A business, through high production and provision of services, aims to increase profit. While on the other hand, an organization whether public or private, aims to increasingly better serve the needs of each person and improve the services it offers.

To achieve the above objectives, in recent years a new term called business intelligence (BI) has been developed. This technique aims to help a company increase the activities it carries out to improve performance, decision-making and its predictive analysis. This new-for-the-time term combines fields such as business analytics, data mining and virtualization with the help of data tools and infrastructure, with the aim of driving a business, small or large, to make sound decisions based on the data it has in its possession. Business intelligence (BI) begins to be strongly present in the business sector since the 1950s with the digital revolution of the time, showing since then the pivotal role it would play for businesses. In the following decades several companies, which today are technological giants like IBM and SAP, adopted this technique as their business strategy to prevail over their competitors. To reach today where hundreds of companies embrace this technique. [3]

Figure 1: Business intelligence in architecture and technology

Some of the special benefits that businesses and organizations get when using it BI include:

- Increased efficiency of business processes.
- Insights into customer behavior and purchasing patterns.
- Accurate tracking of sales, marketing and financial performance.
- Delete reporting criteria based on historical and current data.
- Instant notifications about data anomalies and customer issues.
- Analytics that can be shared in real-time across departments.

In the past, business intelligence tools were mainly used by data analysts and IT users. Now, self-service BI platforms make business intelligence available to everyone, from executives to business teams. Today, business intelligence is used in many industries such as financial services, healthcare, manufacturing, and transportation.

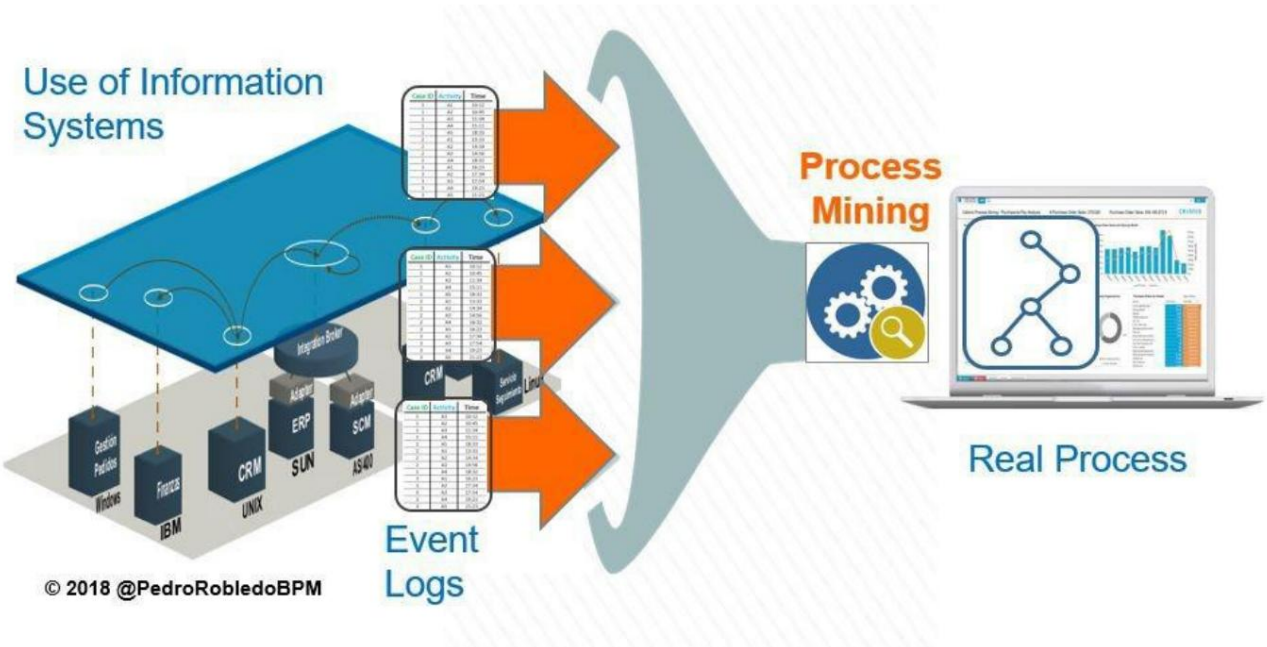


Figure 2: The course of Process mining

2.2 BUSINESS PROCESS MANAGEMENT (BPM)

But as business intelligence was increasingly used to mine and visualize data, analyze and design their more efficient strategy, and search for new better decisions driven by the analysis of the data they had, they noticed that business intelligence had some major limitations. .

In essence, what is done in an organization can be considered as a process. The success of a business is directly related to how well it manages its business processes. Effective management requires an understanding of the quality and timeliness of how these processes are performed. A major limitation of business intelligence and data discovery tools is that measurements and analyzes have no connection to the business environment, making it difficult for the user to interpret and give meaning to these measurements. These obstacles reveal the need for an automated process intelligence (PI). This term marks the next step in business intelligence and involves the collection of data for the purpose of business process management (BPM) and workflows to increase process efficiency. . [4]

In business process management we are dealing with a discipline that uses various methods to discover, model, analyze, measure, improve and optimize business processes. A business process coordinates the behavior of people, systems, information and things to produce business results in support of a business strategy. Processes can be structured and repetitive or unstructured and variable. Although not required, the technologies are often used with BPM. BPM is key to aligning IT/OT investments with business strategy.

As an approach, BPM views processes as important assets of an organization that must be understood, managed, and developed to communicate and deliver value-added products and services to customers or clients. This approach is very similar to other total quality management or continuous improvement process methodologies. [5]

With the element of intelligence that this term includes, it signals the need for an additional level of transparency that must be added to fully understand what needs to be done in a business to preserve the reliability and quality of its processes. Process intelligence (PI) aims to improve the services and goods offered by a business through the essential understanding and analysis of processes

Essentially, this branch deals with the processes (actions) of a business, observing and analyzing them, to find what is needed, so that they become even more efficient.

2.3 DATA MINING

Another very important industry that began to develop slowly was data mining, although it was unknown until the 1990s, it experienced rapid growth in the following years until today it has become gigantic. data mining had come to study the data that existed from the processes. Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a wide range of techniques, you can use this information to increase revenue, reduce costs, improve customer relationships, reduce risk, and more.Within this technology, mining is also combined with other sciences to be able to improve the results it produces such as statistics (the numerical study of data relationships), artificial intelligence (human intelligence displayed by software and/or machines) and machine learning (algorithms that can learn from data to make predictions).As the years go by data mining technology continues to evolve to keep up with the limitless possibilities of big data and affordable computing power. [6]

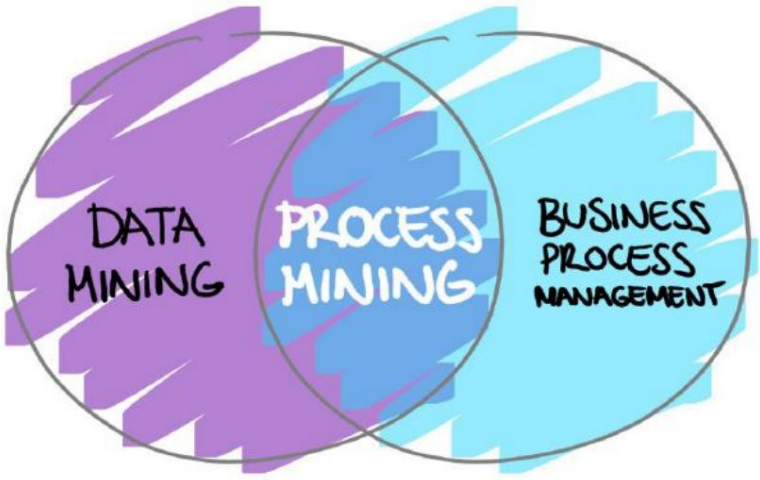


Figure 3: Connection between data mining and business process management

2.4 MINING PROCESSES

The connection between data mining (data mining) and business process management (PI) is made by process mining (PM), which is also the main topic of this thesis. This fledgling branch stems from the process intelligence we mentioned above and has as its purpose the in-depth analysis, monitoring and improvement of real processes that take place every day in millions of businesses around the world. We could say that the main difference between the "father" of this technique, business process management, that started this branch and its "child", process mining (PM) is that the latter deals with the deeper causes that led to a situation, while the former, with the techniques it includes, can simply identify the problem without is able to fix it.

2.4.1 INTRODUCTION

The field of process mining emerged in the late 1990s at the Eindhoven University of Technology with the pioneering work of prof.dr.ir. Will van der Aalst. Over time the discipline matured and today, there are over 35 commercial process mining vendors and thousands of organizations successfully implementing process mining. The IEEE Task Force on Process Mining and the open source ProM framework have played a key role in the development of the industry. [7]

In more detail the events that followed, in chronological order to make the mining process this is today, are as follows:

TIMELINE OF MINING PROCESSES

ÿ 1999 The beginning of process mining research at Eindhoven University

ÿ 2000 Creation of the alpha miner algorithm for process model extraction

ÿ 2001 Creation of the heuristic miner algorithm

ÿ 2004 Evidence-based compliance testing

ÿ 2005 Decision mining

ÿ 2006 Mining Organisms

ÿ 2007 First process mining company (Futura Pi)

ÿ 2009 First working groups in process mining and creation of Fluxicon process mining platform

ÿ 2010 Aligned compliance check

ÿ 2011 Creation of first process mining ledger and creation of Celonis process mining platform

ÿ 2014 Coursera MOOC Process Mining

ÿ 2016 Release of book “Process Mining: Data Science in Action”

ÿ 2018 Over 30 companies specializing in process mining and its tools. Celonis software is the flagship in the field.

ÿ 2019 ICPM 2019 :First Process Mining Conference

2.4.2 PROPERTIES AND APPLICATIONS

Process mining as a field is able to answer 4 major questions

ÿwhat actually happened in a process (task)

ÿdeviations from its original design

ÿat which points is there congestion (bottleneck)

ÿprocess improvement

Process mining is widely used in a wide range of fields. Its biggest contribution is in important areas such as:

Business: as is the main subject of the diploma, process mining is applied quite a bit in the business sector. Its techniques optimize the processes that take place in a business. Algorithms help to better analyze and study tasks, while in this way performance can be evaluated and possible points that slow it down. Process models offer a visual presentation of the processes, drawing a first conclusion on the degree to which it agrees with the log file. In recent years the giants and later all companies establish such areas , in order to evaluate the mode of operation.

In the field of businesses and organizations, process mining can offer benefits such as: ÿ Potential
for savings, seeing exactly where the waste and loss is in a business

ÿ Complete knowledge of the processes taking place, when they start and when they end, what their course will be and if there is any problem in them.

ÿ Discovery of weak points and vulnerabilities of our system, which can create problems

ÿ Observing how processes are connected and how they interact with consumers, creating patterns for analysis

Health: in recent years, with the internet we can offer medical care to people with needs, which in the past was quite difficult. For example, the patient's history can be found digitally, prescriptions can be made online, even monitoring and treatment can be done electronically, especially in the last months with the pandemic, attempts were made to do everything electronically. With the prevailing situation, millions of such procedures are done every day. Since we have many procedures that are performed, you should evaluate them, search for possible problems and improve the services that are available. For these reasons process mining in healthcare is essential and all organizations/businesses should apply these techniques.

Human activity: every day people around the world perform many activities, from the simplest to the most complex. Everyone has their own activities that are performed daily, while many people are looking for ways to make the activities more efficient and in less time. Interest in this issue exists from many experts, who by applying process mining techniques, are able to analyze activities and behaviors and improve them, contributing to the well-being of citizens. This will overall affect the whole society, helping the its overall development and the development of its standard of living. It is of course worth noting that process mining in people's activities also serves the advertising purposes of businesses, with the aim of offering personalized products and services to citizens.

Transportation: every day people use vehicles to get from one place to another. This is done either by their own vehicle or by paying for transportation. Millions of such processes happen every day in many different places, at different times and with different media. As an industry it is gigantic and you encounter it every day in the lives of many people. The amount of data collected by

this activity and with many variations, depending on the medium and the features it has. Process mining as a field can filter these elements, find recurring patterns, pathogens and other issues that arise. With the techniques it has it can show the issues and help resolve them, thus offering better services for those concerned.

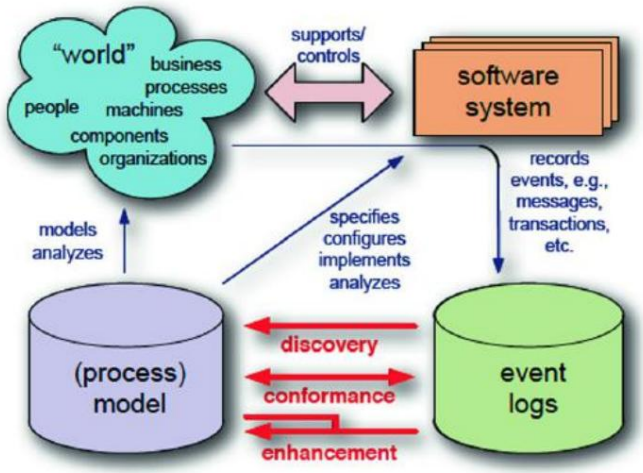


Figure 4: Operation of process mining in everyday life

Applying Process Mining to an organization offers the following capabilities:[8]

-Automated discovery of process patterns, exceptions and process instances(cases) along with key frequencies and statistics.

-Automated discovery and analysis of customer interactions, as well as alignment with internal processes.

-Monitor key performance indicators using real-time dashboards. -Improve existing or previous

process models using additional data from saved records.

-Combining different process models that interact with each other into a single process mining dashboard.

- Support for visualizing how processes contribute to business value (such as business operating models) - framing processes.

-Effective collaboration between Business and IT.

- Standardization of business processes.

-Improving operational excellence by optimizing processes.

2.4.3 PROCESS MINING TOOLS

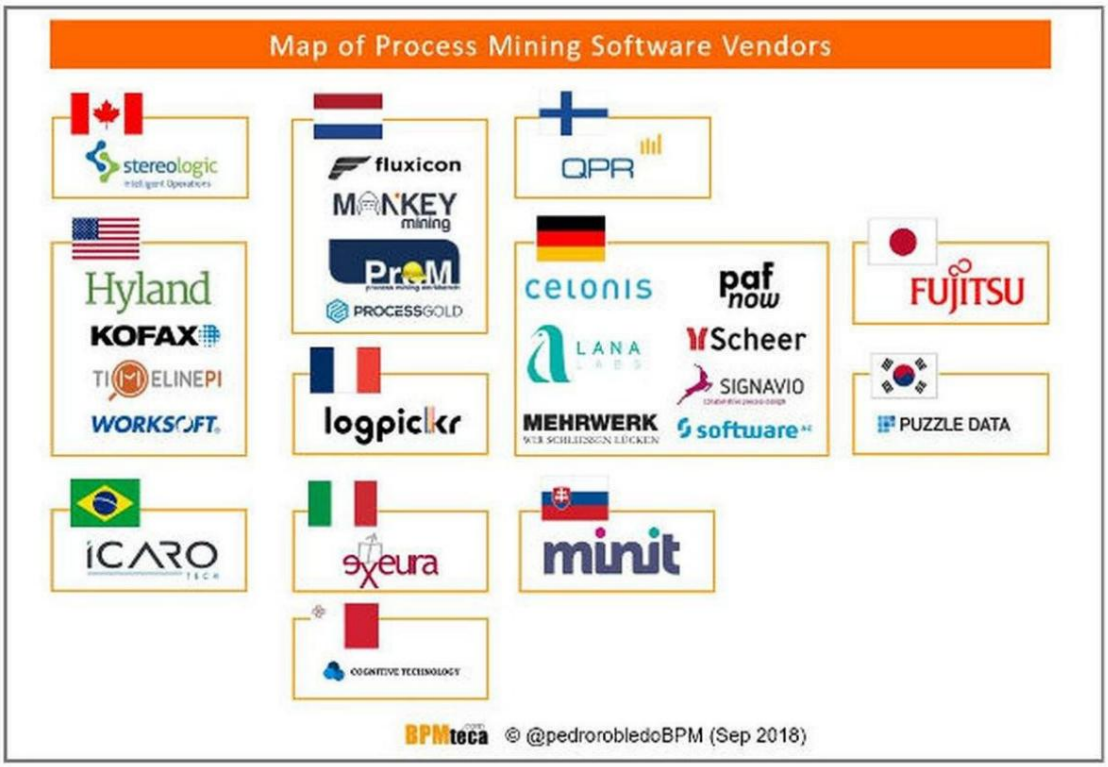


Figure 5: Major developers of process mining software around the world

For process mining we can use various tools and software, both open source and paid, which may have more features. Above we see various options, even the Prom Tools we use in our examples, but also alternatives if we prefer something else. Europe accounts for the largest market size due to the wide acceptance of innovations and upcoming analytics technology among the large number of Process Mining software vendors operating in the region. Germany and the Netherlands are the leading countries contributing to the process analytics market in Europe. North America shows the highest growth rate, representing the largest opportunity in the process analytics market. [9]

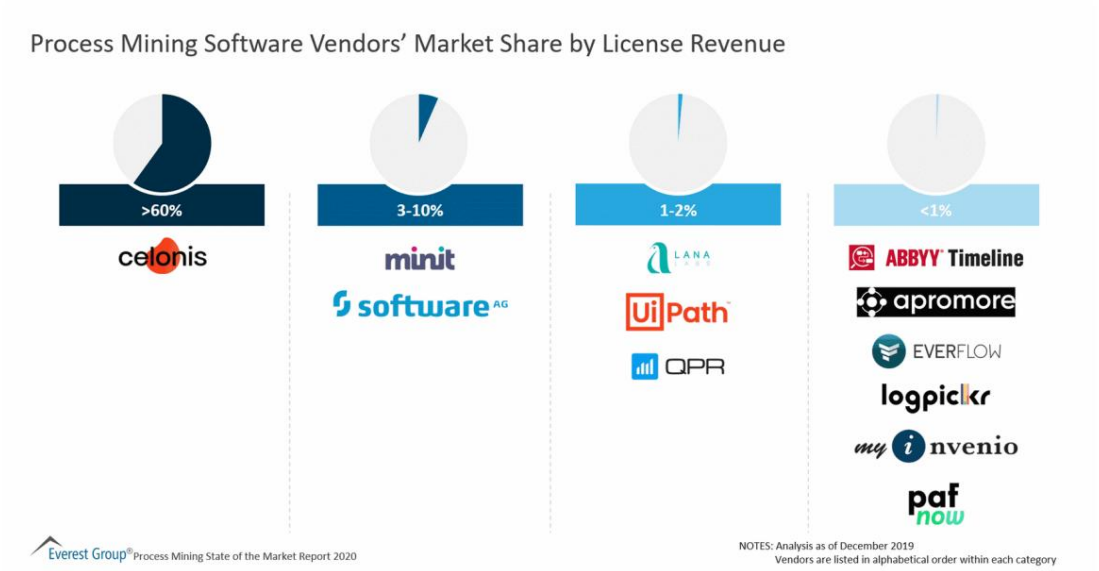


Figure 6: Market share of process mining software vendors

Finally, it is worth noting that in recent years more and more states are beginning to invest in this sector. The prospects it offers are enormous and the benefits are not insignificant for companies and states therefore. In Greece, efforts have begun to implement such techniques and in public sector, with the aim of improving the way work is carried out. Recently there has been an increase in the effort to develop the sector, through significant funding, making the process analytics market quite attractive. As is logical, many new positions for data analyst, data scientist and data engineer have been created by gathering interest from all disciplines. As a science, it combines many fields such as IT, programming, finance and can be applied in any of the areas we mentioned above, for example in our daily life, our work, our movements, even the way of entertainment which we are interested in. Process mining is often misconstrued as a field related to data science. Process mining should be considered as a bridge between data science and process science. By the year 2018, almost 30+ commercially available process mining tools were in the picture. The year 2019 was set for the first process mining conference. Today we have over 35 vendors offering tools and techniques for process discovery and compliance auditing. Despite the young age of this industry, it is the fastest growing in the IT sector and is reaching out to stakeholders from all over the world to engage with it. .

Figure 7: The process analytics market

2.5 PROCESS MODEL AND MINING ALGORITHMS

Process mining to achieve its goals implements a process model, which presents the expected operation of a process and what we expect to get as a final result from it. More specifically, the process model is the graphical representation of business processes or flows work. Like a flowchart, the individual steps of the process are drawn up so that there is an overall overview of the tasks in the process in the context of the business environment. match the logs we have, So at a glance, by looking at it we can check if the actual end stage of a process is the same as the process model predicted.

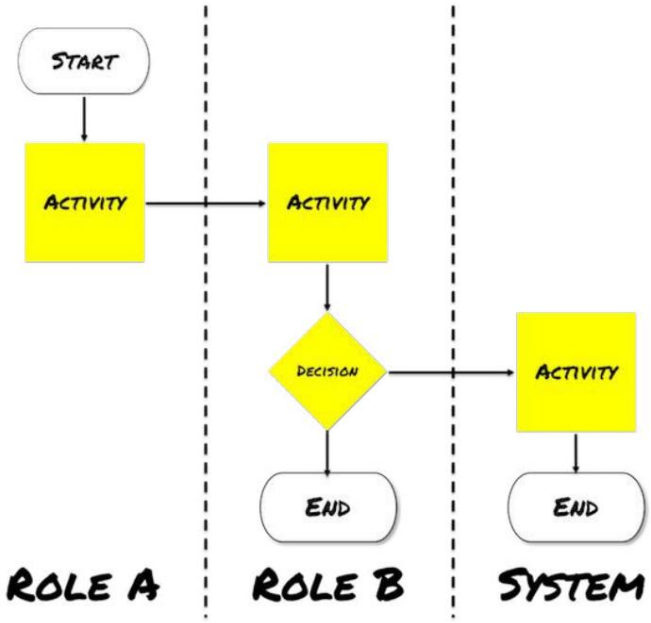


Figure 8: Flow chart or process model structure

In the field of process mining, as we mentioned before, the process model plays an important role. It is the visual representation of the process diagram or flow of actions. To produce this model in the world of process mining, we must have special algorithms , where with the use of these the process tree will be produced. Usually in the tests we use four such algorithms, in order to cover all the different requirements we have in each test. Specifically, most of the time we work with the Alpha miner, Heuristic miner, Inductive Visual miner and Fuzzy miner. Each one of them has different characteristics and offers possibilities, where maybe the others don't offer. In each case we have to look at the problem we are facing and choose the optimal algorithm for this job.

Above were presented the 4 algorithms we use in Prom Tools. Let's mention some facts about them, the alpha miner was the first algorithm implemented for this use, specifically in 2000. The ÿ-algorithm or ÿ-miner is an algorithm used in process mining, aiming to reconstruct causality from a set of facts at our disposal. Implemented by van der Aalst, Weijters and Mÿruÿter. The goal of Alpha miner is to turn the log file into a workflow network based on the relationships between the various activities in the event log. An event log is a multiple set of traces (cases), and a trace is a sequence of activity names (events). Several extensions or modifications of it have been introduced since then, which will be listed below. Many of the process mining techniques have been influenced by the original ÿ miner.

One of them is the Heuristic miner which was implemented a little later around 2001. It is an improved version of the original algorithm and was the second process mining algorithm, closely following the alpha algorithm. Developed by Dr. Ton Weijters, who used a heuristic approach to deal with many problems with the alpha algorithm, making this algorithm much more suitable in practice. One of the advantages is that a heuristic network can be converted to other types of process models, such as a network Petri for further analysis in ProM. It is good to prefer when we have real data with not too many different events or when you need a Petri net model for further analysis in ProM.

Another algorithm that is widely used in Prom is the Inductive Visual miner algorithm, which is a fairly improved version of the two previous algorithms. It gives us the possibility of live monitoring of the process model, seeing in real time the flow of task execution. This algorithm is going to for the most efficient one, in most cases, offering several options for parameterization and a better presentation of the final process model. In our measurements, it has been used a lot due to the ease of handling and the parameterization it offers in live time.

Finally we have Fuzzy miner where here we have the newest algorithm in the genre. Fuzzy miner is one of the newest process discovery algorithms and was developed by Fluxicon co-founder Christian W. Gnther in 2007. It is the first algorithm that directly addresses the problems of large number of activities and highly unstructured behavior. Fuzzy miner uses importance/correlation metrics to interactively simplify the process model to the desired level of abstraction. Compared to him

Heuristic miner, can also leave out less important activities (or hide them in clusters) if you have hundreds of them. It is preferred in cases where we have complex and unstructured log data or when we want to simplify the model in an interactive way. [10]

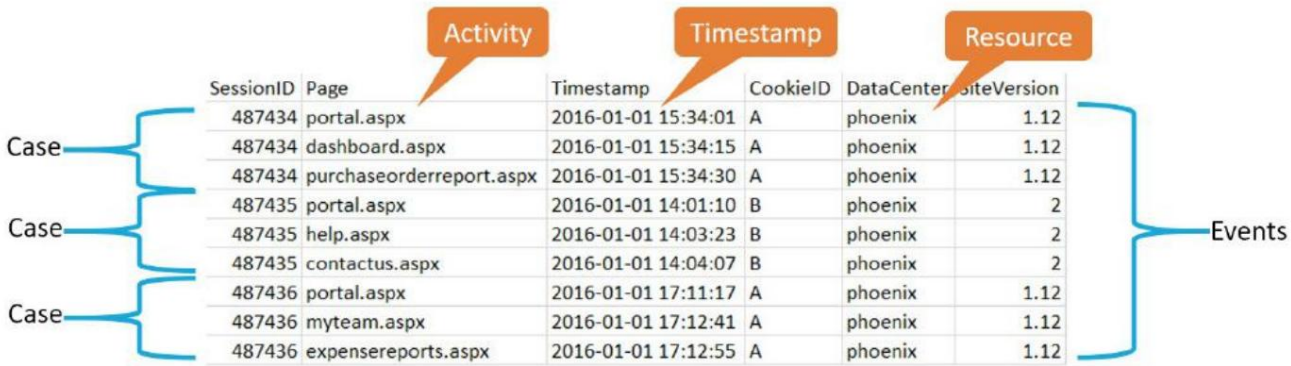
	Alpha Algorithm	Heuristics Miner	Genetic Miner	Fuzzy Miner
Integrity		✓	✓	
Uniqueness	✓	✓	✓	✓
Exceptions		✓	✓	
Real-Time	✓			
Model Integrity	✓	✓	✓	✓
Abstraction level				✓
Structured model				✓

Figure 9: Comparison between mining algorithms

2.6 EVENT RECORDS

When a process is performed in the company, this process, after it is completed, is recorded with its basic elements, in specific files that include all the processes of a company. These files, which we will meet below, are called event logs. It is a a key resource in modern systems that helps provide information about processes taking place and includes information about each process. Such as the name of the activity, its code (each activity has a different code and no two processes can have the same) , the time it took to complete, and various other properties that vary by action. Some other actions may be the person or organization that performed the process, whether it was completed, and how much it cost. There may be other properties in order to they make the tasks comparable to each other, so that conclusions can be drawn more easily. These files store a huge amount of data which is particularly useful in expert data recovery efforts and in cases of analyzing and monitoring the actions of a business, in order to determine the performance and changes that would benefit it.

2.6.1 FILE STRUCTURE



SessionID	Page	Timestamp	CookieID	DataCenter	SiteVersion
487434	portal.aspx	2016-01-01 15:34:01	A	phoenix	1.12
487434	dashboard.aspx	2016-01-01 15:34:15	A	phoenix	1.12
487434	purchaseorderreport.aspx	2016-01-01 15:34:30	A	phoenix	1.12
487435	portal.aspx	2016-01-01 14:01:10	B	phoenix	2
487435	help.aspx	2016-01-01 14:03:23	B	phoenix	2
487435	contactus.aspx	2016-01-01 14:04:07	B	phoenix	2
487436	portal.aspx	2016-01-01 17:11:17	A	phoenix	1.12
487436	myteam.aspx	2016-01-01 17:12:41	A	phoenix	1.12
487436	expensereports.aspx	2016-01-01 17:12:55	A	phoenix	1.12

Figure 10: Example of an event log with traces and processes, each containing attributes

Regarding the structure of these files, we must mention that they are structured with the logic that every time a process runs, we have a trace of it. Each trace in a log file can consist of a sequence of activities that are executed one after the other the other. One sequence can be differentiated from another with the same processes, in the event that the actions run at different times and in a different order. In other words, we conclude that in such cases with too many references, the order in which the actions were taken and which action followed another plays a huge role.

In general, each event log consists of 3 main elements:

ÿ Case id: each event should refer to a case (ie, process instance). It is used to separate cases and to be able to easily study a case.

ÿ Activity: every event must be related to an activity. Events refer to instances of activities, i.e. occurrences of activities in the corresponding process model.

ÿ Timestamp : events within a case must be ordered in time order of execution. Furthermore, timestamps are not only needed for time order: they are also crucial for measuring the process and comparing it with others.

Then it can also consist of more secondary details such as: ÿ Responsible (resource) : the person or organization that has undertaken a process and must carry it out. Each process/case has a person responsible for it.

ÿ Completion status : whether a process has been completed successfully or unsuccessfully, if its operation has been interrupted or if it has been canceled completely.

ÿ Costs : the amount of money spent to execute a case and its processes, usually set at the beginning but can change along the way [11]

It is worth noting that all modern systems we use have such files. For example, Microsoft Windows, with over a billion users every month, has such log files, which are available for users to study and in case of a problem to find out what happened. Similar tools are also found on other platforms, aiming to investigate events and discover the root of the problem. In most cases, these work automatically, with the push of just one button, presenting the user with all the events of his system.

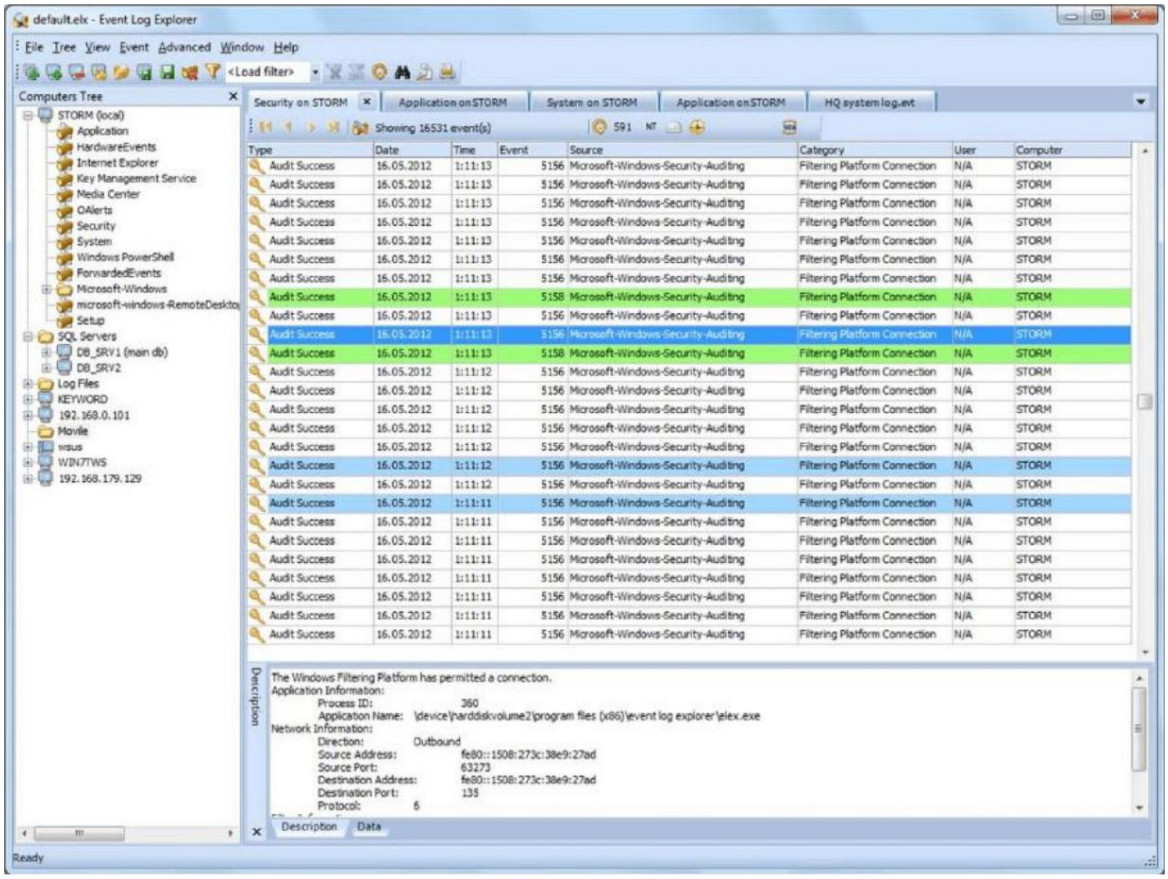


Figure 11: Image from Event Log Explorer with OS events Microsoft Windows

2.6.2 FILE STANDARDS As far as the diplomatic example is concerned, log files have the XES Format (an XML standard for event logs) which aims to provide a general standard of information for the presentation and processing of data in tools and applications.

XES (extensible event stream) was introduced in 2010 and is the successful successor of the MXML standard, which appeared in 2004. MXML is an XML-based user interface language first introduced by Macromedia in March of 2004. Application developers use MXML in conjunction with ActionScript to develop rich web applications, with products such as Apache Flex. Using MXML it is possible to store event logs using an XML-based syntax. MXML has a standard format for storing timestamps, resources, and process state. Additionally, one can add arbitrary data items to cases and procedures.

The XES standard defines a grammar for a tag-based language whose primary purpose is to provide information system designers with techniques and tools

for recording system behaviors via event logs. An XML schema describing the structure of an XES event log/stream. Additionally, a core collection of so-called XES extension prototypes that provide semantics to certain attributes as recorded in the event log/stream included in this standard [12]

Figure 12: Structure of an XES file

The following images show the structure of XES files and each element that we can extract, observing the skeleton of the file. Then the structure of the XES files of the municipalities through Prom Tools is shown.

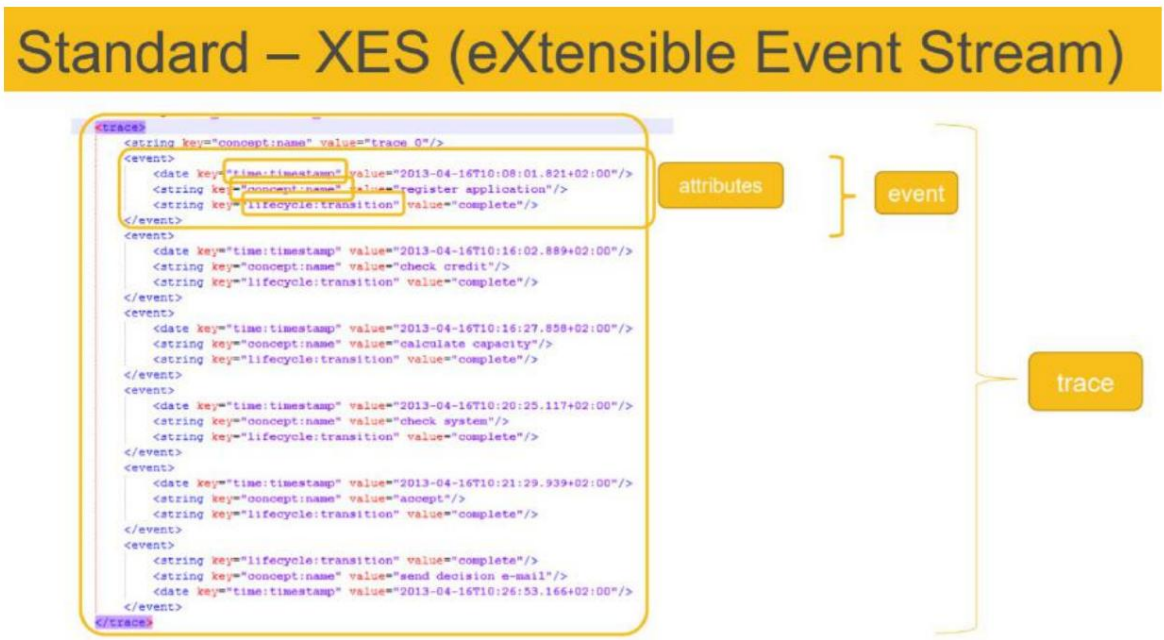


Figure 13: Structure of a trace with processes and attributes

Above we see what I mentioned above, that in each file we have many traces or people (traces) where, in turn, each of them includes many processes (events). Each event has enough elements (attributes) to be separated and to be comparable with other events. It should also be noted that at the beginning the extensions required for the proper operation of the file are loaded. Then the elements of each trace are listed and then the elements of each event. Finally, some classifiers are also listed, acting as criteria for comparing 2 or more events or traces.

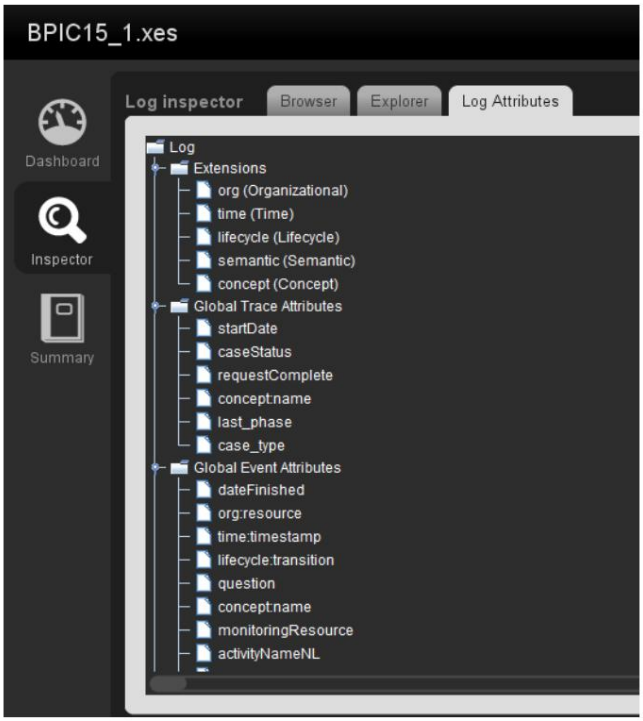


Figure 14: Structure of the log file of our example

In the files we studied in the diploma we have the above template that can be seen through the Inspector of Prom, at the beginning the extensions are loaded, then we see the elements of the trace, the elements of the events and finally the classifiers which compare the data with each other.

Based on the above and then with the next photo we can see the list of all the attributes we have in the event logs, both for cases and for the processes they include (events). This way we better understand the relationships between of traces and processes and how they interact.

variable	description
Case ID	a unique case number,
Activity	a short description in English
Resource	a number representing an employee; 72 different persons
Complete Timestamp	a date, and sometimes a time of completion of this activity
(case) IDofConceptCase	often empty
(case) Includes_subCases	yes/no, often empty
(case) Responsible_actor	a number representing an employee; 64 different persons
(case) SUMleges	the amount of money paid for the permitapplication, often empty
(case) caseProcedure	normal or extended, often empty
(case) caseStatus	Open (O) or Closed (G)
(case) case_type	a number that is the same for all activities in the log
(case) landRegisterID	a number, often empty
(case) last_phase	the last activity in the log for the case
(case) parts	a short description of the type of permit (building, demolishing, advertising,etc)
(case) requestComplete	true or false
(case) termName	not explained, often empty
action_code	a code for the type of activity, indicating position of the activity in the order of the process, eg 01_HOOFD_250_1
activityNameNL	a short description of the activity in Dutch
concept:name	not explained
dateFinished	a date, and sometimes a time of completion of the last activity for this case
dateStop	not explained, usually empty
dueDate	not explained, usually empty
lifecycle:transition	technical formality, Complete, to relate timestamp to the end of the activity
monitoringResource	a number representing an employee; 79 different persons

Figure 15: List of all attributes and their interpretation of our example

We see that each case and each event has some basic characteristics such as ID (identifier), name, resource (employee), start and end date of this, in order to be able to observe and compare them with each other to draw basic conclusions about their behavior. There are also other characteristics such as whether it is complete or no, which organization monitors its progress, the amount of money required and others. In general, each process consists of a case, a name, a time stamp and any other elements that exist.

2.7 CHALLENGES TO LOCAL GOVERNMENT

Of course, as is well known, the public sector and local government do not literally belong to the businesses we know, with the goal of profit through goods and services. Their main concern, through the guidelines of the state, is to carry out work and offer services, to improve the standard of living of citizens and serve their basic needs. For this particular reason when we discuss process mining in the public sector there are some challenges that we should take into account.

A serious issue in municipalities is the execution times and the organizations that perform them. Process
mining through its tools should be able to reduce the times and costs required to process tasks for municipalities. Through the analysis and study of the tasks we can determine which tasks are unreasonably time-consuming and harmful and to modify them. Even with the above tools it can be seen which organizations-employees are not performing satisfactorily and with the appropriate management of them, productivity can increase.

Another challenge is finding, merging, and studying event logs. When extracting data suitable for process mining, several challenges must be addressed: data may be split across sources, data may have gaps, a log file may contain inaccurate values, files may contain information from different layers.

Another challenge is to create appropriate representative benchmarks, consisting of rational data and conclusions, to help study and improve the tools and algorithms we have.

Another important challenge that should be solved through process mining is the balancing between quality criteria such as fitness, simplicity, precision and generalization, these are four competing criteria. quality dimensions. The challenge is to
find process models that score well on all four dimensions and not just on one or two of them. This also shows how complete the picture presented is.

As we all understand, the municipalities are not isolated from each other, of course such a thing would not be wise in terms of mutual assistance and the faster processing of obligations by them. In reality, as in our case, the municipalities share common tasks and cases, while as can be seen from the measurements below, many employees work in more than one municipality. This makes it a challenge for mining processes, as it will have to take these issues seriously and manage them appropriately.

Completing still the most essential issue of process mining is after all the analyses, measurements and conclusions to be able to improve the process, even for the unfamiliar users. It should through reliable tools and user-friendly algorithms, to improve the usability and understanding of mining platforms, so that they can be managed with ease and confidence by all users. [12]

CHAPTER 3
3.1 PROCESS MINING METHODOLOGY

In order to fully understand the steps followed in the diplomatic example, the steps that were taken will be presented in detail. In the first part, we would like to present in a more general step the instructions that should be followed, in cases where we seek process mining in any case. The steps present a generalized path of process mining that will be possible to apply in many different cases. Whether we have the processes of a municipality or a company, whether we have the actions of an individual in his daily life, process mining with its tools it can help us study behaviors, find recurring patterns and study them. These actions are relevant to any platform and tools we use and can very well be applied to any environment we are in for process mining. Next will be presented the main guide of mining that can be followed in whatever situation we are in and want to analyze the processes happening around us.

~~Import files~~

The first step almost always involves importing the event logs that we have and want to analyze. The files include processes that took place in a company, an organization or in a person's life from morning to night. The data is collected from various sources. of our daily life and have some basic characteristics, such as an identifier, a name and a time when it was executed. They can also have other properties such as whether the process has been completed and others. Since we have collected the input files, we must insert them into the desired platform we use. To do this, however, it should be in an appropriate standard that supports the software we have. Usually in the event log files we have the XML, XES, MXML or CSV standards that support the recording of events and respectively the platforms manage such type of files. If our file has another standard, we can use, if our platform has one, an add-on to convert the file to the desired standard. After the first step is completed, we have the input file loaded on the platform we use and ready to analyze it and manage it.

~~Event log visualization~~

Since we have the file ready in our software, the next step will be to observe and analyze the data that the data virtualization offers us. This step presents all the event log data such as the traces that exist, the processes which include other elements. During data virtualization we can have a first look at the files and study the first ones

information we extract. This way we have the ability to study the behavior and patterns we encounter and fully understand them. Then we can use the data we found to process the files appropriately. Each process mining platform offers a variety of virtualization techniques depending on the data we want to extract. There are techniques that show us a statistical picture of the file with the number of traces and processes, their average and the start and end dates of the file. We can if we want to find some frequently repeated work or rarely work to find an inhomogeneity and study its behavior. It is still possible to find the correlation of events over time to see how they behave over time in diagrams for complete understanding. In any case regardless of the information we are looking for, the techniques virtualization offer the possibility of studying and analyzing the information, where later they will help us draw conclusions and process them. Finally, all these techniques are applied without any processing or filtering of the file requested by the user, however with the use of these tools facilitates the subsequent filtering that needs to be done, as it uses data extracted from virtualization.

Discover process model

Another step to follow is the creation of the process model, in which all the processes executed in the event log are presented. The process model describes the flow of work or activities, usually in graphical form, that contribute to achieving a specific goal. Process models are usually used to represent and analyze a series of activities that occur repeatedly and on a regular basis. Each model, or as you often encounter workflow, starts from a process (start event), includes others throughout and ends with a specific task (end event). For the production of these models, regardless of the platform we work on, the methodology is approximately the same. Once we have the file, we apply mining algorithms, such as alpha miner, heuristic miner and others, to produce the model we are looking for. These algorithms differ from each other and we use whichever one we need in each problem, while there are many of them in all process mining platforms. By applying such algorithms to the platform we use, we take as input the event log file and produce as output a model with all the processes that exist. In this way we produce a representation, even live in real time, of the processes of the file, having the possibility to study them, to detect errors or ambiguities that are not visible, seeing the file and to correct them .We can still see how well the original file we have agrees with the process model produced by detecting discrepancies. Many times we choose to first perform the filtering on the file, which is the next step to be presented, and then apply to processed file, a mining algorithm. With this technique we also observe at a graphical level how the file behaves in the filter we apply.

Filter event log

The last step we mentioned above and the most important one we apply to the event log is filtering. Once the event logs are extracted from the data, the next step is to filter them. Filtering is an iterative process. The purpose of this process is to process the data based on some selection criteria that we have and produce the desired output file. All process mining platforms have a variety of plug-ins, which are applied to the file and the filter appropriately. We can deal with the cases or the processes that comprise them. We have the opportunity to study points of interest or inhomogeneities in the data, such as frequently repeated processes and cases or correspondingly rare cases. Applying the previous steps and tools we analyze and identify anomalies, bottlenecks and errors in the event logs. In this step having become aware of such issues, we process the data to precisely study the problems to see how they behave individually, without the rest of the information .In this way we better understand how they work and what patterns they develop. We end up finding the root of the problem and the causes that created it, with the next step to correct it and improve the processes that are performed.

3.2 PROCESS MINING WITH PROM TOOLS

In order to fully understand the steps followed in the example of the 5 Dutch municipalities, the actions taken will be presented in detail with instructions in Prom tools. In this particular example, process mining techniques and algorithms were applied to the Dutch municipalities, however the same steps with minimal modifications can be to be satisfactorily applied to other municipalities, local government bodies and even private companies. The research was done using the Prom Tools ver 6.11 platform.

ProM is an extensible framework that supports a wide variety of process mining techniques in the form of plugins. It is platform independent as it is implemented in Java and can be downloaded for free. ProM 6 is distributed in parts, which offers maximum flexibility. Initially, ProM 6 is distributed as a downloadable package using the open source GNU Public License (GPL). This means that you can download and install ProM 6 without restrictions, but any software that uses the kernel must be downloaded and installed. [13]

3.2.1 IMPORTING EVENT RECORDS

First of all inside the Prom tool, as long as we have the required java(jdk,jre) version, after opening it, we are in the first tab, where we can in the workspace to

see all the files we imported, the saved and our favorite files. Then we have the possibility with the import option to import the desired file in XML format.

A file can also be in csv, mxml or XES format. If we import a file in csv format, we need to convert it to XES format using "Convert CSV to XES" Addendum. In this case we will map the columns of the csv to the new resulting template.

Prom Workspace

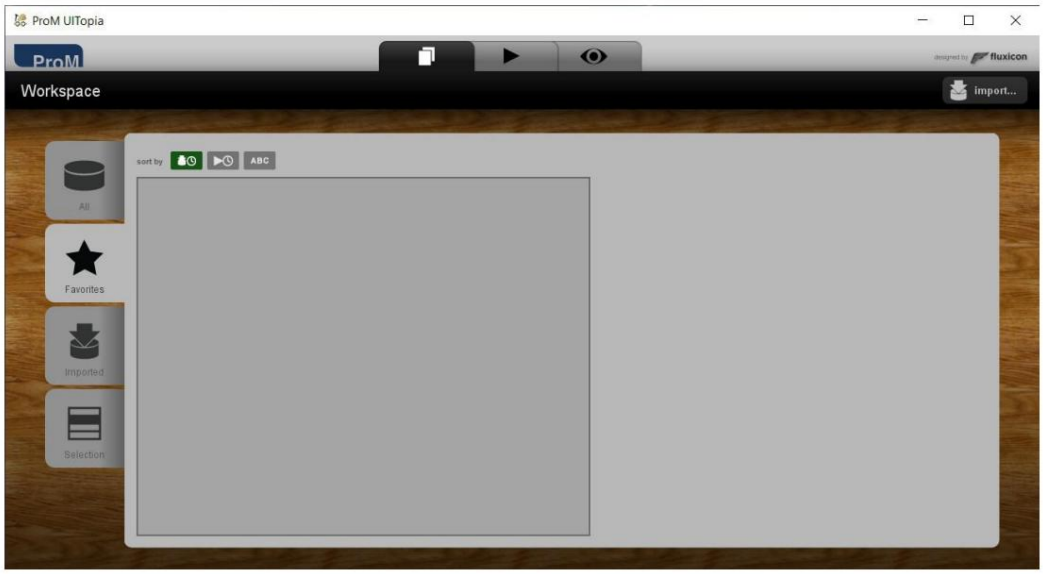


Figure 16: The Prom Tools environment

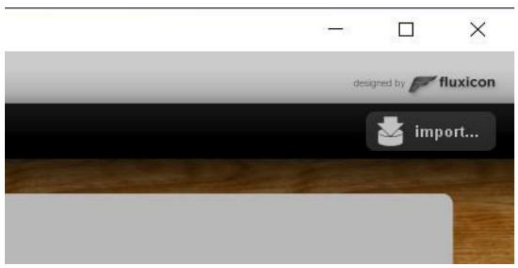
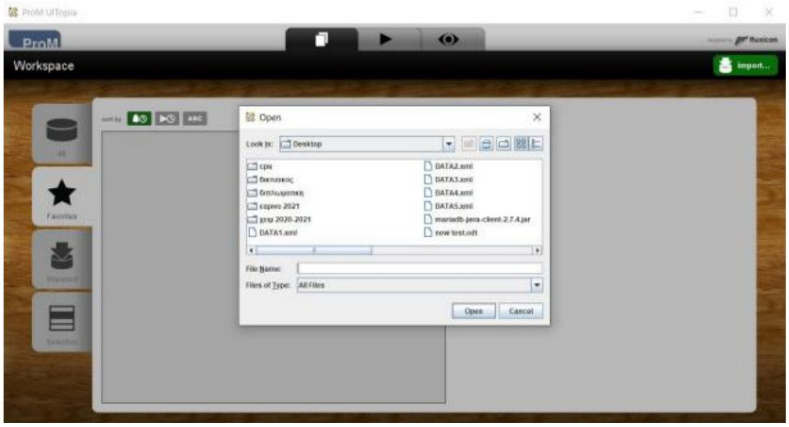


Figure 17: Option to import a file into the software

With the import option, the tab will appear with all the available files on our disk to import any we want and with any of the available plug-ins for import we need. Prom as a tool has over 1,000 plug-ins to use.



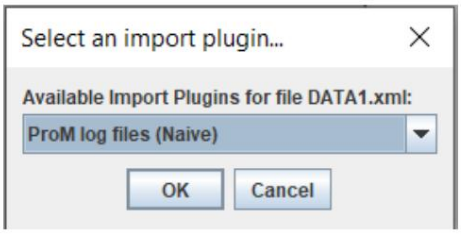


Figure 18: Selecting the desired file from our system

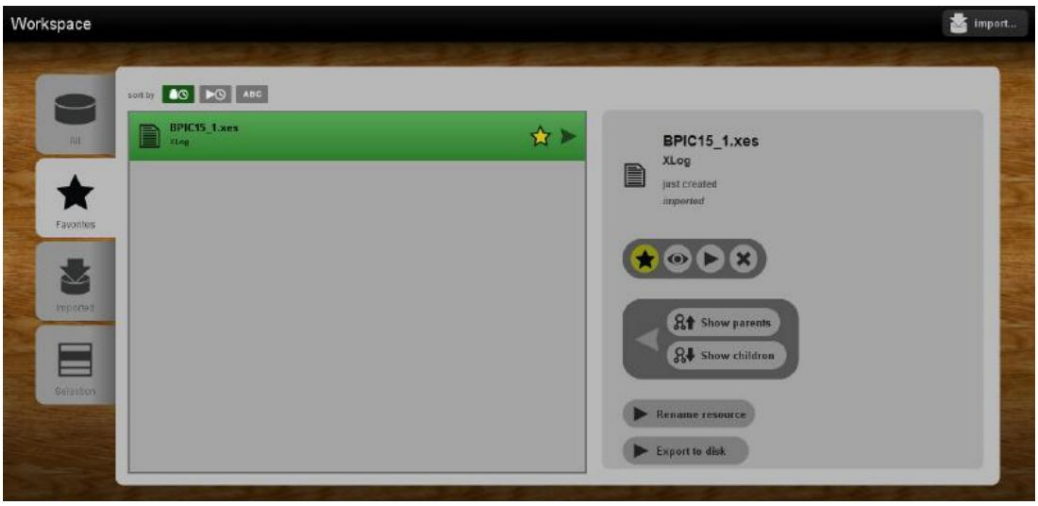


Figure 19: Home page

The Workspace tab shows the file we just placed in the application and we can edit it. Each file we place has some mandatory elements such as case id, process name (event name) timestamp. If we have other parameters the we normally use for further analysis and filtering of the and files.

The next step is to visualize the data and study it from the event log. The available file visualization techniques are in the upper right corner.

Figure 20: List of available data virtualization techniques

3.2.2 EVENT LOG VISUALIZATION

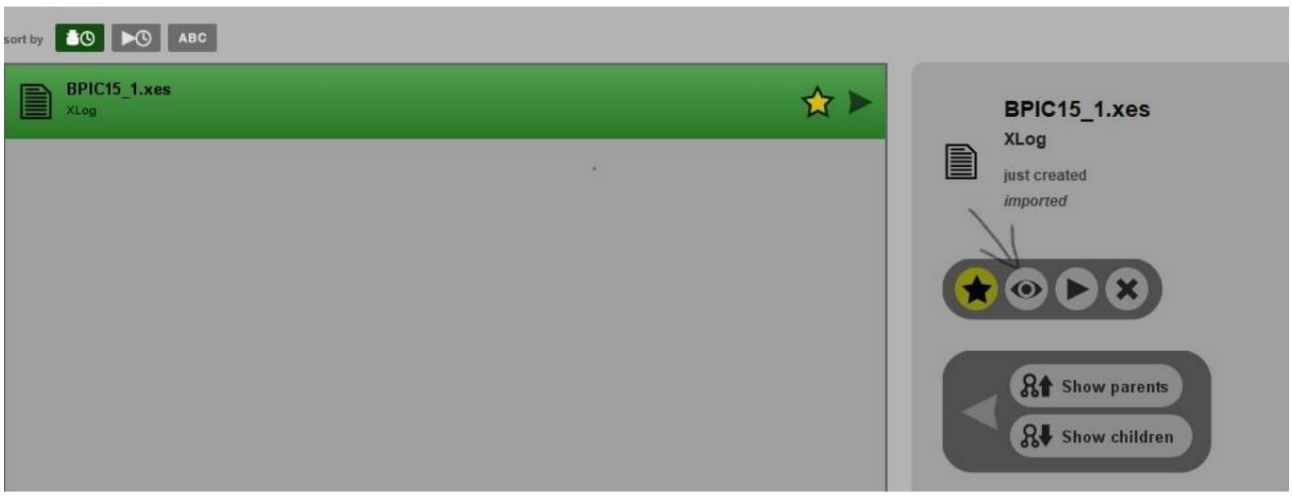


Figure 21: List of files and available actions

Once this step has been done, before filtering anything we can select the view resource button and observe the information that the dashboard tab gives us, to draw some conclusions. This step presents the representation of the file. Then we are able to window to see basic information about the file such as the number of cases, activities and organizations participating in them. Through the graphic display we see the maximum, minimum and average values for the processes per employee. We also note the number of different cases we have, as well as the start and end dates of the file.

Log visualizer/dashboard

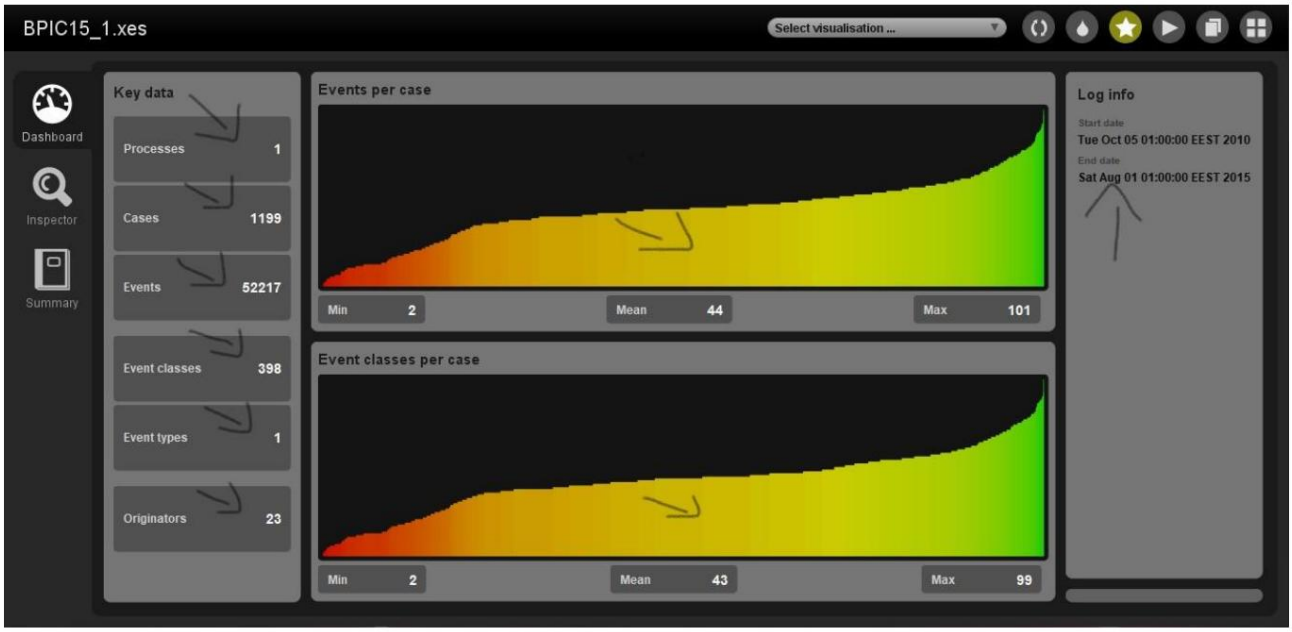


Figure 22: Illustration of event log elements

In more detail from the image above we derive the following information:
ÿ How many processes are included in the event log (obviously one)

ÿ How many cases we have in the event log ÿ How many events
are recorded in the event log (an event can be recorded multiple times)

ÿ How many categories/types of events we have observed in the event log ÿ How many types of events (eg start, finish)
ÿ How many organizations/employees performed processes on the given file
ÿ How many processes were executed per case (including the maximum, minimum and average value of processes per case)
ÿ How many categories of processes were executed per case (including the maximum, minimum and average value of processes per case)
ÿ The first and last observed time process, which gives the time interval the event log

Inspector

In the following windows through the Inspector we have 3 different approaches: browser, explorer and log attributes. Each one gives different details about the cases, processes and the workers.

Browser

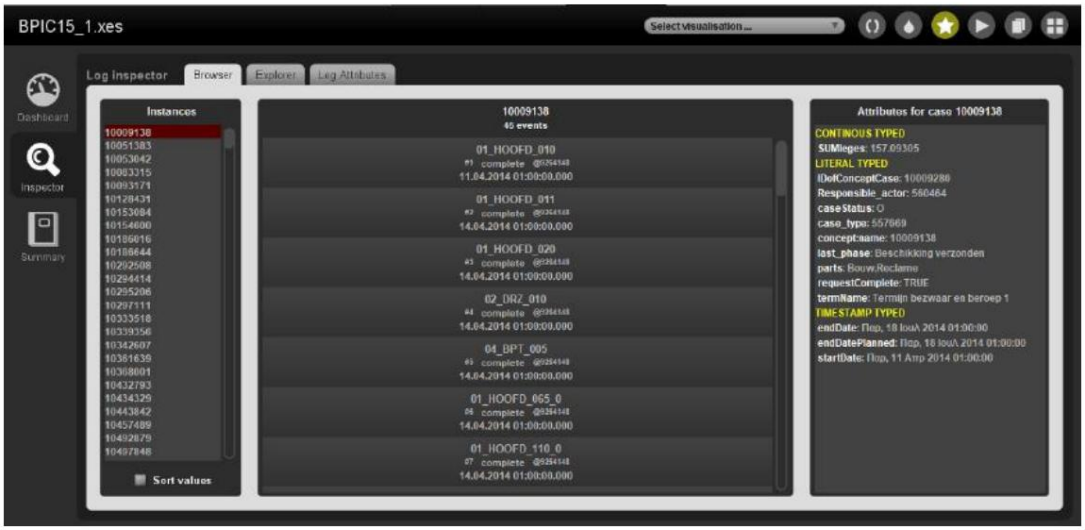


Figure 23: Browser tab in the inspector

The Browser helps us to see all traces (and their number) and their execution order of events included. The information for each event includes the name (code), the transaction status time stamp (inception the integration) and (timestamp, when it happened). While overall for each trace with the processes of it, we can see detailed information about it.

For example we can see:
CONTINUOUSLY TYPED:

ÿ SUMIleges (amount paid for license application)

LITERALLY TYPED:

ÿ IdofConceptCase (case id)
ÿ Responsible actor (responsible organization/employee)
ÿ caseStatus (case status open or closed)
ÿ caseType (a number the same for all cases in the file)
ÿ conceptname (process identifier)
ÿ last_phase (the last process in the file)
ÿ parts (description for the type of process, e.g. building, maintenance)
ÿ requestComplete (case completion or not)
ÿ termName (full name of the case)

TIMESTAMP TYPED:

ÿ endDate (completion date)
ÿ endDatePlanned (planned completion date)
ÿ startDate (start date)

Explorer

Figure 24: Explorer tab in the inspector

In the next tab we can study information about the length of each case and the events of each case. The length shows the amount of tasks performed, while they are colored based on the frequency of their appearance in the file. The more frequent the event the color is darker green, the rarer the event is the darker orange. Also, hovering over the event displays information about the case number, process name, status, employee, frequency of occurrence, and timestamp.

Figure 25: Explorer-elements tab for the trace

Log attributes

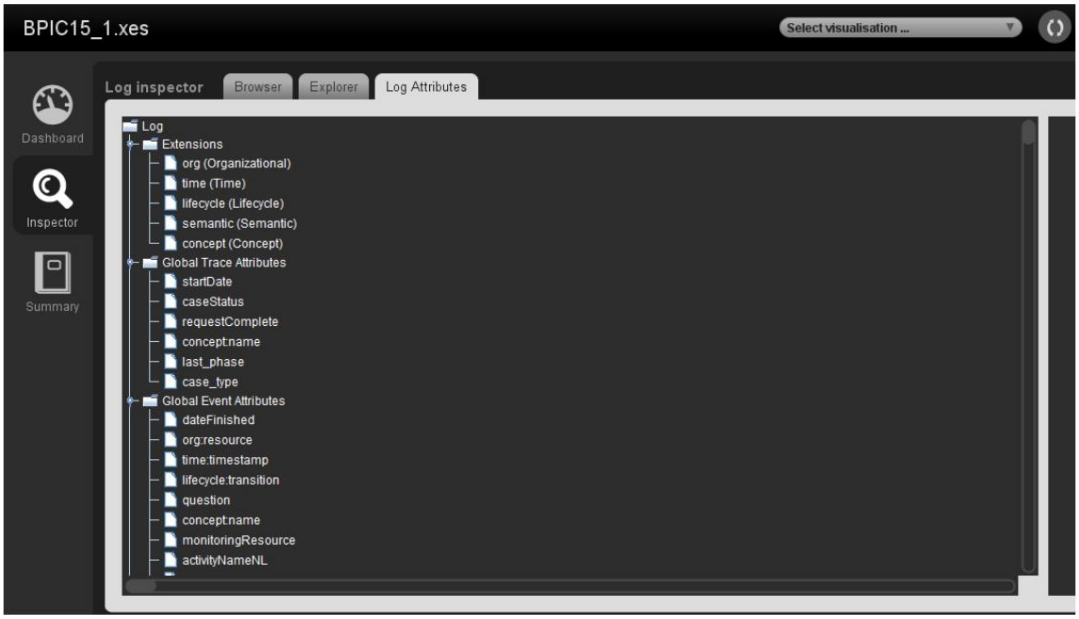


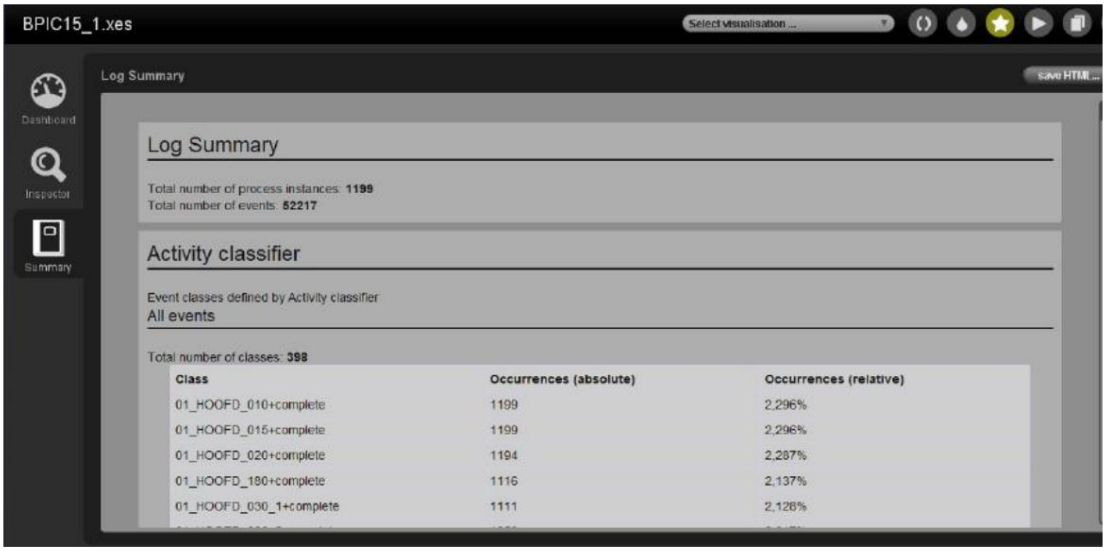
Figure 26: Log Attributes tab

The log attributes tab provides information on the file structure regarding the extensions that have been used, the attributes of both cases and events and the classifiers that have been used.

By observing the image we understand
ÿ the extensions that have been loaded in the
log file. ÿ the attributes of the cases
(trace) ÿ the attributes of the processes
(event) ÿ the separators (classifier) of the file, based on which the data is compared

A separator or classifier is used to describe the meaning of an action. For example in our files we have 4 classifiers one for the code of the action/action, one for the name of the action in English and another for the same name in the original language of the files in Dutch. The last classifier concerns the employee who carries out the work.

Summary



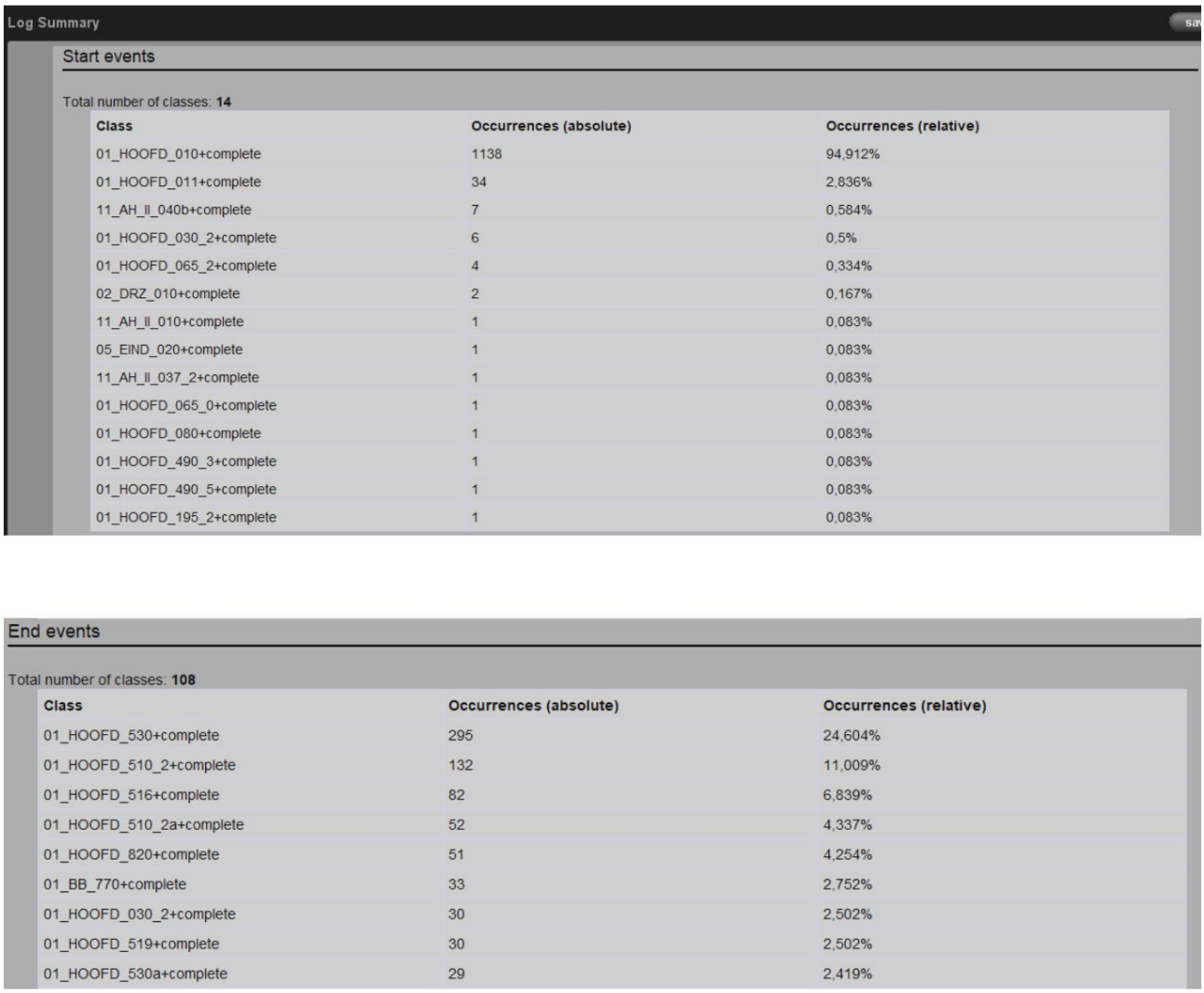


Figure 27: Log Summary tab

In the Summary section we can, as is easy to understand, see the overall picture of the event log file.

From this tab we derive a wealth of quantitative data for accuracy. We can observe: ÿ the total number of cases in the
event log ÿ the total number of actions

While with the use of the classifiers for each case we have: ÿ
for each event: its name, the number of occurrences and its percentage ÿ for
each process, the events with which it starts and specifically the name, the number of occurrences and its percentage ÿ for each process,
the events with which it is completed and specifically the name, the number of occurrences and the percentage of

It should be noted that in the actions, the column with their name also includes the status of the action. If the action name is followed by the phrase "+complete", this indicates that this process has started and has been completed normally. Otherwise, in the case that followed by the phrase “+start” indicates that it has started but not finished yet. This is a very useful tool, as we are able to see which processes in the file have not finished yet. Contributing to understanding the issue and focusing on vulnerabilities with we have to deal with every time, this possibility plays a major role in the analysis of the problem. By studying it we are able to draw some conclusions, which as we will see and later with filtering we can study them in depth.

Explore Event Log (Trace variants/Searchable/Sortable)(Log Enhancement)

One of the most basic data visualization techniques we have is using the explore event log tool. Using this we are able to see, classified by various criteria such as length or sequence, all the traces with the actions that include. It also shows the number of repeated traces with the same event, helping to understand how the actions behave. We can observe the patterns that form and with what frequency of occurrence. Finally for a file we are able to understand if it contains traces with many processes or not.[14]

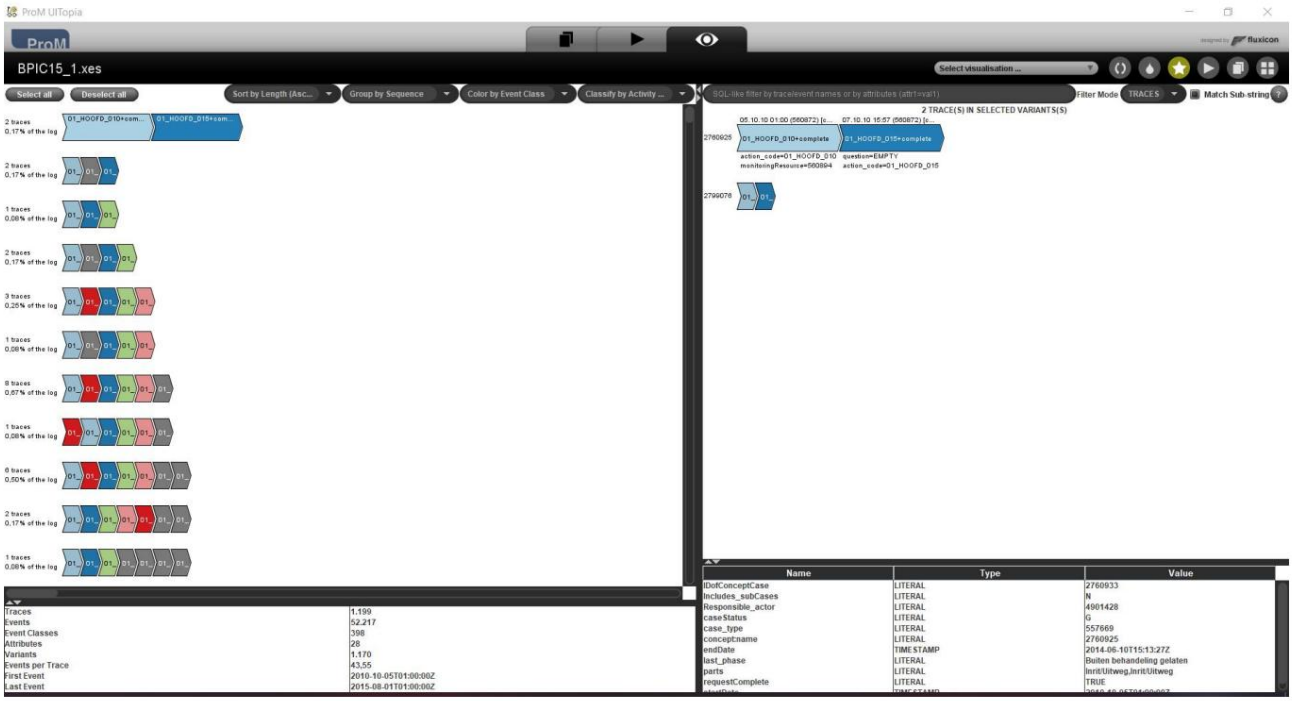


Figure 28: Explore Event Log diagram

After classification we see the traces with a small number of processes, where the biggest vulnerabilities in the systems appear and should be further analyzed. By observing them we can see what patterns form between them and how they behave, to discover unexpected behaviors. By filtering appropriately we are able to comment on their status and how they create delays (bottleneck) in other processes.

Figure 29: Case breakdowns

Selecting a pattern of traces shows us information about their parameters, which traces consist of the same tasks as well as data about the nature of the tasks themselves.

Traces	1.199
Events	52.217
Event Classes	398
Attributes	28
Variants	1.170
Events per Trace	43,55
First Event	2010-10-05T01:00:00Z
Last Event	2015-08-01T01:00:00Z

Figure 30: Information about the file

The bottom part of the tab includes data about the nature of the file, such as: ÿ the number of tracks ÿ the number
of actions ÿ the number
of different categories for the
actions that exist ÿ the parameters of the file for both traces and actions ÿ the
number of patterns detected ÿ the average of actions per trace, useful
to see if we have dense traces in the file ÿ
the date of the first event ÿ the date of the last event

Dotted Chart

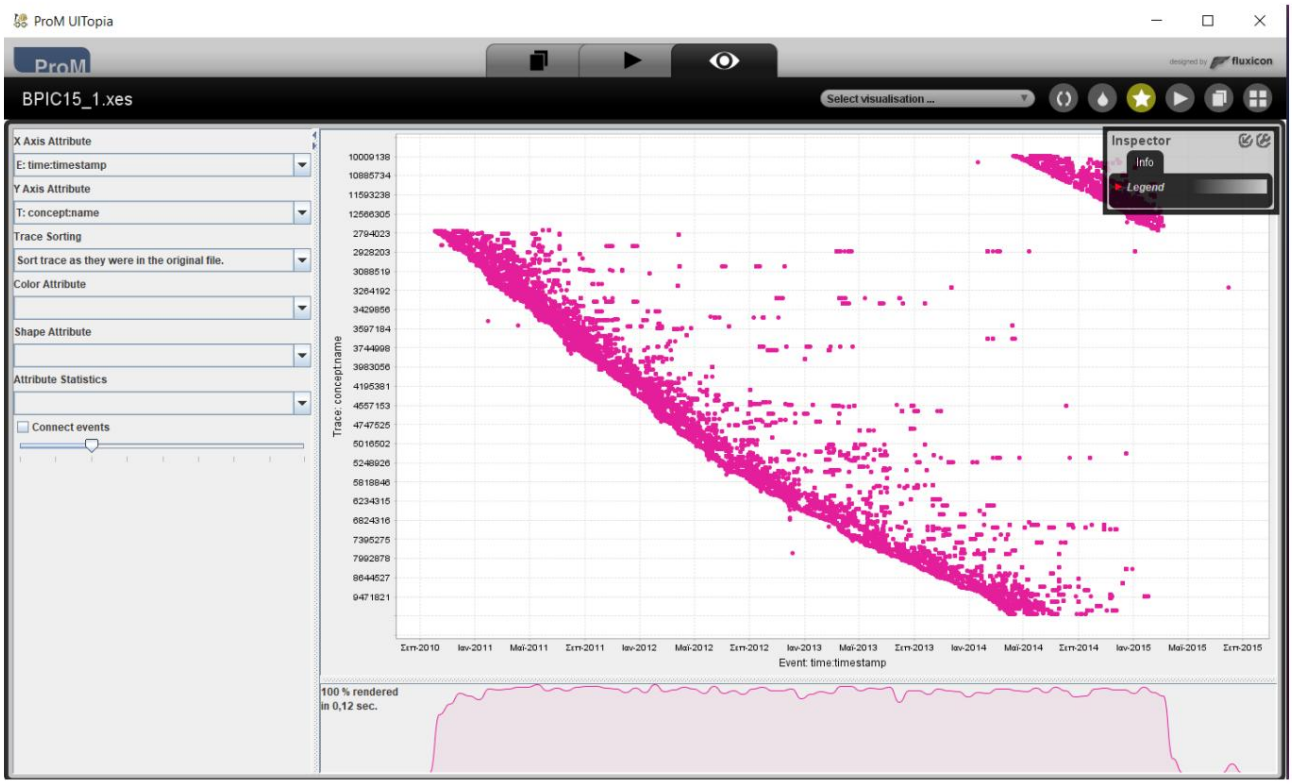


Figure 31: Dotted Chart Environment

With this representation we are able to show the correlation of cases and processes in relation to time. It is particularly useful for analysis and exploration

of the event logs. The vertical and horizontal axis show various variables of our choice, where we can parameterize them depending on the research we are doing. We usually use the horizontal axis for time to show the temporal correlation of events. We still have the possibility, for the sake of better presentation of results, to edit the colors in the diagram and the way they appear.

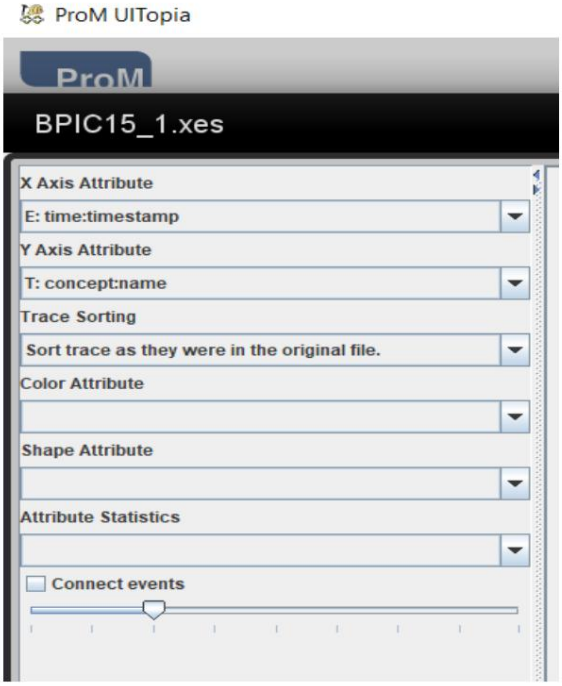


Figure 32: Dotted configuration Chart

Events are always shown as dots, while time shows the starting point of the events as well as the last event that was executed. The time waveform can show the moments where the actions were at their peak and at which points they had declined. Through this we see the total events and traces of · them, which we have found so far, at the time when they were executed. As a chart format it is able to provide a reliable look at the file and the events, as well as the patterns formed over time.

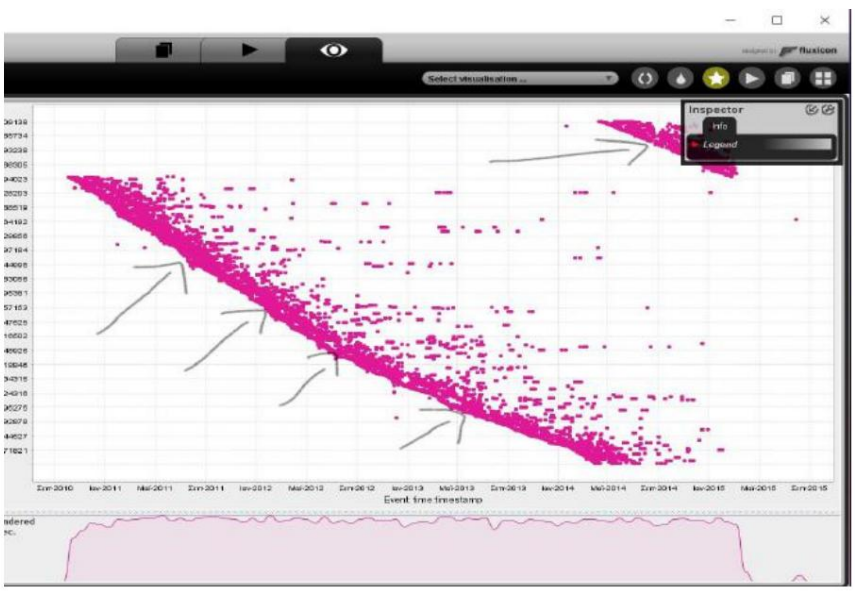


Figure 33: Example Dotted Chart

In the above marked points we see the moments of time where there was a lot of process execution, as we have dense sections of dots that represent processes. In the points where the dots are more sparse we have fewer processes, therefore less performance.

3.2.3 PROCESS MODEL

As we mentioned above more theoretically, there are 4 types of algorithms that we use to create the process model from each event log file. In our example we work with alpha, heuristic, inductive visual and fuzzy miners. Each one has a different utility and can benefit in various situations. Through their use, we are able to generate the process model, which shows the workflow in each file and see how well it agrees with the data we have. It is worth mentioning that we apply such an algorithm either to the original file input, or in the filtered file, to see how it behaves for a specific event or trace. In the continuation of the diploma we will deal with both cases. [14]

Alpha miner

We will start with the alpha miner as it is the first such algorithm that was made. However, the algorithm is not considered a very practical mining technique as it has problems with noise, rare/ missing behavior and complex paths that make the process difficult. The alpha algorithm it is simple and many of its features have been incorporated into more complex and powerful algorithms, which are quite recent. The alpha miner takes an event log as input and produces a Petri net. The first action alpha miner does is to scan the traces to sort the relationships between the events. While using the relationships to create a table of step distances. There are several variants of alpha miner in Prom Tools, namely alpha+, alpha++, alpha#, alphaR. The alpha algorithm recognizes four relationships between activities 1) “>” or otherwise direct, where it checks for each pair of activities if the A is directly followed by B in a trace (A > B)

2) “ÿ” or otherwise sequence, if A is directly followed by B, but B never immediately after A, we conclude that A is in sequence with B

3) “||” or parallel, although A is directly followed by B, and B is immediately after A, we conclude that A and B are parallel. In other words, both are executed

4) “#” or otherwise directly unrelated, i.e. when A is never followed by B and when B is never followed by A. Then the two events have no connection

ÿ Output :petri net
ÿ When to use: In cases where we want to get a first glimpse of the data, draw some early conclusions, and then apply specialized techniques. Alpha miner is simple from a scientific point of view but for real-world logs life, it is almost never the right choice. It will not work properly. It will give a result, but it will not be reliable.[15]

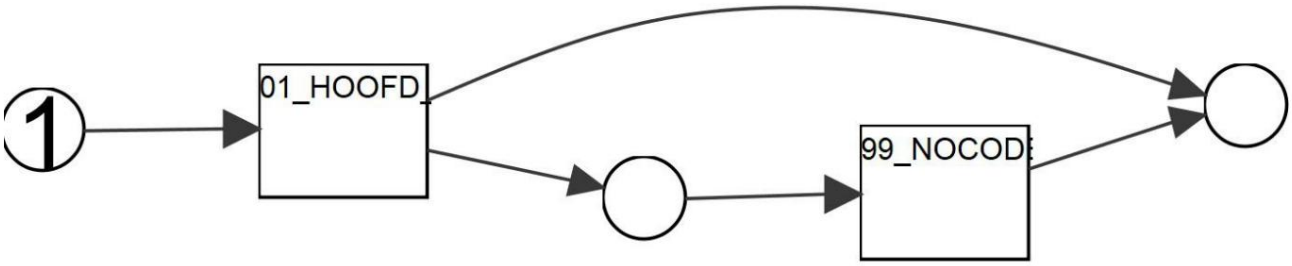


Figure 34: Alpha miner example

Heuristic miner

The heuristic miner is in practice an improved version of the alpha miner we mentioned above. It was created a little later and has the function of taking the frequencies of events and sequences, taking them into account when building the process model. In this way it can filter a noisy behavior or a rare behavior and be able to detect short paths, allowing single actions to be skipped. Unlike alpha where it struggled in cases of noise or rare behavior. The thinking behind this is to cut out rare paths from the activity graph .

The heuristic recognizes four types of relationships, like alpha, but also incorporates the frequency of occurrence of each trace.

ÿOutput:Petrinet
ÿ When to use: When you have real data with not too many different events or when you need a Petri net model for further analysis in ProM [16]

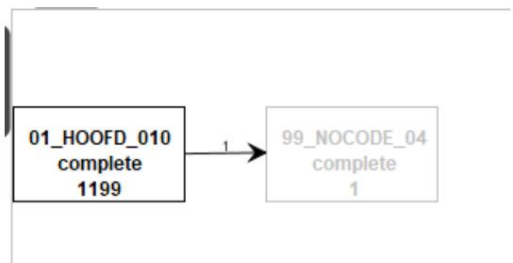


Figure 35: Heuristic miner example

Inductive Visual miner

As far as the inductive algorithm is concerned, we are dealing with the most efficient of the four, perhaps the best choice we have, when we want to create the process model from an event log. It largely applies the divide and conquer technique), where it splits the file, builds parts of the overall process model and processes each one separately. This gives it speed and flexibility in testing, which the others lack. Of course it always depends on the end goal we want to achieve but in general lines offers huge advantages such as: ÿ the model is always able to accurately represent the processes of the event log ÿ it has the ability to handle cases of noise, rare behavior and
cases with low occurrence patterns ÿ it is quite configurable and allows the user to several experiments depending
on its purposes ÿ it offers a live presentation of the model in real time, giving the opportunity to observe exactly what is happening and
at what time ÿ the output it produces, usually a file, can be easily edited and filtered without any conversion, giving high file
portability

ÿOutput:Petrinet
ÿ When to use: In problems where we want to deal with noisy behavior from data. When we want to find the most obvious split in the event file, identify the operation and process each one separately.

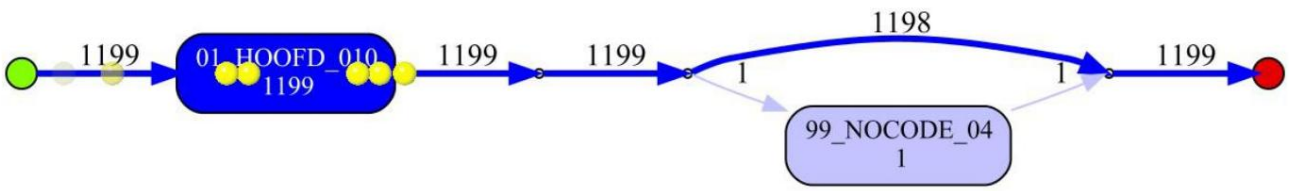


Figure 36: Example of Inductive Visual miner

Fuzzy miner

The fuzzy miner is a variation of the previous algorithms, as it does not work in the traditional way. Contrary to the rest of the process model creation algorithms, it does not manage petri nets but graphs. In this case, the creation

of the model, the importance of the processes and the relationships between them play an important role. Still in contrast to the other algorithms, the possibility of removing and editing the file elements is offered, with the result that the processes and their relationships are edited or removed in groups.

The importance of this algorithm arises from the need to have techniques for managing vague and rare events. All algorithms until then were lagging behind in this part, until the moment when Fuzyy miner was created. It gives the possibility of handling such situations with unstable data and solving these issues.[17]

ÿOutput:FuzzyModel ÿ
When to use: When you have complex and unstructured log data or when you want to simplify the model interactively

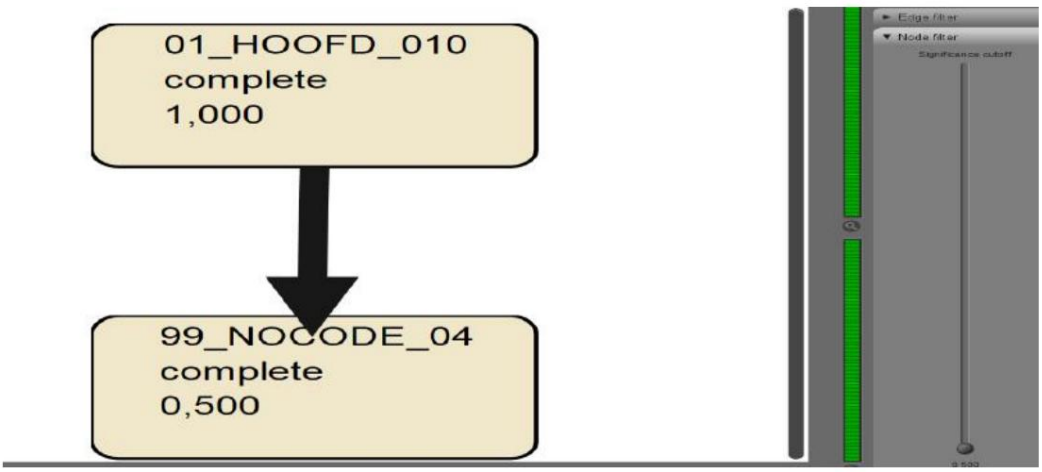


Figure 37: Fuzzy miner example

3.2.4 FILTERING EVENT LOG

Filtering the event log file is probably the most important step in process mining, as this is how we enter the practical part of this technique. The data that a file contains should come from different sources to cover a multitude of elements from different cases and patterns. In our example the records of the municipalities should include actions from many workers, different sectors, times and regions. With the main purpose to ensure a diversity of data, to be as representative as possible of the

most of the work. This is needed because there may be errors or even the measurements do not include a large number of samples. The main goal is to filter event logs that contain errors and need to be fixed. It is still useful to do in cases of unusual behavior, such as very large or small values or some other inhomogeneity that we have noticed and want to analyze.

With these criteria we usually apply filters to the event logs either to clean them of errors or invalid values, or to limit the data to a smaller scale, which we will later study its behavior, patterns and process model. the Prom Tools platform in most cases, we filter to remove some traces, remove or add some events and edit or customize some of them.

Then the next step would be to filter the input file with one of the available plug-ins that Prom Tools has. Some of them are filter log by attributes, filter events based on attribute values and filter log on events attributes values. any of the above reliable plug ins we can edit the file based on a particularity (or inhomogeneity) that we encountered before, such as the most common or rare activity, which municipality performs the most tasks or which the least. It should be noted that in the filter log by attributes there is also the trace length option to see which and how many cases in each municipality consist of several activities. In the input option we place the input file, in the actions tab the desired plug in and finally the output tab contains the final result.[18] So in the Actions tab we have the choice of input file and desired plug-in Actions

Figure 38: Actions tab

So with this technique, the results appear only for the issue we want to search for and we can see individually how it behaves. For example, we can search for a specific event or any attribute we are interested in.

Then after we have the filtered file with our own criterion, we can apply again a miner that has Prom Tools. In diplomacy we deal more with Alpha, Heuristic, Inductive Visual and Fuzzy miner to produce the corresponding process models and see which one of them applies better and agrees more with the data set we have. We can finally apply some visualization technique to the filtered output file, such as explore event log or dotted chart, to observe the new elements.

CHAPTER 4

4.1 STUDY OF DIPLOMACY

In the previous parts, we mentioned that the number of tasks that take place every day is enormous. Both at the individual level and at the collective level, actions are continuously and unceasingly carried out, whether they have to do with professional obligations or personal interests of each one. The processes occur throughout the duration of our life and everything we do is considered an action. From waking up in the morning to the actions we do in the workplace, all of these are processes. As you can see, every person consists of hundreds of tasks in a single day.

Especially in the work sector that interests us the most, there are many tasks and each of them carries a lot of data, which must be managed and analyzed. A business or an organization must know the processes that run daily, be aware of their performance and the ways they can contribute to its improvement. Especially in the branch of local government that we dealt with in diplomacy, an effort must be made to improve the work and therefore the services offered to the citizens of the country. There are many data files, with elements that collected from municipalities with activity tagged data and timestamp information that can be used as event logs. Using process mining techniques, we can discover the process model, analyze the tasks and how the people performing them work. Analyzing this data, we can draw conclusions about them. We can find out for each activity how many times it is performed and by which employees, if it is performed often or rarely. We still find the municipalities that perform the most tasks at a collective level and which municipalities have the most efficient employees. On the other hand, in cases of low performance, we are able to find where there is a performance bottleneck and with the appropriate tools to deal with it. We have the ability to create patterns, based on the conclusions we draw, so that we can to anticipate sensitive points before they happen and to be ahead of developments and not behind them.

The tasks that are performed every day have many parameters and vary from process to process. In many cases we have to deal with thousands of tasks in many different domains. They consist of several properties, each describing a different domain, while in many examples we are dealing with tasks that they are constantly changing and we should be able to study them.

In the study we used incident logs from 5 Dutch municipalities between 2010 and 2015. The records support the XES standard we mentioned above and were collected by the municipalities and by the competent bodies. The incidents concern processes/tasks performed in depth 5 years from the municipalities and present at a daily level the actions that took place. Each task had different characteristics, presenting various processes in the file such as tasks

construction or maintenance of premises and facilities, administrative work and work to serve the needs that existed. In these cases we see the different behavior that exists in each process and municipality, as well as the patterns that develop throughout the 5-year period.

The purpose of this thesis is to use process mining techniques to analyze features and information that can be revealed from the event log. We aim to study and describe the processes that are performed, to investigate the behaviors and patterns we encounter to identify pathologies or inhomogeneities in the data and with the use of the appropriate tools to resolve them.

We will use ProM Tools as it is very powerful open source software, it is extensible and used in research work. We will discover the processes with various algorithms and compare the results. Our aim is, through the research, to provide answers to major questions that concern the field of processes

The steps we will need to follow in each event log to achieve our goal are as follows:

ÿ Import event logs

ÿ Event log visualization

ÿ Discover process models

ÿ Filter event logs

4.1.1 IMPORTING EVENT RECORDS

The first step is to import, through the import option, the files of the municipalities that we will have to study and edit. These are files with xml format. The files appear with names ÿ DATA1.xml (1st municipality) ÿ DATA2.xml (2nd municipality) ÿ DATA3.xml (3rd municipality) ÿ DATA4.xml (4th municipality) ÿ DATA5.xml (5th municipality) Below are the dashboard/log visualizer tables that we have mentioned in the first place.

1st municipality

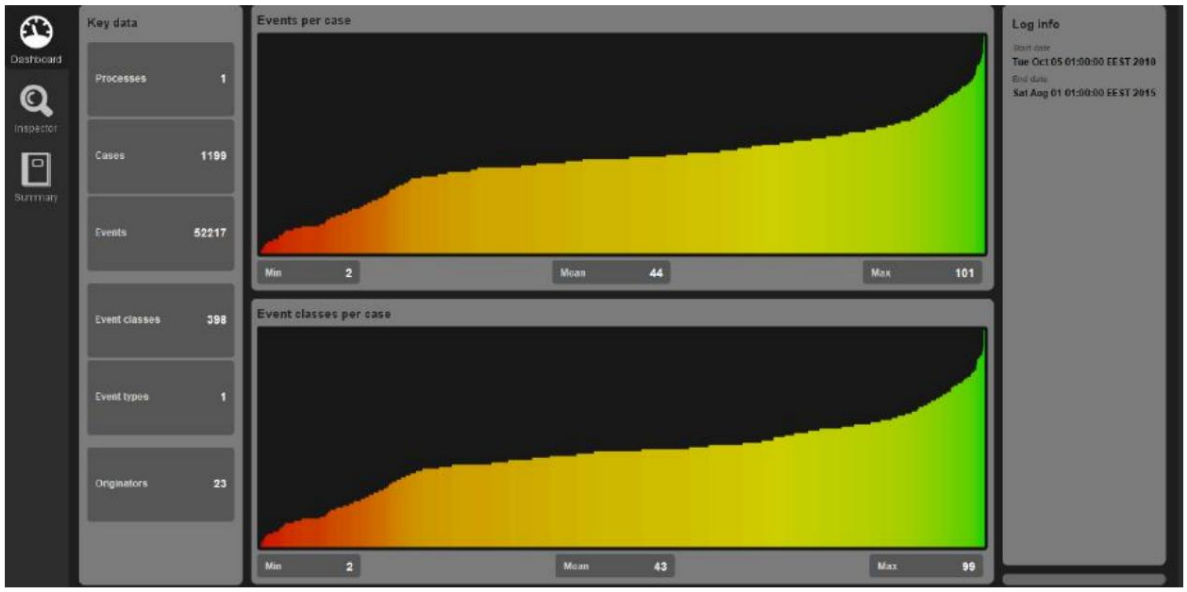


Figure 39: Dashboard of the 1st municipality

From the above image we get the following information:

Processes 1
Cases 1199
Events 52217
Event classes 398
Event types 1
Originators 23
Start date 05/10/2010 01:00:00
End date 01/08/2015 01:00:00

2nd municipality

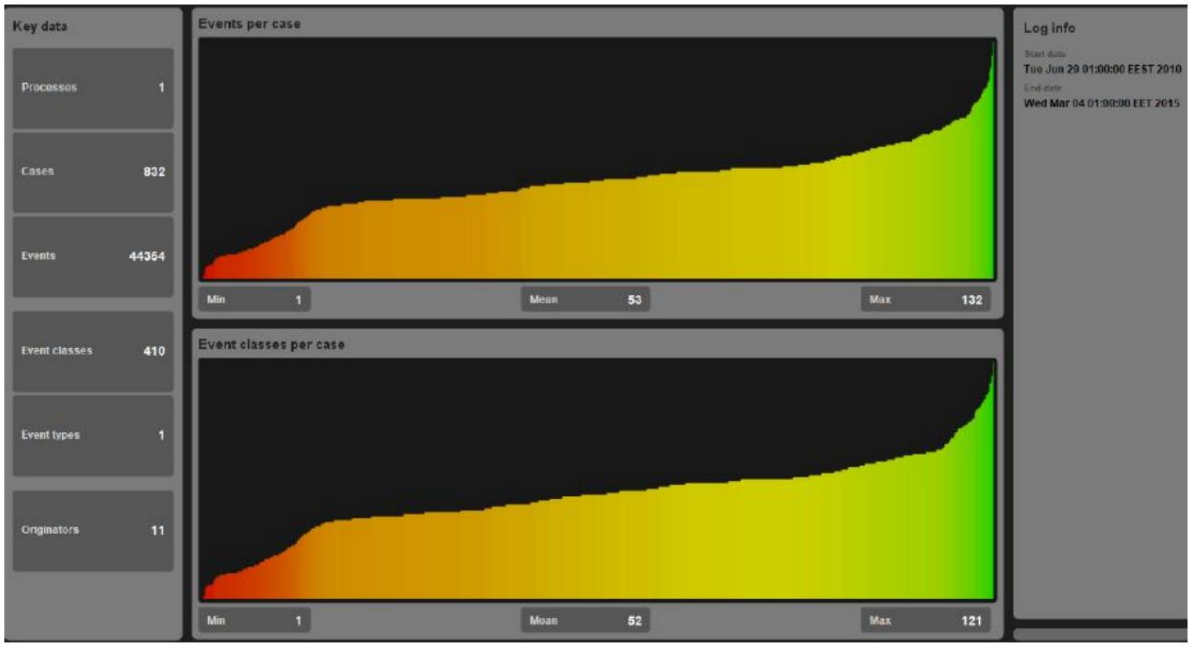


Figure 40: Dashboard of the 2nd municipality

From the above image we get the following information:

Processes 1
Cases 832
Events 44354
Event classes 410
Event types 1
Originators 11
Start date 29/06/2010 01:00:00
End date 04/03/2015 01:00:00

3rd municipality

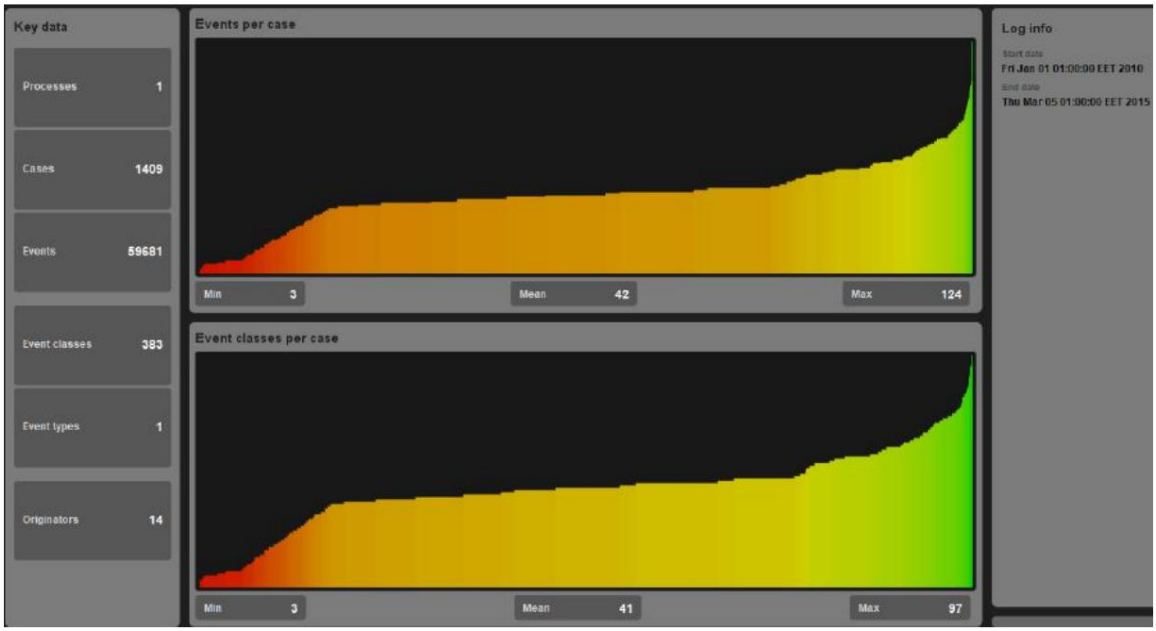


Figure 41: Dashboard of the 3rd municipality

From the above image we get the following information:

Processes 1
Cases 1409
Events 59681
Event classes 383
Event types 1
Originators 14
Start date 01/06/2010 01:00:00
End date 05/03/2015 01:00:00

4th municipality

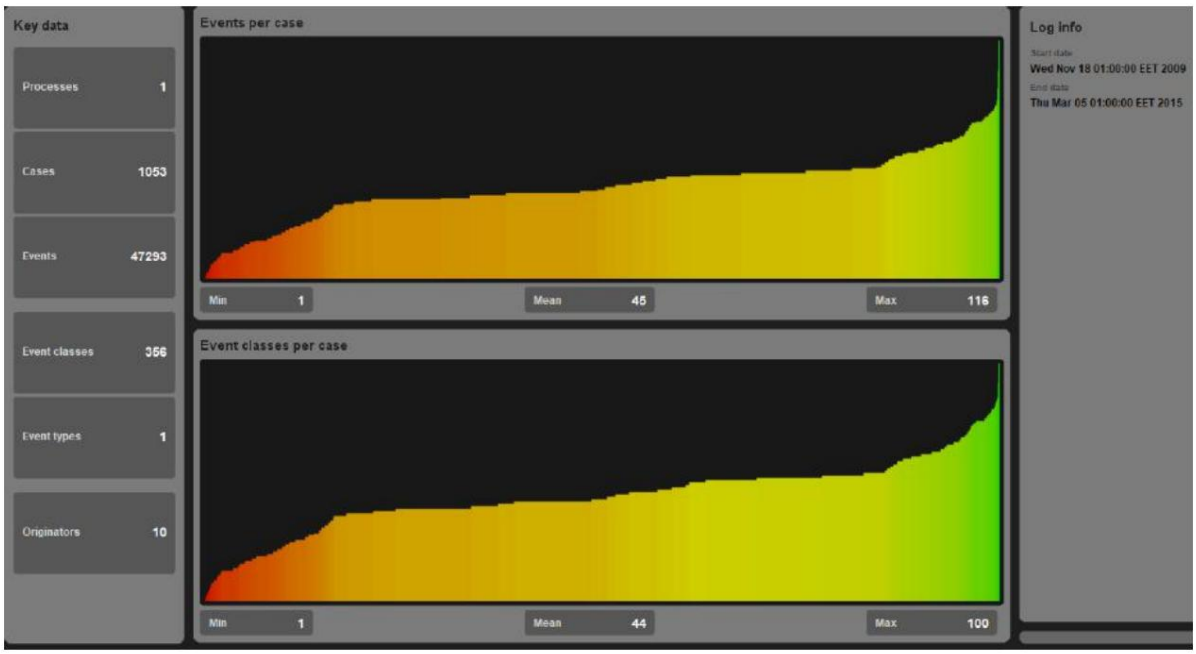


Figure 42: Dashboard of the 4th municipality

From the above image we get the following information:

Processes 1
Cases 1053
Events 47293
Event classes 356
Event types 1
Originators 10
Start date 18/11/2009 01:00:00
End date 05/03/2015 01:00:00

5th municipality

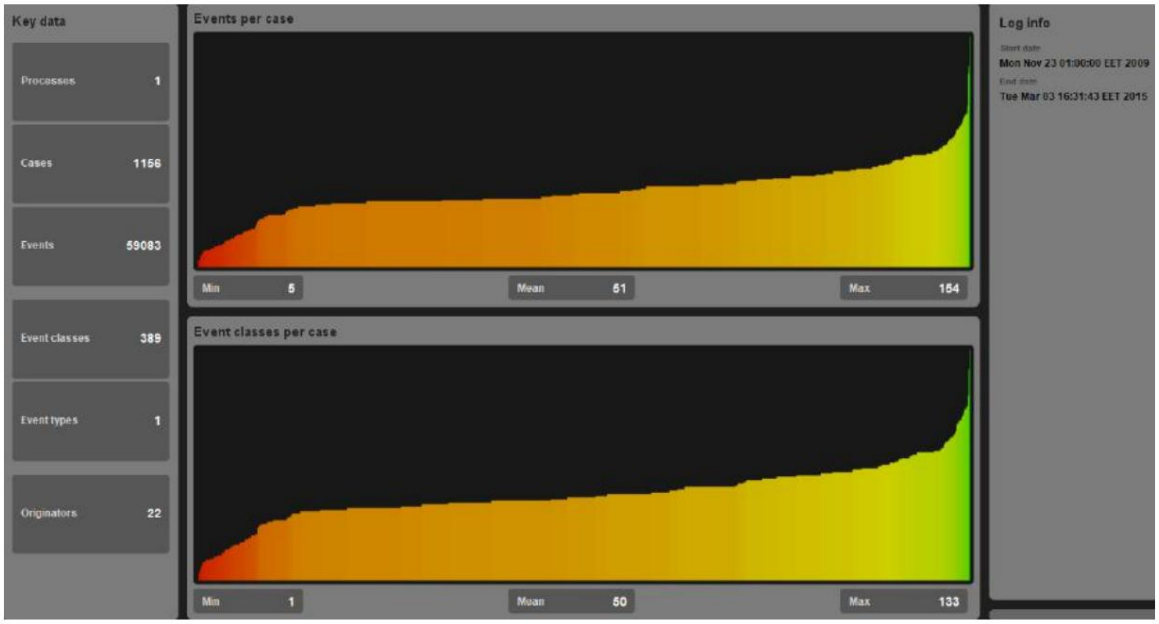


Figure 43: Dashboard of the 5th municipality

From the above image we get the following information:

Processes 1
Cases 1156
Events 59083
Event classes 389
Event types 1
Originators 22
Start date 23/11/2009 01:00:00
End date 03/03/2015 01:00:00

From the measurements we draw some conclusions about the behavior of the municipalities and the processes that are carried out all this time. More specifically we have the 1st, 3rd and 5th municipality with several additional procedures over a period of 5 years. All three municipalities are above of 50,000 procedures and their cases exceed 1,000.0 1st and 5th municipalities have several employees, while the 3rd has fewer, which means that it performs several procedures with fewer employees. The 2nd and 4th municipalities have even fewer, respectively with the fewer procedures they perform. Regarding the start and end dates of processes in the event logs, the following table summarizes them the data.

MUNICIPALITY	START DATE	END DATE
1st	5/10/2010	1/08/2015

2nd	29/06/2010	03/04/2015
3rd	1/01/2010	5/03/2015
4th	18/11/2009	5/03/2015
5th	23/11/2009	3/03/2015

We find that all data files end in 2015, specifically 4 of the 5 in March, while the last one in August of the same year. We still see that the 4th and 5th municipality start recording in the archives earlier, specifically end of 2009, unlike the rest starting in 2010. Especially the 1st municipality starts almost 1 year after the first ones, the recording of the log files and the 2nd municipality about 7 months after. Impressive if we consider the high performance and productivity of the 2nd municipality, in in relation to the period of time he had at his disposal.

Following the previous measurements, I also calculated an average number of processes per day for each municipality, knowing the number of tasks and the days available to do them perform.

We arrive at the statistics: 1st municipality ÿ 7.25 events per day 2nd municipality ÿ 9.22 events per day 3rd municipality ÿ 7.80 events per day 4th municipality ÿ 7.42 events per day 5th municipality ÿ 9.06 events per day

Concluding that the 2nd and 5th municipalities had the highest performance, as the results show that more than 9 tasks were performed daily, in contrast to remaining 3 municipalities ranging in 7 tasks per day.

Before we move on to measurements with Prom tools, because we will encounter many processes with their codes and they should be completely understandable, I quote one table with the most popular processes with its name in English next to the code that have already. The purpose is to help us understand a little better the processes that take place and be able to manage them better, not just with a random code assigned to them.

Thus we have the following:

EVENT CODE	EVENT NAME
01_HOOFD_010	register submission date request
01_HOOFD_011	OLO messaging active
01_HOOFD_015	phase application received
01_HOOFD_020	send confirmation receipt
01_HOOFD_030_	enter send date acknowledgment
01_HOOFD_050	applicant is stakeholder

01_HOOFD_061	start WABO procedure
01_HOOFD_065	create procedure confirmation
01_HOOFD_040	forward to the competent authority
01_HOOFD_060	regular procedure without MER
01_HOOFD_065	send procedure confirmation
01_HOOFD_090	publish
01_HOOFD_101	registration date publication
05_EIND_010	terminate on request
16_LGSV_010	calculate provisional charges
06_VD_010	extend procedure term
01_HOOFD_540	suspend term
01_HOOFD_110_	create subcases completeness
01_HOOFD_110	treat subcases completeness
01_HOOFD_180	procedure change
15_NGV_010	request further information
01_HOOFD_195	phase application receptive
01_HOOFD_196	procedure change after completeness
01_HOOFD_200	send letter in progress
01_HOOFD_250_	create subcases content
01_HOOFD_250	treat subcases content
01_HOOFD_250_	completed subcases content
01_HOOFD_190	request complete
09_AH_I_010	article 34 WABO applies
01_HOOFD_370	assessment of content completed
01_HOOFD_265	phase advise known
01_HOOFD_380	grounds for refusal
13_CRD_010	coordination of application

As we can see above, processes with code HOOFD are the most common in the 5 municipalities, while many of them also exist in the 5 municipalities. For example, the municipalities have common processes code HOOFD

-send confirmation receipt with code 01_HOOFD_020
-procedure change with code 01_HOOFD_180
-enter senddate decision environmental permit with code 01_HOOFD_510
- register submission date request with code 01_HOOFD_010
-phase application received with code 01_HOOFD_015

And also process code AWB45.[19]

We still have to mention that the task codes have another one semantics.The first part of the code '01_HOOFD_xxx' indicates the domain in which the task is entered, while the trailing 'xx_xxxx_001' indicates the order in which the activities are performed, where the first digit often indicates a phase within one process. In our diploma example we have 29 different types of assignments, 500 different codes for tasks and 354 types of activities.

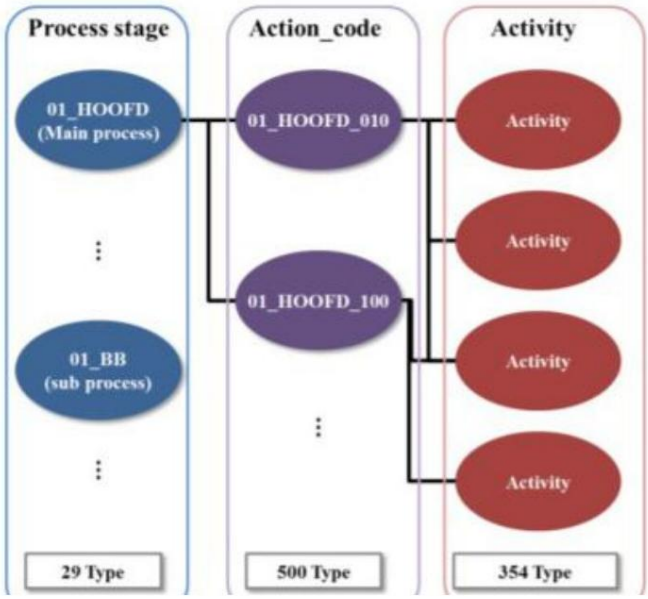


Figure 44: Analysis of process codes

4.1.2 EVENT LOG VISUALIZATION

INSPECTOR/EXPLORER

Then with the next Inspector window we can see the cases of each file, what processes they consist of and the attributes they have. Then by using the explorer we have the possibility to observe the cases and the number of processes included in each one. Still, as we mentioned above, for each process we can see information about it and based on its color, how often it appears.

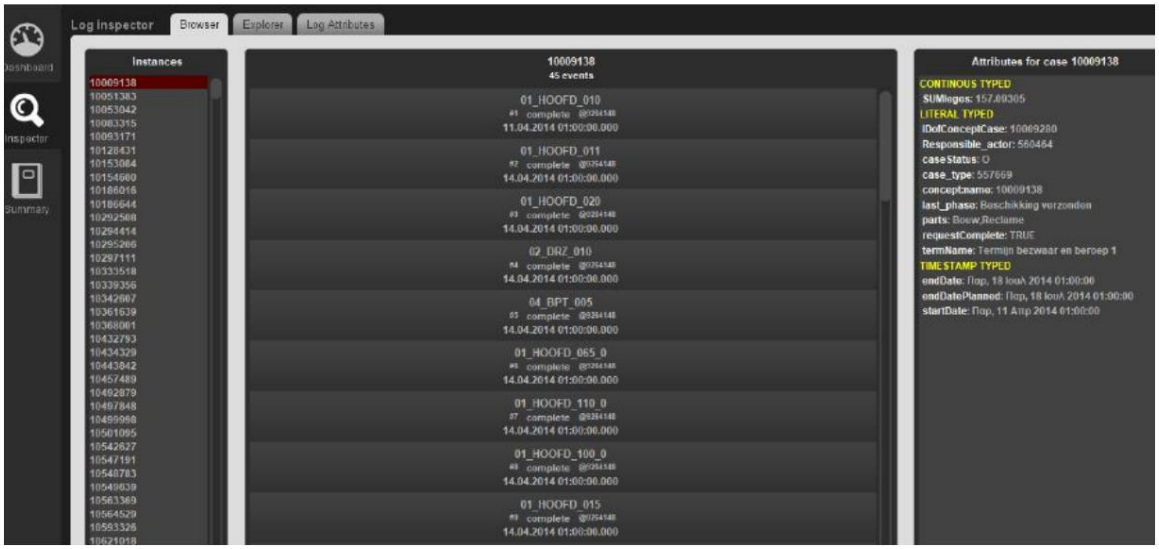


Figure 45: Browser tab of the inspector for the municipalities of the study

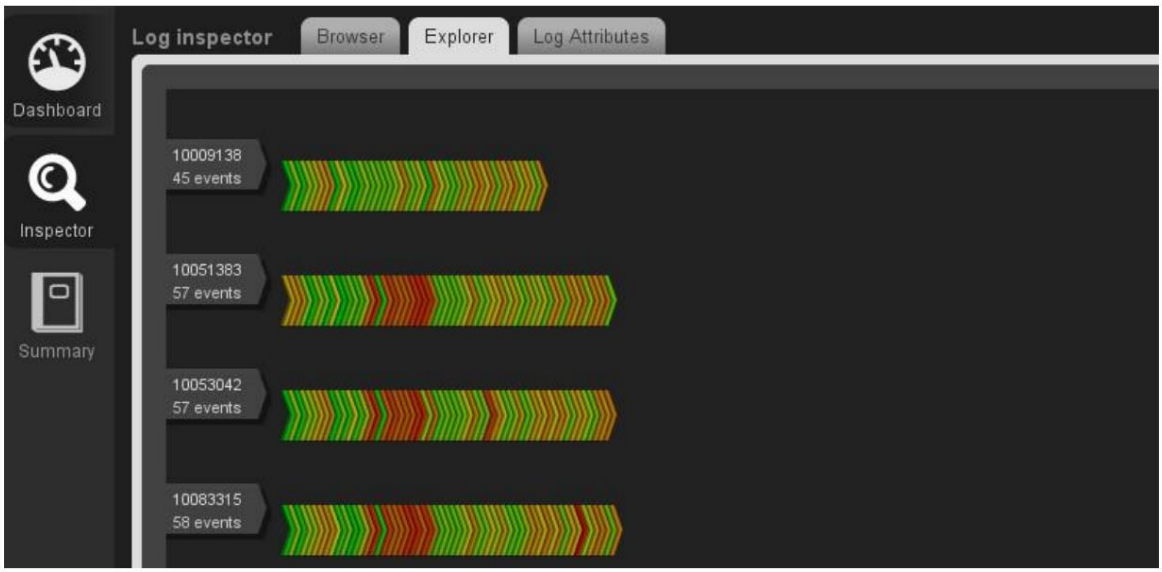


Figure 46: Explorer tab of the inspector for the municipalities of the study

In the above examples we observe all the traces with their procedures through the Inspector and on the right side the attributes of each trace. From the example we see the following attributes for the case with code 10009138 with 45 procedures, where the first is the action with code 01_HOOFD_010

CONTINUOUS TYPED:

ÿ SUMIleges (amount paid for license application)

LITERAL TYPED:

ÿ IdofConceptCase (case identifier) ÿ Responsible actor (responsible organization/employee) ÿ caseStatus (case status open or closed) ÿ caseType (a number that is the same for all cases in the file)

ÿ conceptname (process identifier) ÿ
last_phase (the last process in the file) ÿ parts
(description for the type of process e.g. build, maintain) ÿ
requestComplete (completion or not of the case) ÿ
termName (complete name of the case)

TIMESTAMP TYPED:

ÿ endDate (completion date) ÿ
endDatePlanned (planned completion date) ÿ startDate (start
date)

While then in the Explorer we see about the cases, the events they contain and how often they appear. For example in the aforementioned trace 10009138 we see actions with a light green to yellow color, which means that it consists of relatively popular events. On the contrary, the next trace 10051383 includes actions in red, i.e. with a small execution number.

LOG ATTRIBUTES

The next window gives us information about the extensions, global trace attributes, global event attributes and classifiers that have been used. It shows us the way the data in the files is structured.

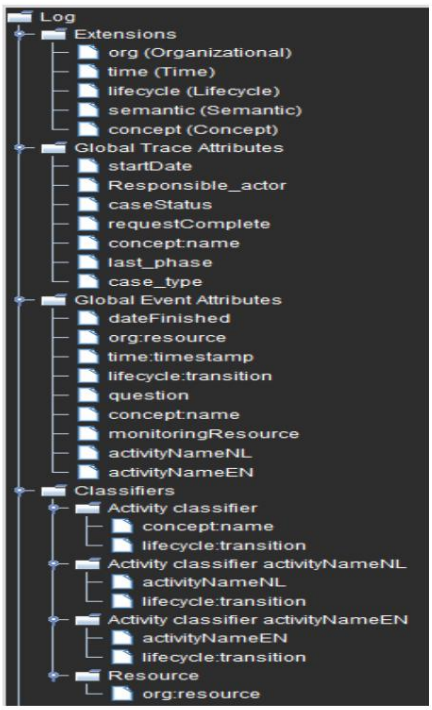


Figure 47: Log Attributes tab for the structure of the study files

ÿ Extensions for the structure, timing, life cycle, semantics and assumption of each trace and action present in the event log

ÿ characteristics for each track: start date, responsible employee, status case open or closed, completion or not of the case, process identifier and the last process in file
ÿattributes each for employee, process: completion date, organization timestamp, control status, name in completion,process id,responsible Dutch and English
ÿ Classifiers: based on the code of the action (name and status), based on the name at Dutch(name and status), based on name in English(name and status) and with based on the competent official.

Summary

From the Summary tab we can see the number for all activities their appearance in the event log.

1st file

Event classes defined by Activity classifier
All events

Total number of classes: **398**

Class	Occurrences (absolute)	Occurrences (relative)
01_HOOFD_010+complete	1199	2,296%
01_HOOFD_015+complete	1199	2,296%
01_HOOFD_020+complete	1194	2,287%
01_HOOFD_180+complete	1116	2,137%
01_HOOFD_030_1+complete	1111	2,128%

2nd file

Event classes defined by Activity classifier
All events

Total number of classes: **410**

Class	Occurrences (absolute)	Occurrences (relative)
01_HOOFD_010+complete	830	1,871%
01_HOOFD_015+complete	829	1,869%
01_HOOFD_020+complete	828	1,867%
01_HOOFD_180+complete	801	1,806%
01_HOOFD_030_1+complete	792	1,786%

3rd file

Activity classifier

Event classes defined by Activity classifier
All events

Total number of classes: **383**

Class	Occurrences (absolute)	Occurrences (relative)
01_HOOFD_010+complete	1409	2,361%
01_HOOFD_015+complete	1409	2,361%
01_HOOFD_020+complete	1408	2,359%

4th file

Activity classifier

Event classes defined by Activity classifier
All events

Total number of classes: **356**

Class	Occurrences (absolute)	Occurrences (relative)
01_HOOFD_010+complete	1052	2,224%
01_HOOFD_015+complete	1052	2,224%
01_HOOFD_020+complete	1051	2,222%

5th file

Activity classifier

Event classes defined by Activity classifier
All events

Total number of classes: **389**

Class	Occurrences (absolute)	Occurrences (relative)
01_HOOFD_180+complete	1161	1,965%
01_HOOFD_020+complete	1155	1,955%
01_HOOFD_015+complete	1155	1,955%
01_HOOFD_010+complete	1154	1,953%

Figure 48: Log Summary tab of the 5 municipalities

From the above photos we can see for each file, which was the most common task that was executed many times, the number of its appearances and what percentage of frequency it had.

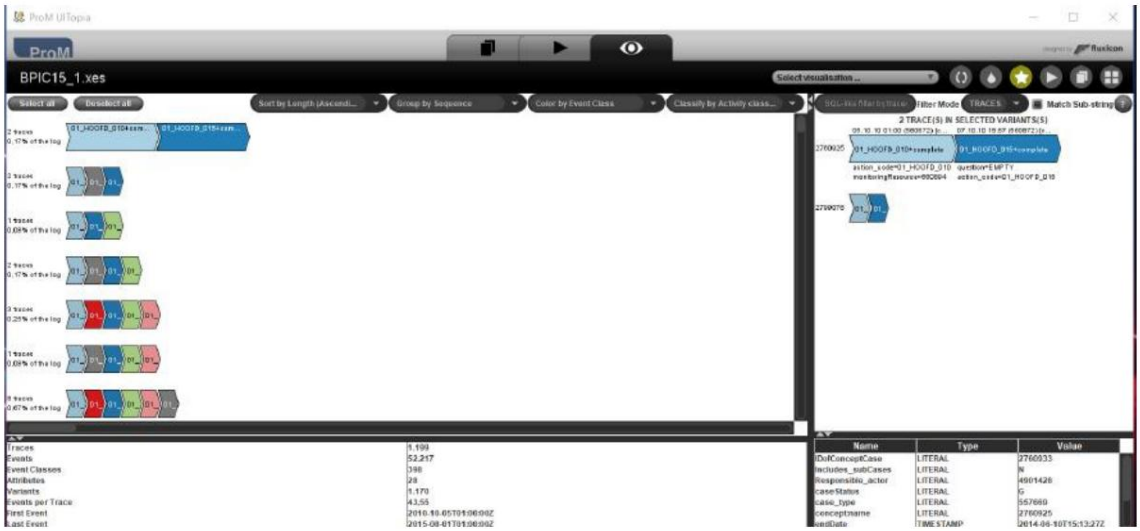
It is clear from the results that all the files have a common element, especially files 1,2,3 and 4, have as the most popular work, the same work in all 4 municipalities. Specifically, it is the work with code 01_HOOFD_010, which it is in a complete state, in the first place of appearances in the 4 municipalities. Even in the 5th municipality, for a few units, it is not in the first place, but you find it in the first four works. This makes us think about this work and about its behavior ,for this

later through filtering the files, we will deal more with it to see how it works and the patterns it develops. Accordingly, we can search for the rarest jobs or employees and study them. Even if we want through the status of the jobs listed next to every action, we have the possibility to check whether a process is completed or not.

Essentially, through the virtualization techniques we are referring to, we try to observe the data and draw some conclusions, which we will later filter in more detail. Likewise, process mining as a branch tries to find some pathogenicity or inhomogeneity in the data, where it can later be processed and to understand how it occurs. In the case that it is a problematic behavior, he will be able to improve it, increasing the efficiency index.

Explore Event Log (Trace variants/Searchable/Sortable)(Log Enhancement)

Also as we mentioned, a very useful visualization tool of Prom is the Explore event log which shows all the traces and the events contained in each of them. By sorting these elements, based on the length, we can see which and how many traces have the fewest events and which have the most. It is a measure that can show us in which cases we have the most efficient workers and in which the least efficient ones. 1st file



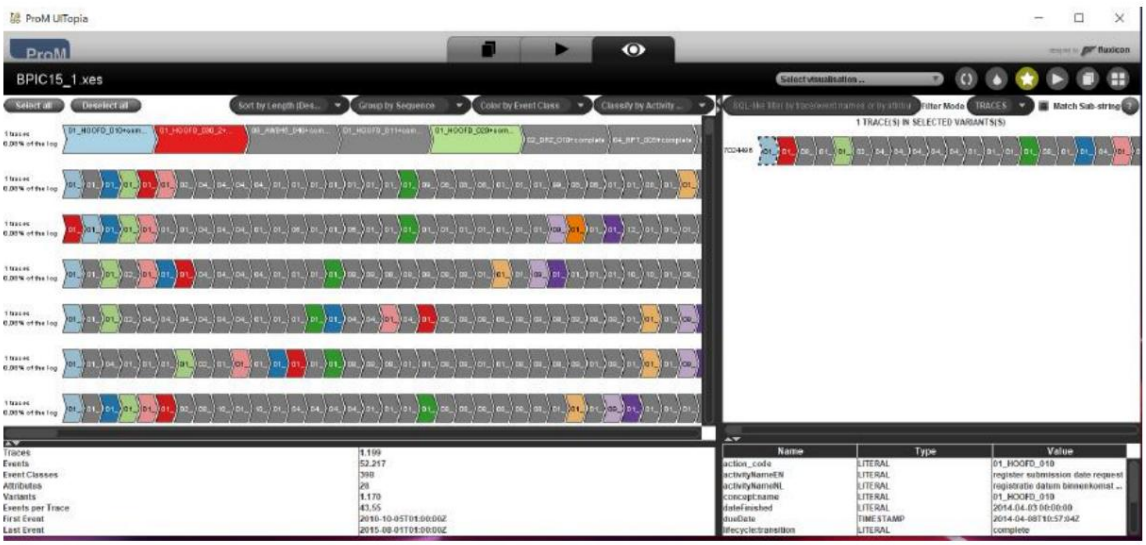
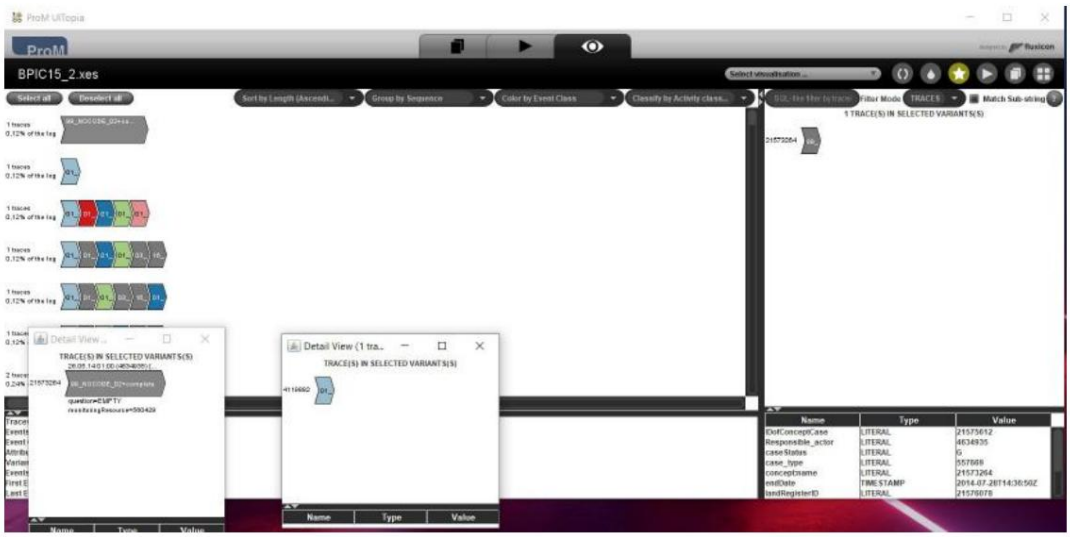


Figure 49: Explore Event Log of 1st municipality

We see that we have 2 cases with the smallest number of activities and to be precise 2 events each. Then we notice that the next traces are scaled up gradually, increasing the number of events by one unit.

2nd file



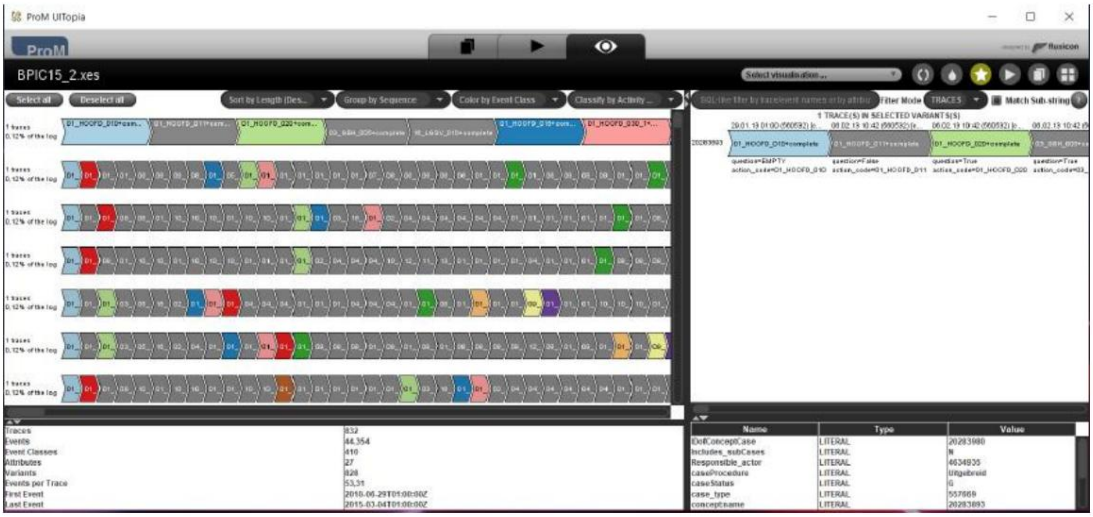


Figure 50: Explore Event Log of 2nd municipality

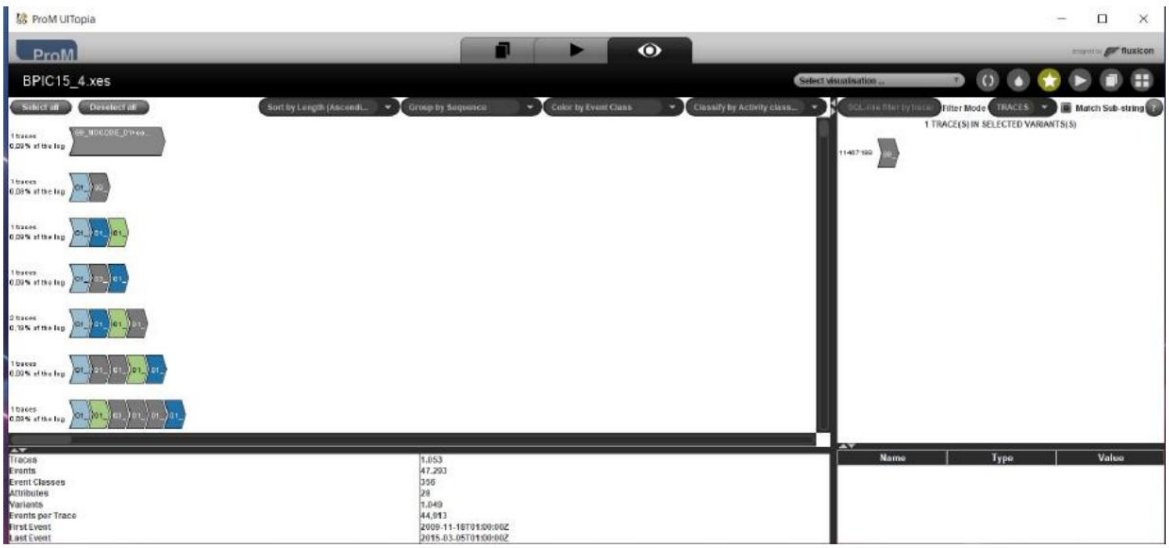
Here we realize that we have 2 traces with only 1 event implemented which makes them of less efficient traces and the 2nd municipality should focus there. Despite these immediately subsequent traces include 5 or more events, making us think about the problem it lies in the first 2 traces while then the situation smooths out.

3rd file

Figure 51: Explore Event Log of 3rd municipality

The 3rd municipality starts with better results as there are 4 traces with 3 events and all subsequent traces include 4 or more events.

4th file



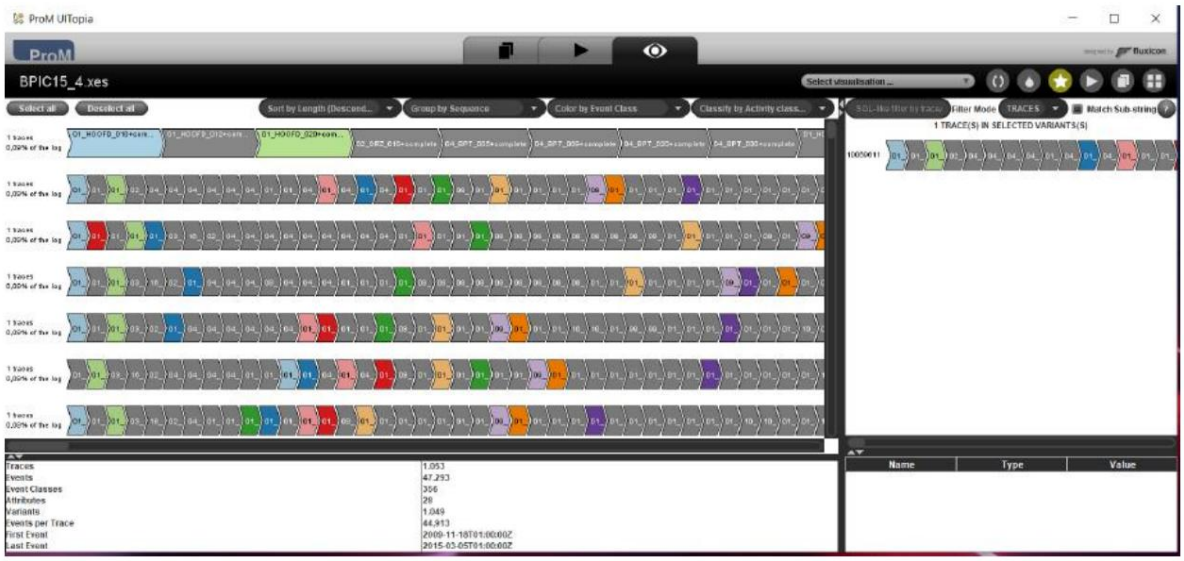


Figure 52: Explore Event Log of 4th municipality

In the 4th municipality we find a corresponding case with the 2nd municipality where we see a trace with only 1 event and then we notice a gradual increase in activities. But without catching them performance of the 2nd municipality where after the first worrying traces, the next ones had many activities, 5 and above specifically.

5th file

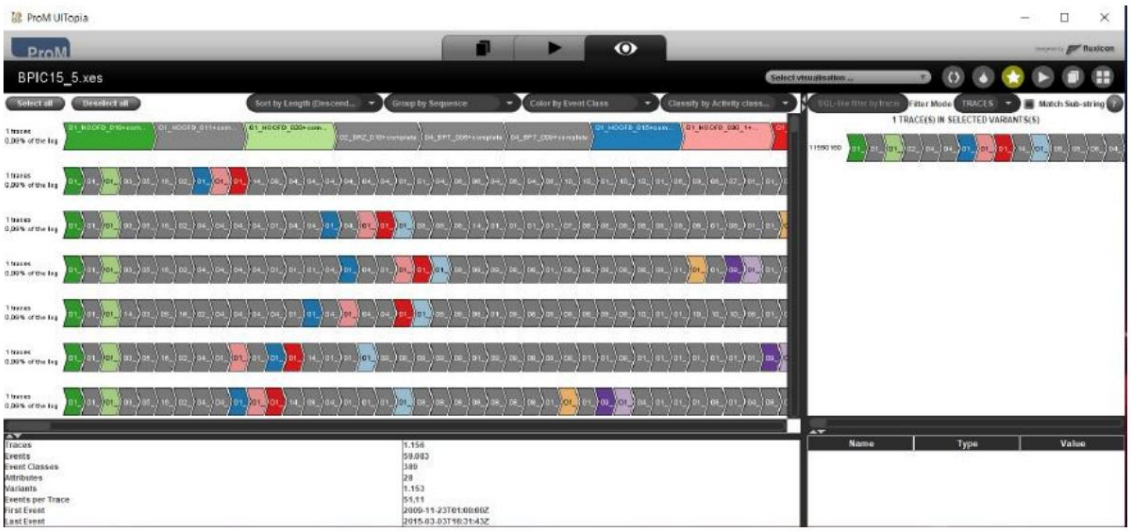


Figure 53: Explore Event Log of 5th municipality

In the 5th and last file we find perhaps the best performance from the 5 municipalities. In particular, the traces with the fewest events start from 5 events and increase quite a bit, in contrast to the rest of the municipalities where several traces had much less than 5 events each. Considering the multitude of events performed in the municipality, with almost 60,000 events, we notice that it has the best performance and production so far.

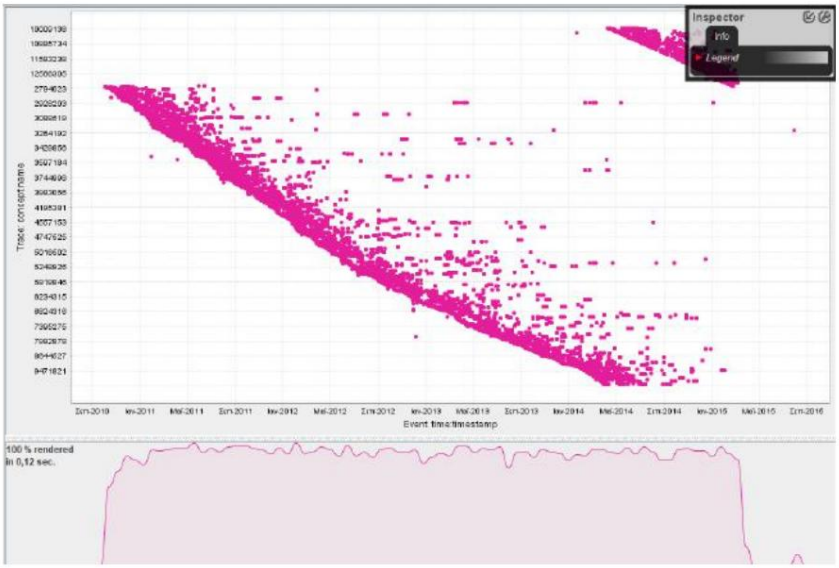
Then by observing the same plug in that we had above we can see for each file what is the average of activities per case. This helps our research for each file of the municipalities by seeing, with the average always, which municipality performed the most tasks .

So we observe for: 1st file 43.55 events per trace ÿ 53.31 for
2nd file ÿ 42.357 for 3rd
file ÿ 44.913 for 4th file ÿ
51.11 for 5th file

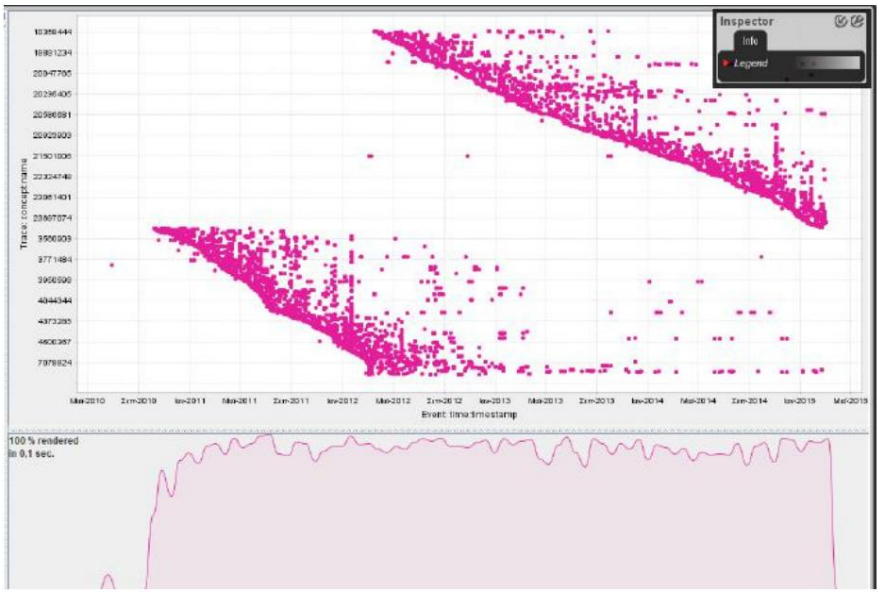
In this case, we see the fairly high average of the 2nd and 5th municipalities, as they exceed 50 events on average. Also pointing out that the 5th municipality has brought us remarkable results in many measurements. Then municipalities 1, 3, 4 have lower performance, as they range around 43 events average per trace. Not so good performance if we consider that the 1st and 3rd municipalities exceed 50,000 events. While on the other hand the 2nd municipality with only 44,354 events collected the highest average .[20]

Dotted Chart

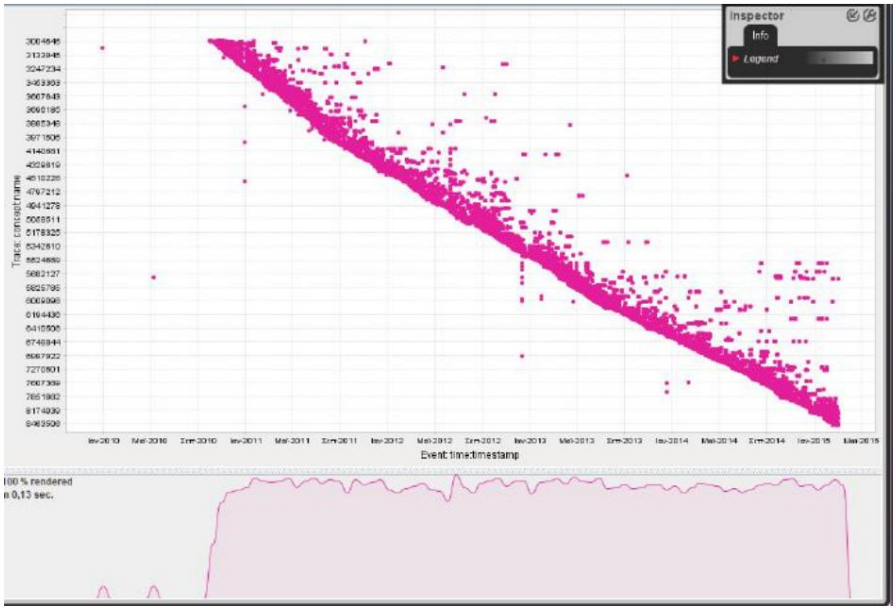
In addition, a useful virtualization technique offered by Prom Tools is the dotted chart. This technique shows the time marking of the events and in which periods we had high or low performance in the tasks. The denser the chart, the greater the performance. 1st file



2nd file



3rd file



4th file

5th file

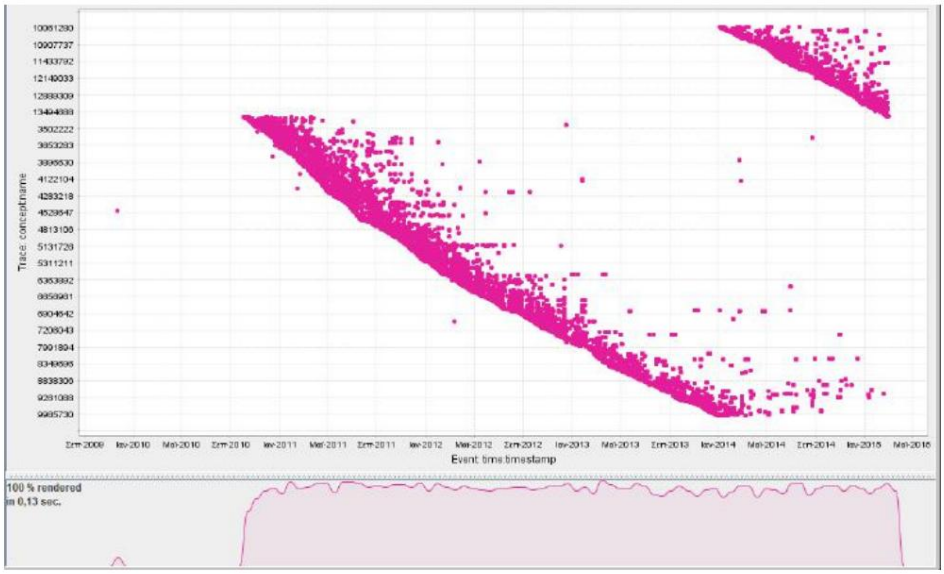


Figure 54: Dotted Chart of municipalities

The above dotted charts show the cases on the vertical axis and the time on the horizontal axis, showing the activities performed by each employee in relation to time. The measurements we see start for all files from the end of 2009 and thereafter in 2010 they increase. The exception is the 1st municipality which contains several works already from 2009. We see in files 1, 2 and 5 denser diagrams, which contain many more works compared to the rest of the municipalities. Still the waveforms under the diagrams show the numerous cases of municipalities 1, 2 and 5, which remain at a high index throughout the 4 years, in contrast to the rest of the municipalities that reach a peak of their performance and then decline.. They maintain a more uniform performance throughout the 4 years, keeping it at quite high levels. A noteworthy point seems to be that in all the records the peak of the work occurs around the first 6 months of 2013, where it seems to have been the most productive period for the municipalities. We still have to point out that in in all municipalities we have cases, which start their procedures later than expected, some after the middle of the 4th year. Through the above diagrams we can again observe the behavior of the actions and draw some conclusions about them, so that subsequently filter these processes and study them.

4.1.3 PROCESS MODEL

As we mentioned above in Prom Tools we have several plug-ins available to create the process models. With this technique we apply an algorithm we want to the file and see the produced process model. This way we are able to understand if the data we have so far agree from the measurements, with the representation of the task diagram. Applying these algorithms we pass

essentially from the theory we have seen so far, in practice where we can observe the way the processes work and how they interact with each other.

Many times in our tests, especially when dealing with a respectable amount of data we seek to first filter a file based on some criterion and then apply an algorithm to the resulting file. Usually on the initial input, the process model produced is hard to understand , because it contains enough information. For these reasons, we usually first filter a file and then use the desired algorithm. This way we see if the elements we got from the filtering are also evident in our model.

The plug-ins we have available for the mining algorithms are the following: ÿ
Alpha miner for the alpha algorithm **ÿ**
Mine fr a heuristics net using heuristics miner for the heuristic algorithm
ÿ **Mine with inductive visual miner** for the inductive algorithm
ÿ **Mine for a fuzzy model** for the fuzzy algorithm
Alpha

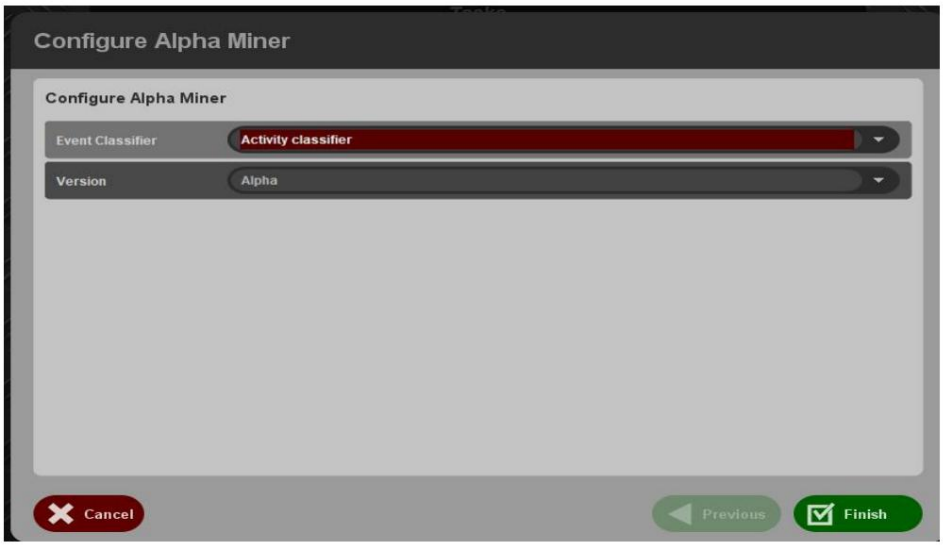


Figure 55: Alpha miner plug-in

Alpha miner is not a very good algorithm for process mining, but some are later modern algorithms, so we will use it. it gives us the activity and as a version Alpha. Prom has other variants like (alpha+, alpha++, alpha#, alphaR) depending on what we want and the resulting process model is presented as a petri net.

Heuristic

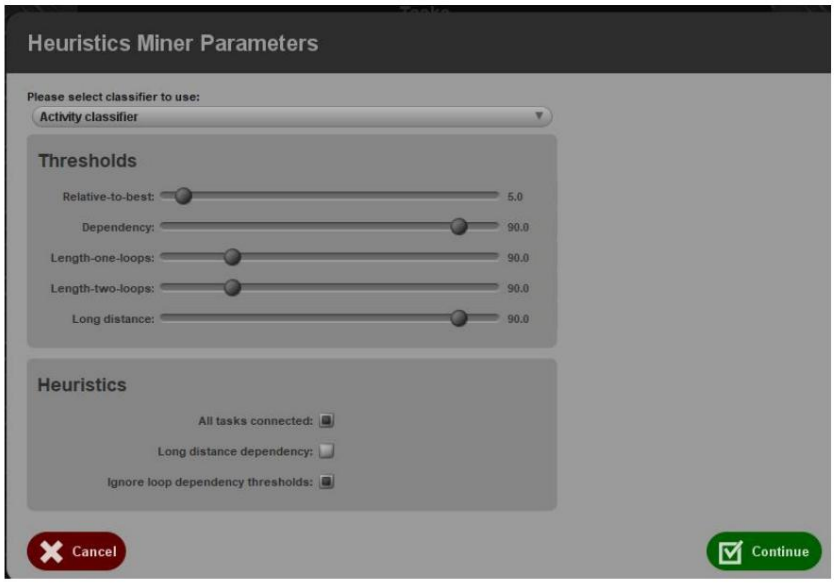
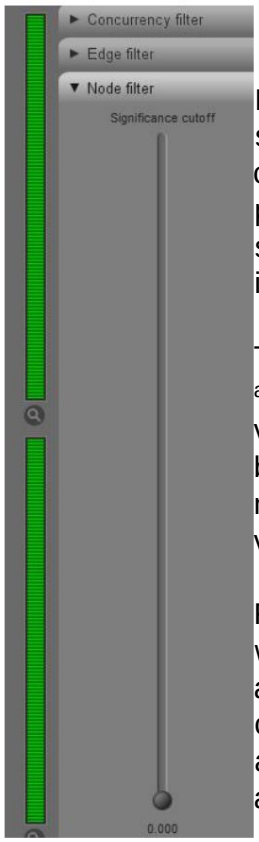


Figure 56: Heuristic miner plug-in

As we discussed earlier the heuristic miner takes frequencies into account and can thus filter out rare behavior. It is also able to detect loops and skips. We have several parameters that we can control and most notably the "Thresholds" parameter "Relative to the best" which indicates that only events that are observed more times than the certain threshold value will be considered. Specifically only events that have been observed more than 5 times will be included in the generated model. We still check the threshold values for the length of loops and distances between processes and whether the events will be connected or omitted from the model. Finally as before we can to choose the desired separator, depending on the test we are doing.

Fuzzy miner



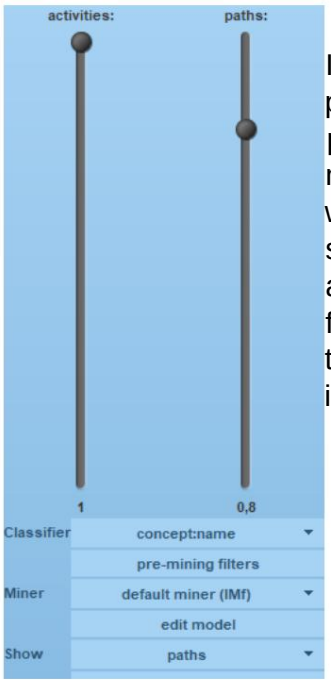
In the case of Fuzzy miner we have 3 types of filters, the node filter sets the number of different processes that will be presented in the chart.We can check the cutoff percentage and so every process with an occurrence rate lower than the threshold that set, it will be removed from the model. Otherwise it will be included in the generated graph.

The edge filter the amount of edges, affects of and the option which will be included in the final graph. The given in the value affects the percentage of edges that will be included in the chart. The higher this value the more edges will be included. While correspondingly the smaller n value the fewer edges in the graph, maintaining a proportional relationship.

Finally the concurrency filter has its usefulness in the synchronization that we want to apply. It can filter the conflicts that arise between two or more activities. With the term collision refers to the situation where 2 nodes (or activities) are connected bidirectionally to each other, which can cause loops and dead ends in our model.

Figure 57: Fuzzy miner plug-in

Inductive visual miner



In the inductive visual miner we can configure various parameters prices, such as the correlation between prices for activities and paths that will appear. By increasing these values we place more and more activities in the diagram, while decreasing them we make the model simpler separator we will use and which element (attribute) will appears. We can also add some extras filters to mine the model and process it, after the user. Finally, we check if it will display additionally information, except for activities and paths.[21]

Figure 58: Inductive Visual miner plug-in

4.1.4 FILTERING EVENT LOG

Through the Prom Tools platform we have access to several available plug-ins for filtering the event logs. Noticing any inhomogeneity or vulnerability from previous measurements, we can now study it thoroughly, through the available tools. As we mentioned above, we have the ability to we deal with tasks with a high frequency of occurrence or, correspondingly, with tasks that are performed quite rarely. We can apply mining techniques to municipal employees to find out in which cases we have the busiest employees. We will also study the behavior of municipalities in terms of employees to find the municipalities with the highest productivity. In several tests after applying the filter and producing the new file, we will apply process modeling algorithms to see the new workflow. These algorithms can also be applied to the original input file , to take a first look at the file and the running processes.

As we mentioned, Prom Tools has several plug-ins for filtering the event log file. In the following measurements, we apply the Filter log by attributes plug-in, based on the most common activity with code 01_HOOFD_010 in the 5 municipal files. The data is displayed with LogDialog/Dashboard and then using alpha, heurstic, inductive visual and fuzzy miner to represent the process model.

Figure 59: Actions tab for plug in selection

We select filtering on the trace that includes this activity, based on the code of the action and select the desired code from the available list and press Continue, to display the results.

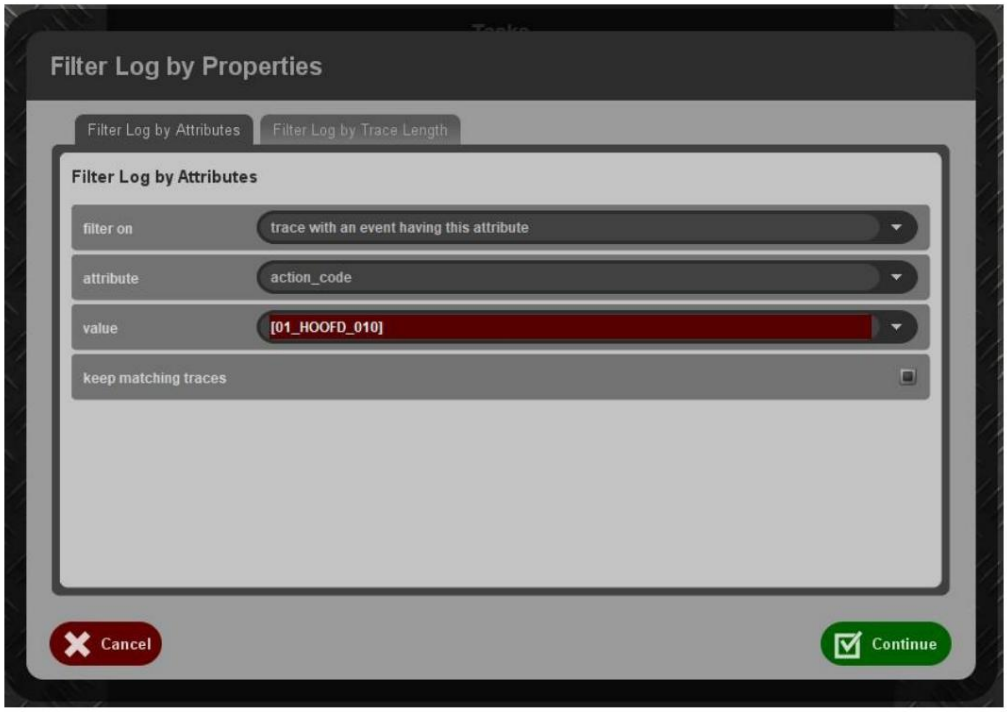
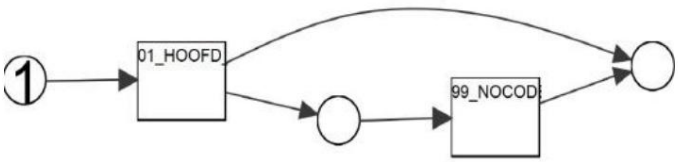


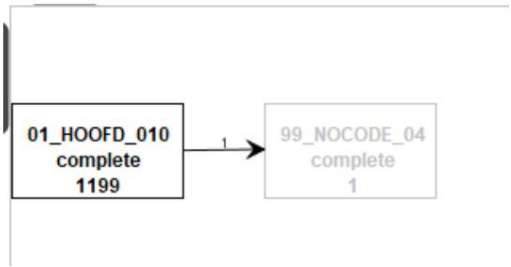
Figure 60: Filter Log By attributes plug-in

1st file
Dashboard

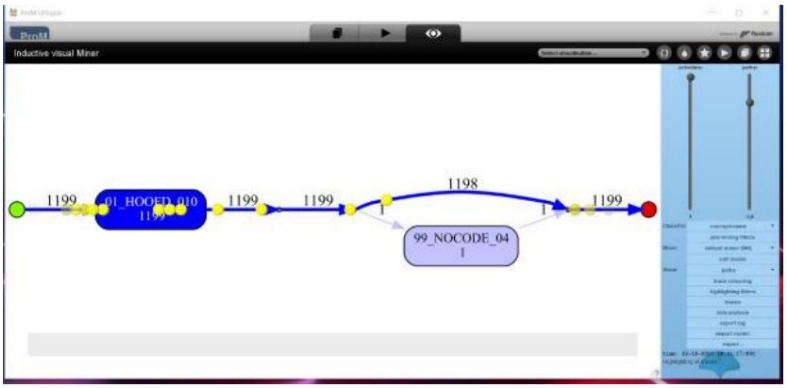
Alpha



Heuristic



Inductive



Fuzzy

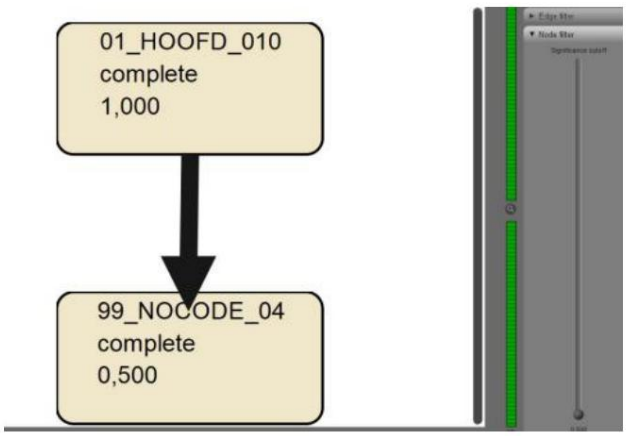
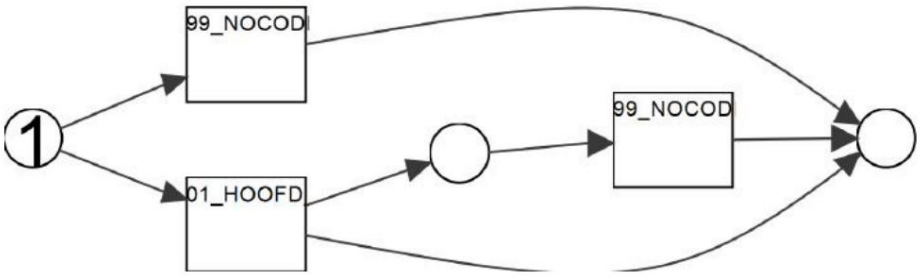


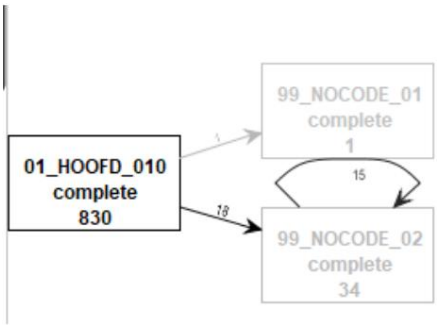
Figure 61: Dashboard and Process model of 1st file for task 01_HOOFD_100

2nd file
Dashboard

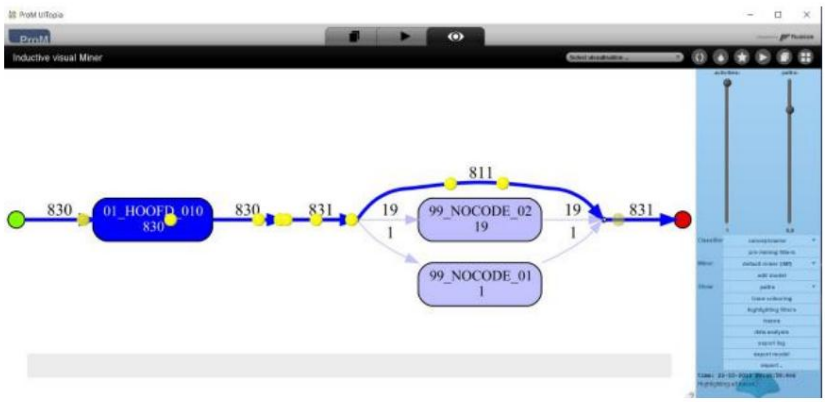
Alpha



Heuristic



Inductive



Fuzzy

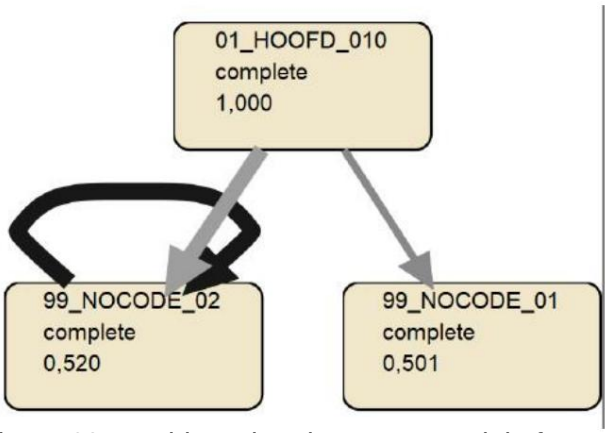
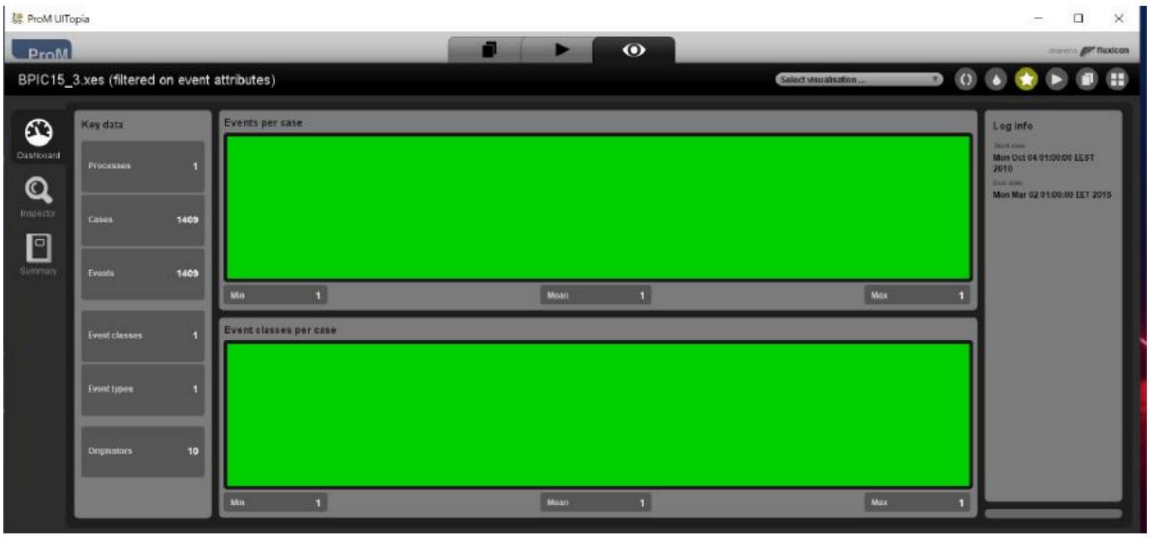
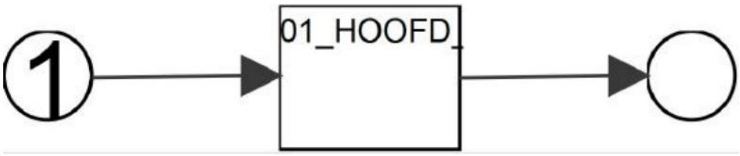


Figure 62: Dashboard and Process model of 2nd file for task 01_HOOFD_100

3rd file
Dashboard



Alpha



Inductive

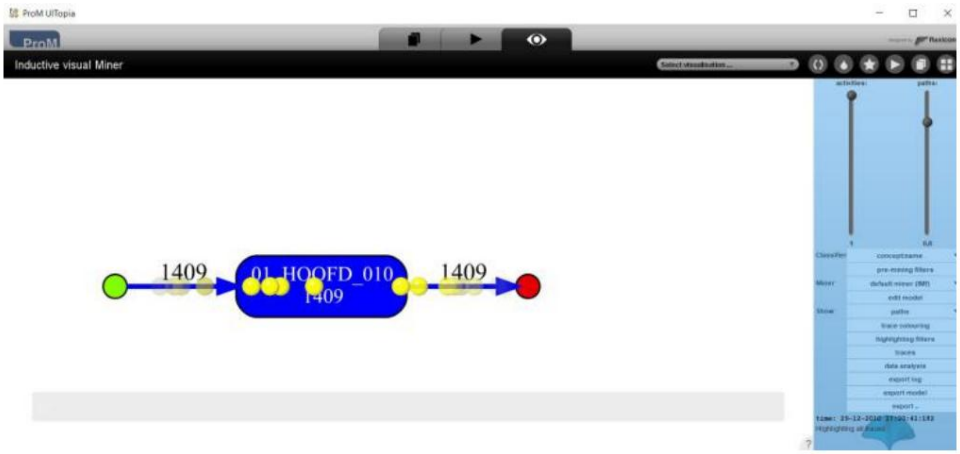
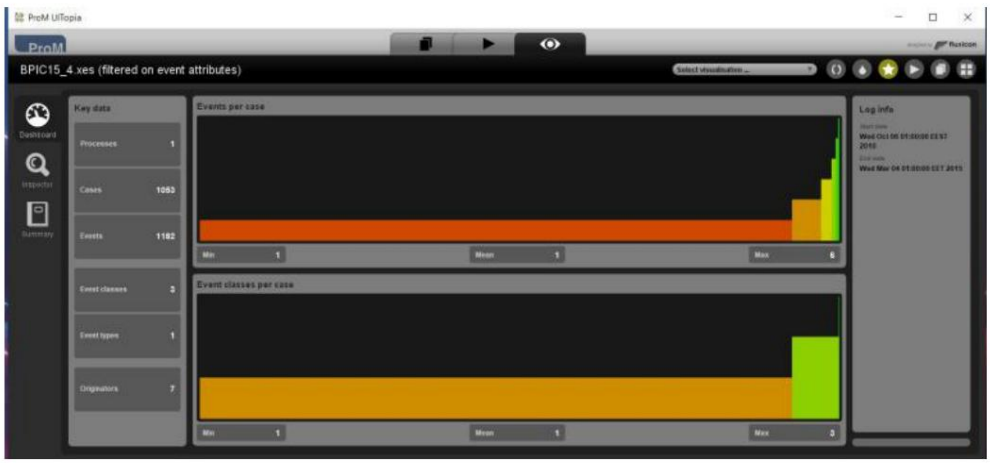
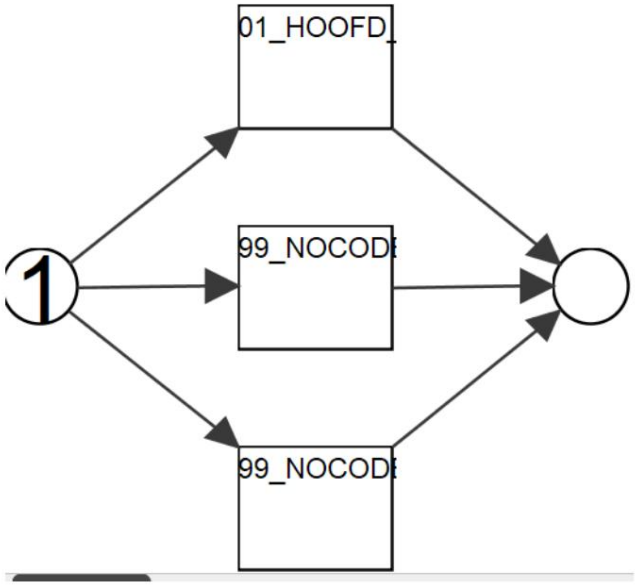


Figure 63: Dashboard and Process model of 3rd file for work 01_HOOFD_100

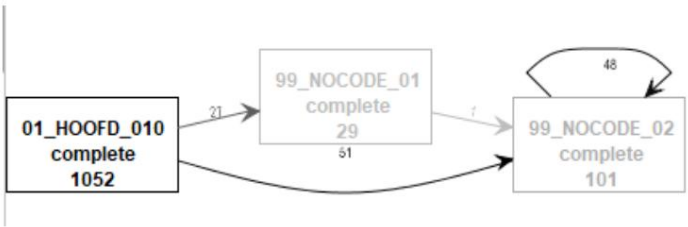
4th file
Dashboard



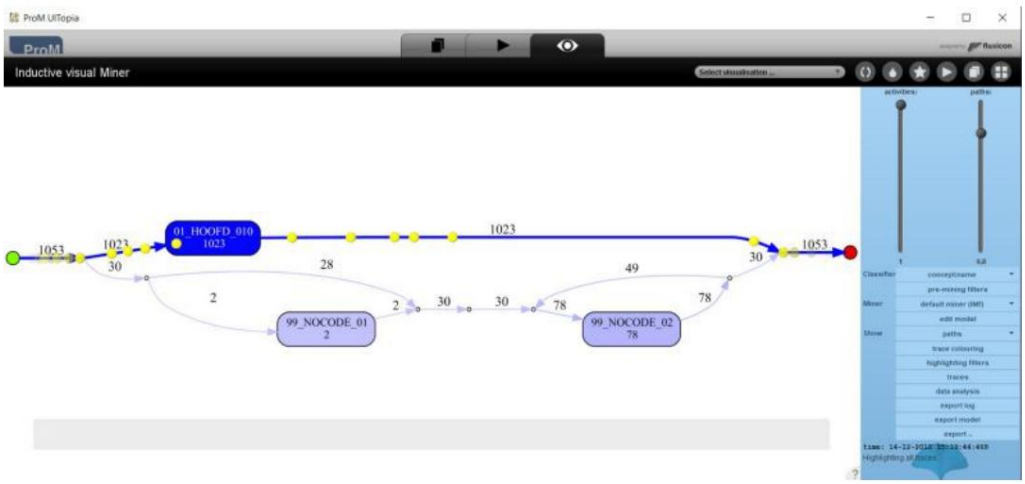
Alpha



Heuristic



Inductive



Fuzzy

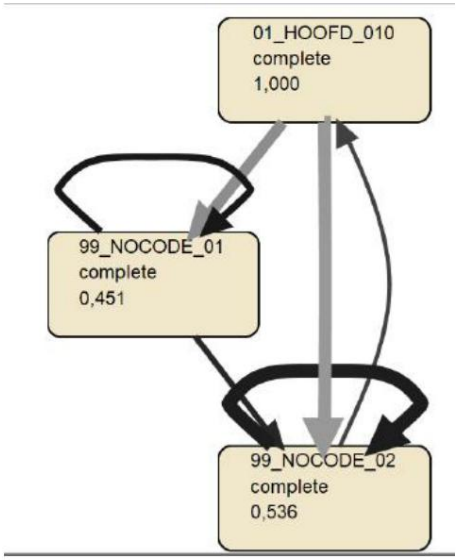


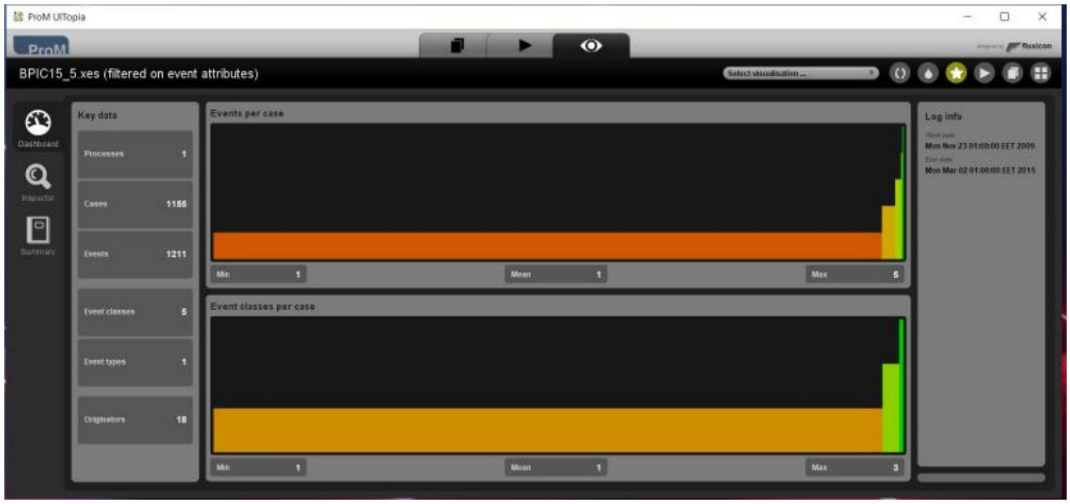
Figure 64: Dashboard and Process model 4th file for work 01_HOOFD_100

We notice that in the 4 files the activity 01_HOOFD_010 appears an average of 1,122 times which makes it one of the most popular activities in the municipalities. In file number 3 it has the highest number of appearances 1409 to be precise, while on the other hand in file 2 it has the fewest appearances numbered 830.

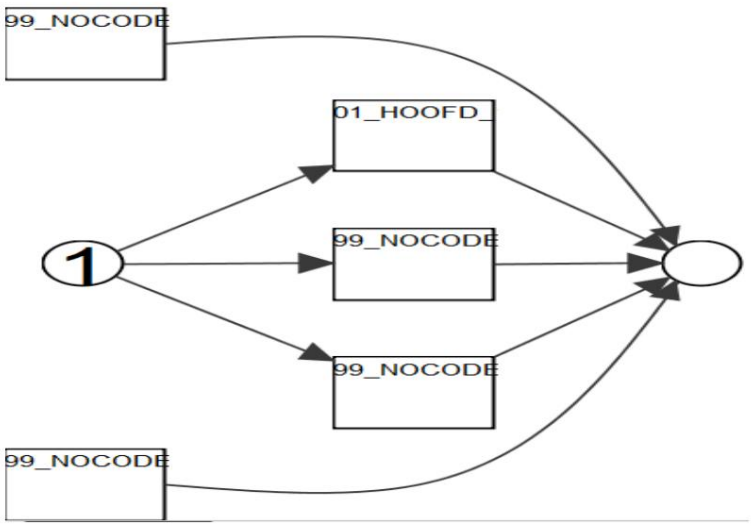
We also note that in the 3rd file, through alpha, fuzzy and inductive visual miner, the simplest and most understandable process model we have seen so far is produced. In addition, in the 3rd file we have the greatest homogeneity in the number of occurrences of the event in each case, for accuracy one at a time. On the contrary, files 2 and 4 present the most complex and difficult to understand process models and the greatest inhomogeneity in the distribution of activity in the cases. Which makes them not so efficient and orderly. We notice that in the produced process models the inductive visual miner algorithm shows the most understandable and illustrative results, which confirms its usefulness.

Then follow the process models for the 5th municipality, for the most popular activity with code 01_HOOFD_180, using the same miners as before. And in this case, however, the activity 01_HOOFD_010 was quite frequent in the executed tasks.

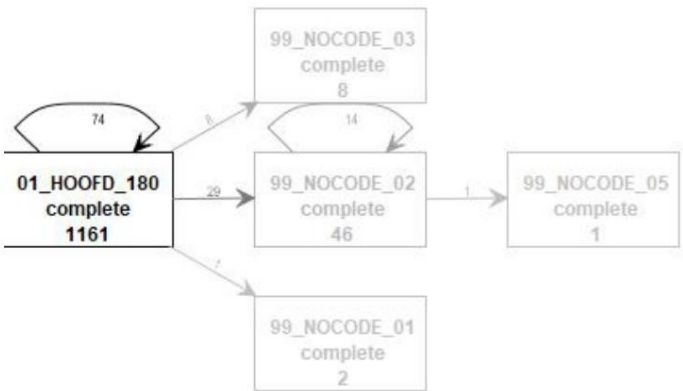
5th file
Dashboard



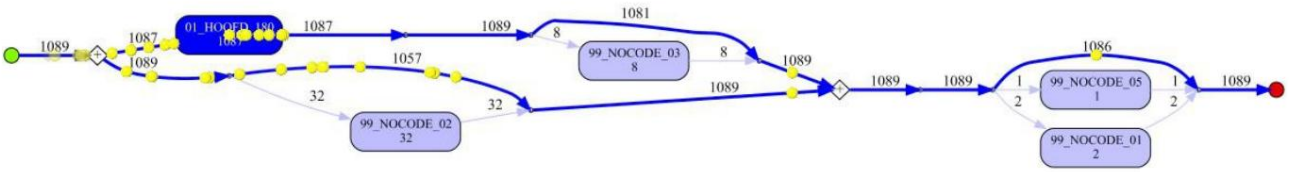
Alpha



Heuristic



Inductive



Fuzzy

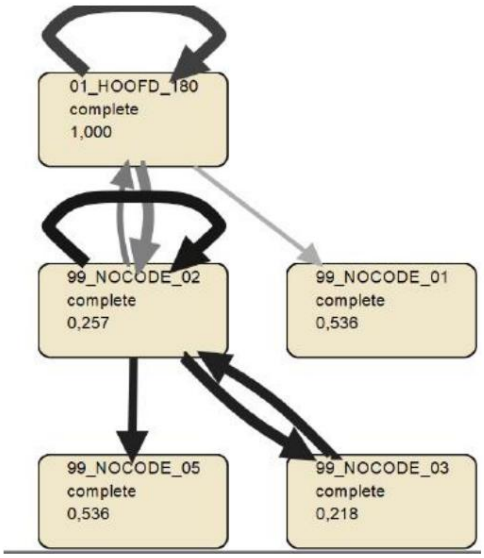


Figure 65: Dashboard and Process model of 5th file for work 01_HOOFD_180

We see that in the 5th municipality the process models produced for the corresponding activity are more complex than in the other municipalities and contain considerably more information. We notice that we have diagrams with loops, two-way relationships and paths that could be omitted. [22]

Following on from the previous measurements, I tried using the filter log by attributes plug-in, specifically in the filter log by trace length tab, to find points of particular interest for each municipality. To begin with, I set a threshold for the traces, those that have the length of activities below 5 and then in the generated file I applied the Inductive visual algorithm, which as we have seen performs quite well, to see the graphic model. The main purpose was to find in each municipality, how many employees we have who undertake and perform many few activities. With it the way we will see how many employees perform under just 5

procedures and we will conclude which municipality has the least efficient employees. The main issue that requires special attention is the cases that have the smallest number of events, which means that some employees (organizations) are not as efficient as we would expect.

1st file (minimum length 2 to 5)

Cases: 11
Activities: 41
Employees: 5

Inductive

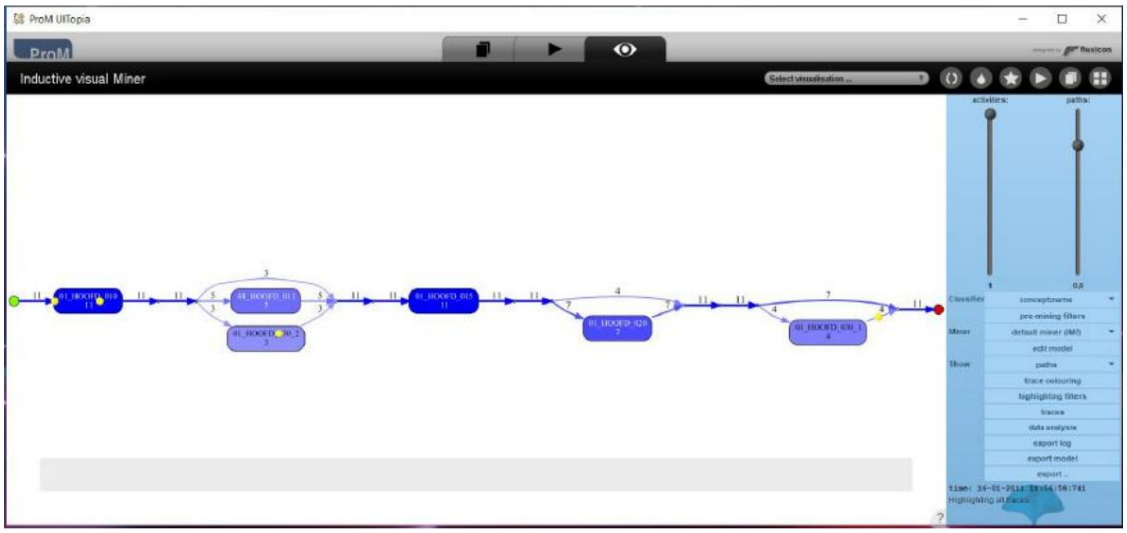


Figure 66: Dashboard and Process model of 1st file for traces with less than 5 actions

2nd file (minimum length 1 to 5)

Cases:3
Activities:7
Employees:3

Figure 67: Dashboard and Process model of 2nd file for traces with less than 5 actions

3rd file (minimum length 3 to 5)

Cases: 32
Activities: 148
Employees: 5

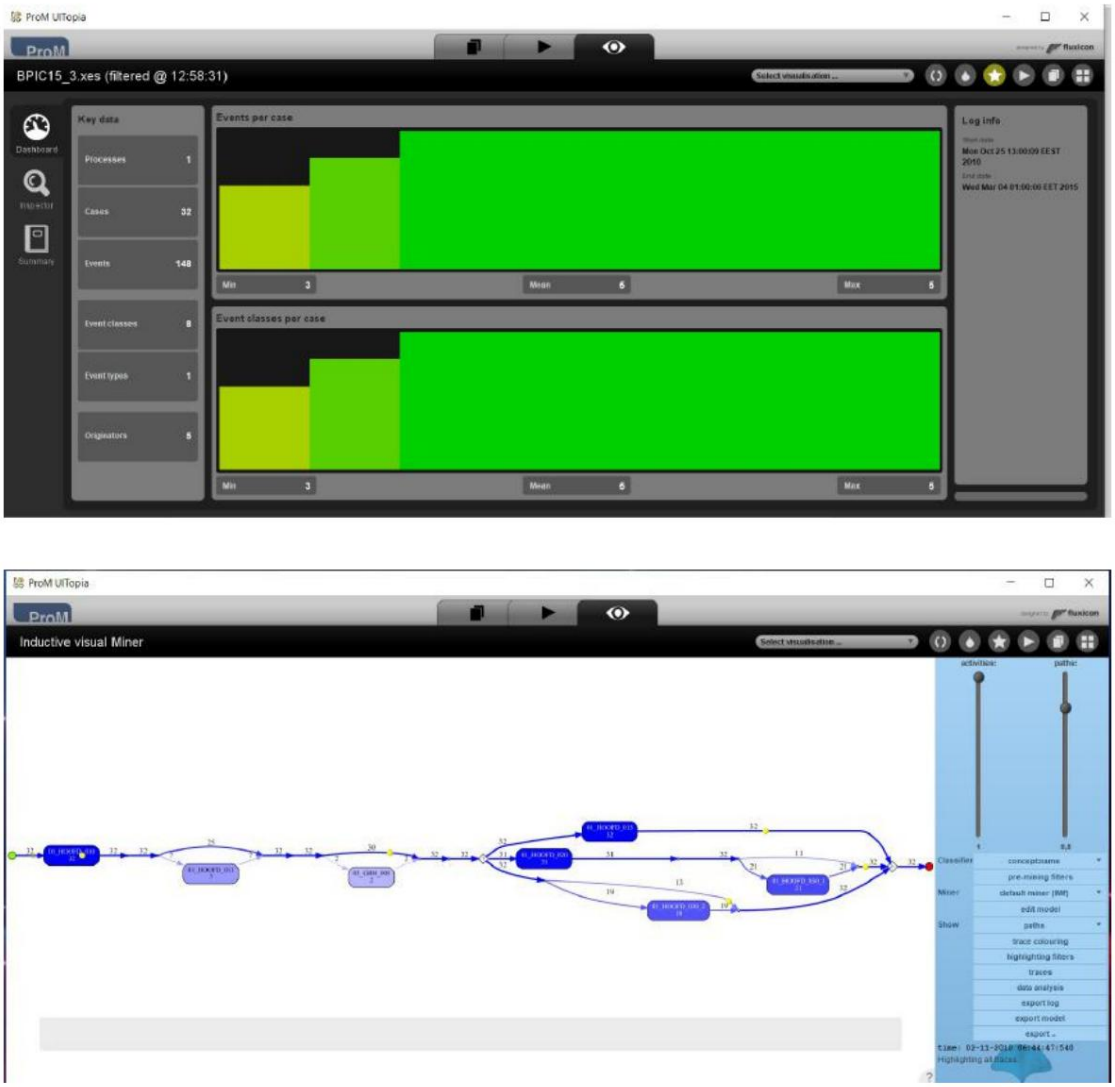


Figure 68: Dashboard and Process model of 3rd file for traces with less than 5 actions

4th file (minimum length 2 to 5)

Assumptions: 7
Activities: 22
Employees: 3

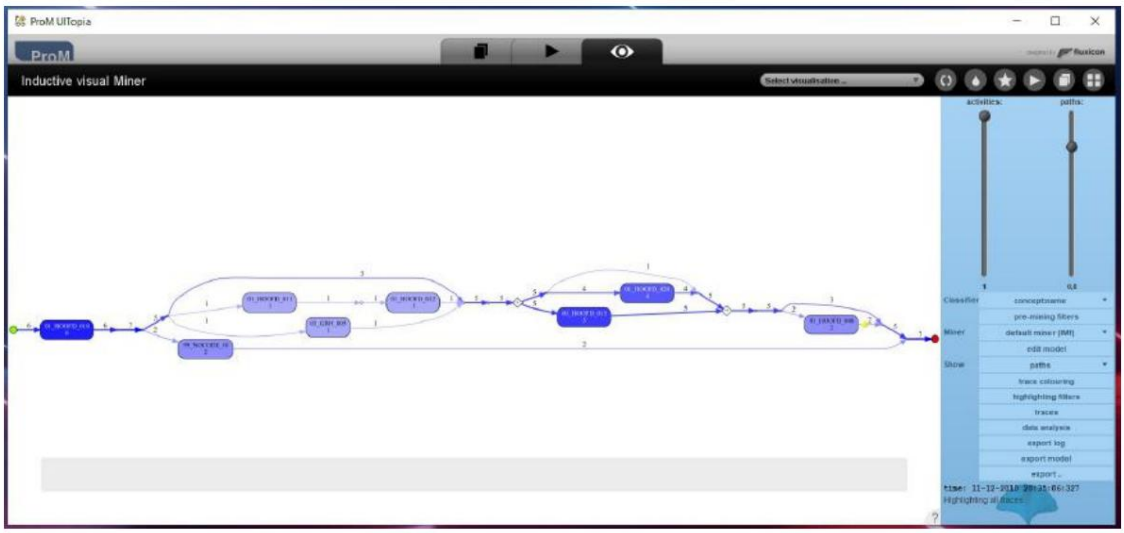


Figure 69: Dashboard and Process model of 4th file for traces with less than 5 actions

5th file (minimum length 5 to 10)
Assumptions: 8
Activities: 58
Employees: 6

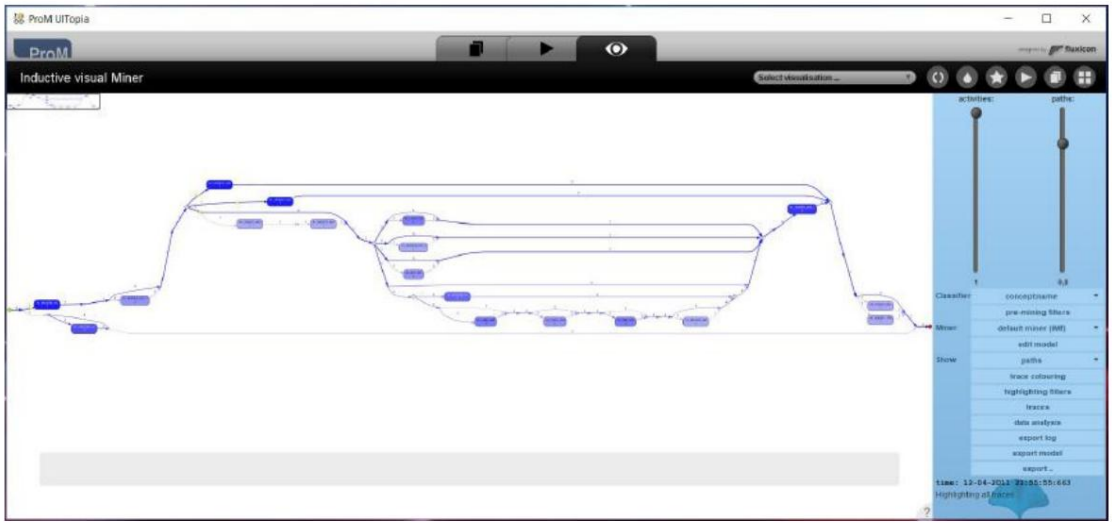


Figure 70: Dashboard and Process model of 5th file for traces with less than 5 actions

Observing the results we see that the 2nd file had few cases with an event number from 1 to 5, which makes it the municipality with the highest performance and productivity so far. Then the 1st and 4th municipality had the next best performance with cases with few events. On the other hand, the 3rd municipality had a disappointing performance this time, with 32 cases having between 3 and 5 events (there was no case with less than 3 events in the municipality). A noteworthy point seems to be the 5th file, the which did not have a case with less than 5 events and therefore the measurement concerned cases with 5 to 10 events, which makes it one of the most productive municipalities, with quite busy agencies and employees. The 2nd municipality that has the fewest procedures produces, as we expected, a simple model without loops and iterations.[23]

In the next measurement we seek using the same filter to see the other side of the coin. More specifically before we saw the municipalities with employees who perform few procedures, while on the other side now we will see the municipalities with the busiest employees. We will apply again the plug in **filter log by attributes** and especially the filter log by trace length, setting the threshold value to 80. This way we will see the results for the traces with more than 80 processes executed. In this case we want to see large percentages , which will mean that each municipality will have quite efficient employees. In this test, after the filtered files are produced, we will use the aforementioned visualization technique, the Dotted Chart diagram. In order to see the statistics of employees with over 80 tasks to their credit, to observe how they behave over a 5-year period and what patterns they develop.

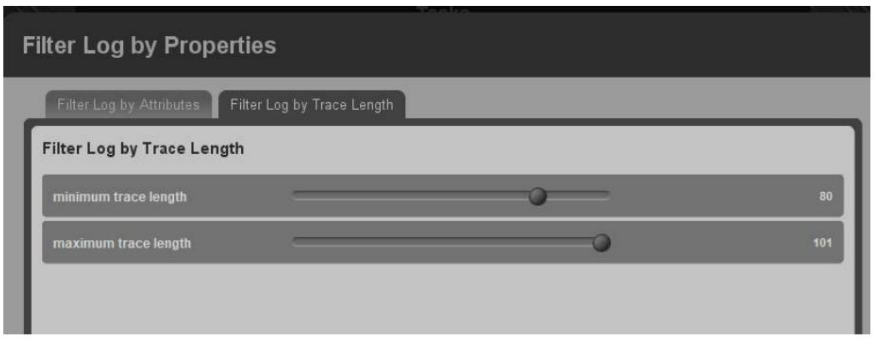


Figure 71: Filter Log by trace length

1st file

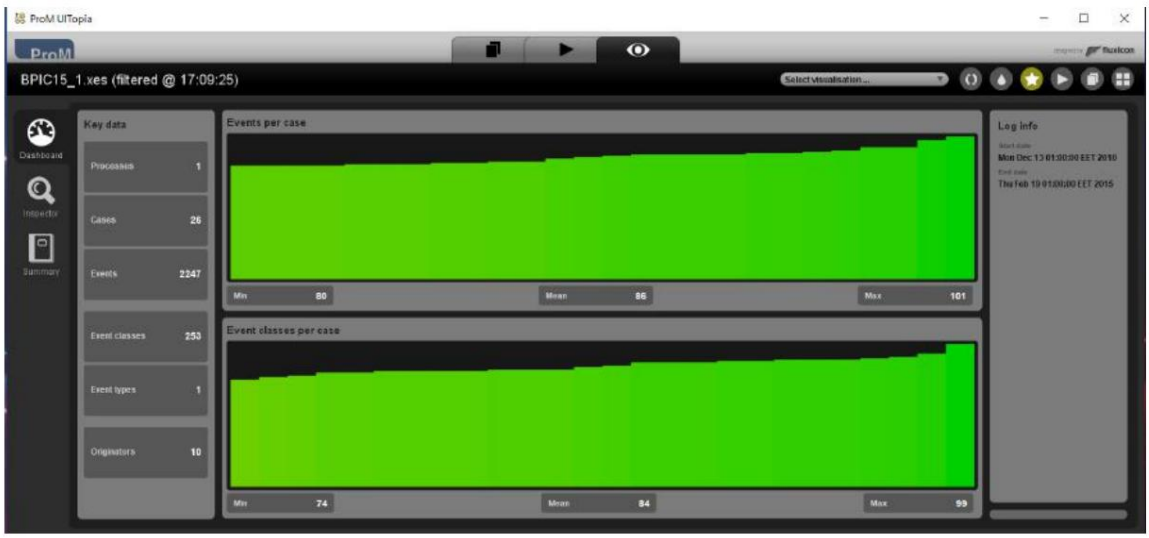


Figure 72: Dashboard and Dotted Chart of 1st file for traces with more than 80 actions

2nd file

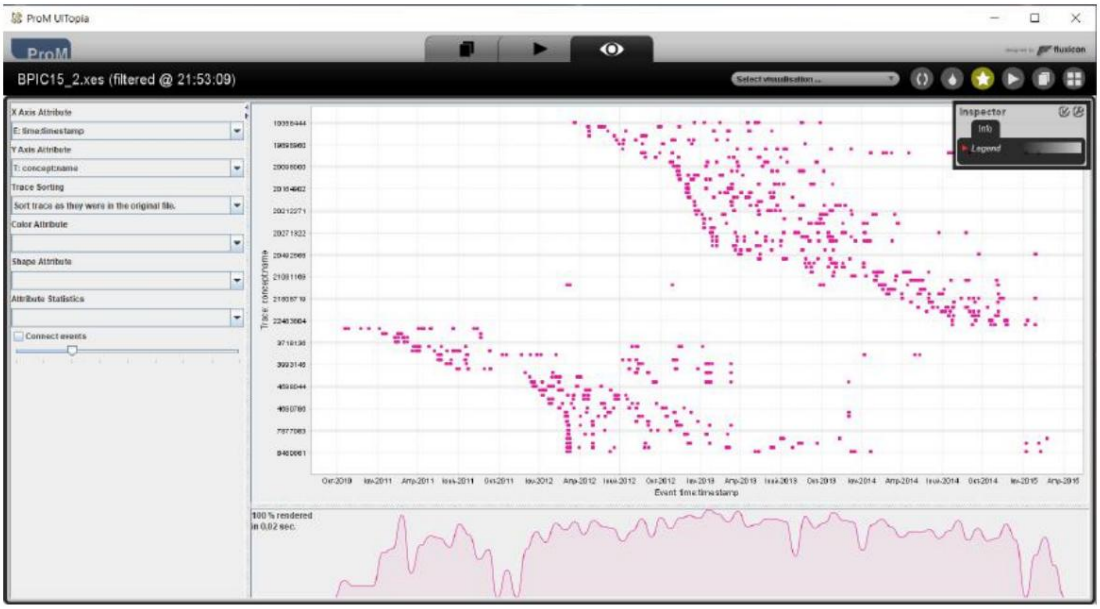


Figure 73: Dashboard and Dotted Chart of 2nd file for traces with more than 80 actions

3rd file

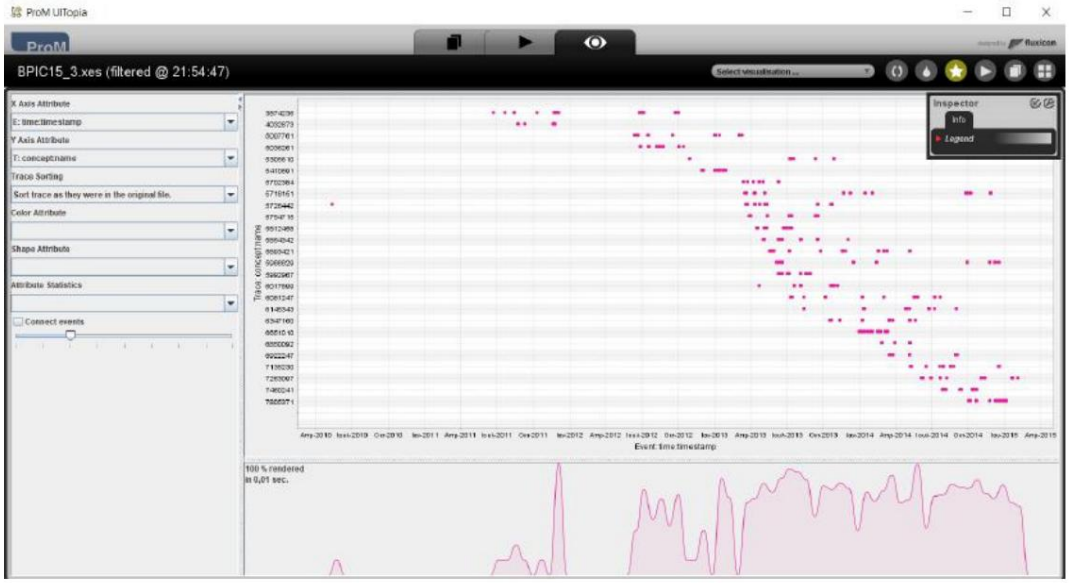


Figure 74: Dashboard and Dotted Chart of 3rd file for traces with more than 80 actions

4th file

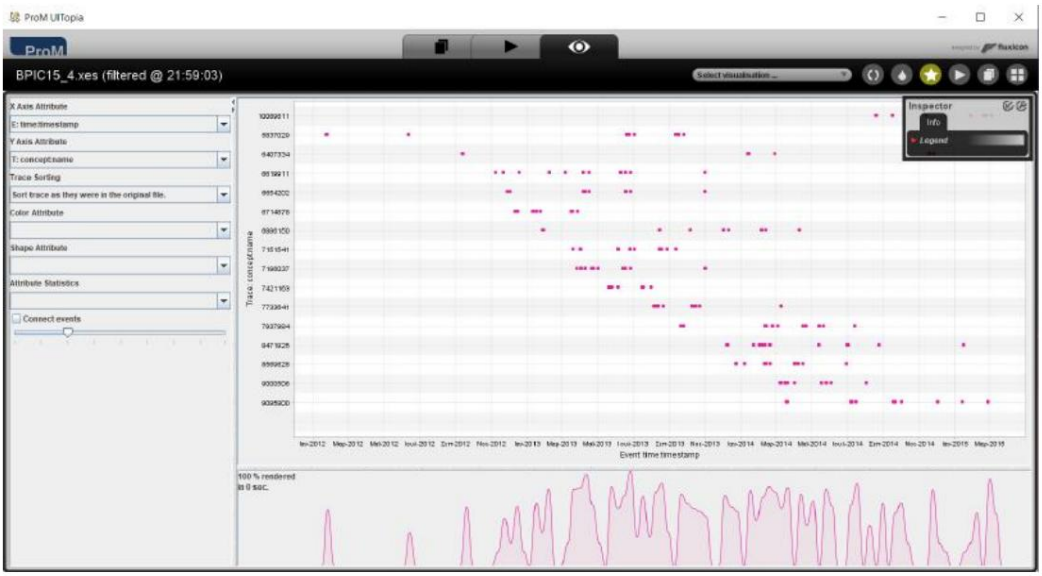


Figure 75: Dashboard and Dotted Chart of 4th file for traces with more than 80 actions

5th file

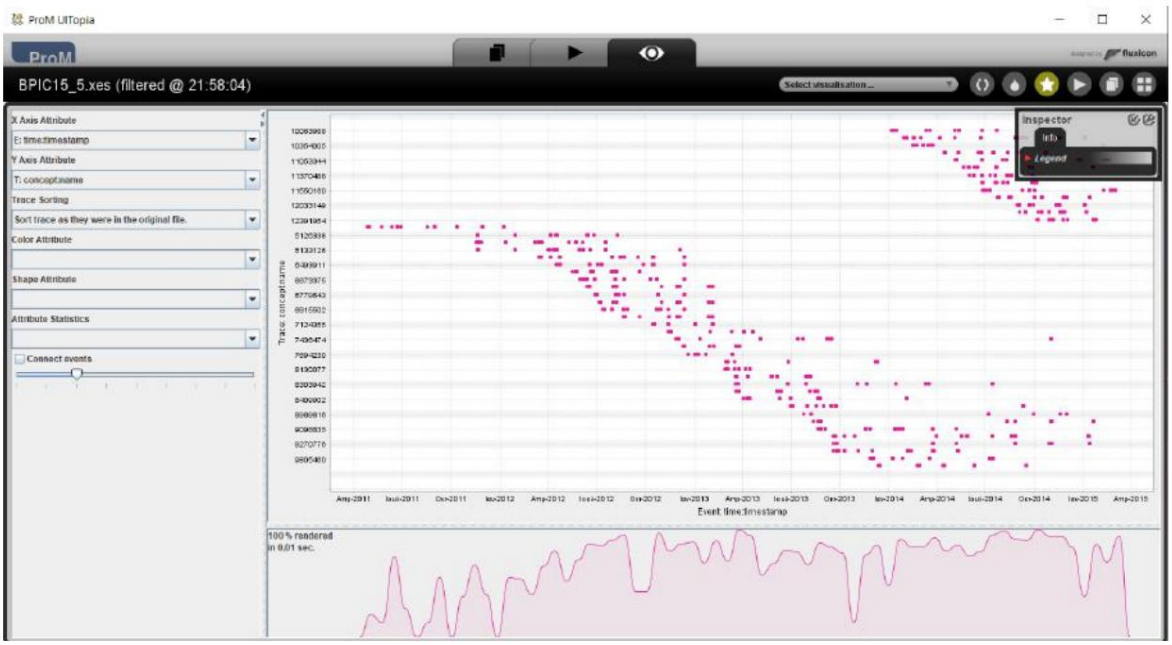


Figure 76: Dashboard and Dotted Chart of 5th file for traces with more than 80 actions

Continuing the tests and in this plug in we have similar results, because the 2nd and 5th municipalities have the most cases with over 80 events, while the other 3 municipalities continue to be the lowest in performance.

More specifically we have: 1st
municipality ÿ cases 26, events 2247
2nd municipality ÿ cases 76, events
6922 3rd municipality ÿ cases 26, events
2313 4th municipality ÿ cases 16, events
1384 5th municipality ÿ cases 46, events 4263

We see the unexpectedly good performance of the 2nd municipality and the quite appreciable performance of the 5th municipality. On the opposite side, the remaining municipalities fluctuate in low numbers, with the performance of the 2nd municipality reaching up to 6 times that of the 4th. It is still attracting great interest that the 5th municipality had a trace with up to 154 events and the 2nd one had up to 132 events. The application of the filter was done with over 80 events up to the upper limit of each municipality. With the following data we see that the 2nd municipality had up to 7 times more events than the corresponding traces of the 4th municipality, which had the lowest performance

this point the high productivity of the employees of municipality 2 and 5, having quite a large number of employees with over 80 performed procedures, which means that in these municipalities there is high efficiency and production,

The above data come to confirm the dotted chart diagrams which show us cases on the vertical axis and time on the horizontal axis, showing the activities performed by each organization in relation to time. We see in files 2 and 5 denser diagrams that contain many more works than the rest of the municipalities. Even the waveforms below the diagrams show the numerous cases of municipalities 2 and 5, which remain at a high index throughout the 5-year period, in contrast to the rest of the municipalities that reach a peak performance and then decline. A notable point seems to be that in all the records the peak of work occurs around the first 6 months of 2013, which seems to have been the most productive period for the municipalities.

As before we dealt with the employees of the municipalities and the processes they perform, I would like in the next measurement to deal with the distribution of employees in the municipalities. In this way we will really see what is happening in the municipalities and how the tasks are distributed to each responsible. This the measurement in order to have a better picture of the processes and their mining, was done using the plug-in Generate log from org:resource perspective in prom tools. The purpose of this filter was to show us for each municipality file how many org:resources there are , for how many events they are responsible for and in general to see in more detail how many processes each organization undertakes both on average and also the organizations with the most and the fewest processes respectively. In this way we will draw conclusions about the most productive organizations, as it is not enough to many tasks are performed, the burden of these should be properly shared among everyone. It does not benefit anyone, let alone the municipality itself, to have too many employees without correspondingly many tasks being performed, nor to perform many tasks with too many employees. In more detail, we have the following information:

1st file

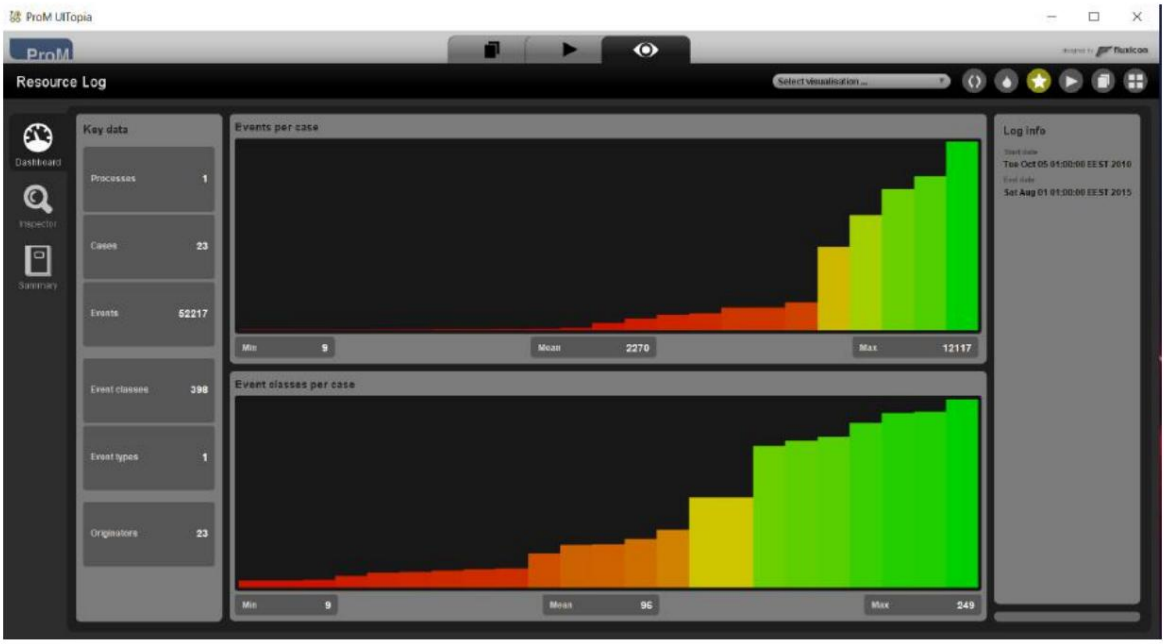


Figure 77: Generate log from org:resource perspective for the 1st municipality

2nd file

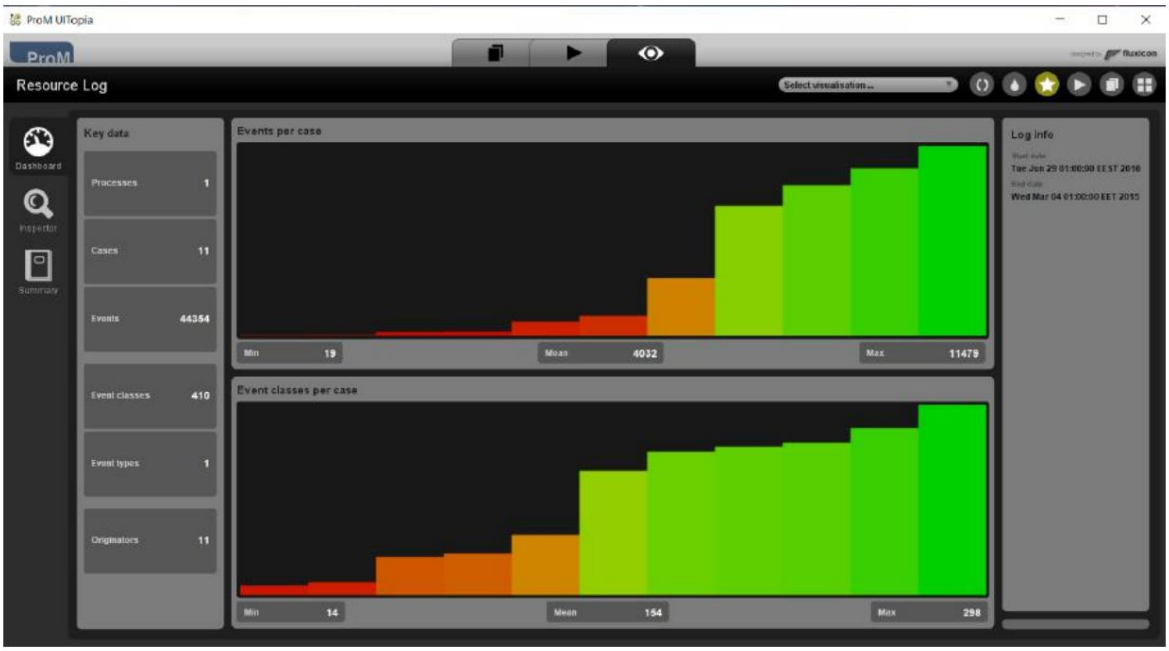


Figure 78: Generate log from org:resource perspective for the 2nd municipality

3rd file

Figure 79: Generate log from org:resource perspective for the 3rd municipality

4th file

Figure 80: Generate log from org:resource perspective for the 4th municipality

5th file

Figure 81: Generate log from org:resource perspective for the 5th municipality

As can be seen in the results above, I have first listed the log visualizer tab from prom tools, where we can see the number of resources, how many events, on average, each resource receives and which resource handles the fewest and most events.

1st municipality ÿ 23
2nd municipality ÿ
11 3rd municipality ÿ
14 4th municipality ÿ
10 5th municipality ÿ 22

With the above, we see that the 1st and 5th municipalities have the most resources, as is logical, because they also have many events to execute, specifically more than 50,000. Especially the 5th municipality is used to high rates of productivity and performance. On the other side, the 3rd municipality with almost 60,000 events we see that it employs only 14 services to execute them, which is admirable for the municipality. While the 2nd and 4th municipalities having the fewest events, under 50,000, expectedly employ the fewest organizations.

Of course, it is necessary to note that in many cases organizations/employees cooperate with more than 1 municipality at a time, so we cannot produce such reliable results.

1st municipality ÿ 2270 mean events, 12117 max events
2nd municipality ÿ 4032 mean events, 11479 max events
3rd municipality ÿ 4263 mean events, 14620 max events

4th municipality ÿ 4729 mean events, 15748 max events
5th municipality ÿ 2686 mean events, 12225 max events

Nevertheless, we must comment that the workers in municipalities 2, 3 and 4 have on average many more tasks, exceeding 4,000, in contrast to municipalities 1 and 5 which are around 2,500 tasks. Also noting that the 4th municipality has an organization with over 15,000 events and that the organization with the fewest tasks in the 2nd municipality performs 19 activities, while in the other municipalities there are employees with 2 or 3 activities.

Furthermore, I would also like to present the employees who are, as I mentioned above, in more than one municipality and how they perform in it. In more detail, we have the following
information: The organization/employee
(org:resource) with a code

- 6 with work in the 1st, 3rd and 4th
municipalities. In the 1st municipality it has 26 completed tasks, while in the 3rd and 4th it has 2 and 3 respectively as we mentioned above, as it is the organization with the fewest processes in these municipalities.

- 560429 with operations in the 2nd and 5th
municipality. In the 2nd it has 19 tasks and in the 5th 7,590 tasks.

- 560530 with works in the 2nd and 5th municipality.
In the 2nd it has 11,479 tasks and in the 5th 683 tasks.

- 560532 with works in the 2nd and 5th municipality.
In the 2nd it has 10,080 tasks and in the 5th 1,317 tasks.

- 560598 with works in the 2nd and 5th municipality.
In the 2nd he has 183 tasks and in the 5th 1,737 tasks.

- 560752 with works in the 4th and 5th municipality.
In the 4th it has 11,948 tasks and in the 5th 1,676 tasks.

- 560849 with operations in the 4th and 5th
municipality. In the 4th he has 764 tasks and in the 5th 154 tasks.

We see above that in 6 of the 7 cases these employees belong to the 5th municipality. While in 4 of these the employees are between the 2nd and 5th municipalities and in the other 2 between the 4th and 5th municipalities. These indications come to complete the point of view on the high productivity of the 2nd and 5th municipality, proving that with the appropriate cooperation

service issues can be resolved in a shorter period of time.

Based on the above measurements that were performed, we can reach some conclusions about the functioning of the municipalities. For example, we notice that municipality number 2 and number 5 have the highest productivity, with a fairly large average number of tasks per day, a fact that makes them perform many processes daily and very quickly. The 5th municipality in the period of 4 years performs the most procedures. In contrast, the low rate of work execution makes the other municipalities late in processing issues. Also the 2nd and 5th municipalities have the most productive employees, who they perform many additional tasks, while at the same time they manage to cooperate with each other for the best execution of tasks. In particular, the 2nd municipality had only two "isolated" employees and all the rest collaborated with others to carry out procedures. The same is the case with the 5th municipality where all the workers contribute to the other municipalities as well, even the 4th municipality, which still has a reduced performance. Especially if we consider that the 2nd municipality's work start date is 6 months later than the 4th municipality, which starts first, and yet performs several additional tasks.

On the other side of the coin municipalities 1, 3 and 4 have employees who do not cooperate with other municipalities, which seems to reduce their efficiency. They have a low average work per week, while especially municipality 1 has the most employees and manages to perform the fewest tasks on average. Finally, the number of employees with more than one number of tasks (threshold in our case 80 procedures) is added, where the employees of the 2nd and 5th municipalities exceed again, showing that the total of them, and not just one portion, are much more efficient and take on tasks with a correct distribution. In the corresponding survey for employees with a small number of tasks (threshold in our case 5 tasks) we see that municipalities 1 and 3 have a large number of employees with very few procedures to their credit. And in this case municipalities 2 and 5 managed to prevail with the 2nd municipality having few cases with less than 5 procedures, while the 5th municipality had no case with less than 5 tasks.

CHAPTER 5
CONCLUSION

With all the above results and data, we see that the number of processes that are performed every day at the professional level of each person is huge and greatly affects the viability and performance of the business and the people in it. The process mining technique with its tools can analyze the processes taking place, study their structure and mode of operation and draw conclusions about them.

Event logs from municipalities and companies in general offer us a wealth of information to study and process. Analyzing them using mining techniques, we can see how procedures are performed, if in any case we have low performance from a municipality or employee and why specifically, whether the process models match the data in the files and perhaps cases where there may be an unjustified waste of resources for the tasks.

With the process mining tools we applied in this dissertation, we draw useful conclusions for the questions raised at the beginning of the presentation. Such as what are the most common procedures in municipalities, where it is proven that the work with code 01_HOOFD_100 (register submission date request) is in 4 of the 5 municipalities, the most frequently executed task. Even in the question about the most productive workers we see that those in the 2nd and 5th municipality perform several tasks in relation to the rest, ensuring high productivity in the respective municipalities. especially in The 2nd municipality, which also has a few employees. Pointing out based on the results, that many workers in the 2nd municipality also perform work in the 5th municipality and vice versa. Showing that in cases where there is cooperation and mutual help, better results are achieved. However, on the other hand, we see the remaining 3 municipalities with lower production, in relation to the human resources they have and the minimal cooperation they seem to have with each other. This conclusion is confirmed by the long time required for the preparation of works in municipalities 1, 3 and 4, as and the low average they hold. In the measurements we saw from the beginning the statistics of the municipalities and in agreement with the visual results (diagrams, process models) produced along the way, they confirm the conclusions we drew.

Process mining for the analysis and evaluation of business processes in local government is a major issue with a wide scope of research that can help us better understand how processes are performed and contribute to their improvement in order to deliver a satisfactory performance. for municipalities and people.

BIBLIOGRAPHICAL REFERENCES

[1] [Online] https://en.wikipedia.org/wiki/%CE%95%CF%80%CE%B9%CF%87%CE%B5%CE%AF%CF%81%CE%B7 %CF%83%CE%B7

[2] [Online] https://en.wikipedia.org/wiki/%CE%9F%CF%81%CE%B3%CE%B1%CE%BD%CE%B9%CF%83%CE%BC %CE%BF%CE%AF_%CE%A4%CE%BF%CF%80%CE%B9%CE%BA%CE%AE%CF%82_%CE%91%CF%85%CF%84%CE %BF%CE%B4%CE%B9%CE%BF%CE%AF%CE%BA%CE%B7%CF%83%CE%B7%CF%82

[3] [Online] https://en.wikipedia.org/wiki/Business_intelligence

[4] [Online] https://www.gartner.com/en/information-technology/glossary/business-process management-bpm#:~:text=Business%20process%20management%20(BPM)%20is,support %20of %20a%20business%20strategy

[5] [Online] https://www.investopedia.com/terms/b/business-intelligence-bi.asp#:~:text=Business%20intelligence%20(BI)%20refers%20to,performance%20benchmarking% 2C%20and %20descriptive%20analytics.

[6] [Online] https://en.wikipedia.org/wiki/Data_mining

[7] [online] http://www.processmining.org/introduction.html

[8] A Study of Process Mining-based Business Process Innovation, Sungbum Parka, Young Sikkang

[9] Process Mining plays an essential role in Digital Transformation, Pedro Robledo
BPM process management expert

[10] [Online] https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/

https://en.wikipedia.org/wiki/Alpha_algorithm

[11] [Online] http://www.processmining.org/event-data.html

[12] Towards Cross-Organizational Process Mining in Collections of Process Models and Their Executions, JCAMBuijs, Boudewijn F. van Dongen, Wil MP van der Aalst

[13] [Online] https://www.promtools.org/doku.php

[14] [Online] https://eventlogxp.com/

[15] [Online] https://www.futurelearn.com/info/courses/process-mining/0/steps/15636

[16] Heuristic mining: Adaptive process simplification in education, Publisher: IEEE

[17] [Online] https://www.futurelearn.com/info/courses/process-mining/0/steps/15645

[18] Discovery and analysis of the Dutch permitting process, Liese Blevi and Peter Van den Spiegel

[19] Benchmarking of Municipalities with Process Mining Techniques Reveals Opportunities for Improvement, Ube van der Ham

[20] Handling of Building Permit Applications in The Netherlands: A Multi-Dimensional Analysis, PM Dixit , BFA Hompes, N. Tax, SJ van Zelst

[21] [Online] **https://research.aimultiple.com/process-mining-algorithms/**

[22] An Exploration and Analysis of the Building Permit Application Process in Five Dutch Municipalities, Niels Martin, Gert Janssenswillen, Toon Jouck, Marijke Swennen, Mehrnush Hosseinpour, Farahnaz Masoumigoudarzi

[23] Process Mining, Wil Van Der Aalst