2019

Robotic Process Automation

AN ASSESMENT OF PROCESS DISCOVERY TECHNIQUES WITH THE PURPOSE OF FINDING RPA ELIGIBLE PROCESSES

VINCENT GORIS

Master Thesis Business informatics

An Assessment of Process Discovery Techniques With The Purpose of Finding RPA Eligible Processes

Vincent Goris

Department of Information and Computer Sciences

2019

Supervisors

Inge van de Weerd

Hajo Reijers

Leo van Neutegem

Utrecht University

Acknowledgements

First of all, I would like to thank a number of people who have helped me tremendously to make this project possible. First of all I would like to thank my supervisors Inge van de Weerd, Hajo Reijers and Leo van Neutegem. I thank Inge van de Weerd and Hajo Reijers for providing the guidance I needed to conduct proper experiments and for helping me to understand the topic of RPA. I want to thank Leo van Neutegem for supporting me throughout the process, keeping me motivated and providing me with the connections I needed to get the data that were needed for this research. I would like to thank InfoSupport for helping me finish my education and going above and beyond to help me supplying data for my research. I want to thank Andres Jimenez-Ramirez for allowing me to use the Screen Scraper tool and I thank the various experts that have helped me to validate my results. Finally, I would like to thank Wilma for proofreading this thesis and help me improve the readability of the text and correct typing mistakes.

I sincerely hope you will enjoy reading my thesis!

Table of Contents

1: Introduction

Robotic Process Automation is a process where simple tasks that are performed by humans are automated by employing 'software robots' to do the task. Using Robotic Process Automation (RPA), mostly rule-based business processes are automated, enabling the user to spend more time on other and more valuable types of work (Boulton, 2017). Typical processes that can be automated using RPA are structured, rule based tasks such as copying information from one form to another or processing transactions (Lacity & Willcocks, 2018). However, some sources claim that RPA goes even further and might even perform knowledgeable tasks that could formerly only be performed by knowledge workers (Institute for Robotic Process Automation, 2015; Kirchmer, 2017).

RPA can be of huge benefit for a company. Applying RPA correctly leaves time for employees to focus on more interesting and complex tasks (Asatiani & Penttinen, 2016) instead of losing valuable time on performing mundane and boring tasks. Other advantages are fewer human errors (Lintukangas, 2017) and a 24/7 continuity of service (Accenture, 2016). RPA is easier to implement when compared to traditional automation methods. This is because business processes do not have to be changed, but are instead replaced by a robotic worker (CapGemini Consulting, 2016). This is considered a more bottom up approach instead of top down when changing business processes (Agaton & Swedberg, 2016; CapGemini Consulting, 2016; Lacity & Willcocks, 2018; Lacity, Willcocks, & Yan, 2015).

1.1 Problem statement

At present time, not all RPA projects succeed. A survey by Deloitte found that out of 400 firms, 63% did not meet delivery deadlines for RPA projects and 30 to 50 percent of initial RPA projects fail (Trefler, 2018). When an RPA project fails, it is not usually the technology that is to blame but either the use case or candidate process that is the root cause of the failure (Rutaganda, Bergstrom, Jayashekhar, Jayasinghe, & Ahmed, 2017). Companies are blinded by the promised opportunities of RPA which result in not carefully picking a suitable process. Attempting to use RPA on unsuitable activities might result in a software robot that is unable to perform all the required activities (Burgess, 2018). The root cause of many failed RPA projects can thus be traced back all the way to the process identification phase at the start of an RPA project.

Currently, the initial phase of RPA, picking the right processes, takes up a lot of time for the developers (Agaton & Swedberg, 2016). In many cases, the process identification and picking is done through interviews, observation of employees and manual labour, creating an overview of the existing business processes in a company (Asatiani & Penttinen, 2016). These processes are then analysed and one or multiple processes are chosen to be automated through RPA. This takes up a lot of time and is prone to errors. Therefore, the following two problems are identified:

1. It is difficult to find and pick the right candidate process for an RPA project
2. Finding potential candidate processes is time consuming and often a manual process which could be improved upon

Having a more structured and tested method to identify processes in an RPA project, could shorten the time to delivery and could reduce the error rate, resulting in more successful RPA projects (Agaton & Swedberg, 2016). In this study, I will look at different Process Discovery techniques and assess them in an RPA context in order to attempt to find a suitable method for process identification in an RPA context.

1.2 Research Questions

In order to assess different Process Discovery methods in the context of Robotic Process Automation, the Design Cycle proposed by Wieringa (2014) will be used. Design science research focusses on improvements upon a stated problem. This is done by designing an artefact which can either be found in literature or created by the author, to improve a problem context. The Design Cycle is an ongoing cycle where the problem context is investigated, design artefacts are created and used to improve upon the problem context, and artefacts are implemented. The aim is to improve upon the problem context and the cycle will be run until a satisfactory result is obtained (Wieringa, 2014).

Aims and objectives

This research aims to improve the identification of processes that are suitable for RPA *by* testing different Process Discovery methods in a practical setting *that* should be able to find the right process properties that are needed to discover RPA suitable processes *in order to* improve the accuracy of the chosen processes and shorten the time for discovery.

Research questions

To address the main objective, the following main research question is formulated.

Which Process Discovery method is best suited to find processes eligible for RPA?

This main research question is broken down in multiple sub questions.

Sub question 1: What Process Discovery methods have already been used in RPA projects?

This question will focus on methods for Process Discovery that have already been used in the context of RPA. The found methods will be explained and evaluated in order to achieve a clear understanding of the strengths and weaknesses in Process Discovery for RPA.

Sub question 2: Which other Process Discovery methods could be used in an RPA project?

RPA is still a young field of research. Not all Process Discovery techniques that could be applied to RPA, have yet been applied. This question aims to find and evaluate techniques for Process Discovery that have not yet been used in an RPA project. The main example are Process Mining techniques that can be used to automatically identify business processes within an organization using software logs. In the context of the problem statement, Process Discovery methods could be of benefit because of their systematic and accurate nature (van der Aalst et al., 2012a). Process Mining techniques are also deemed faster than finding business processes manually (van der Aalst et al., 2012a), therefore tackling the full scope of the problem statement.

Sub question 3: What are the requirements for a Process Discovery method in the context of RPA?

In order to be able to test which Process Discovery method found in SQ1 and SQ2 will perform best for RPA projects, a set of requirements should be set which will be used to evaluate the methods. This should be done before the tests are performed to ensure a fair assessment. These requirements will consist of functional requirements which are tailored to the RPA setting and non-functional requirements which are based upon the problem statement and usability in a practical setting.

Sub question 4: Which of the Process Discovery methods (from SQ 1 and SQ2) performs best in a practical setting?

This question aims to evaluate the methods that are found in this research. This will be done by executing each method on the same part of an organization. In theory, the result of each test should contain similar business processes. The result of these tests are evaluated using the requirements stated in SQ3. The evaluations can be compared in order to decide which Process Discovery method performs best in a practical setting.

Scientific Contributions

The aim of this research is to improve the success rate of RPA projects by ensuring the process that has been chosen for RPA, will be a suitable process. Contributions to this goal and the scientific base have been as follows. The first contribution to the scientific base is that different methods for Process Discovery have been tested in the context of RPA, including Process Discovery methods that have never been used in RPA projects before. To be able to decide if the methods perform well, a set of requirements has been created tailored for Process Discovery in the field of RPA. These requirements can be used in future research to test methods for Process Discovery that have not yet been covered in this research (van der Aalst, Bichler, & Heinzl, 2018). The final contribution is that through this research multiple Process Discovery methods have gotten clear guidelines for using them in RPA, where before this research there was only one method that had been explained thoroughly (Agaton & Swedberg, 2016). The methods as explained in this research, are ready to be used in other studies.

2: Research methods

In this section the execution of the research using the Design Cycle by Wieringa (2014) will be explained. Design in the context of Information Science is considered to be both an iterative process and a resulting product (Hevner & Chatterjee, 2010). Design Science is defined by Hevner (2010) as *'a paradigm in which a designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing to the body of scientific evidence. The designed artefacts are both useful and fundamental in understanding that problem'.* The method for Design Science as proposed by Hevner (2010) puts emphasis on creating new artefacts and by doing so, contribute to the body of scientific evidence. Wieringa (2014) has a similar but slightly different definition of Design Science*: 'Design Science is the design and investigation of artefacts in context. The artefacts we study are designed to interact with a problem context in order to improve some-thing in that context'*.

For this research, I use the Design Cycle as proposed by Wieringa (2014). This method puts emphasis on creating a solution within a specific context, which is relevant for this research because goal is to assess existing methods in the new problem context of RPA. Another approach of the Design Cycle is proposed by Hevner (2010). However his method puts focus on creating new artefacts to approach a problem (Hevner & Chatterjee, 2010), while this research aimed to test existing method of Process Discovery in the context of RPA. Another argument for choosing Wieringa (2014) over Hevner (2010) is that Wieringa (2014) provides solid guidelines for conducting an experiment in a real world setting, which is important for this research in order to answer sub question 4.

The Design Cycle consists of three phases and is a continuous cycle, which therefore enables continuous improvement for a developed method (Wieringa, 2014). The cycle starts with a Problem Investigation where the problem is identified and elaborated upon. In the second phase, Treatment Design, a solution for the problem is proposed. This phase is also used to set requirements for the treatment. The third phase is the validation of the treatment. In this third phase, the proposed treatment is tested and compared to other treatments. If the requirements are sufficiently fulfilled, the treatment will be implemented in the fourth phase of the Design Cycle. When this is done, a new Problem Investigation could be performed and the cycle could start again if necessary.

This research used the Design Cycle, which is part of the Engineering Cycle. The full Engineering Cycle was not used because implementing an RPA software robot was outside the scope of this research. The activities that have been executed during the Design Cycle are explained below. A summarized version of the activities can be found in Figure 1.

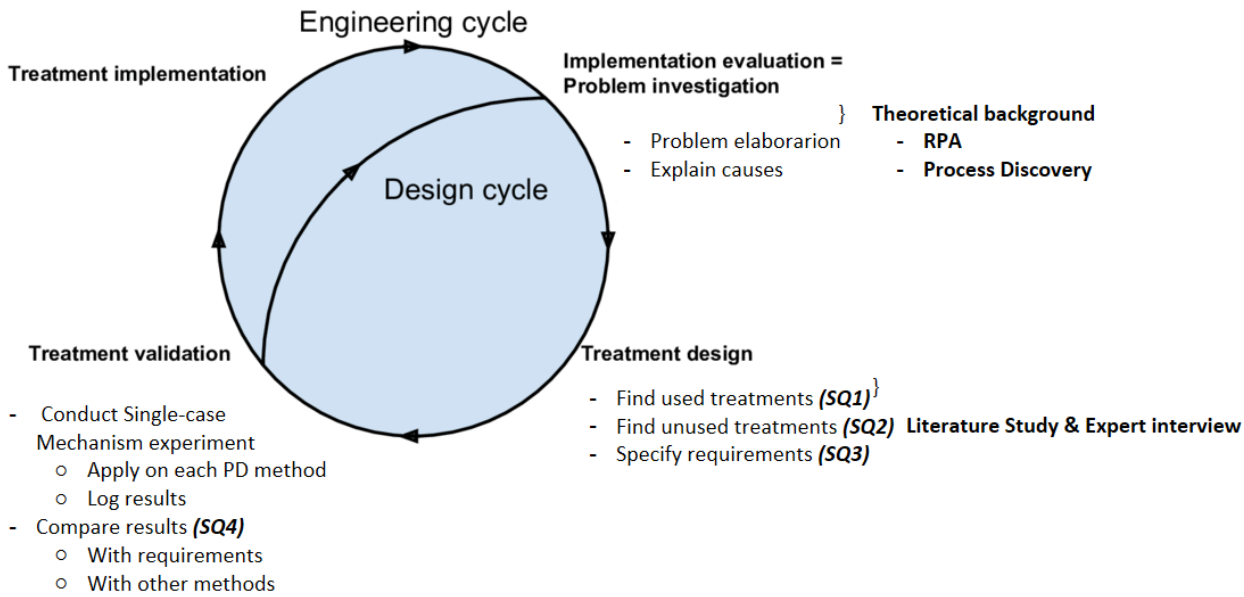


Figure 1: the Design Cycle as used in this research, indicating how the cycle is used to answer all the research questions

Phase 1: Problem Investigation:

The Problem Investigation is covered in the theoretical background of this research. In the theoretical background, RPA is defined and elaborated upon. This resulted in a better understanding of the topic. The theoretical background did also focus on Process Discovery, because Process Discovery is an important part of the problem context.

Phase 2: Treatment Design

First, we have looked into literature to find Process Discovery methods that have already been used in an RPA project. Second, Process Mining techniques have been listed and evaluated as possible treatment for Process Discovery in RPA. This was also done through literature research. The methods that have been found were formalized using a Process Deliverable Diagram (Van De Weerd & Brinkkemper, 2009). Using a PDD, both the activities and deliverables have been modelled and are therefore more clear, which makes the third phase of this research more reliable.

The requirements for the treatment have been created. The requirements are be based upon the process characteristics of RPA processes. The process characteristics were needed for a proper judgement of the RPA eligibility of a process. There are also non-functional requirements that adhere to the problem statement (concerning time to delivery and practical use). This was initially done through literature study and followed by an expert interview to validate the created set of requirements.

Phase 3: Treatment Validation

The Treatments that are described in Phase 2 have been tested in a practical setting. This was done through a single-case mechanism experiment (Wieringa, 2014). '*A single-case mechanism experiment is a test of a single case in which the researcher applies stimuli to the case and explains the responses in terms of mechanisms internal to the case*' (Wieringa, 2014). Single-case mechanism experiments are considered useful for validation research and for tests in the 'real world'. The problem case (finding processes) is exposed to stimuli (a single treatment) and the responses are observed. This was done by creating an experiment protocol for which various checklists from Wieringa (2014) have been used. The observed responses were logged in an 'experiment log' and are used to explain what happened.

3: Validation Design

In this section, the execution of the experiments and validity of this research will be elaborated upon.

Context

The case on which the methods will be tested, will be introduced in this section. As mentioned in Section 1.2, the aim of this research is to test different methods to find the right processes to use in an RPA project. To be able to fairly compare the different methods that are described in section 5, they will be tested on a single case.

The company where this thesis was written (InfoSupport), has a wide variety of applications running within the organization. However, only one of these applications turned out to have to proper event logs that are needed for the Process Mining technique, which is one of the methods chosen to test in this research. This application is used inside the HR recruitment department. The HR recruitment department uses Microsoft CRM for their daily processes. They mainly use it for recruitment purposes. All the event logs that are available for this research, will therefore be about the recruitment process of the company.

The main focus of HR recruitment is on finding new employees. Their business is everything from first contact with a potential new hire, up until the moment the new employee is signed on. Everything HR does within this period of time, is logged within the CRM. This research will focus on this process and will try to use different Process Discovery methods to find out if the process, or parts of the process, are RPA eligible.

Conditions

The conditions of this research, will be the methods that are tested. As explained in section 2, these methods were be found in a literature study. The methods for Process Discovery that have been used in this research are fully explained in section 5.

Procedures

The Design Cycle as explained in section 2 provided the order as to which the different elements of this research have been executed. First, a literature study has been conducted to find methods for Process Discovery. A PDD of each method has been constructed to use as a guideline for the experiments. Before the start of the experiments, the requirements have been set and evaluated with an expert in the field of RPA. After this, each experiment has been conducted. The Process Discovery methods that involves Process Mining have been performed first, followed by the Process Discovery method that involves interviews. All experiments have been executed according to the PDD's that have been created in section 5. The execution of each of these experiments resulted in an visualization of the recruitment process at InfoSupport. Then for each requirement has been checked if a corresponding element to this requirement could be found on the visualization created by the experiment. If the tested methods for Process Discovery were adequate, could be derived from these checks. As a final step, the procedure and findings of this research have been presented at InfoSupport in a 1.5 hour session, where domain experts of the recruitment process and people with experience in Process Mining were present. The feedback provided through this session has been included in the Discussion, but no critical errors have been found.

Measures

The measures used to decide which method for Process Discovery in RPA performs best, are all created in this research. The measures are the requirements that are set for Process Discovery in RPA and the visualization of the process that has been created in each experiment.

Threats to Validity

In preparation of the experiments, a preparation table was created to help guide the experiments and to limit threats to the validity of this research. A full overview of all the answered questions, as stated by Wieringa (2014), can be found in Appendix 5. A short summary of the most important threats to the validity of this research and how these are limited, is given here.

To guide the experiments and to make sure the methods that were tested, were properly executed, a PDD was created of each method (Van De Weerd & Brinkkemper, 2009). Using the PDD during the execution of the test should help me as a researcher to correctly execute the methods, and thus help to ensure fair and reliable results.

The requirements created in this research are essential to decide which method performs best in the conclusion. To make sure these requirements are accurate and fair, they have been verified by an expert in the field. This has helped me in removing some redundant requirements and deciding which requirements are more important than others.

The order in which the experiments have been executed, was also established upfront, to minimize knowledge of results of one method to influence the execution of the other methods. It was decided that the methods using Process Mining were executed first and the method using interviews was performed last. This order of execution ensured best that knowledge about the process obtained through the execution of a method, would not interfere with the execution of the next method.

The last threat to the validity of this research, is that I, as researcher, prior to this study had no experience with Process Mining. To make sure this lack of knowledge would not interfere with the results of this research, an online Process Mining course has been followed on Coursera. In addition to this, practice with Process Mining with the PROM tool has also been done and a course on how to use the screen capture tool (for more information see section 5) was. Finally, the entire process of executing the experiments has been explained to colleagues at InfoSupport, who have knowledge of the recruitment process or experience with process mining. This step was used to verify the execution of this research and explore any missed steps, or parts of the execution that could have been done better.

4: Related Literature

In this section, related literature to RPA and Process Discovery will be explained to come to a better understanding of the topic. This knowledge will be used later in this thesis to ensure that the Process Discovery methods that are evaluated, perform well in the context of an RPA project.

4.1: RPA

The core of RPA is to replace human tasks in business processes by software bots that will do the tasks for them (Accenture, 2018; Bruno, Johnson, & Hesley, 2017; CapGemini Consulting, 2016; Lacity & Willcocks, 2018; Wright, Witherick, & Gordeeva, 2017). '*RPA deals with smart software to do high-volume, repeatable tasks that usually take humans an unbearable length of time to accomplish and which they typically find mundane to perform*' (Institute for Robotic Process Automation, 2015).

Using RPA business processes are automated from the bottom up, instead of from the top down. RPA technology uses the same interface as a human user would use for executing a business process. This implicates that business processes remain the same, although they are not performed by humans anymore, but by software (Agaton & Swedberg, 2016; CapGemini Consulting, 2016; Juntunen, 2018; Lacity & Willcocks, 2018; Lacity et al., 2015). RPA can be applied to a complete process, or also to small parts of the process. The barrier to implement RPA is low, because RPA can be performed on small processes or tasks while not changing the existing business process (Willcocks & Craig, 2015).

Generally speaking, there are two kinds of software robots that can be developed using RPA technology: unattended and attended.

Unattended software bots are designed to stand alone and automatically execute tasks in the background. The time it takes to integrate and automate unattended bots is usually larger, because the complexity is sometimes underestimated. Another risk with unattended software robots is that they become part of a stack of apps within an organization which can be hard to keep track off (Trefler, 2018).

An attended RPA software bot is set up to run alongside a human controller. Some tasks of the process are performed by the RPA software robot, while the more complex tasks are still performed by a human employee. This is a way of preventing over complex RPA projects, while still being able to improve business processes and increase employee satisfaction (Trefler, 2018). Attended software robots seem to be favourable over unattended software bots, because implementation is less risky (CapGemini Consulting, 2016).

RPA Process Properties

Many sources state that the nature of RPA is to automate tedious and repetitive tasks. This should leave human employees to spend their time on more complex tasks. It is more difficult to define the nature of the tasks however. It is also claimed by Accenture (2016) and CapGemini (2018) that tasks that are most suited for RPA are simple and should follow clear business rules. Other sources claim that RPA is rapidly evolving and could even be seen as an elevation of Artificial Intelligence and Expert Systems (Institute for Robotic Process Automation, 2015). Properties of processes that are stated most often as RPA suitable are listed here and explained below:

- Low complexity of tasks (Lacity & Willcocks, 2018; Lacity et al., 2015)
- High number of repetitions (Lacity & Willcocks, 2018; Slaby, 2012; Wright et al., 2017)
- Multiple systems are involved (Institute for Robotic Process Automation, 2015; Lacity & Willcocks, 2018; Rutaganda et al., 2017)
- The process follows clear business rules (Accenture, 2018)

- Stable environment (Fung, 2014; Slaby, 2012; Sutherland, 2013)
- Limited need for human intervention (Fung, 2014; Slaby, 2012)
- Structured data (Accenture, 2016, 2018; Wright et al., 2017)

Low Complexity

Cognitive capabilities of RPA solutions are low. Therefore tasks that are suited to be automated by using RPA solutions should not be too complex. Robots are not yet able to think as humans do and are for now best suited to perform simple tasks (Agaton & Swedberg, 2016). Business processes as a whole are often too complex, because many decision points, actors, events and activities can be involved, and, when combined in certain ways, can lead to many different outcomes (Dumas, La Rosa, Mendling, & Reijers, 2013). Automating only a part of a business process, let's say only a single decision point, will lead to a decrease of complexity and will thus increase the chance of successfully implementing RPA.

High Number of repetitions

RPA solutions are best suited for processes or parts of processes that have a high number of repetitions. This is not so much because an RPA solution cannot be successfully implemented on a process that only rarely occurs, but it is deemed hardly cost effective to automate processes that occur not that often. This does not mean that automating rare processes is always cost in-effective, however since RPA should focus mostly on simple processes rather than complex processes to be successfully implemented, it is considered a best practice to implement RPA on often occurring processes. By automating processes that often occur, the business case will be stronger because more FTE can be saved. It is difficult to determine exactly when a process is 'often occurring'; some sources state that the higher the number of occurrences, the greater the chance of success (Wright et al., 2017), while other sources state that the number of repetitions does not have to be extremely high in order for the project to be successful (Institute for Robotic Process Automation, 2015; Rutaganda et al., 2017; Willcocks, Lacity, & Craig, 2015). It probably depends on the particular case if the number of repetitions of a process is a leading indicator of how successful the RPA project might be.

Clear business rules

When a process has structured business rules, implementation of RPA is usually more successful. RPA solutions are bad at handling unexpected situations (Slaby, 2012). All possible outcomes need to be programmed into the robot beforehand, because when a situation occurs that is not predefined a human will need to intervene. A successful implementation of RPA should need as few interventions as possible to increase efficiency and thereby profitability. When a process has clear business rules, the rules can be implemented in the software robot during the developing of an RPA software robot (Accenture, 2018). The processes have to be rule based, predictable and replicable on such level that it is definable all the way down to key-stroke level in order for the process to be properly configured in the RPA software (Burgess, 2018) . Implementing the business rules beforehand will decrease the need of adjustments and human intervention and thus increase the success of the RPA implementation.

Multiple involved systems

RPA is exceptionally effective in situations where multiple systems are involved. Situations where an employee has to access multiple systems during the completion of a task are strong candidates for RPA. Such processes are also referred to as 'swivel-chair' processes and are basically copying and pasting information between different IT systems (Geary, 2016). Such processes are prone to human error and are therefore a strong candidate for automation in general. Traditional IT solutions would require application or data layer integration, which is considered to be much more complex and will increase costs, risks and chance of failure (Slaby, 2012). RPA solutions are considered to be much more lightweight and do not require these complex integrations, because the RPA software bot works on top of the front-end user interface, therefore diminishing the risks that are involved with traditional IT solutions (Strömberg, 2018).

Another motivator to consider processes that involve the use of multiple systems, is that these processes are particularly vulnerable to human errors (Fung, 2014). After software robots take over the execution of the process, such errors should not occur anymore (Sutherland, 2013).

Stable environment

Every time the context in which the software robot operates changes, the software robot will have to be reprogrammed. Software robots are not good at handling changes, and with each reprogramming the costs of the project will increase and risks of failure will occur more often. Therefore it is best for RPA projects to operate in a stable environment, where changes will occur as minimal as possible. Doing this will increase the chance of having a high return on investment (Rutaganda et al., 2017).

Limited need for human intervention

A process considered for RPA should not contain parts that require human judgement to make a decision. Current RPA solutions are not yet ready to make complex decisions, only decisions that are rule based. If there are many possible exceptions within a process, an RPA solution will become more costly to implement, the process to create and test the software robot will take longer and the chance of errors will increase. It is claimed that AI solutions can handle these complex decisions that pose a problem for RPA (Slaby, 2012; Sutherland, 2013). However no examples of actual RPA implementations that are supported with AI are available.

Structured data

Data that are used for processes, should be structured in order to automate the process using RPA. RPA needs to be modelled using simple if/then rules, which can be best applied to structured data (Wright et al., 2017). The importance of using structured data is not mentioned by scientific sources, only by companies. It is even stated by the company Accenture (2016) as the first factor to consider about a process when implementing RPA. When a process does not use structured data, an RPA process will simply not be successful.

RPA Advantages

If RPA is performed in the right processes with the right properties as described above, there are many possible benefits for a company. It differs as to which of these benefits are most attractive for a company, depending on the company goals and vision. The most stated benefits are listed and explained below:

- Cost reduction (ACCA Global, 2015; Fernandez & Aman, 2010; McKinsey, 2016; Willcocks et al., 2015)
- Quality improvement/error reduction (CapGemini Consulting, 2016; Institute for Robotic Process Automation, 2015; McKinsey, 2016; Willcocks et al., 2015)
- Focus on value adding work (Lacity & Willcocks, 2018; Lamberton & Brigo, 2017; Strömberg, 2018; Tarquini, 2018)
- Relatively easy to implement (Asatiani & Penttinen, 2016; Burgess, 2018; CapGemini Consulting, 2016; Lintukangas, 2017)

Cost reduction

Cost reduction can be achieved in two different ways. First, and maybe the most evident way of reducing costs, is savings on FTE. If RPA is applied on suitable processes, examples of two human controllers controlling 300 software robots that have a similar output as 600 people would have, are not uncommon(Lacity et al., 2015; McKinsey, 2016). This leads to massive reduction in costs. RPA technologies have proven to cut the cost of human related recourse spending by 20% to 50% (ACCA Global, 2015; Fernandez & Aman, 2010; Institute for Robotic Process Automation, 2015). In short, RPA reduces the cost of manual tasks by being more efficient when executing a business process (Fung, 2014).

Using RPA, another way to reduce costs is on outsourcing. The tasks that are typically candidate for outsourcing, are also tasks that are strong candidates for RPA. The advantages of implementing RPA over outsourcing, is that the tasks are performed in-house, which reduces complexity of managing remote operations (Slaby, 2012).

Focus on value adding work

RPA automation focusses on simple, mundane and repetitive tasks. By taking these tasks away from human employees, they can focus on more value-added activities that involve personal interaction, problem solving and decision making (Institute for Robotic Process Automation, 2015). RPA robots should be seen as a tool that enables human employees to do work that is more valuable to the company in any sorts of way (Accenture, 2018; CapGemini Consulting, 2016). Employees welcome the technology of automation through RPA because '*they hated the tasks that machines now do, and it relieved them of the rising pressure of work'* (McKinsey, 2016).

Better and faster results

Humans make errors. These errors consist of incorrect data inputs, missed steps or mistakes in how rules are applied to the process (Accenture, 2016; Strömberg, 2018). This might be caused by the dull nature of tasks that are suitable for RPA. Software robots will not make these errors. When RPA is applied to the right processes, tasks are expected to achieve 100% accuracy (Lamberton & Brigo, 2017; Tarquini, 2018).

Not only to quality of the output of the process should increase by reducing the errors, but the speed at which the process is performed will be increased as well. Software robots are able to work 24/7, providing reliability and continuity of service (Accenture, 2018).

Easy to implement

RPA is, especially when compared to traditional IT automation solutions, considered a lightweight implementation. RPA is kept simple by providing easy configuration and a simple/intuitive interface for users (CapGemini Consulting, 2016). Automating processes using RPA technology does not require extensive knowledge of programming, employees are able (with the help of an RPA software provider) to learn how to automate processes within weeks and can automate processes independent in a matter of months (Asatiani & Penttinen, 2016). Some examples even range from three weeks (BluePrism, 2011; Slaby, 2012) to six weeks (Lintukangas, 2017).

RPA risks

As mentioned in the problem statement, not all RPA projects succeed. It is important to understand why RPA projects fail, because a failure could have often been prevented already in the first phase by picking the right processes to automate and presenting a solid business case (Rutaganda et al., 2017). The following pitfalls are explained so they can be better understood:

- Unrealistic expectations (Lacity & Willcocks, 2018; Linden & Fenn, 2003; Trefler, 2018)
- Picking the wrong process
- Getting employees on board

Unrealistic expectations

RPA is still quite a new phenomenon. RPA is now at the peak of its popularity, which comes with overhyped expectations and promises that cannot be fulfilled according to the Gartner Hype Cycle (Linden & Fenn, 2003). At this stage of the hype cycle, one should not join in, because everyone is joining in. Unrealistic expectations are partly caused by RPA vendors, who spend more on advertising and marketing instead of actually improving and building new automation capabilities (Lacity & Willcocks, 2018). The problem of unrealistic expectations in RPA is also referred to as 'the big RPA bubble' (Trefler, 2018) or 'RPA washing' (Lacity & Willcocks, 2018).

Picking the wrong process

One of the main reasons for the failure of RPA projects is due to picking a process that is not eligible for RPA. This is partly due to unrealistic expectations as described above. Companies think RPA can solve anything and try to automate processes that are too complex for RPA. This results in a software robot that does not work properly, because RPA is not yet able to handle complex processes. It is advisable to start with small and easy processes or tasks to ensure immediate success (Institute for Robotic Process Automation, 2015). In this way, the company is able to get familiar with RPA technology and is able to better understand what and what not to automate with RPA technology. As already mentioned in the problem statement, there are hardly any described methods as to how to find the right process to automate. It is known what properties typical RPA processes possess, these properties are described by many sources and are summarized in section 4.1 of this research. It is interesting to note that although the process properties of RPA seem to be well known, there are still errors being made when judging if a process is suitable for RPA. This indicates the necessity of this research and implies that having a link is vital between the known process properties of RPA and the method that is used for Process Discovery.

Getting employees on board

Some RPA projects fail because employees of a company view RPA technology as a threat. This could be either due to the fear of losing your job because software robots will replace your position, or IT departments that think of RPA technology as simple or dangerous (Willcocks & Craig, 2015). In one example, an RPA pilot was flagged as a security breach by the IT department, because the process had become so efficient and fast that the IT department thought they had been hacked. This almost caused the person in charge to lose his job, while in fact the pilot was extremely successful (Willcocks & Craig, 2015). This illustrates the importance of communicating well and making sure everyone within an organization sees the value of RPA.

4.2: Business Processes

In order to be able to understand how processes work in the context of RPA, it is important to first establish the basics of what a business process entails. In this section, theory about (discovering) business processes will be linked to theory on RPA. A business process is defined as follows: '*a collection of inter-related events, activities and decision points that involve a number of actors and objects, and that collectively lead to an outcome that is of value to at least one customer*' (Dumas et al., 2013).

Business Process Management

Business Process Management (BPM) is about managing and improving business processes within an organization. It is defined as *'Body of principles, methods and tools to design, analyse, execute and monitor business processes, with the aim of improving their performance'* (Dumas et al., 2013). The goal is either to make business processes more cost efficient, less time consuming, or to reduce error rates. (Dumas et al., 2013). Robotic Process Automation is a method to make business processes or tasks less time consuming and more cost efficient, and it should reduce error rates as explained in section 2.1. The relation between BPM and RPA is, that BPM has its focus on managing the entire chain of events of a process as a whole, while with RPA the focus can be on a single task where the most profits might be achieved within a process. RPA could therefore be seen as a valuable tool within a broader context of BPM, where RPA complements BPM (van der Aalst et al., 2018; Willcocks & Craig, 2015).

How BPM works in practice is explained through the BPM lifecycle as can be seen in figure 2. Walking through the full cycle will help a company to identify and improve business processes. The cycle is continuous, meaning that BPM is never really finished but seeks to always improve upon business processes (Dumas et al., 2013). The execution of an RPA project could be very well compared with the Business Process Management Lifecycle. The BPM Lifecycle starts with Process Identification, followed by more in depth Process Discovery. This is followed by Process Analysis, where insights on the process are gathered. The processes that are chosen are then redesigned. The new process is implemented and will be monitored and controlled. These steps should have changed the way the company works, and the cycle starts again at Process Discovery of the BPM lifecycle (Dumas et al., 2013).

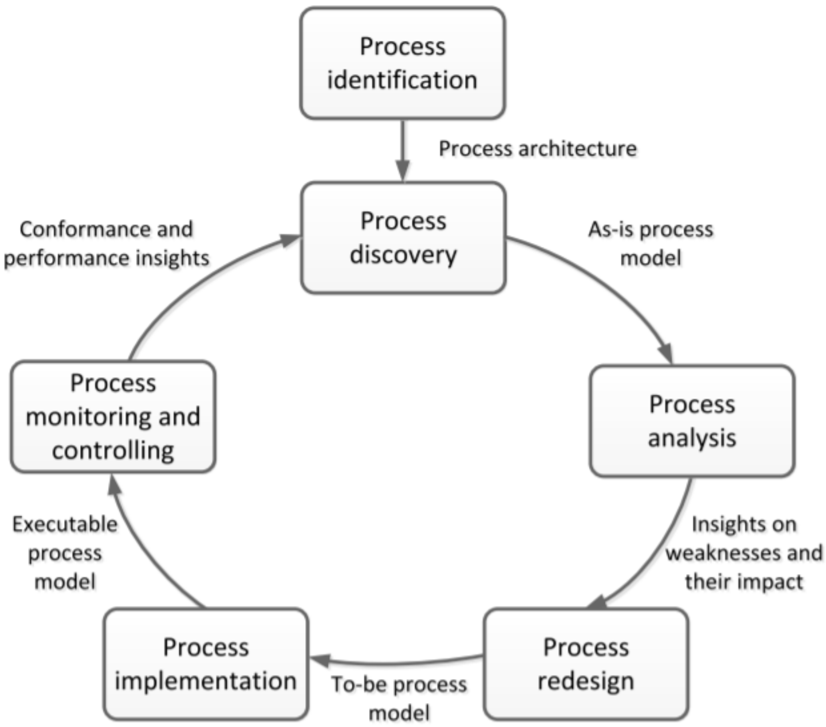


Figure 2, the BPM lifecycle

In my research, the first phases of the BPM Lifecycle are most relevant and will be explained in full below. The other phases of the BPM lifecycle and their relation and differences to the execution of an RPA project will be explained here.

- **Phase 1 (Process Identification)**: identifying processes for an RPA project and for BPM goes quite similar. The goal is to get an overview of the processes of a company, and a first choice on which processes to improve, can be made.
- **Phase 2 (Process Discovery):** When a selection of processes has been made, the processes should be discovered more in depth. This is done in the Process Discovery phase of BPM. BPM offers various methods to discover a process more in depth

- **Phase 3 (Process Analysis):** This phase involves analysing the potential impact of a BPM or RPA project. For RPA projects, it is important to be able to present a strong business case to get the potential client on board (Asatiani & Penttinen, 2016). This is more crucial for RPA, as it is preferable to automate non critical processes as opposed to critical processes in BPM. For non-critical processes is it more difficult to understand the necessity of investing in improving those processes. Failing to do so could result in a difficult implementation of RPA. Another suggestion is to start small with an initial RPA project. Only automate a few small and simple processes using RPA, so the client will get familiar with RPA technology and is better capable to avoid pitfalls in future implementations (Rutaganda et al., 2017).
- **Phase 4 (Process Redesign):** The process of creating an RPA software bot is usually quite straightforward and takes little time. An RPA vendor such as Opus Capita or BluePrism has tools ready that are used to model the process and its business rules, which can be done by using software development kits or desktop recorded automation (Geary, 2016). Using such a 'click and play' tool is an important contributor to the short time to delivery of RPA. When tasks become more complex, it takes more time to model all the business rules and outcome in the RPA software robot and increases the risk of errors during construction.
- **Phase 5 (Process implementation):** implementing RPA software bots will cause some disruption within an organization, not because business processes change (they do not), but because processes are executed much more swiftly and employees have to find other tasks. It will also become clear where human intervention is still needed. Errors can manifest themselves in unexpected ways, such when Business Rules are changed or when the user interface is changed (ACCA Global, 2015) .
- **Phase 6 (Process monitoring):** Monitoring the software bots is important. Even stable business processes change over time, and RPA software robots will not automatically adopt these changes. Errors could of course always occur and should be handled accordingly by human operators. For these reasons it is important for a company to properly monitor the deployed software robots when implementing RPA.

There is no literature on how the Process Identification and Process Discovery phase of BPM relate to finding suitable processes for RPA. Because the close relation between RPA and BPM as explained above, the Process Identification phase and Process Discovery phase of BPM will be explained in relation with RPA below.

Process Identification

The goal of the Process Identification phase is to find and log the main business processes. First, the Process Architecture of a company is created. From there, the most vital processes can be selected based on importance, health of the process and feasibility, meaning which processes are most susceptible to successful Process Management. In relation to RPA, Agaton & Swedberg (2016) propose a similar method to identify processes. The selection criteria differ and are more tailored to suit RPA projects. Instead of picking the most vital processes in BPM (Szelągowski, 2018), non-complex and non-vital processes should be picked for RPA projects (Agaton & Swedberg, 2016). There are several levels on which processes can be identified:

- Level 1: process landscape with process groups, grouped in management processes, core processes and support processes.
- Level 2: containing process subgroups of a process group from level 1
- Level 3: containing the main processes of a process subgroup from level 2
- Level 4: containing the tasks of a main process. This is more often done in the Process Discovery phase.

RPA does not necessarily aim to automate the most vital processes. Automating such processes can be a risk. It is better to apply RPA to non-vital processes for the company (Agaton & Swedberg, 2016). A first choice of which processes to explore for BPM is made early, in a stage where processes are only modelled on a high level (Szelągowski, 2018). For RPA it is important to judge a process on a task level, because RPA does not necessarily operate on a process wide level (van der Aalst et al., 2018), which is at level 4 on process modelling in BPM.

Process Discovery

When the most vital processes have been selected, a deeper understanding of those processes can be achieved by modelling them into more detail using the Business Process Modelling and Notation (BPMN) (Dumas et al., 2013). This is only done for a selection of the processes, because modelling a process takes a lot of effort (Asatiani & Penttinen, 2016). Getting the right information in order to correctly model the process can be obtained in multiple ways, including automated methods such as process mining (van der Aalst et al., 2012a; Dumas et al., 2013):

Evidence based

Evidence based Process Discovery can be done in three different ways. The first option is to look at existing documentation that can be related to the business process. This method is never mentioned however in combination with RPA in case studies. The second option is observation, where an individual case of a process is followed in order to understand how the process works, either as observer or as customer. This is similar to an RPA case described by Asatiani & Penttinen (2016) where RPA consultants follow employees of the client company to discover potential eligible RPA processes. The third option of evidence based Process Discovery is automated Process Discovery. This method uses event logs from information systems (Dumas et al., 2013). These logs can be used to re-engineer business processes and can provide the full range of business processes exactly the way they are executed (W. Van Der Aalst et al., 2012). Using event logs has not yet been a method for Process Discovery used in an RPA project. A similar method however has been used, where the company created their own event logs through the use of a keylogger tool (Gartner, 2016). This tool that is used by the company KRYON, is used to track the actions of an employee and mapping these into a complete overview of the processes within a company.

Interview based

With interview-based discovery, the details of a process are discovered through different interviews with domain experts and stakeholders (Dumas et al., 2013). Interviews can be conducted in different forms and in different order with the right domain experts. These decisions are very case dependant (Dumas et al., 2013). This might explain why numerous sources state that interviews have been used in an RPA project (Agaton & Swedberg, 2016; Willcocks & Craig, 2015; Willcocks et al., 2015), however none of those formalized or explained how this process has been conducted. Formalizing this process might not be relevant because it is different in every case. Important to consider in the context of RPA is that one of the potential pitfalls of conducting interviews, is that people tend to describe the normal way of processing and forget to mention exceptions. One of the main factors for RPA is that it is not good at handling exceptions as mentioned in section 2.1. When mapping the processes, these exceptions should be known to be able to properly assess if a process is eligible for RPA.

Workshop based

The last method for Process Discovery as described by Dumas (2013) is through a workshop where multiple participants together discuss on how a process is executed within a company. It-usually takes multiple sessions to come to a detailed process model. Opus Capita uses this approach in their RPA projects, where at the start of each project a workshop is held to find possible RPA candidates

(Asatiani & Penttinen, 2016). The scope of their approach is a bit broader, because it not only focusses on finding business processes but also on getting employees enthusiastic for the project.

A summary of these findings can be found in table 1.

Table 1: summary of Process Discovery methods

Type	Method	Advantage	RPA examples	RPA concerns	Sources
Workshop	Workshop	Input from different levels	Opus Capita	More onboarding	Asatiane & Penttinen (2016)
Interview	Unstructured	Explanation from domain expert	-Research cases -Business cases	Risk of leaving out exceptions, work intensive.	Agaton & Swedberg (2016) Willcocks & Craig (2015b) Willcocks et al. (2015)
Interview	Structured	Explanation from domain expert	No cases found	Could miss vital information on process	Willcocks & Craig (2015) Willcocks et al. (2015)
Evidence	Keylogger	As-is, not as should be	-Kryon with Keylogger	Keylogger is intrusive	Gartner (2016)
Evidence	Process Mining	Data is there, could be fast and straightforward method	Mentioned by UIpath & Celonis	Data does not always encompass entire process	Van der Aalst et al. (2018)
Evidence	Shadowing	See what happens, able to interact	Opus Capita uses shadowing in cases	Work intensive	Asatiani & Penttinen (2016)

The result of any of the above described methods, are multiple processes which are modelled all the way to task level detail. The method proposed by Agaton & Swedberg (2016) works in a similar way, where only the most likely successful processes are modelled. After the processes are already filtered in the Process Identification phase, they are further filtered down, based on an estimation of potential time savings, employee satisfaction, the importance of accuracy and the availability of the process (Agaton & Swedberg, 2016). A full overview of this method will be given in section 5.1.

4.3 Modelling Languages

There are many languages in BPM that can be used for modelling a process. A short overview will be given below, with all the properties of a process that can be modelled using the specific language. These properties should match the properties that RPA cases typically have, so a judgement of RPA eligibility can be made using the model without additional information. Table 2 helps deciding which type of models to use when preparing the experiment in section 5. How the different available models correspond to the important process properties of RPA is summarized in table 2 below. The following sources have been used to create this table: (Chinosi & Trombetta, 2012; Vergidis, Tiwari, Majeed, & Telecom, 2008; White, 2004)

Table 2: an overview of numerous modelling languages and how they visualize RPA process properties

Modelling language	Description	Attributes	Task Complexity	Number of repetitions	Systems involved	Business Rules	Environment	Human Intervention	Data type
BPMN	Standard language for process modelling	Events, Activities, Decisions, actors and outcome	Yes	No	Yes	No	No	Yes	No
BPMN-R	Standard language for process modelling, with an extension extra information on data	Events, Activities, Decisions, actors, data being handled and outcome	Yes	No	Yes	Yes	No	Yes	Yes
Flow Chart	Sequential flow of actions	Everything in single model, gets really big. No subprocesses visible	Yes	Yes	No	yes	No	No	No
DFD	Flow of data from one place to another	Data oriented	No	No	Yes	No	No	Yes	Yes
Petri Net	Graphical oriented language	Number of processes which communicate and synchronize. Also used for simulation.	Yes	Yes	Yes	No	No	Yes	No

Gantt Chart	Activities and task with duration	Can model different layers, usually used in project management for planning purposes	Yes	No	No	No	No	No	No
IDEF	Structural graphical representations of processes or complex systems	What do I do models	Yes	Yes	Yes	Yes	No	Yes	No
Object Oriented Methods → UML	How to deal with objects of certain types	Object, state, behaviour, class, messages	Yes	Yes	Yes	Yes	No	Yes	No
Workflow technique	Flow of tasks between computer and human	Time, definition of tasks, implementation model	Yes	Yes	Yes	Yes	No	Yes	No

5: Treatment Design

In this chapter the setup of the research will be explained. First, the requirements that are used to compare methods for RPA Process Discovery will be stated and explained. Second, the methods that are tested will be stated and explained, using a PDD for each method. Thirdly, the case on which the different methods will be tested, will be explained.

5.1 RPA Process Discovery Requirements

In this chapter the requirements that are needed for a Process Discovery method in an RPA context will be specified. This is done beforehand, as to be able to objectively assess the Process Discovery methods that will be used later on in this thesis. The way in which the requirements will be described, is based on the IEEE 830 standard (IEEE Computer Society, 1998).

Goal of Requirements

As mentioned above, the goal of describing the requirements for a Process Discovery method in the context of RPA, is to be able to assess which method is the best Process Discovery method in this specific context. To do so, functional requirements will be created, based on described typical process attributes that can be found in literature and that have been described in the theoretical background of this thesis. Non-functional requirements will be formulated to ensure that the problem statement of this thesis is fully addressed by including usability metrics in the assessment of the different Process Discovery methods.

Product Functions

The function of the product (the methods that will be tested) is to discover processes in such a way that the eligibility of the discovered process for RPA can be determined. This means that the process characteristics that are needed to judge if a process is RPA eligible, need to become clear through the method that is used for Process Discovery.

User Characteristics

There are two kinds of stakeholders. The first stakeholder is the RPA implementor, meaning the company/employee or consultant who is responsible for creating the RPA solution. This stakeholder will mostly be interested in creating a working RPA solution. This stakeholder is called the ‘RPA vendor’.

The second stakeholder is the client company, which can be seen as the owner of a business process that could be candidate for automation with RPA. This stakeholder will mostly be interested in getting a solution that is worth the investment. This stakeholder will be called the ‘process owner’.

Requirements

The requirements are presented in the form of user stories. They will be divided in functional and non-functional requirements and will be ordered using the MoSCoW technique. Stating which requirements are more vital than others could be important when comparing different methods for Process Discovery after the experiments are finished and therefore this will be decided beforehand. The requirements have been validated by an expert in RPA with experience in RPA projects.

Functional Requirements

The functional requirements are focussed on the first problem as mentioned in the problem statement of this research ‘*It is difficult to find and pick the right candidate process for an RPA project’*. A suitable method for finding business processes, should lay bare the process properties that are relevant in an RPA context, in order to be able to correctly assess its suitability. The functional requirements are therefore focussed on these process properties, which can be found in the Related Literature part of this research. An overview of all the functional requirements can be found in table 3.

Must have

Functional requirements that are a 'must have', are requirements that are considered essential for a successful RPA project. When a Process Discovery method does not comply with one of these requirements, the eligibility of a candidate process cannot be evaluated. The process properties that the requirements refer to, are deemed essential by literature as described in section 4.1 for a working RPA software robot. When the requirements are not met, there is a chance that a process will be picked that will result in a failing RPA software robot.

Should have

The requirements in this section contribute greatly to picking the right candidate process, but are more focussed on a successful RPA software robot rather than a working RPA software robot. When these requirements are met, the chance of picking a candidate process that will result in a working RPA software robot, which will also greatly benefit a potential client, will increase.

Could have

The following requirements could help in certain cases to pick a suitable candidate process, but are not as essential to a working software robot and solid business case as the requirements that have been described above.

Table 3: requirements for Process Discovery in RPA

Name	Description	MoSCoW
Complexity 1	As an RPA vendor, I want to be able to assess the complexity of a task, so that a proper judgement of RPA eligibility can be made	Must
Complexity 2	As an RPA vendor, I want to be able to discover the different tasks that together form the process, so that the complexity of a process can be better assessed.	Must
Complexity 3	As an RPA vendor, I want to be able to discover different paths the tasks of a process may take, to be able to better understand the complexity of the process	Must
Business Rules	As an RPA vendor, I want to be able to know if processes follow clear business rules and what the logic behind those rules is, so that a proper judgement of RPA eligibility can be made	Must
Human Intervention	As an RPA vendor, know if the business logic is such that humans will always need to play a role in the decision making, so that a proper judgement of RPA eligibility can be made	Should
Data 1	As an RPA vendor, I want to know what type data sources and how many are involved, so that a proper judgement of RPA eligibility can be made	Must

Data2	As an RPA vendor, I want to know what the data type is, so that a proper judgement of RPA eligibility can be made	Must
Complexity 4	As an RPA vendor, I want to discover every possible step of a process, so that the software bot can be properly modelled	Should
Repetitions	As an RPA vendor, I want to be able to see the number of repetitions of a process, so that the business case of RPA can be assessed	Should
Process Stability	As an RPA vendor, I want to assess the environment and stability of the process, so that a proper judgement of RPA eligibility can be made	Should
Involved Systems	As an RPA vendor, I want to be able to see if multiple systems are involved in the process, so that a better judgement of the business case of RPA can be made	Should
Involved Employees 2	As a process owner, I want to be able to know which employees are involved in a process, so that impact on HR can be assessed	Could

Non-functional requirements

The non-functional requirements will focus on the second part of the problem statement*: 'Finding potential candidate processes is time consuming and often a manual process which could be improved upon'.* In order to be able to decide which method for Process Discovery is most practical, the following requirements will be used to assess which method performs best in the case of this research. All the requirements in this section are put in the 'should have' category, because none of them are essential for creating a working RPA software robot, but they are more convenient for the person that is responsible for finding a candidate process. The non-functional requirements can be found in table 4.

Table 4: non-functional requirements

Number	Name	Description	MoSCoW
NF1	Accuracy	As an RPA vendor, I want the discovered processes to be accurate, so that the RPA project can run smoothly	Should
NF2	Process overview	As an RPA vendor, I want to find as many processes as possible, so that the chances of finding suitable processes for RPA are increased	Should
NF3	Time to decision	As an RPA vendor, I want to find processes as quickly as possible, so that the RPA implementation process is speeded up	Should

NF4	Effort	As an RPA vendor, I want the processes elicitation phase to require as less work as possible, as to further increase efficiency	Should

5.2 Proposed Methods for RPA Process Discovery

The following methods are chosen to test in this research:

- Finding processes through interviews and describing them in a BPMN-R model as described by Agaton & Swedberg (2016). This is a method that exists already and has been executed
- Process Mining techniques with the use of event logs, with a focus on techniques for discovery and performance (van der Aalst et al., 2012b). Process Mining is an existing technique, however no sources could be found that relate RPA and the use of event logs in Process Mining
- Process Mining techniques using a screen capture tool. This technique for Process Discovery is similar to using Process Mining based on event logs, however the data source differs. This method has been used by various companies in practice and it is claimed to yield fast and effective results (Gartner, 2016)

5.2.1 Interviews

The method is thoroughly described in the PDD that can be seen in figure 3. A short description will be provided here for clarification of the model. This manual method is based on using unstructured interviews with employees of the client company to discover processes to perform RPA on. In literature on RPA, it is often mentioned that interviews are used to discover processes suitable for RPA. However, only one source explains how the discovery of processes in an RPA setting should be executed (Agaton & Swedberg, 2016). Because this paper is the only paper that thoroughly explains how the Process Discovery should be executed, the method they provide, has been chosen to be followed and test in this research.

The first step of this method is to find all high-level processes within a company through interviews. These can be divided in core, management and support processes. For all of these high-level tasks, an assessment on the complexity and essentiality for the company is made. Processes that are not too complex and not essential for the company are preferred candidates for RPA according to this method. The strongest candidates are chosen and move on to the next phase of this method.

In the second phase, the most eligible processes are discovered in more detail. This is done in two steps. In the first step, the RPA consultant should discover through interviews the following process properties:

- Satisfactory for employees where unsatisfactory processes are good RPA candidates
- Quality and accuracy importance, where processes that need high quality and accuracy are good RPA candidates
- Possibility on time savings, where processes where time could be saved are good RPA candidates
- Availability and flexibility, where processes that need to be available at all time are good RPA candidates

Based on these metrics, the strongest RPA candidates are chosen and the others are not considered anymore. For the second step, the strongest RPA candidates are modelled into a BPMN-R model. One or more processes that meeting the criteria best, are chosen as final candidates for RPA.

The ‘R extension’ of BPMN is used because it allows for specific modelling that is relevant for RPA. The R extension allows for features to be modelled concerning: (Agaton & Swedberg, 2016)

- Data quality and data source
- Shows different systems that are used during the process
- Allows for a specific description of decision points in the process

In chapter 3, table 2 it is shown that BPMN-R yields the most promising results, concerning the relevant process properties of RPA. Therefore, this modelling language is used to model the final RPA process candidate.

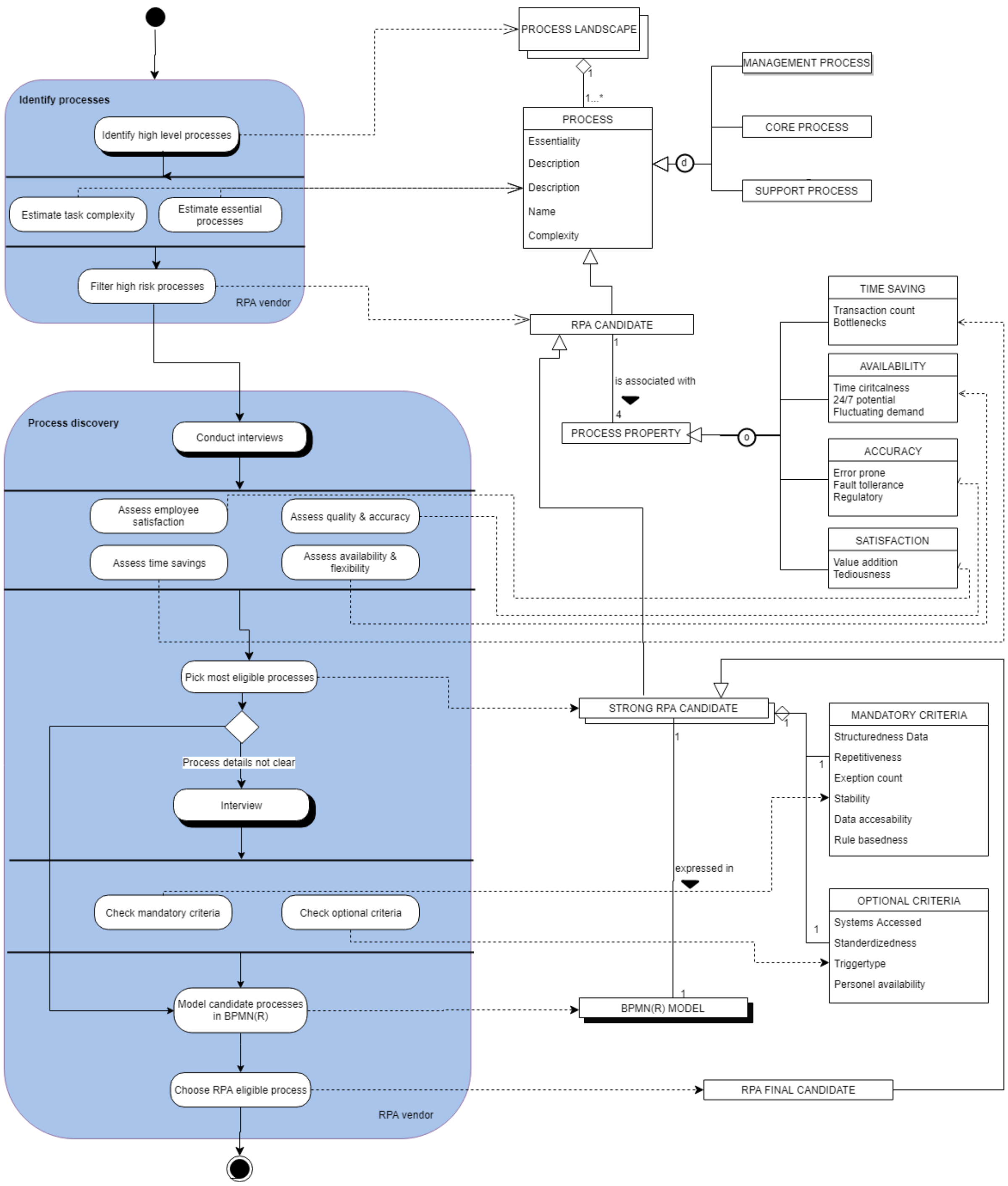


Figure 3, a PDD of Process Discovery through interviews

5.2.2 Event Logs

The second method tested in this research, is using event logs for Process Mining. Process Mining is able to recreate a model business processes through the use of event logs. When these logs are available, it could be possible that the manual process of conducting interviews is no longer necessary, which could speed up the phase of finding RPA eligible processes. The PDD that explains this method can be seen below in Figure 4.

As can be seen in the figure, this method seems much less complex. It consists of two main phases, where the first phase is collecting the data and the second phase is applying Process Mining techniques. The main advantage from using Process Mining techniques is that multiple types of models can be made of the same process, without having to do much extra work. This could give a better view on multiple angles of the process, which could contribute to a better understanding of the process and thus better judgement of RPA eligibility. In table 2, an estimation has been made if a certain modelling language is able to visualize the RPA process properties. Table 2 helped in deciding which modelling languages the Process Mining tool should use to create a visualization. Using the information from table 2 and the available modelling languages in PROM, a decision has been made to create the following models in order to be able to discover as many RPA process properties as possible.

- Petri Net: This should help finding bottlenecks in the process. As explained before, this is important for finding a business case.
- BPMN: A BPMN model visualizes many of the features that are important for choosing the right RPA process. In table 2, other languages show to perform also good in this matter, however BPMN is chosen, because BPMN is readily available in the Process Mining tool.
- Dotted Chart: This is a very easy visualization method that shows potentially interesting patterns in the data. This method was not covered in table 2, but was advised by an expert in the field because the patterns could potentially reveal bottlenecks or other interesting events.

Using these different models, an estimate of which processes are most eligible for RPA should be made, which will result in the final product of this PDD.

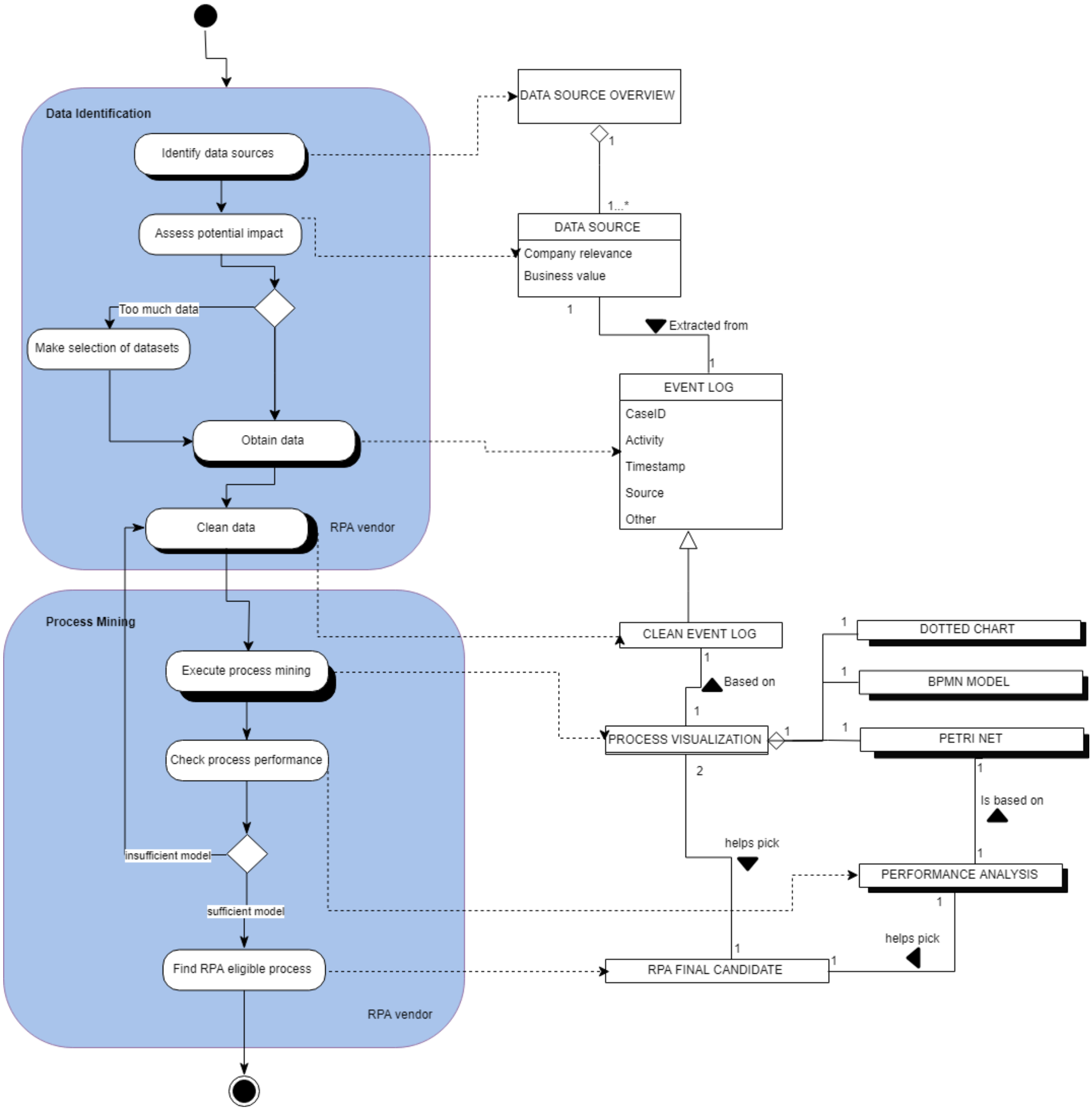


Figure 4, a PDD of Process Discovery for RPA using event logs for process mining

5.2.3 Screen Capture

The third and last method that is proposed in this research, is the use of a screen capturing piece of software. This software will follow and log all actions that users performs on their computer, and it is, in some instances, able to automatically decide which processes are most eligible for RPA (Gartner, 2016). This method is rather similar to the method of Process Mining. The major difference is the manner in which data are collected, Process Mining uses existing event logs while the data for this method are created by using a screen capture tool. The screen capture tool that will be used for this research, has been built by the University of Seville (Jimenez-Ramirez, Reijers, Barba, & Del Valle, 2019).

The similarities of this method to Process Mining using event logs are very well visible in the PDD. The main differences are in the data collection, where for this method the data have to be created by using the tool and that requires some more extensive preparation and cleaning, compared to Process Mining using event logs. The PDD for the method using the screen capture tool can be seen in figure 5. In the first phase, the screen capture software will be set up at the desktop of one or several employees. The screen capture tool should run for a number of days or weeks to gather enough data on all processes. In the second phase, the data have to be processed and cleaned. Depending on the nature of work, the data have to be interpreted and different processes have to be separated. The screenshots that the software will capture have to be analysed and processed by calculating the differences between each frame. This can be used to help to group the activities of a process. Lastly, the data have to be cleaned. Redundant activities that are not a part of a process need to be removed, such as Facebook visits that the key logger software will inevitably capture. When the data have been properly cleaned and analysed, the same techniques will be applied just like in the second method in this research. But, because the data are different, the results could very well be different as well.

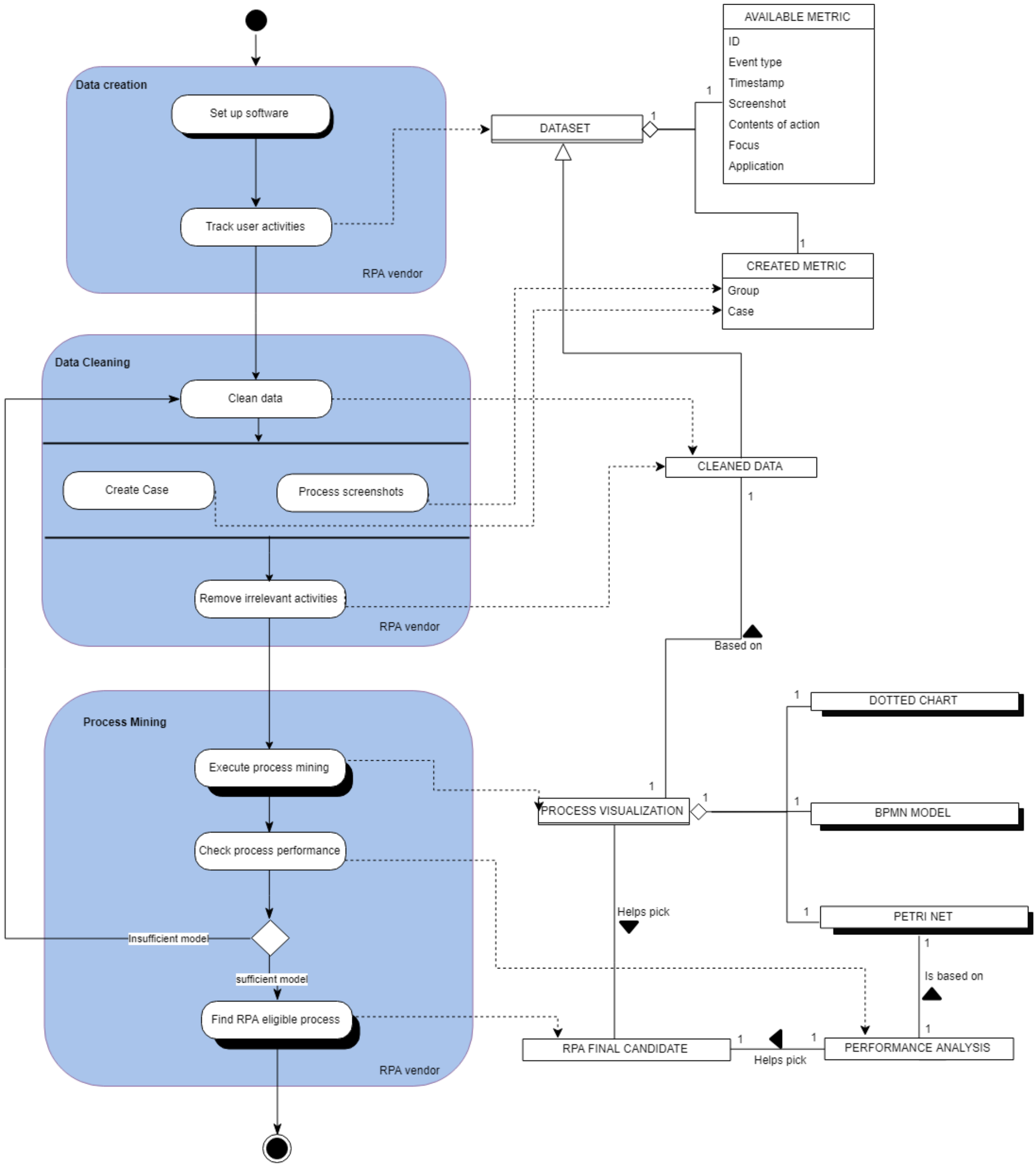


Figure 5, a PDD for Process Discovery in RPA using a screen capture tool as data source for Process Mining

6: Results

In this section, the results of executing the three different Process Discovery methods that are tested in this thesis will be explained. The execution of the method and interesting or unexpected findings will be elaborated upon first. Then the resulting visualization of each method will be presented and relevant features will be outlined. Finally, the potential cases for RPA that each method has produced, will be stated.

Event Logs

Data identification

The first step of executing this method, as explained in section 5.2.2 and figure 3, is to find possible data sets within an organization. Within this case study, the only available dataset within the organization was found within the recruitment department. This puts a limit on potential processes that can be found, since only in parts of the organization where data are available one is able to find potential RPA eligible processes when using this method.

The data obtained contain one year worth of logs from the HR Recruitment CRM. The progress of each candidate is stored in here. The dataset originally consisted of 65.843 rows, 99 different activities and does contain a timestamp, activity and process ID which are necessary to be able to perform Process Mining (van der Aalst et al., 2012b). Cleaning this dataset resulted in a data source containing 54.261 rows and 27 activities. The complete data cleaning process is described in *Appendix 1.*

Process Mining

As proposed in the PDD, the dataset has been loaded into PROM and a BPMN, Petri Net and Dotted Chart have been created. The figures are added in *Appendix 3.* When looking at the BPMN figure, it seems like a really nice and well readable 'lasagne layered' process (W. M. P. van der Aalst, 2011). A closer inspection of the Dotted Chart reveals two unexplainable vertical lines where much logging occurs. At close inspection, these lines are responsible for thousands of logs that have been created in a matter of minutes. It is impossible that those have been created by an employee, however their origin is unknown. It was decided to remove these rows from the dataset, leaving only 19.368 rows in the dataset. The analysis was performed again, resulting in figures that consisted of an unreadable 'spaghetti process'(W. M. P. van der Aalst, 2011). In order to make the process more readable, a filter has been applied that only allowed candidates who made it past the first step of the solicitation process to be included in the dataset. This resulted in the figures that can be seen in *Appendix 4.*

These figures are still difficult to read. Trying different settings in PROM did not help, and looking at these figures it is still difficult to understand what is going on. The main concern with the created visualizations, is that there is no recurrence of certain steps of the process. PROM only allows for a certain step to be visualized once. When looking at the original data, an occurrence of the process usually starts with the entry of 'persoonsgegevens'. The BPMN as can be seen in Appendix 4, does not show this step at the start of the process. To verify if 'persoonsgegevens' should not occur at the start of the process, another attempt to visualize the process has been made using DISCO. This tool is more user friendly compared with PROM and allows for quick filtering on the number of activities and paths you want to be visualized.

The used settings for DISCO are the same settings as used for PROM. The filter on candidates making it to the first round, has been applied in DISCO as well. This resulted in figure 6 which can be seen below. Figure 6 shows that the entry of ‘persoonsgegevens’ does occur at the beginning of the process. It also shows that it occurs multiple times within the process, which is expected when looking at the data.

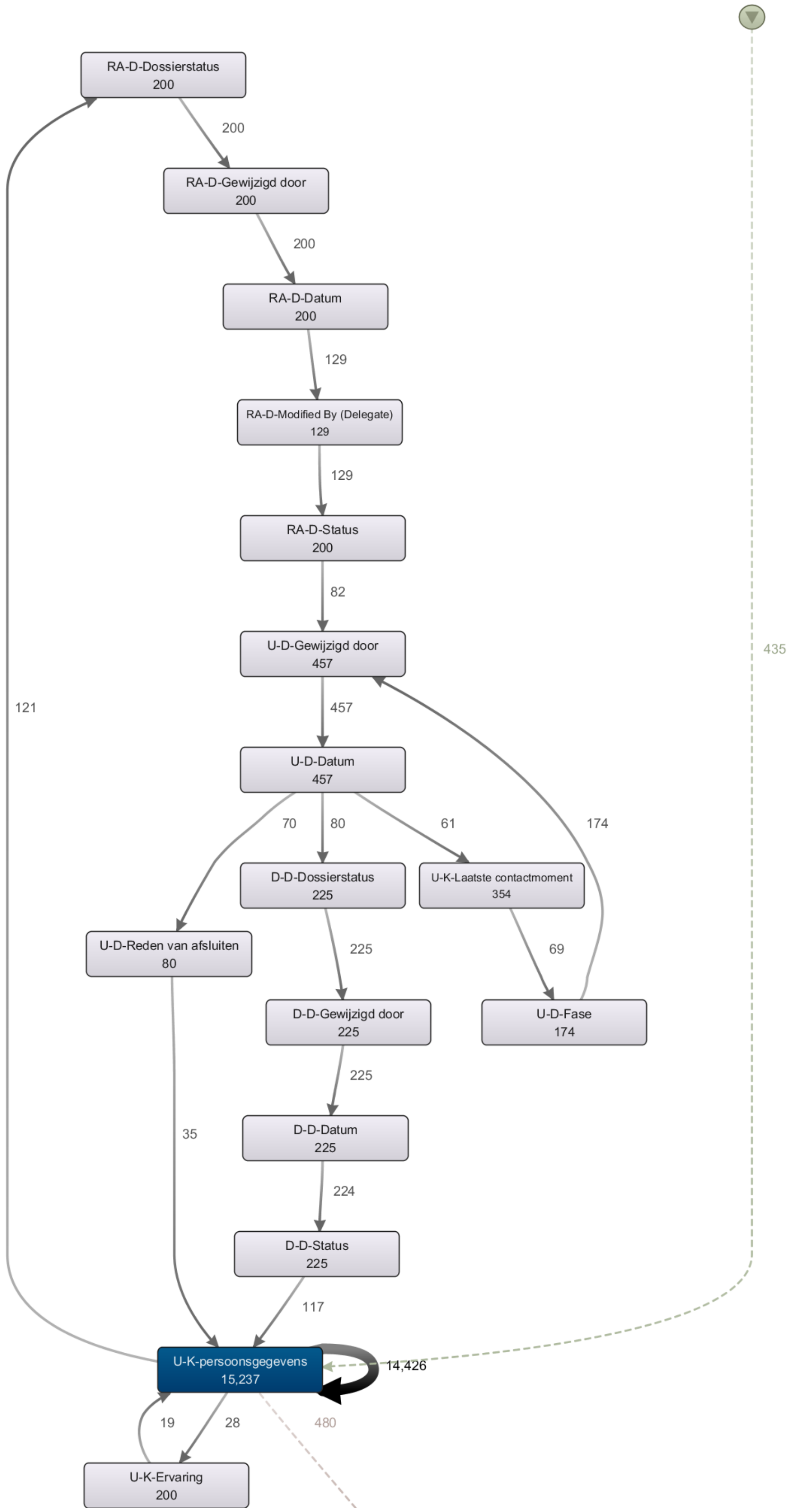


Figure 6, a process diagram created with event logs through Disco

The process starts with entering personal data into the system. This is often immediately followed by entering more data. Then the state of the dossier is updated, which means that a candidate is starting the acquisition process. Then a few updates occur, which seem to be similar patterns recurring three times. Then a final conclusion is reached, which can lead to different activities. This is followed by more data entry, after which the process ends.

DISCO allows for performance analysis, which used the timestamps to visualize the flow of the process. This shows that most activity occurs around updating the personal info, which could also have been derived from the high frequency of that activity, compared to the other activities.

The process that would most likely be a good candidate for RPA, is the ‘persoonsgegevens’ activity where personal information is entered into the system. This activity occurs very often, which makes it more profitable to automate, and the data that are handled consist of small things like addresses and names. Only from this visualization it is unsure whether or not this is a viable candidate. From this visualization we cannot determine where the data are coming from and it is still guesswork to decide if the activity is really not complex.

Screen Capture

The second method that has been tested, is Process Mining with data created by a screen capture tool.

Data creation

The use of this tool immediately caused some problems. This tool is installed at the computer of an employee, and the screen capture tool logs what is typed, and makes a screenshot after every change that is made. The data that are to be accessed to perform the recruitment process, are personal. Because of the GDPR law, these data must be secure and cannot be shared amongst third parties or be used for a purpose that the owner of the data has not agreed to (Zarsky, 2017). Therefore a different solution had to be found to be able to test this method.

In order to still be able to use the screen capture tool in collaboration with the HR department, it was decided to create fake user profiles with fake data to be able to come as close to the realistic situation as possible. An HR employee created the fake profiles and went through the process within an afternoon. This is a shorter timespan than usual, but allowed for a complete dataset which allows for the entire process to be visualized.

Data cleaning

This resulted in a dataset containing 499 actions, that together formed 2 instances of the recruiting process. To be able to properly process this data, the screenshots that have been created of each action have to be processed as stated in figure 5. This is done by deciding how much images differ from each other and put them into categories according to these differences. For this research, this process had to be done manually. Because dividing a total of 320 images into categories is too difficult to do manually, it was decided to focus on a single case of data entry. This resulted in a dataset of 97 rows, including 20 different activities. After assigning each row to an activity, based on the screen shots, a name was assigned to the activity to improve the data visualization.

Process Mining

In the next step, the data have been visualized, using DISCO. In the PDD from figure 5 it was mentioned that PROM would be used, but for analysing the event logs DISCO proved to be more useful and accurate, because recurring activities are modelled multiple times. To be able to compare the two data sources in an equal manner, it was decided to also use DISCO on the screen capture data. This has resulted in a visualization of the data entry part of the process, which can be seen in figure 7.

Every step of the process is modelled, as can be seen in figure 7. The activity 'data entry' occurs often and is looped with the activity called 'E-mail'. With the additional information provided by the screenshots , it is clear that this is simple copy and paste work and a clear example of a 'swivel chair' process. In a later phase of the process, 'Data Entry' occurs again. It is not clear where the data that are entered in the later part, come from. Maybe the recruiter still had the information by heart. The screenshots show that the information is still of a structured nature. Using this information, it is very likely that the process of entering of the data that are displayed in this visualization, can be a strong candidate for RPA.

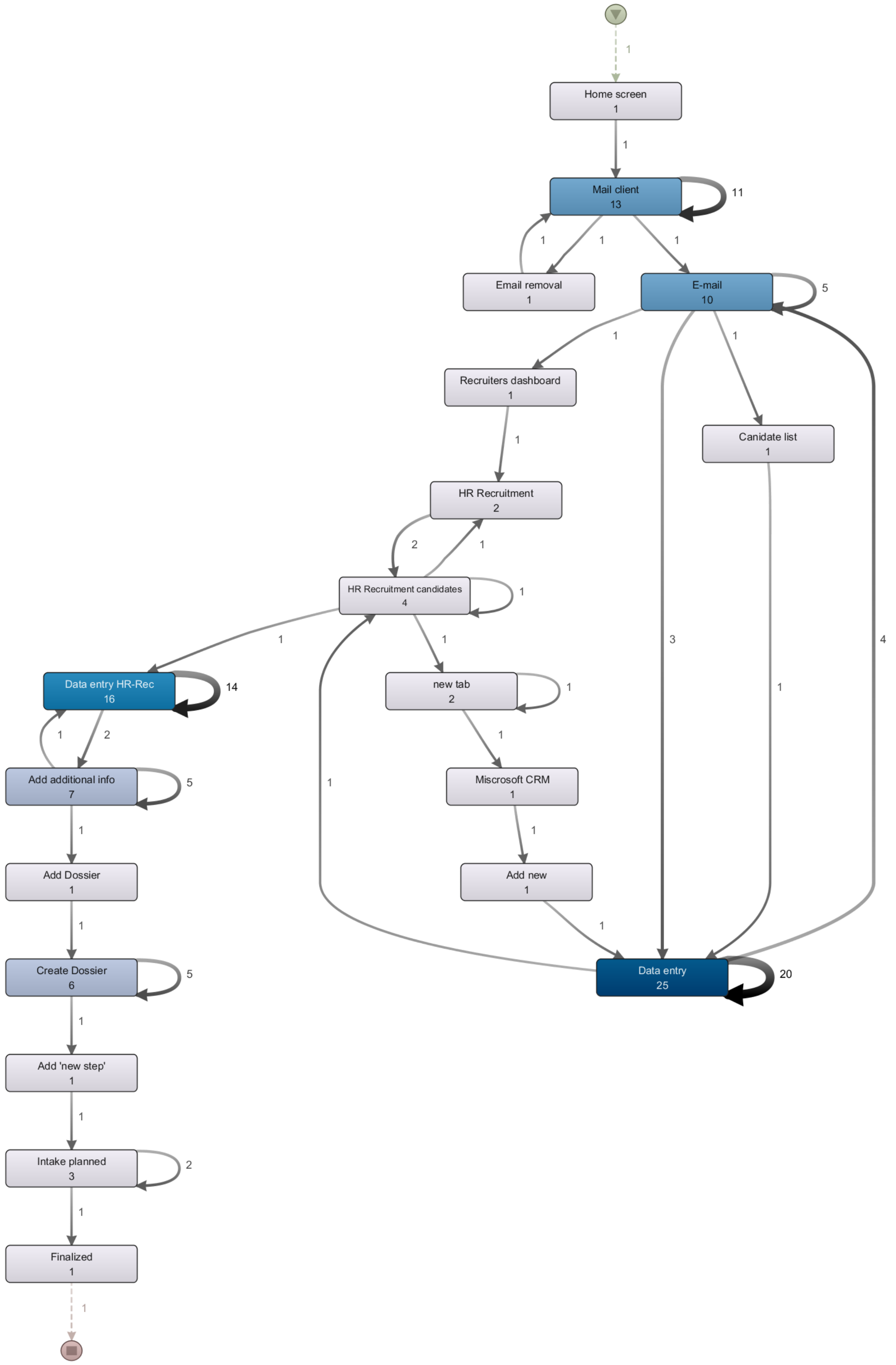


Figure 7, a process diagram created using data from the screen capture tool in DISCO

Interviews

The last method that has been tested, was attempting to find RPA eligible processes through interviews with employees of the company.

The execution of the interviews did differ from the PDD as shown in section 4.2.1. The event logs were only available at the Recruitment Department, and therefore this method could not be executed companywide but only within the Recruitment Department.

To model the recruiting process correctly, two interviews have been conducted. The first interview has resulted in a preliminary model of the process in general in BPMN. Three more in depth models have been created in BPMN-R to go as much detail as possible as described by this method.

In a second interview, the created models have been verified and updated. Some minor errors have been changed to correctly represent the actual process. This resulted in a high level overview of the process in figure 8. All the activities with a + sign have been modelled in depth in BPMN-R, to be able to decide if there are parts than can be automated with RPA. Three models have been created, of which the first can be seen below in figure 9. The other models are included in *Appendix 6.*

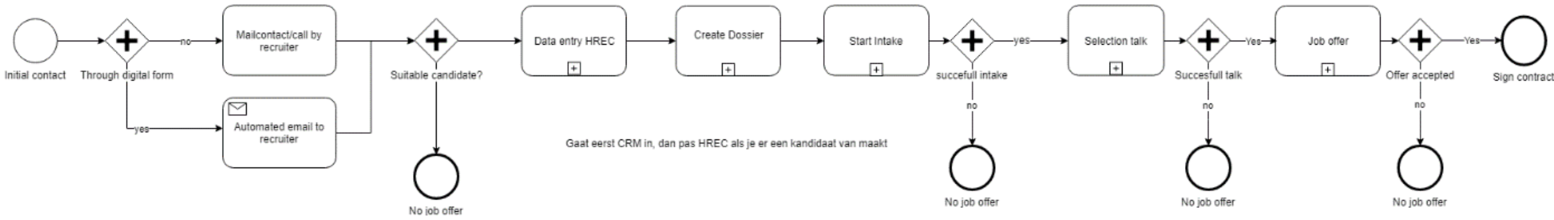


Figure 8, an overview of the recruitment process

In figure 9 the administrative work is shown that has to be done in order to start the intake process. This involves data entry into both the CRM and HREC systems. The data that are entered are also included in the model, together with the source. When the source is a generated email, the data are structured. When the data are from another source, the type of data is unspecified as it can be presented in many types from many different sources. It is important to note that after the verification, it turned out that the order of data entry was inaccurate. It was specified by the employee that the data were first entered into HREC, then alter into the CRM. This turned out to be false, and to be the other way around. This error was discovered through the other models from the screen capture method.

Out of this visualization, two strong candidates for RPA can be determined. First of all, the part of the process where data are entered when there is an email generated. The data that are entered are digital and structured. It is a process of entering data from an email into a system, which is a process that is not complex at all. This is a perfect example of a ‘swivel chair’ process. It is only a single task within the process, but would be a good candidate for RPA. The second part of the process that can be automated, is the process of finding an available spot in the company agenda to schedule a meeting. The task is simple, finding free spots where all parties from the company can attend, and email the candidate with a few propositions. This is a simple process. It might not be worth the investment to automate, because the return of investment might be too low, but still could be a good candidate to consider. Finding free spots in the company agenda occurs more often in the other

BPMN-R models, as can be seen in *Appendix 6.* By using interviews, two potential RPA candidates have been found.

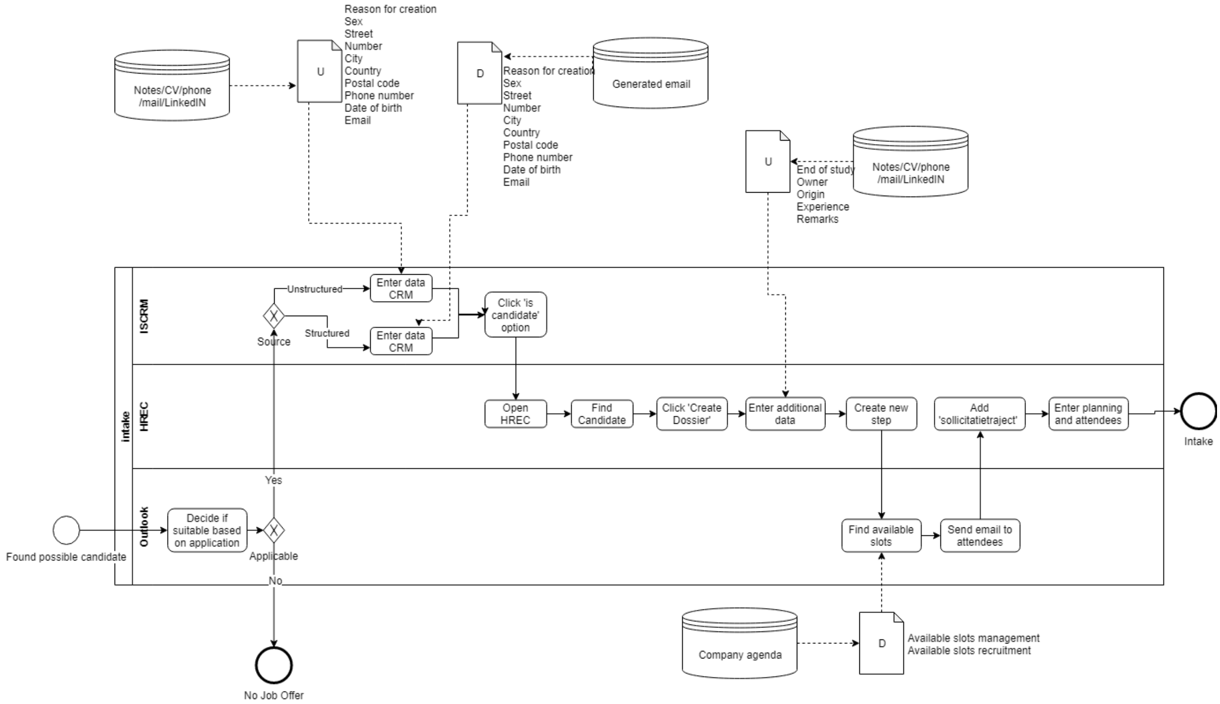


Figure 9, a BPMN-R model of the recruitment process from first contact until the intake

Requirement Fulfilment

In this section, an overview of how the set of requirements has been met by each method will be presented. The satisfaction of the requirements will be shown in two ways. First, an overview of how each RPA process property is met in relation to each tested method will be shown. Second, an overview of how each method performs, regarding the importance of the requirements, which was indicated with a 'must have', 'should have' or 'would have'. A full list with the passing of each individual requirement is shown in *Appendix 7.*

As can be seen in table 5, the method for Process Discovery that uses interviews, meets 13 out of 16 requirements. The requirements that are missing, are mainly non-functional requirements. This is because, as already mentioned in the problem statement, performing interviews is an error prone and time consuming method. Almost all functional requirements have been fulfilled. It has failed in the 'high number of repetitions' process property, because employees find it difficult to know an exact number as to how often a process occurs. They can provide an estimation, which can also be useful, but cannot be as accurate as in the other methods.

The method that involves event logs, does meet 6 out of the 16 requirements. A lot of aspects that are important for RPA, are missing in the visualizations. Business rules, stability of environment, the need for human intervention and the structuredness of the data of the process are still unknown after creating the visualizations using event logs. It possible with this method to learn something about the complexity of the process, however only 2 out of the 4 requirements of this subject are met. This is mainly due to the fact that an activity that is logged in the HREC can still be very vague. The only information available on activity is the activity name. For instance the name 'persoonsgegevens'. We know this is about entering data into HREC concerning personal information, however we do not know where these data are coming from, if they are delivered in a structured way or which are entered into the fields and why. There is still plenty of room for interpretation, which is why many of the requirements are not met when performing interviews.

Using the screen capture method to gather data for Process Mining, meets 12 out of 16 requirements. In contrary to applying Process Mining using event logs when the data come from the screen capture tool, the tasks of the process are logged in much more detail. Every single click is logged, thus a good assessment of the complexity of the task can be made. It is still impossible to know the business rules that are behind a certain decision. Another process property that cannot be discovered, is how the environment acts. Using screen capture for Process Mining performs best on the non-functional requirements, when compared to the other two methods.

Table 5, satisfaction of the requirements of each method by process property

Process Quality	# of requirements	Interview	Event Logs	Screen Capture
Low complexity of tasks	4	4	2	4
High number of repetitions	1	0	1	1
Multiple systems are involved	2	2	1	2
The process follows clear business rules	1	1	0	0
Stable environment	1	1	0	0
Limited need for Human intervention	1	1	0	0
Structured data	2	2	0	2
Non Functional	4	2	2	3
Total	**16**	**13**	**6**	**12**

Another way of looking at how the requirements are met, is sorted by their importance. For an RPA project to be successful, some process properties are more important than others. As explained in section 2.1, some process properties indicate if an RPA robot can actually perform the task it is given, while other process properties indicate if there is an actual business case. If there is no business case, a robot might work but the automation is not worth the investment. In section 4.1, the requirements have been sorted accordingly, where 'must have' requirements indicate if creating an RPA robot is viable, and 'should have' or 'could have' indicate if there is a business case. This is summarized in table 6.

For the method that uses interviews, all 'must have' requirements are met. The screen capture method meets almost all 'must have' requirements, except this method is not able to see if a process follows clear business rules. The Event Log method only meets one of the six 'must have' requirements.

Table 6, satisfaction of requirements from each method by importance

Requirement	Interview	Event Logs	Screen Capture
Must	6/6	1/6	5/6
Should	4/5	2/5	3/5
Could	1/1	1/1	1/1
Non-Functional	2/4	2/4	3/4

7: Conclusion

In this section, the conclusion of this research will be presented and explained. The answers to each sub question will be given, after which the final conclusion of the research will be explained

Sub question 1: What Process Discovery methods have already been used in RPA projects?

This research has found two methods that have already been used to discover possible processes in RPA projects. The first of these methods is using interviews to discover possible processes for RPA. Many sources state that interviews have been used, but only the paper of Agaton & Swedberg (2016) explained in detail how this should be done to ensure reliable results. The second method that has already been used, is deploying a screen capture tool to track the operations of an employee. This tool is used by a company called Kryon, and during the course of this research a paper was published where a screen capture tool was proposed to create data for Process Mining (Jimenez-Ramirez et al., 2019). The tool proposed by Jimenez-Raimrez et al. (2019) has been used in this research.

Sub question 2: Which other Process Discovery methods could be used in an RPA project?

Using event logs for Process Mining is a method for Process Discovery that has not been used for RPA projects yet. Using event logs is a well explained method, but has never been used in the context of RPA before. Because using event logs for Process Mining is a well-established method for Process Discovery, this method was to be tested in this research as well.

Sub question 3: What are the requirements for a Process Discovery method in the context of RPA?

Process properties of typical RPA processes are well known and stated by different sources. I used these process properties and translated them to requirements for Process Discovery in RPA. The requirements have been verified by an RPA expert and can be found in Table 3.

Sub question 4: Which of the Process Discovery methods (from SQ 1 and SQ2) performs best in a practical setting?

Based on the met requirements, as presented in the previous section, conducting interviews performs best of the three tested methods. Most requirements are met when performing interviews, as can be seen in table 5. When performing interviews 13 out of 16 requirements are met, followed by 12 out of 16, when performing Process Mining with data from a screen capture tool. Using Process Mining with data from event logs only met 6 out of 16 requirements.

In the ‘must have’ category, conducting interviews meets all 6 out of 6 requirements, using the screen capture tool for process mining comes really close and 5 out of 6 ‘must have’ requirements are met. Using event logs seems least favourable, as only 1 out of the 6 ‘must have’ requirements is met. The ‘must have’ requirements are essential to choose a viable process for RPA and thus can be concluded that using event logs for Process Mining is not a suitable method. Using interviews meets most requirements and especially all of the requirements that are essential for a successful RPA robot. Using screen capture for Process Mining comes really close, but does, according to the requirements, not perform as well as conducting interviews.

Main question: Which Process Discovery method is best suited to find processes eligible for RPA?

Based on sub question 4, the method using interviews to find processes for RPA performs best. However, with interviews as the most successful method, the problem statement of this research is still not fully addressed. The goal was to find a method that was more reliable, faster and documented. The method using interviews has now clearly been documented through a PDD, however there are still issues with the reliability and the speed of this method. As mentioned in the problem statement, and even experienced in this research, the interviews can lead to inaccurate models of the process, even though the models are verified with employees. The speed issue is also not resolved, because the process of these interviews is extensive and takes up a lot of time of both the RPA consultant and the potential client.

The screen capture method does resolve the issues from the problem statement better. The process is discovered in great detail, as every task that is performed will be documented in full detail. This means that also deviations of the main process are documented, where employees tend to forget noticing such occurrences in an interview. No errors should occur. Depending on how much data cleaning is needed, this method can speed up the Process Discovery phase. It is however difficult to be exact on how much time can be saved.

It could be argued that improving the issues from the problem statement is a good trade-off with reducing the number of requirements that are met by using screen capture instead of interviews. However, it turned out that the application of the screen capture tool has an important flaw. In this case, it was not allowed to run under normal conditions, because it would leak personal data. Due to the nature of the screen capture tool, where every key stroke is logged and screen shots are made after every event, there is no possibility of properly anonymizing the data. This means that in cases where personal data are handled by an employee, this tool cannot be used.

A combination of Process Mining techniques and interviews could result in a situation where almost all of the requirements are met. Performing interviews should be the foundation of any method, because it is the only way to get to know the decision making behind a process. It is also often needed to clear up certain aspects of a process model, as the example stated in section 5.2.2, where it is shown that only the name of a process or task is often not enough information to get to know details, such as the complexity of the process. Both the use of event logs and the screen capture method are situational and therefore making them unreliable as base methods for Process Discovery in RPA.

However Process Mining, using either screen capture or event logs as data source, can provide a useful addition upon the method of using interviews. We propose an alteration to the PDD as presented in figure 3, where the method using interviews is explained. When a process is chosen at the end of the Process Discovery procedure, an extra step can be added for validation. I propose that the Process Mining techniques that are explained in this research, can be used to verify if a process chosen for RPA, is indeed a viable process. By adding Process Mining, errors in the manually made models could be detected and more requirements regarding the business case of an RPA project will be met. By adding Process Mining, more certainty regarding the success of a project can be provided, compared to a situation where only interviews are used. Using the screen capture technique to create data for Process Mining would be the preferred option, since this method proved to be more useful than when event logs are used. This is however a method that can only be used in certain situations, because of the privacy issues that have been discovered in this research. If privacy can be an issue, using event logs for Process Mining would still give relevant extra information. There will be a possibility to find errors in a modelled process, and extra information will be provided on the business case for the RPA project. The updated PDD of the interview method will be included in *Appendix 8.* This PDD is accompanied by two PDD's of the Process Mining techniques, which have been shortened and tailored to function as a verification of a chosen process.

8: Discussion

8.1 Contributions

This research has provided better understanding of what is needed to accurately perform Process Discovery in an RPA project. This is the first time that different methods for Process Discovery have been tested in the context of RPA. The requirements for RPA Process Discovery that have been created in this research, can be used in future work to test other methods for RPA Process Discovery. Most interesting have been the findings in the research that were accidental. At the start of the research, it was not expected that finding event logs could be an issue. It has turned out that not all systems within a company keep logs of events, and that the quality of the logs might not be good enough to adequately perform Process Mining. Another unexpected finding was regarding privacy. Because of the recent introduction of the GDPR law, privacy at the moment is a very relevant topic with. The experiments in this research have shown that, when personal data are involved, companies might be hesitant to share the data that are needed to perform Process Mining. This was especially relevant with the screen capture tool, where anonymising data can be really difficult. This research has discovered these issues, they may be addressed in future research.

In the conclusion of this research, a method for Process Discovery for RPA has been proposed which is largely based on conducting interviews as proposed by Agaton & Swedberg (2016). This thesis proposes the use of Process Mining techniques to further improve the reliability of the method by Agaton & Swedberg (2016) in the later stages of the Process Discovery. Using Process Mining to verify if a chosen process has been modelled correctly and to improve the business case of the RPA project, should increase the success rate of future RPA project.

8.2: Limitations

During the execution of this research, a number of limitations have been involved.

First, the methods that are tested as explained in section 5 and 6, have not been tested to their full extend. The event logs that were needed for Process Mining, could only be found within the HR Recruitment Department of the company where the research took place. Therefore, the tests have been limited to this part of the company.

Another limitation was that the screen capture tool was tested in a controlled setting, rather than an uncontrolled setting. This could have made data analysis easier than it actually is, because the time frame was smaller, and fewer deviations of the process could have occurred. An attempt has been made to limit the threat to the validity of this research, by documenting this part of the experiment extensively.

Being limited in the execution of the research, however, also gave some valuable information. It proved that finding data for Process Mining can be a severe limiting factor when attempting to find RPA eligible processes. This was not expected when formulating the research questions and not an expected finding, but really relevant when methods turn out to be only able to be used under certain circumstances.

Lastly, the problem statement mentioned that one of the problems that occurs when finding processes manually, is that it is very time consuming. Accurately measuring which of the tested methods would be less time consuming, proved to be very difficult. At best an estimation could be made. However, this is very dependent on the data that are acquired and the amount of cleaning that is needed. Also the size of the process could play an important role. These factors made it very difficult to find a clear answer to this question, based on just this one experiment for each method. In the end, it turned out that time consumption is not that relevant in deciding which method performed best, because the method that used manual work showed to be of higher quality for RPA.

8.3: Future Work

For future work, it might be interesting to discover as to how RPA acceptance is involved in the Process Discovery phase. This research has not included the potential resistance from employees that is faced when performing an RPA project (Strömberg, 2018). Workshop based analysis for Process Discovery could help to improve the acceptance for RPA, and could therefore be considered as a part of the Process Discovery (Asatiani & Penttinen, 2016). Future research could possibly investigate if the first round of interviews, as proposed in this research and by Agaton & Swedberg (2016), could maybe be replaced by a workshop session, as to simultaneously find potential processes while getting all the employees enthusiastic about the RPA project.

Another suggestion for future research is to find if a combination of interviews and Process Mining is indeed a viable solution as suggested in the Conclusion of this research. When Process Mining becomes an important part of RPA projects, more research should be conducted on when it can or cannot be used. For this research, limitations that have been found include privacy issues and data availability. This research was only on a single case, and therefore other limitations can exist, which need to be investigated.

References

Aalst, van der, Adriansyah, A., Alves De Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2012a). *LNBIP 99 - Process Mining Manifesto*. Retrieved from https://link.springer.com/content/pdf/10.1007/978-3-642-28108-2_19.pdf

Aalst, van der, Adriansyah, A., Alves De Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2012b). *LNBIP 99 - Process Mining Manifesto*. Retrieved from https://link.springer.com/content/pdf/10.1007/978-3-642-28108-2_19.pdf

ACCA Global. (2015). *The robots are coming? Implications for finance shared services*. Retrieved from www.accaglobal.com

Accenture. (2016). *Getting Robots Right How to avoid the SIX most damaging mistakes in scaling-up Robotic Process Automation*. Retrieved from https://www.accenture.com/t00010101T000000__w__/au-en/_acnmedia/PDF-37/Accenture-Robotic-Process-Auto-POV-Final.pdf

Accenture. (2018). Robotic Process Automation: The future of technology in financial services. Retrieved February 14, 2019, from https://www.accenture.com/no-en/insight-financial-services-robotic-process-automation

Agaton, B., & Swedberg, G. (2016). *Evaluating and Developing Methods to Assess Business Process Suitability for Robotic Process Automation*. Retrieved from http://publications.lib.chalmers.se/records/fulltext/255664/255664.pdf

Asatiani, A., & Penttinen, E. (2016). Turning Robotic Process Automation Into Commercial Success – Case OpusCapita. In *Journal of Information Technology Teaching Cases* (Vol. 6). https://doi.org/10.1057/jittc.2016.5

BluePrism. (2011). *The Role Of IT In Business-Driven Process Automation*. Retrieved from www.forrester.com/consulting.

Boulton, C. (2017). What is RPA? A revolution in business process automation | ComputerWorld Hong Kong. Retrieved February 14, 2019, from https://www.cw.com.hk/digital-transformation/what-rpa-a-revolution-business-process-automation

Bruno, J., Johnson, S., & Hesley, J. (2017). Robotic disruption and the new revenue cycle: robotic process automation represents an immediate new opportunity for healthcare organizations to perform repetitive, ongoing revenue cycle processes more efficiently and accurately. *Healthcare Financial Management*, *71*(9), 54–62. Retrieved from https://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=googlescholar&v=2.1&it=r&id=GALE%7CA506318586&sid=googleScholar&asid=59be3228

Burgess, A. (2018). *The executive guide to artificial intelligence : how to identify and implement applications for AI in your organization*.

CapGemini Consulting. (2016). *Robotic Process Automation-Robots conquer business processes in back offices*. Retrieved from https://www.capgemini.com/consulting-de/wp-content/uploads/sites/32/2017/08/robotic-process-automation-study.pdf

Chinosi, M., & Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards & Interfaces*, *34*, 124–134. https://doi.org/10.1016/j.csi.2011.06.002

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of Business Process Management*. https://doi.org/10.1007/978-3-642-33143-5

Fernandez, D., & Aman, A. (2010). IMPACTS OF ROBOTIC PROCESS AUTOMATION ON GLOBAL ACCOUNTING SERVICES. In *Asian Journal of Accounting and Governance* (Vol. 9). Retrieved from http://ejournal.ukm.my/ajac/article/view/25271

Fung, H. P. (2014). Criteria, Use Cases and Effects of Information Technology Process Automation (ITPA). *School of Management*. https://doi.org/10.4172/2168-9695.1000124

Gartner. (2016). Automated Business Process Discovery (ABPD). Retrieved April 19, 2019, from https://www.kryonsystems.com/process-discovery/

Geary, P. (2016). *Choosing Your Lane: A Guide to the Three Types of Robotic Process Automation Technology*. Retrieved from https://assets-eb99.kxcdn.com/uploads/resources/white-papers/BluePrism-Choosing-Your-Lane-A-Guide-to-the-Three-Types-of-Robotic-Process-Automation-Technology.pdf

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice*. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=89w-scN7_8MC&oi=fnd&pg=PR6&dq=Design+Research+in+Information+Systems&ots=-t1PolGAu_&sig=pnqLpYI0hNNCAljX4GuAgnajU5M#v=onepage&q=Design Research in Information Systems&f=false

Institute for Robotic Process Automation. (2015). *Introduction to Robotic Process Automation*. Retrieved from https://irpaai.com/wp-content/uploads/2015/05/Robotic-Process-Automation-June2015.pdf

Jimenez-Ramirez, A., Reijers, H. A., Barba, I., & Del Valle, C. (2019). *A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle*. https://doi.org/10.1007/978-3-030-21290-2_28

Juntunen, K. (2018). *Influence of contextual factors on the adoption process of Robotic process automation (RPA)*. Retrieved from http://www.teknik.uu.se/student-en/

Kirchmer, M. (2017). Robotic Process Automation – Pragmatic Solution or Dangerous Illusion? *BTOES Insights (Business Transformation and Operational Excellence Summit Insights)*.

Lacity, M., & Willcocks, L. (2018). Innovating in Service: The Role and Management of Automation. In *Dynamic Innovation in Outsourcing* (pp. 269–325). https://doi.org/10.1007/978-3-319-75352-2_9

Lacity, Willcocks, & Yan. (2015). Are the robots really coming? service automation survey findings. *Pulse Magazine*, (17), 14–21.

Lamberton, C., & Brigo, D. (2017). *Impact of Robotics, RPA and AI on the insurance industry: challenges and opportunities*. Retrieved from https://www.semanticscholar.org/paper/Impact-of-Robotics%2C-RPA-and-AI-on-the-insurance-and-Lamberton-Brigo/d5830f82353c9e2827bf462f1627a5cd2bf1bbc3

Linden, A., & Fenn, J. (2003). *Gartner Understanding Gartner's Hype Cycles Understanding Gartner's Hype Cycles Understanding Gartner's Hype Cycles*. Retrieved from https://www.bus.umich.edu/KresgePublic/Journals/Gartner/research/115200/115274/115274.pdf

Lintukangas, A. (2017). *Improving indirect procurement process by utilizing robotic process automation (Vol. Masters)*. Retrieved from http://www.doria.fi/bitstream/handle/10024/143385/Lintukangas Annika%25 2003082017.pdf?sequence=2&isAllowed=y

McKinsey. (2016). *The next acronym you need to know about: RPA (robotic process automation)*. Retrieved from https://assets-eb99.kxcdn.com/uploads/resources/white-papers/WillcocksMcKinseyNov2016The-next-acronym-you-need-to-know-about-RPA-robotic-process-automation.pdf

Rutaganda, L., Bergstrom, R., Jayashekhar, A., Jayasinghe, D., & Ahmed, J. (2017). Avoiding pitfalls and unlocking real business value with RPA. *Journal of Financial Transformation*, *46*, 104–115. Retrieved from https://ideas.repec.org/a/ris/jofitr/1600.html

Slaby, J. R. (2012). *Robotic Automation Emerges as a Threat to Traditional Low-Cost Outsourcing | 1 ROBOTIC AUTOMATION EMERGES AS A THREAT TO TRADITIONAL LOW-COST OUTSOURCING Cheap, easy-to-develop software robots will eventually supplant many offshore FTEs*. Retrieved from www.hfsresearch.com%7Cwww.horsesforsources.com%7Cbpo.horsesforsources.com

Strömberg, K. (2018). *Robotic Process Automation of office work: benefits, challenges and capability development*. Retrieved from https://aaltodoc.aalto.fi/handle/123456789/32211

Sutherland, C. (2013). *Smarter Than You Think: How Technology Is Changing Our Minds For the Better*. Retrieved from The Penguin Press website: www.hfsresearch.com%7Cwww.horsesforsources.com%7Cbpo.horsesforsources.com

Szelągowski, M. (2018). Evolution of the BPM lifecycle. *Communication Papers of the Federated Conference on Computer Science and Information System*, *17*. https://doi.org/10.15439/2018F46

Tarquini, T. (2018). Practical Robotics in Insurance - The Future is Here Already. In *The InsurTech Book* (pp. 231–235). https://doi.org/10.1002/9781119444565.ch52

Trefler, A. (2018). The Big RPA Bubble. Retrieved February 14, 2019, from https://www.forbes.com/sites/cognitiveworld/2018/12/02/the-big-rpa-bubble/#4dc3d91568d9

Van De Weerd, I., & Brinkkemper, S. (2009). *Meta-Modeling for Situational Analysis and Design Methods*. Retrieved from https://6e1d9175-a-62cb3a1a-s-sites.googlegroups.com/site/methodengineering14/schedule/week-10/Meta-modeling for Situational Analysis and Design Methods.pdf?attachauth=ANoY7cpq81ZRBtFr5T579_8h2ZHQh4rQRRX3ZbBP1jk2_LQkpaQi40bERZDZc8Bwh1AZgBnsTB8rE3QrHEsfv-aceSYLZPqkQaRuwyM6nUsjjSyenG87c29ugKKbTHThljMq_osL2dAIibMNEauhcZNTcBqKcjbef5LLxjWeozxp_fKNfRIkZAPwvmXx1Mb_Fi6wVK2p5HR7spMOd2oEQ__-ThRiym3txa-dQHwIfgRYPSb6So9QY3MWaB-jSDhxRRuqoj8nxi9RNezLwNKAWbyo0R34Vcjw6RkFF5sq74oMXaFQT9ExKOeCsUBQnnZGX-NvsaAwCrMN&attredirects=0

van der Aalst, Bichler, M., & Heinzl, A. (2018). Robotic Process Automation. *Business & Information Systems Engineering*, *60*(4), 269–272. https://doi.org/10.1007/s12599-018-0542-4

van der Aalst, W. M. P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. https://doi.org/10.1007/978-3-642-19345-3

Vergidis, K., Tiwari, A., Majeed, B., & Telecom, B. (2008). *Business Process Analysis and Optimization: Beyond Reengineering AUTONOM Integrated through-life support for high-value systems View project Design Rationale and Medical Devices View project Business Process Analysis and Optimization: Beyond Reengineering*. https://doi.org/10.1109/TSMCC.2007.905812

White, S. A. (2004). *Introduction to BPMN*. Retrieved from www.bptrends.com

Wieringa, R. J. (2014). Design science methodology: For information systems and software engineering. In *Design Science Methodology: For Information Systems and Software Engineering*. https://doi.org/10.1007/978-3-662-43839-8

Willcocks, L., & Craig, A. (2015). *The Outsourcing Unit Working Research Paper Series Paper 15/02 Robotic Process Automation at Telefónica O2 Research on Business Services Automation Research Objective*. Retrieved from www.outsourcingunit.org.

Willcocks, Lacity, & Craig. (2015). Robotic process automation at Xchanging. *The Outsourcing Unit Working Research Paper Series*, *2015*, 1–26.

Wright, D., Witherick, D., & Gordeeva, M. (2017). The robots are ready. Are you? Untapped advantage in your digital workforce. In *Deloitte*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/technology/deloitte-robots-are-ready.pdf

Zarsky, T. Z. (2017). *Incompatible: The GDPR in the Age of Big Data*. https://doi.org/10.1093/idpl/ips036

Appendix

Appendix 1

The original dataset consists 7 columns and 65843 rows. The data concerns 4644 cases. The data fields are as follows:

- Created on: time stamp with date and time
- Operation: can only hold two values, ‘Bijwerken’ for when an update has occurred or ‘Verwijderen’ for when an item was removed.
- Action: Holds 5 values that are concerned with how the data was manipulated.
- Entity: can hold only 2 values, either ‘Kandidaat’ or ‘Dossier’.
- Object: holds what field of data has been manipulated. This column can hold many values (99 in total), mostly concerning what type of personal data has been entered or altered in the system or status updates on the potential recruit. Note that the value of these fields is not included in the dataset for privacy reasons.
- Candidate ID: this is a unique number for every potential recruit. In total there is data on 4644 potential recruits in this dataset
- UserID: this is a unique number for the employee that was responsible for entering the data. There are 20 employees of HR involved in this dataset.

The steps that have been taken to clean the data are explained below.

The most important part of cleaning, was to decrease the number of variables that the ‘Object’ column could have. Many of the variables are concerned with the entry of personal information. When doing a quick analysis on the original data set with Disco, it shows that when only 27% of activities are shown, they are mostly concerned with altering personal data. It is obvious that this is a huge part of the workload that shows in the system, because an event log appears for every data field that is entered in the system. However, because of this the nature of the actual process is still unclear. A figure of the analysis of the first dataset has been included in Appendix 2.

To tackle this problem, an additional column has been created which is called ‘ActionDetail’. This column consists of grouped up entries from the ‘Object’ column. The original count of different activities on the ‘Object’ column was 99, this has been decreased by 36 activities by grouping things such as ‘address’, ‘first name’ or ‘last name’ under a variable called ‘personal information’. A full list of the values that have been grouped into ‘personal information’ can be found below.

Every row where the activity was to ‘remove’ an object has been deleted. There are not manual actions but are performed by the software and therefore not relevant for this analysis. Doing so resulted in 54.361 rows of data remaining containing 27 different activities.

The last step of preparing the data for analysing, is to turn the data format into XES and assigning a column as timestamp, a column as activity and a column as CaseID. This is done using the ProM 6 tool by fluxicon. This tool is not particularly easy to use, but does allow for advanced analysis using a wide variety of different algorithms to create different types off visualization of the process. For this dataset, the ‘Created On’ column is assigned as ‘starting time’. The ‘Candidate ID’ is assigned as Case ID, because each candidate represents a different case. Last, the ‘Action’ column, ‘Entity’ column and ‘ActionDetail’ have been assigned together to represent the ‘activity’. This because there are different entity’s (Entity) in the data that can be altered in different ways (Action) which results in certain action within the entity (ActionDetail). This caused the number of different possible activities to increase again to 45 activities. When analysing the data with Disco, the same data preparation happens.

When a first attempt of analysis of the data was performed, it turned out that the diagrams were unreadable because of the 3 columns that together form the activity for this process. This resulted in very long activity names that were unreadable in the diagram. Therefore, the action column has been abbreviated.

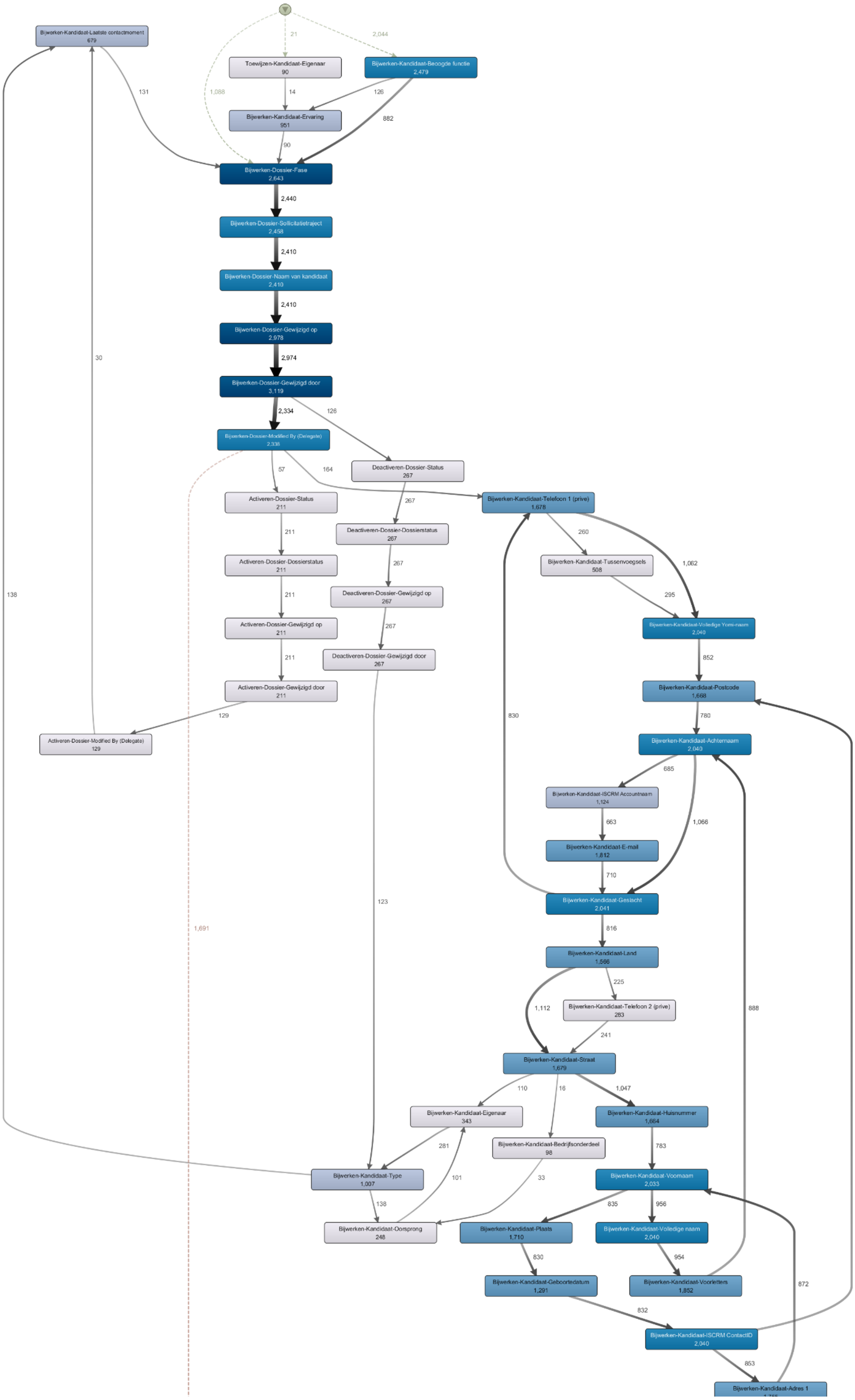
- ‘U’ stands for ‘Update’
- ‘D’ stands for ‘Deactivate’
- ‘R’ stands for ‘Remove’
- ‘A’ stands for ‘Assign’
- ‘RA’ for ‘Reactivate’

A similar action is performed for the ‘Entity’ column. There are only two different entities in this dataset, ‘Dossier’ and ‘Kandidaat’. These are abbreviated to respectively ‘D’ and ‘K’.

Sorted under ‘personal info’

Achternaam E-mail Geslacht Huisnummer ISCRM ContactID Plaats Postcode Straat Telefoon 1 prive Naam van kandidaat	Tussenvoegsels Volledige naam Volledige Yomi-naam Voornaam Adres 1 ISCRM Accountnaam Voorletters Huisnummer Geboortedatum	Land E-mail 2 (prive) Telefoon 3 (prive) Mag recruitmentmateri aal ontvangen Telefoon 2 (prive) Titels voor Adres 2 Titels achter Gestart op	Plaats (weekend) Postcode (weekend) Straat (weekend) E-mail 3 (prive) Regio Voertaal Tags Specialisatie Type

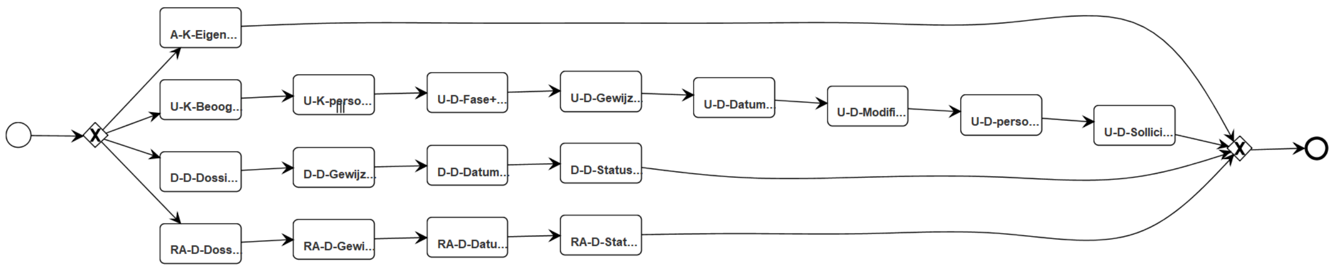
Appendix 2



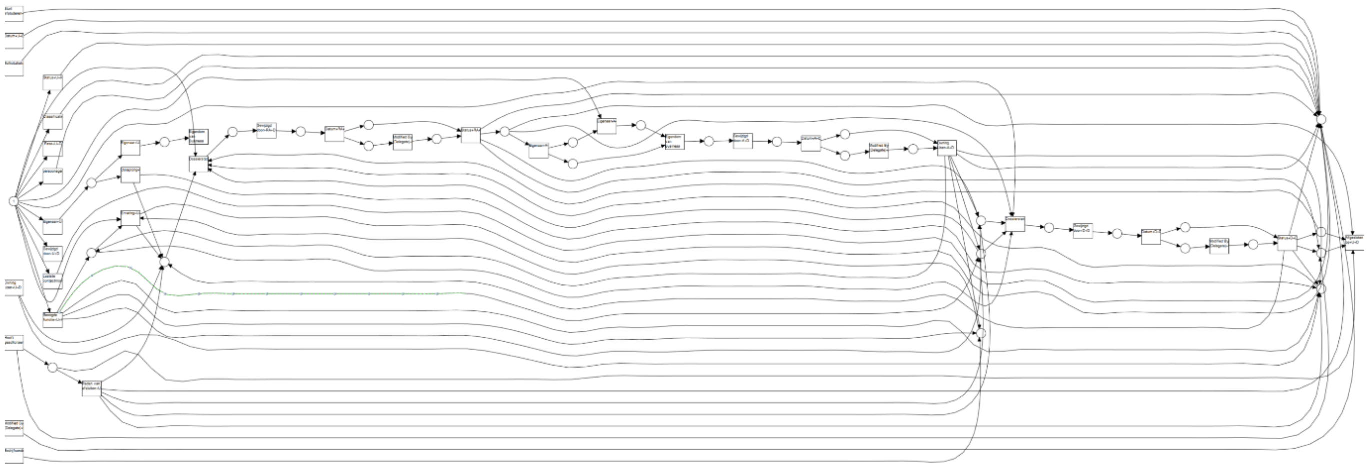
Bijwerken-Kandidaat-Laatste contactmoment
679
Toewijzen-Kandidaat-Eigenaar
90
Bijwerken-Kandidaat-Beoogde functie
2,479
Bijwerken-Kandidaat-Ervaring
951
Bijwerken-Dossier-Fase
2,643
Bijwerken-Dossier-Sollicitatietraject
2,458
Bijwerken-Dossier-Naam van kandidaat
2,410
Bijwerken-Dossier-Gewijzigd op
2,978
Bijwerken-Dossier-Gewijzigd door
3,119
Deactiveren-Dossier-Status
267
Activeren-Dossier-Status
211
Bijwerken-Kandidaat-Telefoon 1 (prive)
1,678
Deactiveren-Dossier-Dossierstatus
267
Activeren-Dossier-Dossierstatus
211
Bijwerken-Kandidaat-Tussenvoegsels
508
Deactiveren-Dossier-Gewijzigd op
267
Activeren-Dossier-Gewijzigd op
211
Deactiveren-Dossier-Gewijzigd door
267
Bijwerken-Kandidaat-Postcode
1,668
Activeren-Dossier-Gewijzigd door
211
Activeren-Dossier-Modified By (Delegate)
129
Bijwerken-Kandidaat-Achternaam
2,040
Bijwerken-Kandidaat-E-mail
1,812
Bijwerken-Kandidaat-Geslacht
2,041
Bijwerken-Kandidaat-Land
1,566
Bijwerken-Kandidaat-Telefoon 2 (prive)
283
Bijwerken-Kandidaat-Straat
1,679
Bijwerken-Kandidaat-Eigenaar
343
Bijwerken-Kandidaat-Huisnummer
1,664
Bijwerken-Kandidaat-Bedrijfsonderdeel
98
Bijwerken-Kandidaat-Type
1,007
Bijwerken-Kandidaat-Voornaam
2,033
Bijwerken-Kandidaat-Oorsprong
248
Bijwerken-Kandidaat-Plaats
1,710
Bijwerken-Kandidaat-Volledige naam
2,040
Bijwerken-Kandidaat-Geboortedatum
1,291
Bijwerken-Kandidaat-Voorletters
1,852
Bijwerken-Kandidaat-ISCRM ContactID
2,040
Bijwerken-Kandidaat-Adres 1

Appendix 3

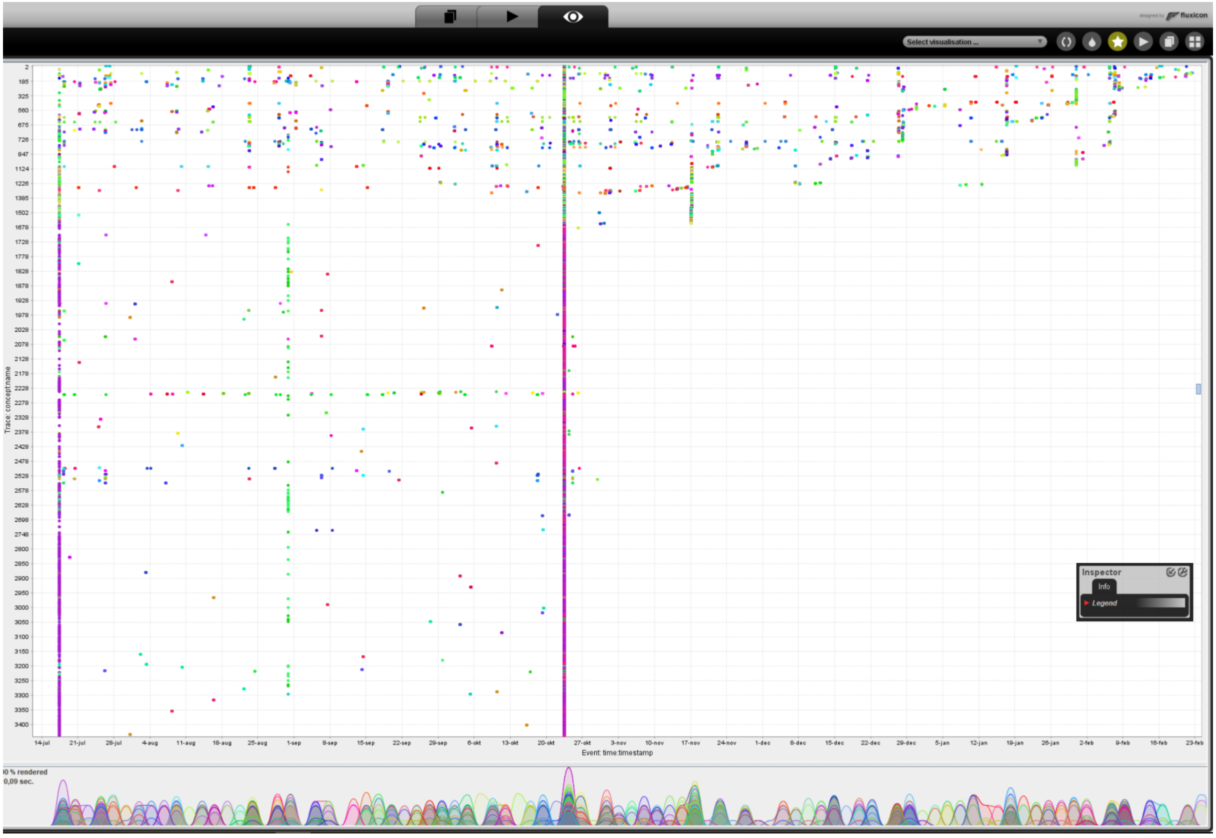
First analysis using PROM



BPMN model in PROM

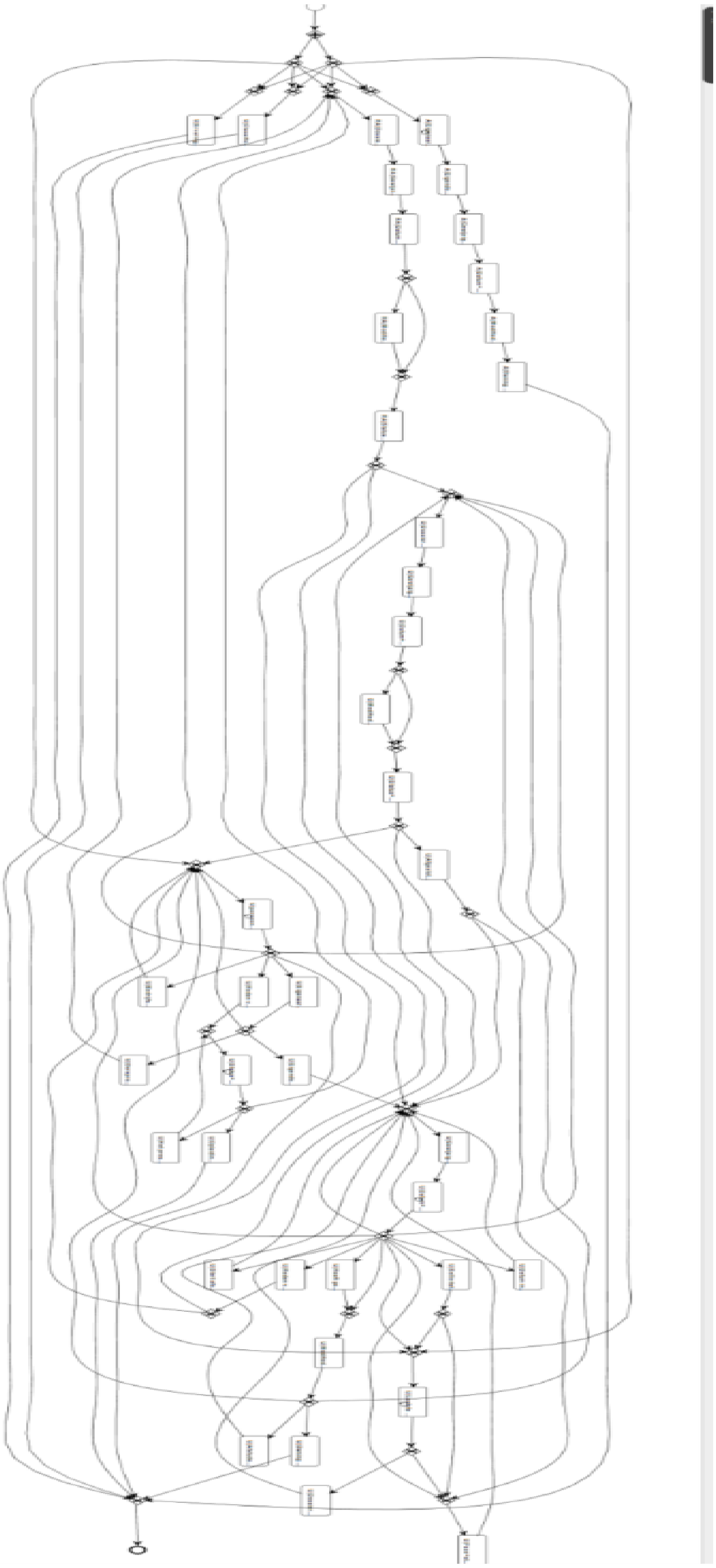


Petri net in PROM

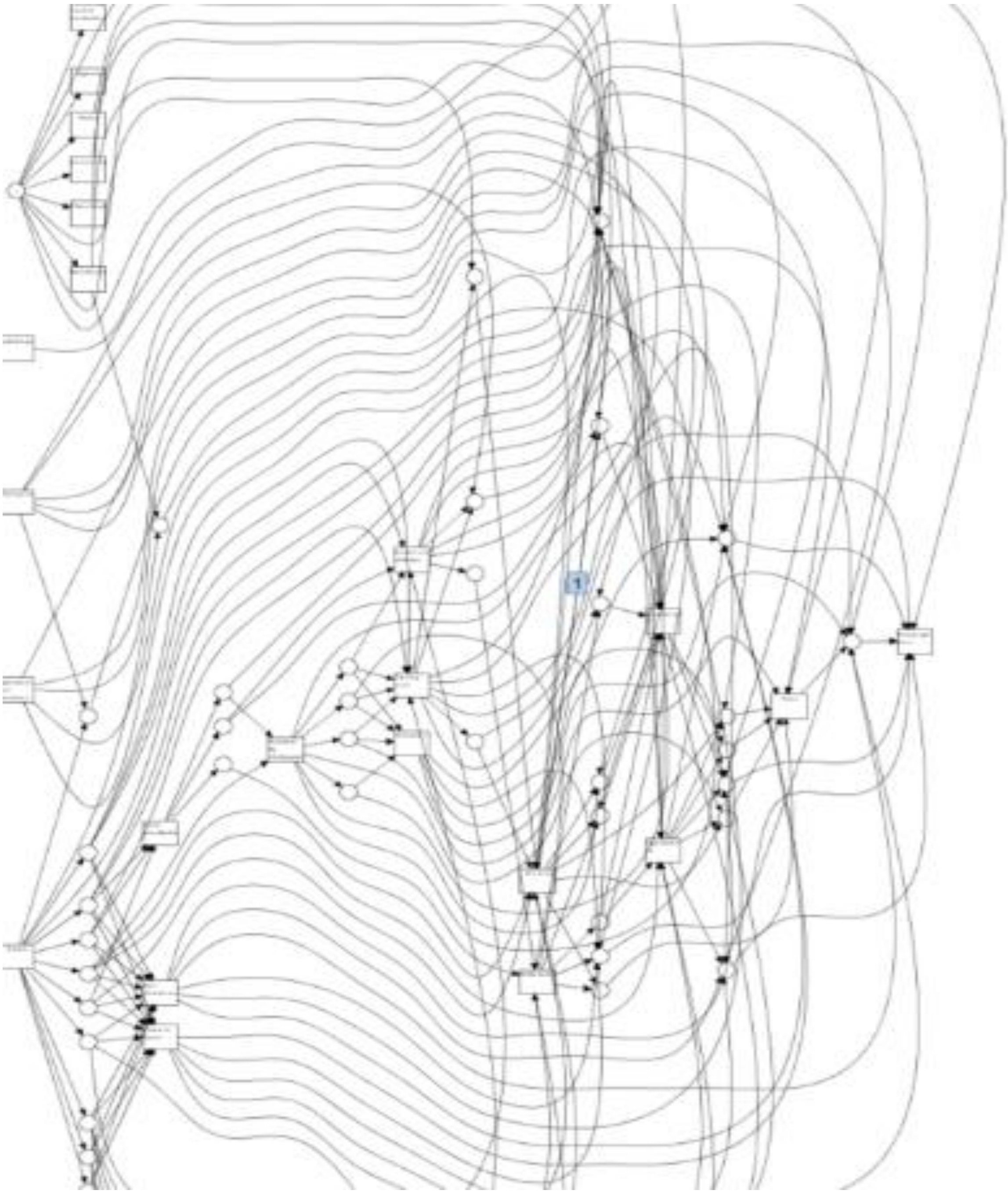


Dotted Chart in PROM

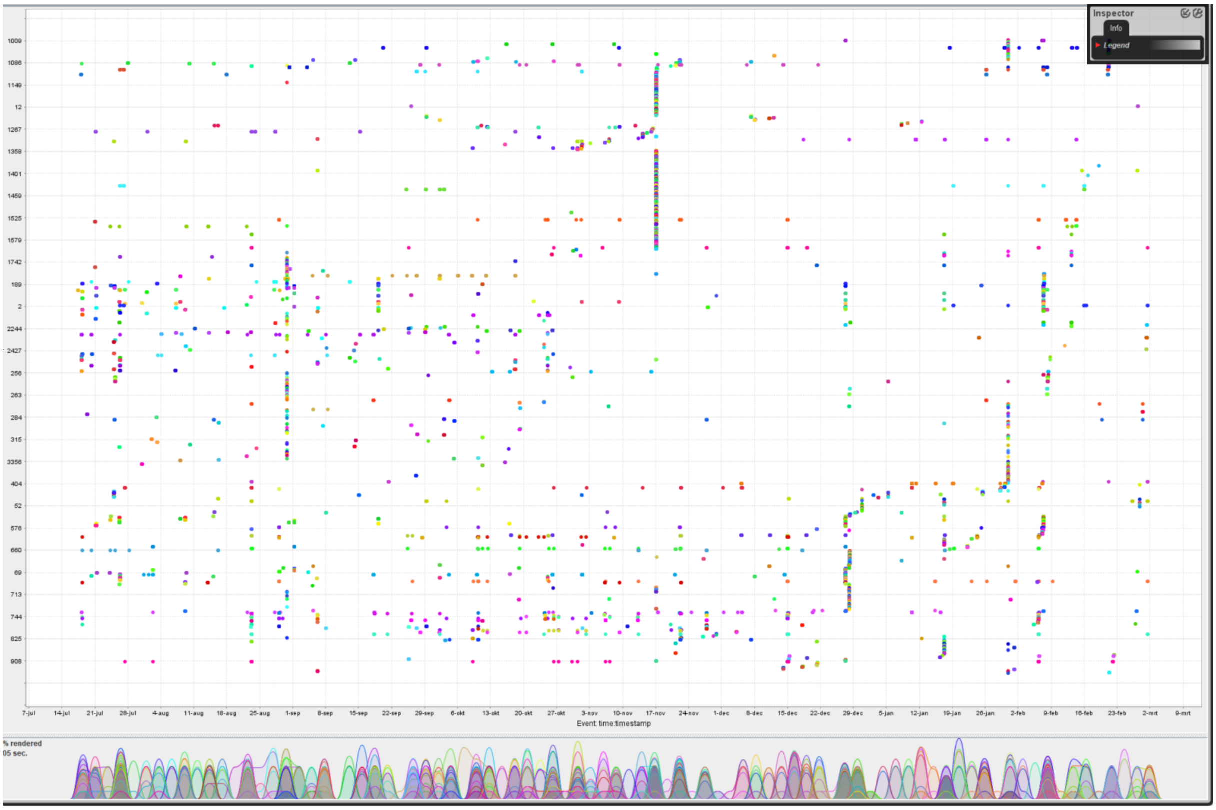
Appendix 4



Unreadable BPMN in PROM



Unreadable Petri Net in PROM



Dotted Chart in PROM

Appendix 5

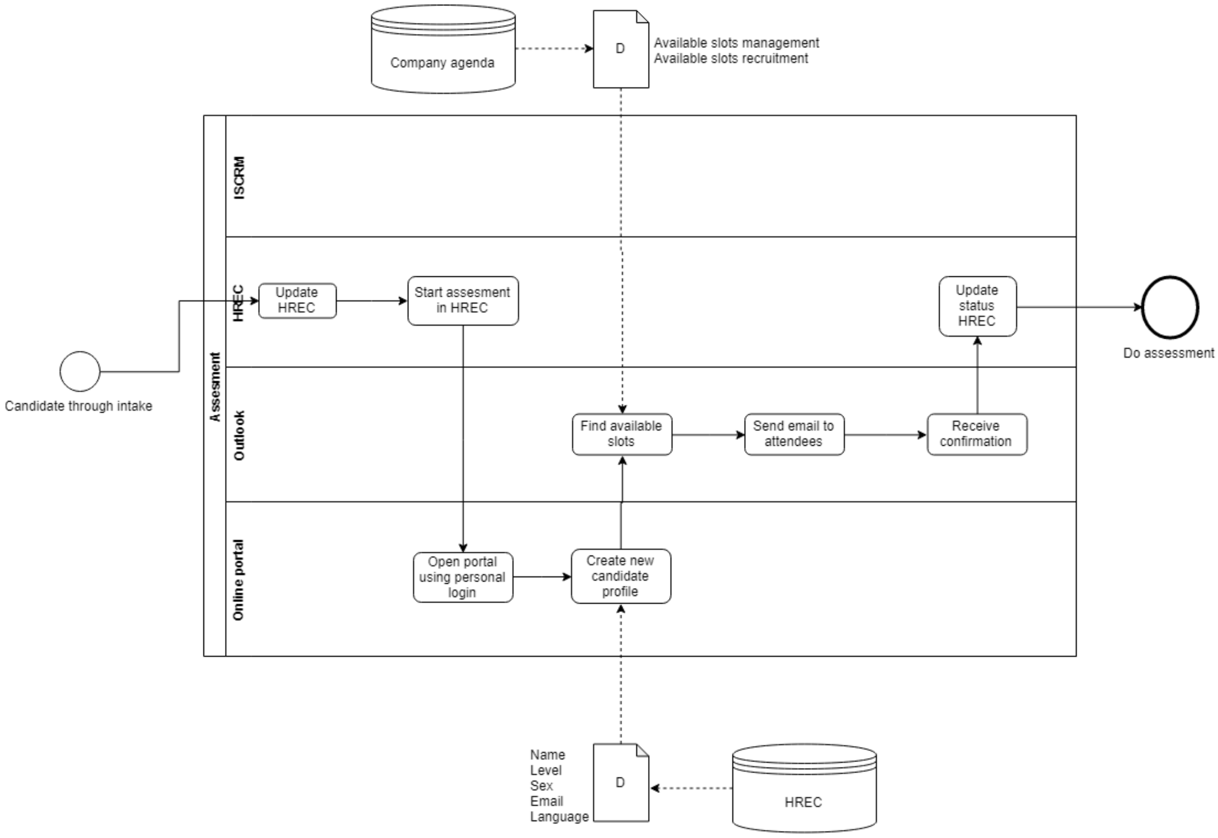
Preparation table

Validation model	Repeatability	The validation model is based on literature, for which all the sources are stated and it therefore should be repeatable.
Validation model	Ethics	This should not be a problem because the validation model is based on literature
Treatment Design	Applied treatments	Three treatments are applied: The first treatment involves interviews with employees to manually create a BPMN and judge if processes are eligible for RPA. A PDD was created up front to act as guideline for this process. The second treatment involves Process Discovery using event logs. For this treatment, another PDD was created that has acted as guideline for the experiment. The third treatment involves Process Mining using a key logger to gather data. Again, for this treatment a PDD has been created.
Treatment Design	Which treatment instruments are used	The PDD acts as guidance for the experiments.
Treatment Design	Treatments allocated to validation models	The validation model for this research are the requirements that have been created and validated upfront.
Treatment Design	Treatment schedule	To ensure the validity of this research, it has been decided that the Process Mining methods should be done before the manual method using interviews, so the results cannot be stained by knowledge of the process.
Treatment Design	Validity treatment design – Inference support	No inference will influence this research. The Event Logs are only of the HR recruitment system and will not include any other data. Interviews will be on a single topic as well and should therefore not suffer from inference. The Key Logger will work in a controlled environment and should therefore not suffer from inference.
Treatment Design	Validity treatment design – Repeatability	The application of the treatment should be repeatable since the treatments are explained and deliberated using a PDD for each treatment. This PDD is used during the experiment to make sure the treatment is executed properly according to the process described by the PDD, and is therefore repeatable.
Treatment Design	Validity treatment design – Ethics	There are no test subjects involved that have to execute the treatments. There are however people involved in the collection of the data that are needed to execute the treatments. There are numerous data sources in this experiment. To execute the treatments, the following data sources are used: - Event Logs The Event Logs are anonymised so no personal information from HR employees or potential candidates can be found in the dataset. This does not affect the results of using Process Mining on this data.

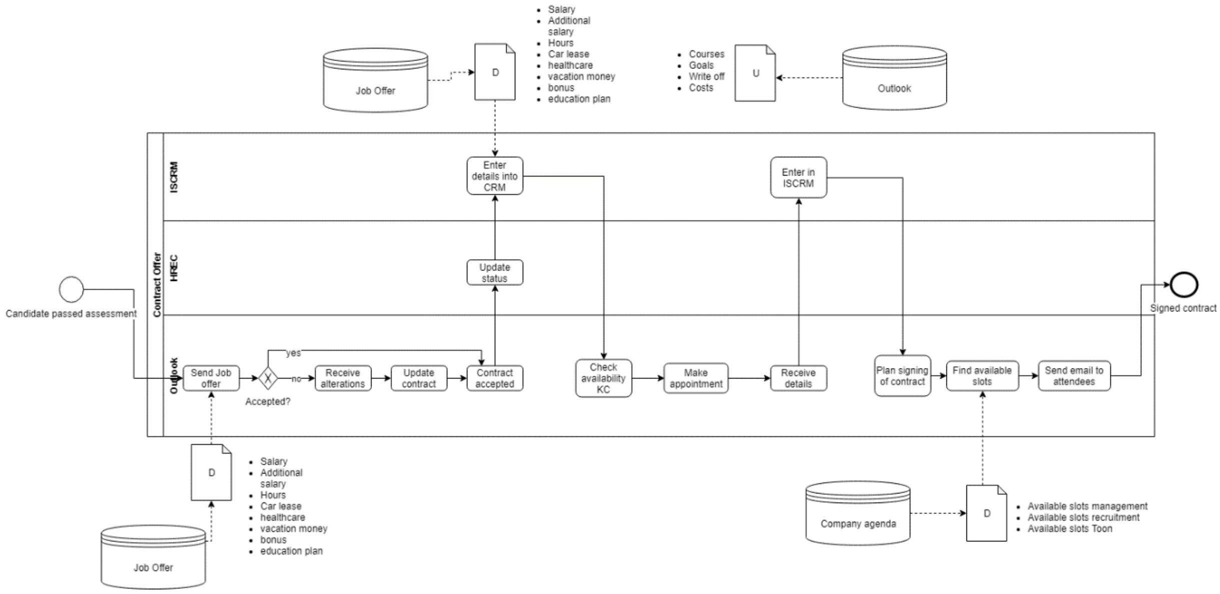
		- Key Logger tool Logging all the data using a Key Logger tool is for HR an unacceptable data breach. The data cannot be fully anonymised because the use of screenshots and therefore, a dummy dataset is created. This is done with the help of an HR employee who will follow every step in the process, using made up potentials. This is done to ensure no personal information will be exposed. - HR employees The HR employees that will help collect the data for this research, are fully aware of why they are being interviewed and how the data that are collected will be used.
Measurement Design	Variables to be measured	This results in various models that will act as measurement. The variable that is measured, is the Recruitment Process from HR. The Recruitment Process will be formalised through the use of different modelling languages. These models will include: - BPMN models - Dotted chart - Petri Net
Measurement Design	Measurement instruments	The instruments: - Notations from the interviews - Excel for data storage - Draw.io to help create the models
Measurement Design	Measurement schedule	The measurement schedule is difficult because it is about ongoing processes. The processes already exist, and are mapped through different techniques.
Measurement Design	Data storage	Data are stored locally on a computer with backup to the cloud to ensure the data cannot be lost.
Measurement Design	Validity - repeatability	The data can be made available so other researchers can verify and retrace the steps that have been undertaken for Process Mining.
Measurement Design	Validity - ethics	No ethical boundaries are crossed
Measurement Design	Validity – Inference support	There are no relevant inferences to the measurement design
Data prep	Will the prepared data represent the same phenomena as the unprepared data?	It will represent the same phenomena, but in a much more comprehensible manner. The unprepared data for Process Mining will likely result in unreadable models. The prepared data should fix this issue.
Data prep	If data may be removed, would this be defensible	Yes this is defensible. It is even a vital part of conducting Process Mining, to clean up and therefore delete or alter parts of the data. This is important because otherwise to fuzzy results will be obtained. The process of cleaning the data will be thoroughly documented on every step.

	beyond reason-able doubt?	
Data prep	Would your scientific opponents produce the same descriptions from the data?	That is difficult to say, because it is difficult to judge how they will clean up the data. However, because the process that is chosen to investigate is quite simple, it should be expected that a scientific opponent should produce the same descriptions from the data.
Data interpretation	Will the interpretations that you produce be facts in your conceptual research framework? Would your scientific peers produce the same interpretations?	To ensure the data are interpreted correctly in the various treatments, and then the results have to be interpreted as well and compared to the requirements, this will be verified by an expert in the field to make sure that the thought process is correct.
Data interpretation	Will the interpretations that you produce be facts in the conceptual framework of the subjects? Would subjects accept them as facts?	This concerns the creation of the models for the treatments. Most care will be put into the treatment using interviews, where the created models are validated by the person that is interviewed to make sure the interpretation is correct.
Support for descriptive statistics	Is the chance model of the variables of interest defined in terms of the population elements?	There are no statistics involved in this research.
Repeatability		Yes, the data on which the models are based are available, and the models will be included in the paper. The comparison and results are based on the models and therefore this research should be repeatable.
Ethics		No ethical boundaries are crossed.

Appendix 6



Company agenda
D
Available slots management
Available slots recruitment
Assesment
ISCRM
HREC
Outlook
Online portal
Candidate through intake
Update HREC
Start assesment in HREC
Update status HREC
Do assessment
Find available slots
Send email to attendees
Receive confirmation
Open portal using personal login
Create new candidate profile
Name
Level
Sex
Email
Language
D
HREC



Job Offer
D
Salary
Additional salary
Hours
Car lease
healthcare
vacation money
bonus
education plan
Courses
Goals
Write off
Costs
U
Outlook
Contract Offer
ISCRM
HREC
Outlook
Enter details into CRM
Enter in ISCRM
Update status
Candidate passed assessment
Signed contract
Send Job offer
yes
no
Accepted?
Receive alterations
Update contract
Contract accepted
Check availability KC
Make appointment
Receive details
Plan signing of contract
Find available slots
Send email to attendees
D
Salary
Additional salary
Hours
Car lease
healthcare
vacation money
bonus
education plan
Job Offer
Company agenda
D
Available slots management
Available slots recruitment
Available slots Toon

Appendix 7

Functional requirements results:

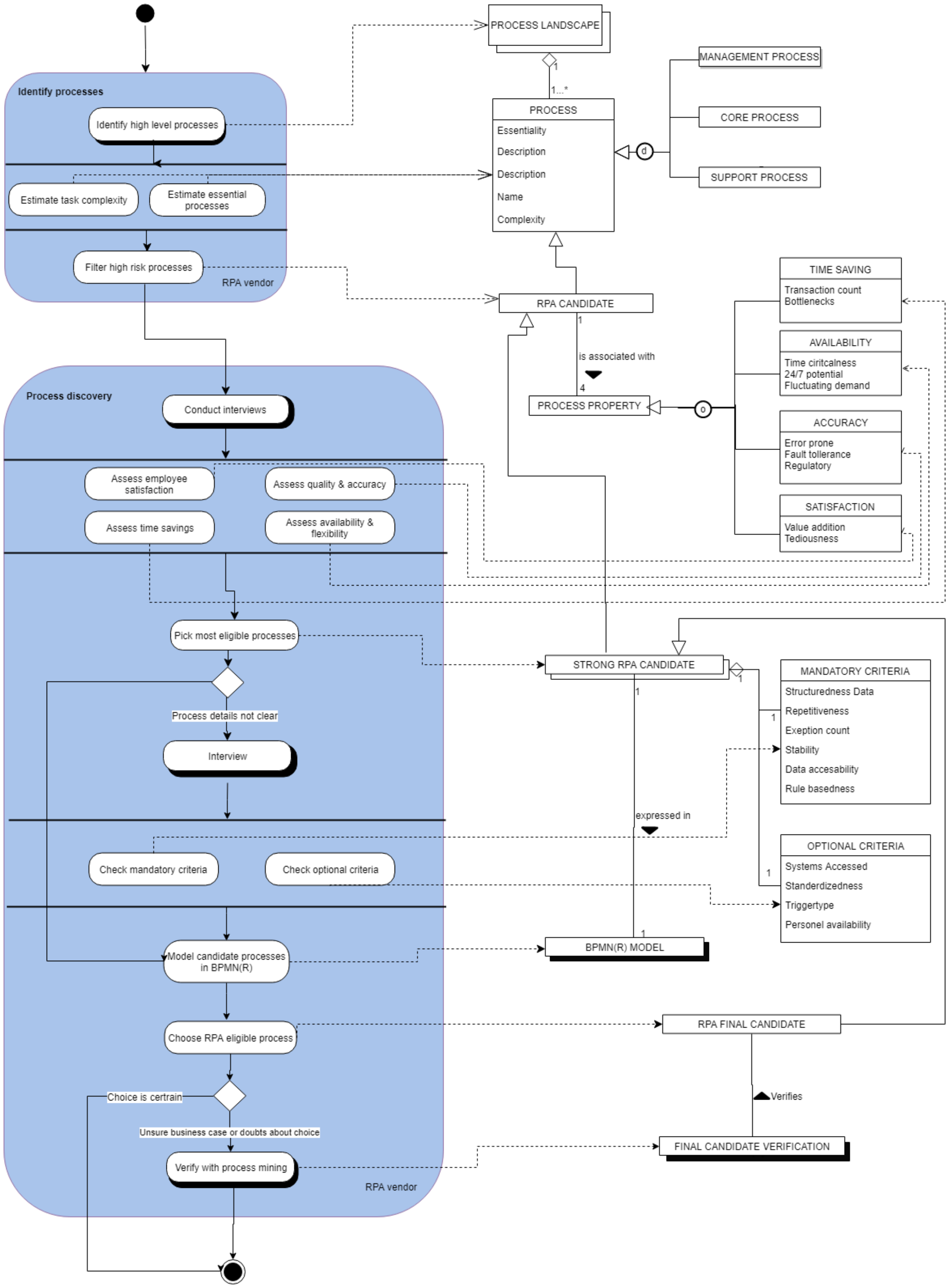
Requirement	Name	Interview	Event Logs	Screen Capture
As an RPA vendor, I want to be able to assess the complexity of a task, so that a proper judgement of RPA eligibility can be made	Complexity	Able	Unable	Able
As an RPA vendor, I want to be able to discover the different tasks that together form the process, so that the complexity of a process can be better assessed.	Complexity 2	Able	Unable	Able
As an RPA vendor, I want to be able to discover different paths the tasks of a process may take, to be able to better understand the complexity of the process	Complexity 3	Able	Able	Able
As an RPA vendor, I want to be able to know if processes follow clear business rules and what the logic behind those rules is, so that that a proper judgement of RPA eligibility can be made	Business Rules	Able	Unable	Unable
As an RPA vendor, I want to know if the business logic is such that humans will always need to play a role in the decision making, so that a proper judgement of RPA eligibility can be made	Human Intervention	Able	Unable	Unable
As an RPA vendor, I want to know what type data sources and how many are involved, so that a proper judgement of RPA eligibility can be made	Data	Able	Unable	Only with digital data
As an RPA vendor, I want to know what the data type is, so that a proper judgement of RPA eligibility can be made	Data2	Able	Unable	Only with digital data
As an RPA vendor, I want to discover every possible step of a process, so that the software bot can be properly modelled	Complexity 4	Able	Able	Able
As an RPA vendor, I want to be able to see the number of repetitions of a process, so that the business case of RPA can be assessed	Repetitions	Unable	Able	Able

As an RPA vendor, I want to assess the environment and stability of the process, so that a proper judgement of RPA eligibility can be made	Process Stability	Able	Unable	Unable
As an RPA vendor, I want to be able to see if multiple systems are involved in the process, so that a better judgement of the business case of RPA can be made	Involved Systems	Able	Unable	Able
As a process owner, I want to be able to know which employees are involved in a process, so that impact on HR can be assessed	Involved Systems	Able	Able	Able if set up correctly
	12	11	5	9

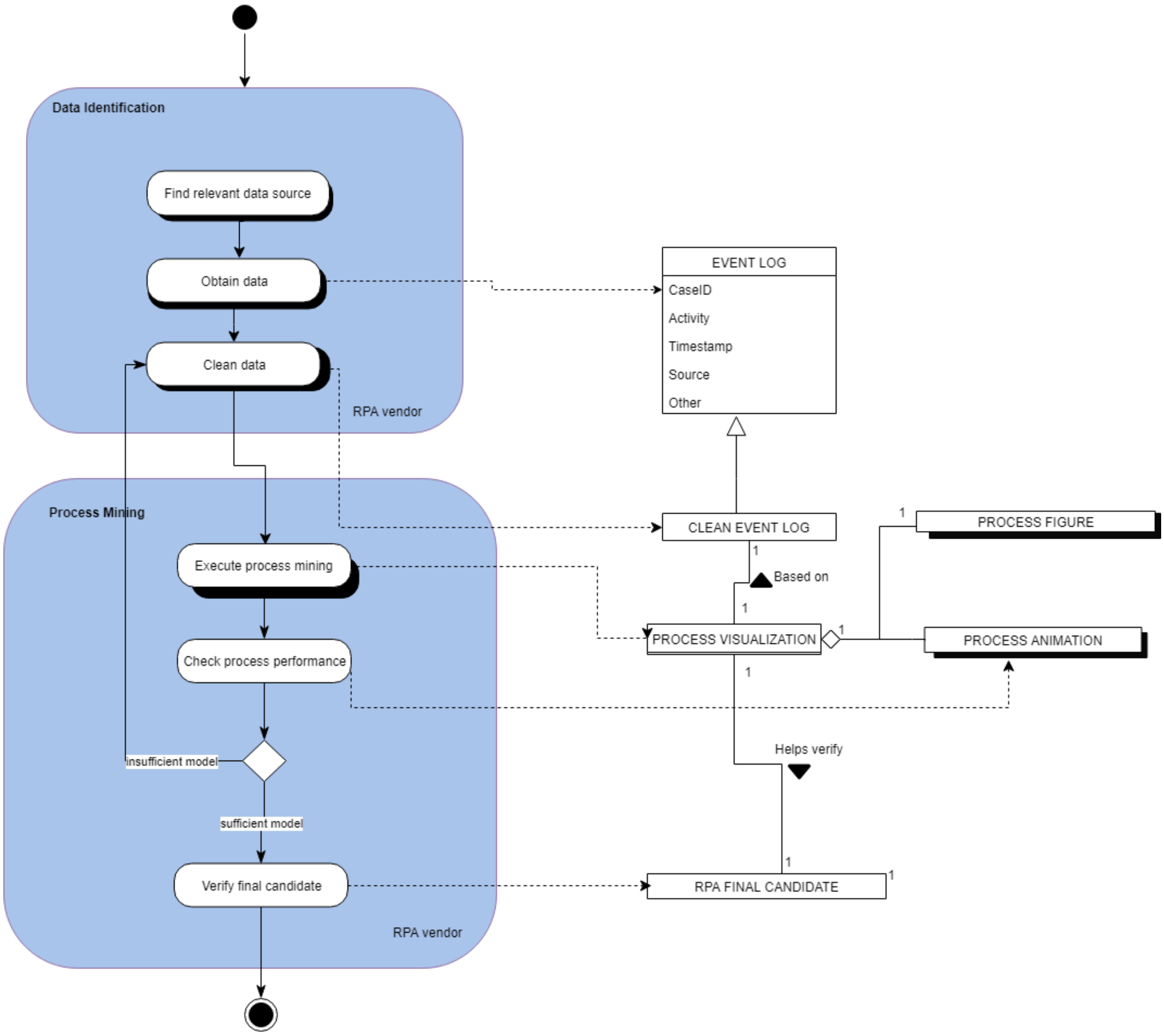
Non Funtional requirements results:

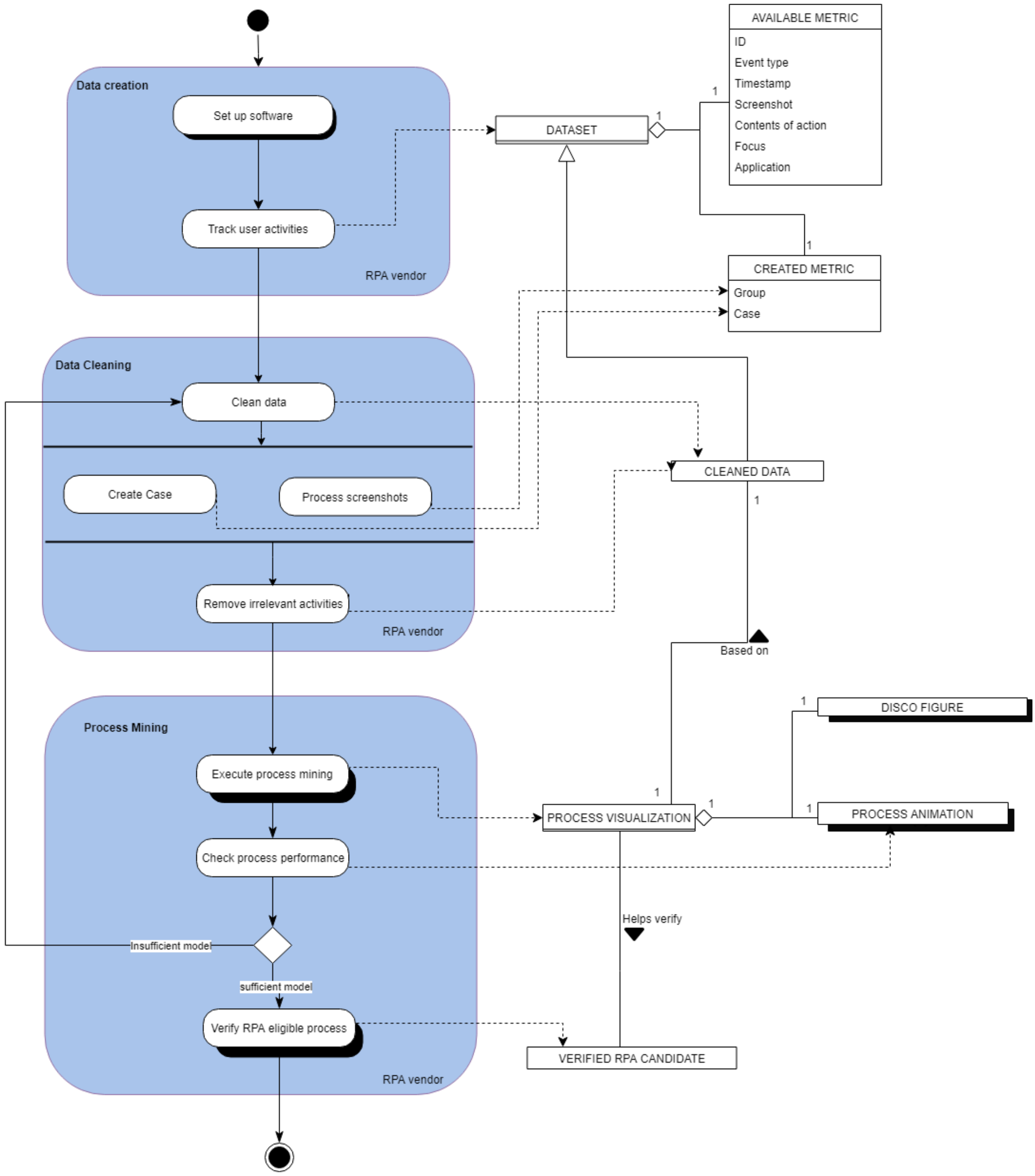
Requirement	Interview	Event Logs	Screen Capture
As an RPA vendor, I want the discovered processes to be accurate, so that the RPA project can run smoothly	Never sure	Able	Able
As an RPA vendor, I want to find as many processes as possible, so that the chances of finding suitable processes for RPA are increased	Able	Depends on data	Able
As an RPA vendor, I want to find processes as quickly as possible, so that the RPA implementation process is speeded up	Unable	Depends on data	Depends on data
As a process owner, I want the most impactful potential cases found, so that the RPA project will be as successful as possible	Able	Able	Able

Appendix 8



PDD Conclusion


Data Identification
Find relevant data source
Obtain data
Clean data
RPA vendor
Process Mining
Execute process mining
Check process performance
insufficient model
sufficient model
Verify final candidate
RPA vendor
EVENT LOG
CaseID
Activity
Timestamp
Source
Other
CLEAN EVENT LOG
Based on
PROCESS VISUALIZATION
PROCESS FIGURE
PROCESS ANIMATION
Helps verify
RPA FINAL CANDIDATE


Data creation
Set up software
Track user activities
RPA vendor
DATASET
AVAILABLE METRIC
ID
Event type
Timestamp
Screenshot
Contents of action
Focus
Application
CREATED METRIC
Group
Case
Data Cleaning
Clean data
Create Case
Process screenshots
Remove irrelevant activities
RPA vendor
CLEANED DATA
Based on
Process Mining
Execute process mining
Check process performance
Insufficient model
sufficient model
Verify RPA eligible process
RPA vendor
PROCESS VISUALIZATION
DISCO FIGURE
PROCESS ANIMATION
Helps verify
VERIFIED RPA CANDIDATE