UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics and Computer Science

# Research and Design of Collecting and Analysing the Customer Journey in a Collaborative Software Tool


J.R. Harms


Master Thesis
April 2020

**Study Programme**
MSc Business Information Technology
**Graduation Committee:**
dr. A.I. Aldea
dr.ir. M.J. van Sinderen
**External Supervisor:**
ir. S.W. Nijenhuis


University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

J.R. Harms: Research and Design of Collecting and Analysing the Customer Journey in a Collaborative Software Tool
Master Thesis, *University of Twente*, April 2020


AUTHOR:
*J.R. harms*

| | |
|---|---|
| Study Programme | MSc Business Information Technology |
| Email: | j.r.harms@alumnus.utwente.nl |

GRADUATION COMMITTEE:
*dr. A.I. Aldea*

| | |
|---|---|
| Faculty: | Behavioural Management and Social Sciences |
| University: | University of Twente, Enschede, The Netherlands |
| Email: | a.i.aldea@utwente.nl |

*dr.ir. M.J. van Sinderen*

| | |
|---|---|
| Faculty: | Electrical Engineering, Mathematics and Computer Science |
| University: | University of Twente, Enschede, The Netherlands |
| Email: | m.j.vansinderen@utwente.nl |

*ir. S.W. Nijenhuis*

| | |
|---|---|
| Company: | Fortes Solutions B.V. |
| Position: | Director |
| Email: | s.nijenhuis@fortes.nl |

# Preface

This thesis concludes my time at the University of Twente. During the bachelor programme Business and IT and master Programme Business Information Technology I always felt at home. I look back at a wonderful time in which I learned a lot, made great new friends and discovered new parts of the world.

I would like to thank my supervisors Adina Aldea, Marten van Sinderen and Sander Nijenhuis for their guidance during the research. Their positive attitude and expertise from their respective fields really helped with writing this thesis. Furthermore, the meetings were always a motivation to move forward. I would also like to thank Fortes for the research opportunity and all my colleagues at Fortes for their input and support during the research. Going to the office was always a pleasure.

Finally, I would like to thank my family, roommates and friends for their support and welcome distractions from time to time.

Rick Harms

# Management Summary

Software companies and specifically *Software as a Service* providers are looking for ways to improve their software. To achieve this, insight in how users use their software can help. The customer journey through an application could give insight in if users follow the expected use cases and use all functionality as expected. Process mining; the area of analysing logs to discover processes, could be a good starting point to discover these customer journeys. Privacy of the user is of course of essence. This research therefore looked at *a customer journey process mining approach that takes the privacy of users into account so that software companies can improve the usability of their collaborative software*.

A systematic literature review on process mining, user behaviour, collaborative software and privacy was used to give an overview of the current stage of user behaviour process mining. Four categories of process mining were identified: *Business Process Mining*, *Service Mining*, *Mining Software Process* and *Mining User Behaviour*. The results from this last category were used in the remainder of this research. Furthermore, the literature review identified a few techniques and tools that can be used for mapping the customer journey in collaborative software. Last, the literature review identified What is needed to guarantee user privacy in terms of the General Data Protection Regulation. The software company should choose between anonymous data, which protects the user better but is less detailed, or pseudonymised data, which is more work to implement because of the privacy measures but gives more details on the user behaviour. Secondary to that, techniques to protect business privacy were discussed.

Based on the literature review, a solution was designed to help software companies with implementing user behaviour tracking. Three methods were created: *Functionality Tracking*, *Customer Journey Tracking* and *Personalised Feedback Tracking*. The first method can be used by small companies with little experience in user behaviour tracking. The tracking is Anonymous, but the results will not include any user journeys. The second method supports customer journey tracking for both collaborative and non-collaborative software. customer journey tracking can also be done anonymously except for the collaborative variant which used pseudonymised data. The third method adds the possibility to give feedback to the user based on

their journey through the application. This is the most advanced variant and therefore only suitable for large companies. All three methods are based on the same concepts which makes it possible to start with functionality tracking and later extend the tracking to customer journey tracking or personalised feedback tracking.

A Prototype implementation was used to show how solution can be used at a software company. Fortes Solutions was used as an example case and anonymous customer journey tracking was added to their application.

The prototype implementation was used in a *single-case mechanism experiment* to show that the tracking actually worked. A scenario based on use cases of the application was made. Participants clicked through the application and this data was then analysed in three different tools (*Grafana* for usage data and *RapidMiner* and *Celonis* for the customer journeys). A workshop was then held at Fortes to show how the method was implemented and what results it gave. This workshop was followed by a questionnaire to determine is the method would be deemed useful by the participants. The results showed a positive attitude towards the method. although the participants were not completely confident about the privacy, the method was considered scalable.

This research showed that it is possible to get insight in user behaviour in existing software by looking at the customer journey with the help of process mining. This was all done without compromising on the privacy of the user or business. The method can be implemented without any prerequisites such as existing logs or permission of the user. The solution can thus be used by any software company. Further research could improve the solution and also examine the feedback to user method.

This research was limited by the fact that not all methods were validated. The case study only considered the customer journey tracking method. Regarding the results on usefulness from the questionnaire it should be noted that the number of participants was limited and that the only interaction with the method was through the workshop.

# Contents

# List of Figures

# List of Tables

# List of acronyms

**BPMN** Business Process Model and Notation

**CSCW** Computer Supported Collaborative Work

**FCC** Fortes Change Cloud

**GDPR** General Data Protection Regulation

**NPM** Node Package Manager

**SAAS** Software as a Service

**SDK** Software Development Kit

**SLR** Systematic Literature Review

**UTAUT** Unified theory of acceptance and use of technology

# Chapter 1

# Introduction

Software companies are transforming in the way they attract new customers. Traditionally, software was sold as a package deal to companies. Customers payed upfront and bought predefined functionality. In newer business models, software is sold as a service, where the added value for the customer is in the service instead of the product itself. Inherently, software companies must become more agile to keep up with the customer needs. To improve their service, software companies should know how their service is used. Are their customers using it as intended? What functions are critical? What functionality are rarely used? The answers to these questions can help software companies to further develop and improve their software.

Fortes Solutions is currently transforming their strategy from sales to customer driven approach. With their product Fortes Change Cloud (FCC), they are offering a full range of online apps that companies use for portfolio and project management. Their mission is to enable change by offering a set of apps that support an agile approach (do agile apps), as well as a set of apps designed for traditional waterfall methods (do waterfall). New customers of Fortes are can use the software as a Software as a Service (SAAS) solution and existing customers are also switching from an on-premise environment to SAAS. This offers opportunities to quicker release new versions of the application. However, to determine which functionalities should be updated, insight in the use of the application is needed.

## 1.1 Research Goals

The goal of this research is to help software companies to improve their software. Software companies such as Fortes Solutions create a product that in their eyes fits the needs of the customer. This is mostly tested by asking the customer for feedback. However, this feedback is limited and in most cases, it does not cover the entire product. This research aims to design a method on how to apply customer journey process mining in collaborative software. With this method, software companies have a step-by-step approach to get insight in the actual use of the software. The generated data can give an overview of the actual use of specific parts of the application. The results of the method could also give insight in how users collaborate in the software.

## 1.2 Scope

This research is intended for software companies that want to investigate how their users behave in their software and want to improve their application based on this. The research offers a step-by-step approach which means that no preceding knowledge is needed. Additionally, it is assumed that the software company currently does not collect data on user behaviour. This research also considers software with a collaborative aspect and for mature software companies this research can help in including personalised feedback to users. The results of this research can be used in agile software development, where feedback from customers is important. The research is scoped for software companies that provide a SAAS solution. This is important since the tracking data should be accessible for the software company.

## 1.3 Research Design

This research is divided in a descriptive research part and a design research part. For the descriptive research part, a Systematic Literature Review (SLR) approach is conducted to answers knowledge questions.

The design research part aims to find a suitable approach for analysing user behaviour in a collaborative software tool. For the descriptive research part, the guidelines of Kitchenham and Charters (2007) are followed.

The design science methodology of Wieringa (2014) is followed throughout the research. Within design science, an artifact is studied in context. The underlying principle is that the context should be understood to understand the design problem. Design science and divides in three stages: problem investigation, treatment

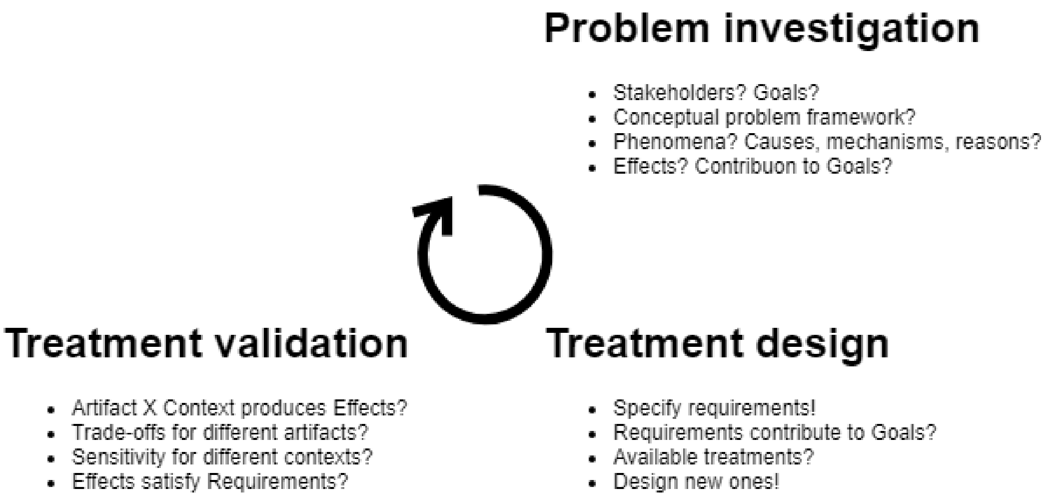



**Figure 1.1:** Design Cycle, adapted from Wieringa (2014).

design and treatment validation. Figure 1.1 shows these stages in context of the design cycle. The design cycle is iterative. After treatment validation it is possible to start again with problem investigation with the input of the validation. For each stage there are knowledge questions and design problems that are applicable for that stage. In Figure 1.1, the question marks indicate a knowledge question, whereas an exclamation mark indicates a design problem. The design cycle is part of the larger engineering cycle which also includes treatment implementation and implementation evaluation. Treatment implementation and evaluation are defined by Wieringa as applying and evaluating the 'final' artifact in the real world. These steps are beyond the scope of most researches and of this research. This research is limited to applying a prototype to a model of the context.

### 1.3.1 Research Questions

The main research question is:

*How to design a customer journey process mining approach that takes privacy of users into account so that software companies can improve the usability of collaborative software.*

The following sub research questions were derived from the main research question:

RQ I What techniques can be used for process mining of user behaviour?

RQ II How can the customer journey in collaborative software best be mapped?

RQ III Which privacy aspects are relevant to collecting user behaviour?

RQ IV How can the tracking of user behaviour be added to existing software?

RQ V How can tracking of user behaviour be applied in a privacy preserving way?

The goal of RQ I is to find out what the current state of art in the literature on process mining user behaviour is. It will discuss what parts of process mining are related and which tools and techniques can be used. RQ II Looks at the literature on customer journey and possible techniques that can be used for mapping the customer journey in collaborative software. RQ III Looks at the privacy part. For example, which parts of the General Data Protection Regulation (GDPR) are relevant and what trade-offs between privacy of user behaviour and usefulness of data should be made. RQ V tries to apply the trade-offs discovered with RQ III so that information on the behaviour of users and the collaboration between users is still useful. RQ IV Focuses on the whole methodology of extracting, storing, processing and visualising (collaborative) user behaviour. The outcome will be in the form of a method for applying customer journey process mining within a software application.

### 1.3.2 Research Methods

Different research methods will be used in this report. First, to answer RQ I, RQ II and RQ III a literature review will be conducted. This review will be discussed in Chapter 2. For RQ V and RQ IV a treatment design is proposed based on the results of the literature review. A prototype will be made for Fortes Solutions. This Prototype will be validated using both A single-case mechanism experiment as described by Wieringa (2014). Finally, expert opinion will be used to validate the prototype and argue about how it would stand in a real-world situation.

## 1.4 Report Structure

This report is structured as follows: Chapter 2 contains the literature review to answer the first three research questions. Chapter 3 proposes a solution design for RQ V and RQ IV. This solution is used in a case study discussed in Chapter 4. A validation of the research can be found in Chapter 5. Finally, in Chapter 6, the research will be concluded.

## Chapter 2

# Literature Review

In this Chapter, the literature review is discussed. The literature review tries to find answers to RQ I, RQ II and RQ III. The literature method will be discussed in Section 2.1. The results can be found in Section 2.2. Section 2.3 discusses the results per research question. Finally, Section 2.4 Concludes the literature review.

## 2.1 Literature Review Method

### 2.1.1 Review Planning

The planning of the review started in July 2018. The review conduction stage was done between July 2018 and January 2019. From January 2019 till Jun 2019 the reporting part took place. All work was done by the author of this review.

#### Search Process

This research focused on scientific databases to find relevant peer-reviewed literature. The following databases were found relevant for this SLR:

– Scopus (https://www.scopus.com)
– Science Direct (https://www.sciencedirect.com)
– Web of Science (https://www.webofknowledge.com)

Scopus is the largest of the three databases. However, it falls short on literature about social sciences. Hence ScienceDirect and Web of Science were added. Initially Google scholar was also considered but this database lacked good filtering options and with the huge amount of results in preliminary searches this database was skipped.

Based on the research questions, the following keywords were used to search in the databases: ('customer journey' OR 'user flow' OR 'user journey' OR 'process mining' OR 'business process discovery') AND ('tools' OR 'technique' OR 'method' OR 'approach'). The keyword search was performed on the title, abstract and keywords of the papers in the databases.

### 2.1.2 Review Conduction

In this section the steps for conducting the SLR are discussed. This covers the inclusion and exclusion criteria (Section 2.1.2), the study selection process (Section 2.1.2), The data extraction form (Section 2.1.2) and the backward reference search (Section 2.1.2).

#### Inclusion and Exclusion criteria

The inclusion and exclusion criteria for this review can be found in Table 2.1. This study looked for literature on the intersection of customer journey and process mining. Papers that discussed both topics were directly added. Papers that had relevant information on one topic in combination with information on techniques, methods or privacy were added as well. Papers not related to computer science were discarded. Only papers from the last five years were considered. Older papers that might be relevant were later added during the backward search. Duplicate papers that were already found in another database or papers that had a significant overlap with another paper from the same author(s) were marked as duplicate and excluded. Studies that could not be retrieved were also excluded.

**Table 2.1:** Inclusion and Exclusion Criteria

| Inclusion Criteria | Exclusion Criteria |
| --- | --- |
| Peer reviewed studies | Studies in languages other than English |
| Journals and conference proceedings | Studies before 2014 |
| Studies that relate to customer journey techniques | Studies not related to the research questions |
| Studies about process mining | Duplicate studies |
| Studies that relate to privacy | Conference reviews, notes, short papers |
| | Inaccessible studies |

### Study Selection

The following process was used for selecting studies:

- Search selected databases using the keywords to find relevant papers
- Apply filters on search results to exclude non-English studies, non-conference or non-journals and studies before 2014
- Merge search results and remove duplicate studies based on title and author(s)
- Exclude studies based on titles and abstracts
- Remove non-accessible studies
- Evaluate studies based on full text
- Add studies based on backward search
- Obtain studies

### Data Extraction

A data extraction form was used to group and select relevant studies that contribute to the research questions. Table 2.2 gives an overview of the eight selected characteristics that were used for analysis in the full text step.

### Backward Search

Backward reference search was used to find papers that are relevant but were outside of the scope of the search results. From the papers found to be relevant, the references were scanned. Relevant papers were added to the final list.

**Table 2.2:** Extracted data

| No. | Extracted data | Description | Type |
|---|---|---|---|
| 1 | Name, Year, Author | General description of the paper | General |
| 2 | Summary | Short summary about the paper | General |
| 3 | Type of study | Case study, literature review | General |
| 4 | Tool | Tools used | RQ I |
| 5 | Privacy | Is the subject of privacy discussed? | RQ III |
| 6 | User behaviour | Is user behaviour discussed? | RQ I |
| 7 | Process mining | Discusses the topic of process mining | RQ I |
| 8 | Customer journey | Discusses the topic of customer journey | RQ II |

### 2.1.3 Synthesis

The full SLR process is shown in Figure 2.1. From the databases Scopus, Science Direct and Web of Science the number of articles found were 2467, 222 and 1026, respectively. In total, 3715 articles were found. The inclusion and exclusion criteria as discussed in Section 2.1.2 were applied. The title, abstract and keywords of the remaining 980 articles were then reviewed and irrelevant articles were excluded. Of the remaining 58 articles, two were not accessible. All others were evaluated using the data extraction form. 33 articles were found relevant based on this step. An additional seven articles were added based on the backward search: S10, S25, S27, S33, S38, S39 and S40. S33 is in the form of a short paper and was initially excluded. However, this paper contained relevant information on customer journey and was therefore added. Likewise, the book S27 was added later because it contains thorough information on process mining from the founder of process mining. All selected papers are listed in Appendix B.

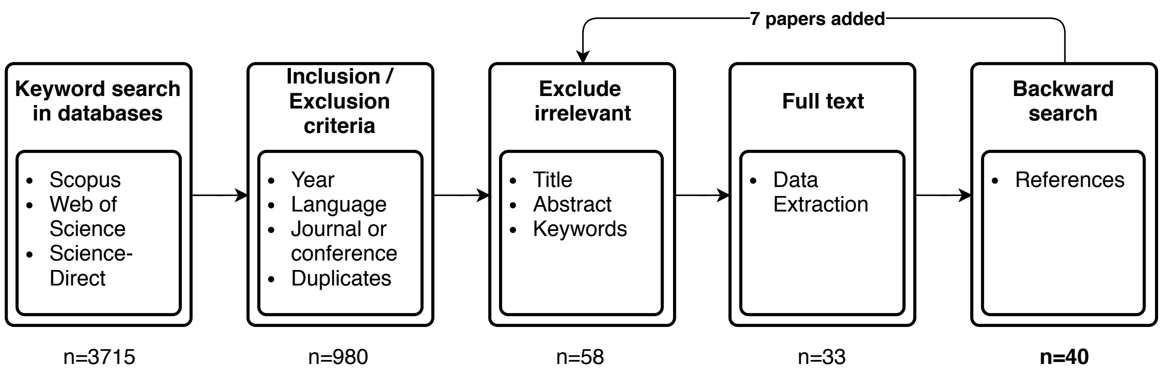



**Figure 2.1:** Study Selection.

The selected papers were added to the data extraction and tested against the criteria as described in Section 2.1.2. The results of the data extraction step can be found in Appendix A.

Figure 2.2 shows the papers found per year. Most papers were published in the last two years. All found review papers were published in 2018. From all papers, only six discuss the topic of privacy, 31 describe some form of tracking or visualising user behaviour. 32 are on the topic of process mining and twelve papers were found related to customer journey. Only four papers were found on the intersection of process mining and user behaviour: S16, S17, S20 and S27. Moreover, the first three papers have overlapping authors and S27 is a book that describes many (theoretical) applications for process mining.

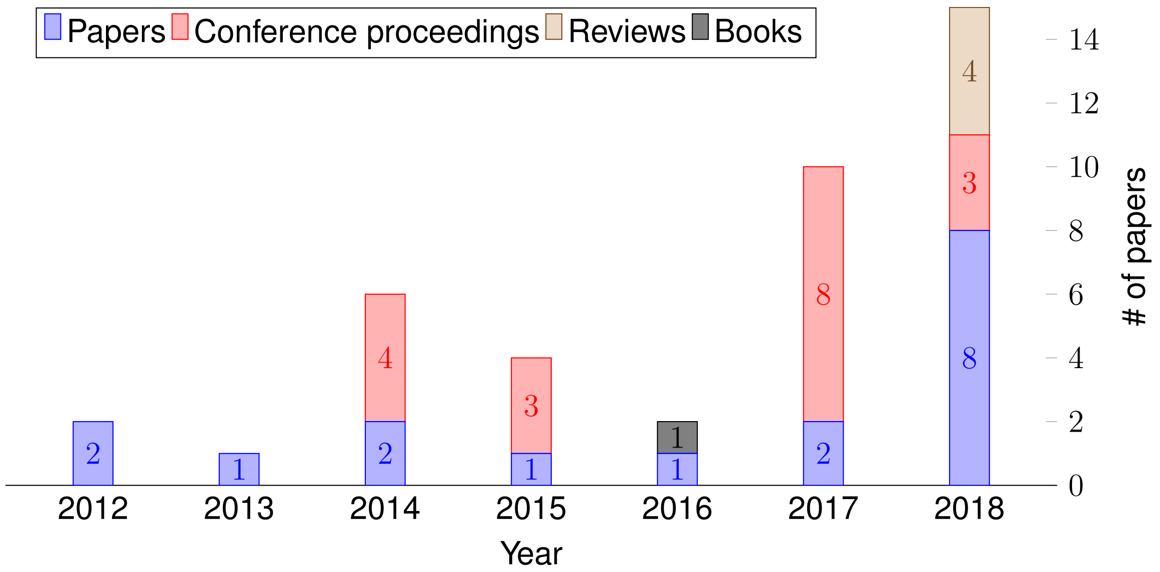



**Figure 2.2:** Relevant publications per year grouped by type.

## 2.2 Results

This section contains the results from the literature review. The results are grouped by research question. Section 2.2.1 discusses RQ I, Section 2.2.2 discusses RQ II and in Section 2.2.3 the results on RQ III can be found.

### 2.2.1 RQ I - What techniques can be used for process mining of user behaviour?

This section answers RQ I. the concepts of process mining are first shown. After that, four categories of process mining that can be used in different contexts are described. Then the eXtensible Event Stream standard is discussed and finally tools on process mining That are presented in the literature are listed.

**What is Process Mining?**

Process mining bridges the gap between process science and data science (S27). Process mining takes advantage of existing log files to analyse the actual business processes. This process can then be checked against the expected business model to find exceptions or bottlenecks.

There are three main types of process mining: discovery, conformance and enhancement (S27). For discovery, event logs are used to discover process models without any prior knowledge about the process. Conformance checks the models from the event logs against the existing process models and looks for any discrepancies. Enhancement extends and improves existing models with information from the actual process event logs. In the literature, Process discovery is the most studied type, followed by conformance checking and then model enhancement (S9). Figure 2.3 shows a schematic overview of the concept of process mining.

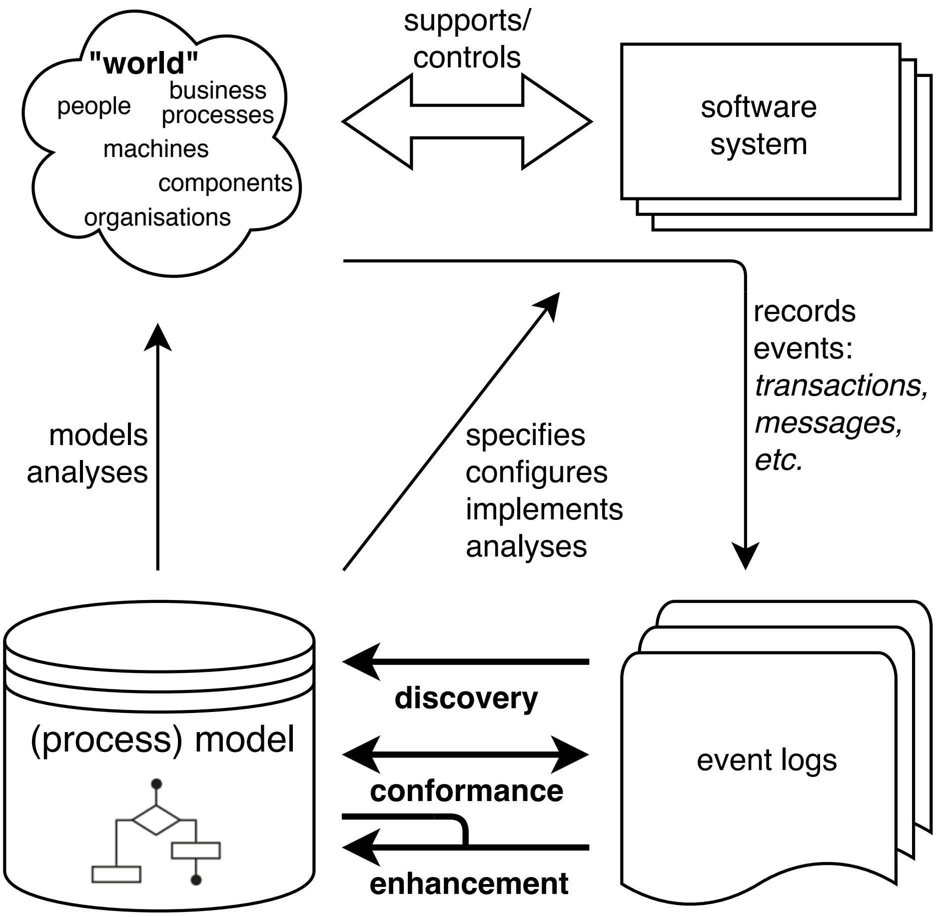



**Figure 2.3:** Concept of process mining: discovery, conformance and enhancement. Adapted from: van der Aalst (2016).

In 2012, van der Aalst et al. from the IEEE Task Force on process mining presented the process mining manifesto (S40). According to this manifesto, there are different perspectives which can be covered by process mining: control-flow, organisational, case and time. The organisational perspective is close to mining user behaviour, since this perspective focuses on hidden information about resources in the log. This can give information on which actors are involved and how they relate to the task and each other (S40).

Process mining is used in a wide range of domains (S9). Although most research on process mining is pure or theoretical (S9), some researchers focus on the application of process mining. Industries such as healthcare and manufacturing seem to get more attention from researchers on process mining. According to Thiede, Fuerstenau, and Barquet, A possible explanation is that various studies use the same publicly available example log files (S14).

#### Process Mining Contexts

Process mining can be used in different contexts. From this literature review, four categories of process mining were identified that can be applied to a specific context: Business Process Mining, Service Mining, Mining Software Process and Mining User Behaviour. The categories are described in the paragraphs below. All are a subset of generic process mining. However, the context and thus the approach and challenges per category differ. The focus in this literature review is the use of process mining in the category Mining User Behaviour.

#### Business Process Mining

Business process mining is used to gather information on the business process model. It is used to analyse the steps or activities companies follow to deliver a product or service. An example is a business process of a web shop that follows the order of a customer from ordering till delivery. In most cases, a basic control-flow perspective, i.e., ordering of activities is shown (S27). Dijkman, Turetken, van Ijzendoorn, and de Vries used process mining to discover exceptions in business processes in different companies (S3). Their results show that it is possible to show the difference in throughput time between normal paths and paths with exceptions. Bolt, de Leoni, and van der Aalst compared different variants of the same process (S2). Their focus is on visualising differences based on a transitions system. Other perspectives as mentioned in Section 2.2.1 can be used in business process mining. For example, a timeline can be used to which activities are popular on which days. Another example is to create a social network graph between resources. These perspectives are mostly combined with the control-flow perspective (S27). In that case, these perspectives are added to the control-flow as labels to show for example which resource is linked to an activity.

### Service Mining

Process mining can also be used in the context of web services. Web services enables different business applications work together within and across organisational boundaries. The monitoring and analysing of activities inside services or interaction between services is called service mining (S38). Two challenges that exist in service mining are “how to correlate instances” and “how to analyse services out of context” (S38). The first challenge refers to the situation in which cases in one service need to be related to cases in another service. These relations might be not one to one (S38), or the relation might not be stored at all (S24). This so-called correlation challenge can be overcome by using information on how many times an event occurred and at what time an event has occurred (S24). However, this technique only of there are no cycles in the model, which heavily limits applying this technique in real-world cases where loops can occur. According to Thiede et al., Research on process mining from a service perspective is still limited (S14).

### Mining Software Process

Mining software process is defined by Dong et al. (S4) as “utilising mining techniques to discover and analyse software process and eventually (semi-) automatically build software process models from raw event data generated.” The context in this case is the development process of software. Data is collected from software repositories, bug reporting tools and other development software. Liu, Van Dongen, Assy, van der Aalst, and Society looked at behaviour of software components from execution data (S22). Jorbina et al. made a dashboard that can be used for the prediction of various indicators at runtime (S21). That is, based on a training set, runtime data can be analysed, and the probability that a case will succeed based on previous cases in the process can be made.

Liu, Zhang, Li, Gao, and Zeng describe a framework for the discovery of software behaviour (S7). The authors used the process mining toolkit ProM as an example on how the framework works. To collect execution logs, the toolkit itself was modified by adding logging classes to the software. These classes collect event logs on starting a case, using a plugin and ending a case. Based on these logs, a graph on plug-in calls and a user behaviour model was created. In their conclusion the authors suggest that instead of manually instrument the software, it might be more accurate to use method-level log events. There is however a gap between low-level method call events and high-level operations like user behaviour. in S8 this gap is addressed by using a training set in which method-calls were labelled manually with the corresponding user action. This set was then used for alignment-based matching to abstract the user operation log. Existing process discovery approaches can be used on this operation log.

### Mining User Behaviour

As discussed in the previous paragraph, mining software process can be used for user behaviour analysis by manually labelling user actions to calls. (S8). Other research shows that it also possible to directly log user actions (S35). In the case study of Rubin, Mitsyuk, Lomazova, and van der Aalst, the behaviour of users in two different systems: a computer reservations system and a travel portal (S35). For the first case, the tool Disco was used. This tool uses a fuzzy mining algorithm. For the second case, the toolkit ProM was used with a fuzzy miner and a heuristic miner. Rubin, Lomazova, et al. suggest embedding process mining for user and system runtime behaviour into the agile development lifecycle (S36). Using process mining, the authors could visualise the behaviour of the user and discuss this with them. This enabled them to monitor the usage of the system in real time, discover bad usage patterns, gather scenarios to create more realistic acceptance tests, discover frequent and critical paths and retrace system failure with concrete events. As noted by the authors, this paper is only a first step for integrating process mining in the agile lifecycle (S36). Details about collecting data, processing the data and visualising the data are not given.

S28 and S13 used fuzzy mining to investigate how users interact with the software. S15 describe a model called Fuzzy Discrete Event System Specification (Fuzzy-DEVS). Event logs from an e-commerce site were extracted using the System Entity Structure method, which enabled them to have a broader concept of the activities in the case study (S15). Gadler, Mairegger, Janes, and Russo modelled the use of a system with hidden Markov models, to show the intents of users (S19). Setiawan and Yahya take process mining to the physical world by monitoring the physical activities of employees either inside or outside of their workplace using wearables (S12). A behaviour model was created using sequential rule mining and considering time constraints. Privacy concerns on logging daily behaviour of employees are not discussed. Padidem and Nalini looked at usage patterns of customers in an e-commerce website (S23). Four distinctive types of customers with different shopping behaviour were identified.

### XES

In 2010 the IEEE Task Force on Process mining introduced the IEEE Standard for eXtensible Event Stream (XES) (van der Aalst et al., 2012). This standard is officially published by the IEEE as an XML schema for describing the structure of an

XES event log/stream (Verbeek & van der Aalst, 2018). There are a handful of extensions already available, such as: concept, organisation, lifecycle and time. Multiple data mining tools already have support for this format. 12 papers in this literature research described using the XES standard. S7 and S11 both used the organisation Extension, which has three attributes: the name of the resource that triggered the event, the role of the resource and the group in the organisation of which the resource is a member. In both cases only the resource attribute of this extension was used. S13, S15, S17, S22 and S29 only mention That they used the XES standard without any details on how they used it.

### Tools

S4 describe four categories of process mining tools and their usage in mining software process: Data extraction tools, Data pre-processing tools, data mining tools and process discovery tools.

Most researchers use a process discovery tool. Table 2.3 shows the usage of such tools in the papers found in this review. The majority uses ProM and Disco. ProM is an open source tool founded by the process mining group whereas Disco is offered for free for academic usage. This might explain the usage of these tools. Commercial tools such as Celonis are not widely used in literature. According to Maita et al. a possible explanation is academic research bias, since most scientific research papers generally do not use commercial tools (S9). 2.3

**Table 2.3:** Use of process discovery tools.

| Tool | Count | Papers |
|---|---|---|
| ProM | 12 | S2 S4 S7 S8 S12 S15 S22 S23 S29 S30 S32 S35 |
| Disco | 8 | S3 S4 S13 S19 S24 S28 S35 S36 |
| DPILMiner | 1 | S11 |
| nirdizati | 1 | S21 |
| interpretA | 1 | S18 |
| CJM-ex | 1 | S16 |

On tool that is specially design for process mining of user behaviour is CJM-ex from Bernard and Andritsos (S16). This tool is further discussed in Section 2.2.2.

Tools from the other categories are not widely discussed in process mining literature. Maita et al. give a few examples on each category for tools used in mining software process S9. Most process discovery tools have however support for a limited form of data extraction and data pre-processing. For example, ProM has built in functionality to translate csv files to XES.

### 2.2.2 RQ II - What techniques can be used for mapping the customer journey in collaborative software?

This section shows the results of RQ II. Section 2.2.2 Describes how the concept of customer journey is explained in the literature. In Section 2.2.2, the usage of customer journey and customer journey mapping is described. This leads to Section 2.2.2, where the concept of personas in customer journey is explained. Section 2.2.2 literature that researched the combination of customer journey and process mining is shown. Finally, in Section 2.2.2, results on collaborative user behaviour is shown.

#### Customer Journey

The topic of customer journey is best explored by Følstad and Kvale (S5). The authors refer to the customer journey as to "obtain a customer viewpoint on the service process" (S5). There are two broad groups for customer journey approaches: customer journey mapping and customer journey proposition. The former looks at the existing or "as is" service process whereas the latter is more about generative design and which leads toward a possible service "to be" (S5).

The term customer journey is mostly found in marketing journals focusing on e-commerce. Heuchert, Barann, Cordes, and Becker propose an entity-relationship-model to describe the customer experience in Omni-Channel management (S6). This model helps to relate the marketing view on customer journey to the technical Information System perspective. This model can be used in future research in mapping the customer journey in an Information System perspective to decide how logging and monitoring of the customer journey can be embedded.

#### Applications of Customer Journey

Wolny and Charoensuksai used mapping of the customer journey in multi-channel shopping (S37). 16 Research diaries on cosmetic shopping were used to research the journey in buying cosmetic products over two weeks. These diaries were then grouped in three distinctive groups (impulsive, balanced and considered journeys) and for each group a customer journey was made. Each map has multiple distinctive stages (such as pre-shopping, information search and purchase) and each stage shows related topics in the form of an image or a short text. Although the considered journey types were visually shown concisely and clear, the manual work is high in all stages (collecting, interpreting and visualising) (S37).

S1, S31 and S26 are more on the analytical side of customer journey. All three papers focus on analysing the path of users towards an online purchase. Ballestar, Grau-Carles, and Sainz describe the case of a cashback site in which users are clustered based on different variables (S1). For each cluster, a description is made on what type of users belong to it. Most clusters seem to be based on the variable role in social network, which determines if the user is a lonely user or if the user is either referred by someone, has referees or both.

Anderl, Schumann, and Kunz propose to classify customers based on the contact origin and brand usage rather than relying on the assumed browsing goal of the customer (S26). The contact origins were in this case based on firm-initiated channels or customer-initiated channels. The interaction effects between different touchpoints were also reviewed and it was shown that some behaviours show increased purchasing propensity (S26). Wooff and Anderson also looked at the click stream data but included time-weighted multi-touch attribution and channel relevance (S31). Instead of using the so-called first click wins, last click wins or even-weighting methods, a bathtub method is proposed, where the first and last interactions are weighted more than those in between (S31). Both researches are done from the view on e-commerce where the goal is conversion of mostly products. Therefore, the touchpoints are merely a means to an end.

### Personas

Earley suggest the use of personas to get into the mind of the user (S33). Creating personas can help understand how users might react. For example, personas can be used to group the experience of users and to create a customer journey map for each persona (S39). Figure 2.4 gives an example of such customer journey map on user experience at a theme park (S39). In this example, the strong and weak points of each section of the theme park were listed, forming a graph that shows the satisfaction level of the persona over the course of the visit. Thus, visualising the service experience (S39).

### Customer Journey Process Mining

Four papers discuss the use of process mining for discovering the customer journey. The first mention of this concept is in the book of van der Aalst (S27). The author discusses how each touchpoint in the customer journey can generate events that make it possible to understand the customer better and create a better service. The author describes how these events could be used to build a customer journey map. However, this is only theoretical, and no examples are given. Bernard and Andritsos propose a customer journey mapping model based on the XES standard (S16). This enables the use of process mining for customer journey mapping by extending the process mining framework. In another paper, the same authors introduce a

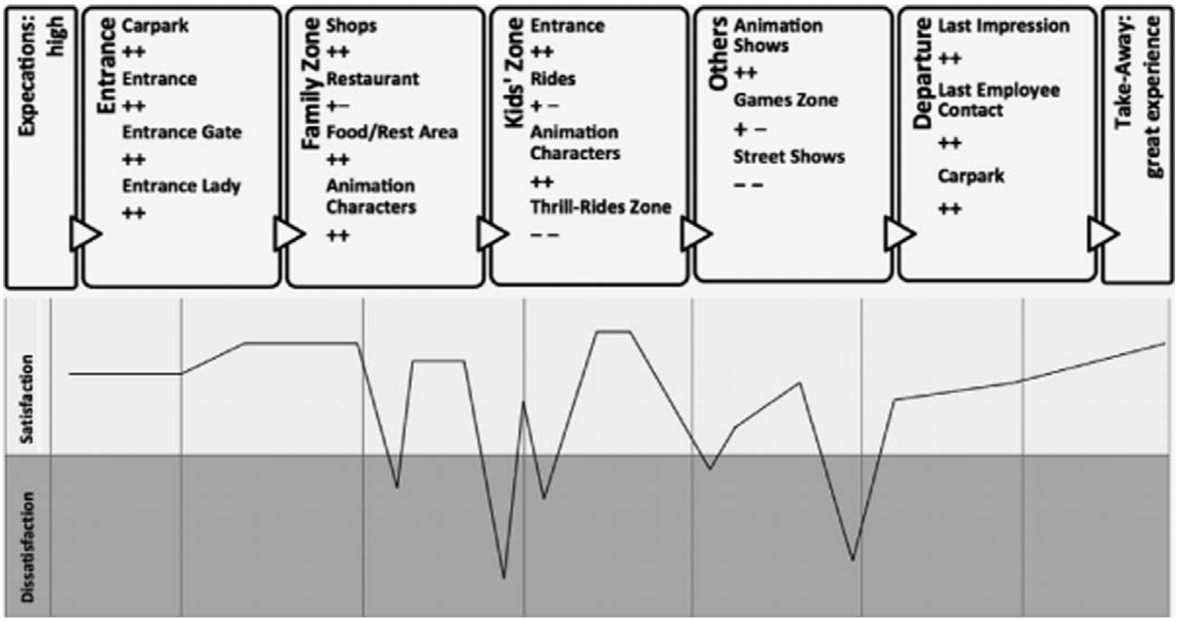



**Figure 2.4:** Customer journey map from (S39).

tool called CJM-ex that can be used for exploring customer journey maps using process mining techniques (S17). The main challenge addressed in this paper is to represent many journeys in an intelligible and efficient manner. This is done using a hierarchical clustering approach, merging activities that are most similar in each iteration (S17). In S20 the authors Harbich, Bernard, Berkes, Garbinato, and Andritsos take a probabilistic approach in to convert event logs in customer journey maps. A combination of Markov models and expectation-maximisation is used for event sequences (S20).

### Collaborative User Behaviour

Krumeich et al. (S34) look at monitoring users their process decisions to create individual process models. This creates knowledge about how people are working and how decisions are taken. The individual processes are generally less complicated and show a more logical flow compared to the 'crowed-based' process model for all users. For the proof of concept, an email-based process miner was used, which extracts information about the business process from the emails based on regular expression information extraction.

Unlike Krumeich et al. (S34), Diamantini, Genga, Potena, and Ribighini looked at the collaborative process (S32). The case study contained data about a research paper on which different authors worked. Data was collected from Dropbox events, svn logs and email and skype conversations. Events were classified by what and who dimensions. A fuzzy mining algorithm was used with the ProM framework. The outcome shows a graph of activities that are related to different actors. This gives an overview on which author was most active on different tasks (S32).

Schönig, Cabanillas, Di Ciccio, Jablonski, and Mendling researched how the process mining framework can be extended with integration of collaborative activities (S11). Teams participating in a collaborative activity were extracted from log files and then the characteristics in terms of skills, roles etc. were uncovered.

### 2.2.3 RQ III - What privacy aspects are relevant to collecting user behaviour data?

The results of RQ III are covered in this section. In Section 2.2.3, research related to GDPR and the collecting of user behaviour is shown. Section 2.2.3 contains papers related to process mining and business privacy. Results on user privacy are shown in Section 2.2.3. The combination of business and user privacy is reported in Section 2.2.3.

#### General Data Protection Regulation

As of the 25th of May 2018, the EU regulation 2016/679, otherwise known as the GDPR, is in act in all member states of the European Union (European Union, 2016). These regulations harmonise the data privacy laws across Europe. These regulations are applicable on collecting user behaviour data since this is information on an identifiable natural person.

Most papers do not address the topic of privacy. Even in cases where users are tracked throughout their daily activities such as in S12. S18 diminishes privacy in one sentence by stating: "the data was anonymised by using pseudonymised patient ids". This is however not considered to be anonymising the data in terms of the GDPR, which makes a clear distinction between anonymisation and pseudonymisation.

Still there is some research that focuses specifically on privacy in combination with process mining or user behaviour. This topic is twofold, one part considers the privacy of businesses, such as protecting confidential business information. The other part is specifically about protecting the privacy of the users.

#### Business Privacy

Process mining is not the core business of most companies. Outsourcing is therefore a relevant scenario for these companies. In such case confidentiality of the dataset is important because sensitive information about the company might leak. Encrypting the data to hide sensitive information is a solution, but it should still be possible to group cases or determine the order based on timestamps. A Paillier cryptosystem can be used which has homomorphic properties which allows to do calculations such as additions without decrypting the data in advance (S29 and S25). Burattin, Conti, and Turato used this method to anonymise business data in the context of process mining (S29). Tillem, Erkin, and Lagendijk used the method to also protect user data (S25). This is further discussed in Section 2.2.3.

Another possibility in which business privacy is relevant is that in the case of collaborative business processes. In this scenario, different companies work together, and each company delivers a part of the process. Irshad et al. discusses this topic and propose a solution in which privacy is preserved when a central repository that supports process mining for generating business processes (S30).

#### User Privacy

For user privacy, the GDPR plays an important role. Mannhardt, Petersen, and Oliveira studied the aspects of privacy and the GDPR in the context of process mining, specifically for human-centred industrial environments (S10). They focused on monitoring the well-being of operators in industrial manufacturing environments (S10). Although this is more privacy invasive than monitoring user behaviour, the challenges are also applicable. Privacy guidelines on using process mining are proposed which can be used as a starting point for further research.

#### Both User and Business Privacy

Tillem et al. propose a solution for guaranteeing privacy of users and software companies when their data is analysed by a third-party process miner that handles the process mining (S25). The process miner only has access to the encrypted data from which activity names and frequencies cannot be derived. To create the output based on the alpha algorithm, the process miner sends the necessarily parts of encrypted data to the software company, which answers with the relevant information for each step. This assures that the software company does not have direct access to the raw data which includes resources and timestamps and that the process miner cannot decrypt data such as activity names or frequencies of activities. This protects the user from the software company and the software company from the process miner. This will only work if the software company has no direct access to the data set, otherwise the privacy of the user will be compromised. The process miner and software company should thus not work together.

## 2.3 Discussion

In this section, the results of the literature review are discussed. Section 2.3.1 discusses the general literature review. Section 2.3.2 discusses the identified categories on process mining, the tools used and the XES standard. Then in Section 2.3.3, the customer journey and collaborative process mining is discussed. This is followed by a discussion on the GDPR and privacy in process mining in Section 2.3.4.

### 2.3.1 General Discussion

This section discusses the combination of customer journey and process mining to get insight in user behaviour. From this review it seems that this combination is not widely found in the literature. This might be since both topics are relatively new. Most papers found in this review date from the last few years. Some theoretical literature is available on the intersection as shown in Section 2.2.2, but research from a practical perspective is missing. The proper steps to include process mining of user behaviour in an existing are not well documented.

Most studies assume that data is already available and only focus on the process discovery part. This is also one of the core concepts of process mining: using already available logs to extract business processes. However, not all companies have these logs available. Especially logs with information on user behaviour are not commonly found. This means that companies that want to start with mining user behaviour have no proper guideline on how to collect and store data.

### 2.3.2 Discussion on RQ I

The results show different context in which process mining could be used. Furthermore, the XES standard and different process mining tools were identified. These are discussed below.

#### Process Mining Context

This literature review identified four distinctive categories on process mining based on the context in which process mining is used. Each category has their own approaches and challenges. Business process mining is the original concept of process mining and studies most. The other three categories are considered descendant based on business process mining. The core ideas are used but placed in a different context. This makes that there is some overlap between the different categories. For example, as explained in Section 2.2.1, Liu, Wang, Gao, Zhang, and

Cheng started from mining software process but combined this with user actions (S8). The result is a form of mining user behaviour.

Challenges can also overlap between different categories. For example, the correlation challenge seen in service mining as described in Section 2.2.1 could also occur in mining user behaviour. Due to privacy concerns it might not be possible to link certain actions to the same user.

Process mining can thus be used for different purposes. The general ideas as described in Section 2.2.1 apply to all four categories. In the future, more categories could be discovered that make use of the concepts of process mining.

The results show that there are already techniques available for process mining of user behaviour. These techniques mostly focus on analysing the available data with process mining tools. The goals differ from improving the feedback cycle in an agile development process (S35) to mapping different types of customers of an e-commerce website (S23). However, the papers do not discuss good strategies to start with user behaviour process mining.

#### XES

For storing event logs, the best format is to follow the XES standard. The basic concepts of process mining are defined in this standard and these are applicable to all categories. The standard can be extended such as S16 did for customer journey mapping. Most tools support the standard and the IEEE task force on process mining issues certification on the standard for process mining tools (IEEE CIS Task Force on Process Mining, 2019).

#### Tools

Section 2.2.1 showed that most researchers prefer the tools ProM and Disco for process discovery. The explanation as given by S9 that this is caused by academic research bias makes sense. However, this implies that although these tools are widely used, they might not be the best fit for all use cases. A good comparison between different tools does not yet exist in the literature. Commercial tools might be better for large organisations that want to implement process mining, because of the support and functionality to handle large datasets.

#### RQ I - What techniques can be used for process mining of user behaviour?

The above sections show that there are techniques available for process mining of user behaviour. However, an implementation strategy for companies is missing. Especially for mining of user behaviour. The XES standard could help with storing event logs since most tools support this. The tools mentioned in the literature are missing the link with real business cases.

### 2.3.3 Discussion on RQ II

from the results it is shown that customer journey mapping and collaborative user behaviour using process mining are already mentioned in the literature. Below these two topics will be further discussed.

#### Customer Journey Mapping

Customer journey mapping is most known in the e-commerce world. The concepts such as touchpoints and personas are however also applicable in the context of user behaviour in applications. By combining this with process mining is should be possible to get a better view on how users behave in an application. S17 made a fist attempt at combining customer journey with process mining. Their research is however limited to a theoretical view. The main challenge of both S17 and S16 was to group customer journeys so that a good representation could be made. Their approach with Markov models and expectation-maximisation seems to work well for this use case.

#### Collaborative Process Mining

The Process mining manifesto already discusses the possibility of discovering roles in organisations with the help of process mining (S40). This literature review also identified a few papers that discuss this possibility (Section 2.2.2). In CSCW systems, the insight in how different users work together can help to identify possible bottlenecks in the software, for example, where one user must wait on another user before the next action can be executed. Another scenario is that users execute tasks that they are not supposed to do, such as an admin user that misuses their power or users that bypass the process in place to get to the next step. These scenarios were not identified in this literature review.

#### RQ II - What techniques can be used for mapping the customer journey in collaborative software?

Customer journey mapping should make the translation from the e-commerce world to applications so that it can be used for getting insight in application usage and can help contribute with improving applications. Process mining can help delivering the tools for collecting and analysing the data.

### 2.3.4 Discussion on RQ III

Privacy in literature is not widely discussed. Based on the literature that is available, some concepts that might help with privacy aspects in combination with processing user behaviour data are discussed.

#### General Data Protection Regulation

In the process mining manifesto, maturity levels of events logs are given. Only the top-level mentions privacy and security. However, in sight of the recent GDPR, this should be the case for all privacy sensitive data. In fact, when collecting data such as the behaviour of users, the goal of collecting this data must be determined beforehand and users must give their consent. The only exception is if the data collection is completely anonymous. van der Aalst, 2016, p. 290 argues that process mining makes use of existing logs and that therefore privacy and security issues already exist. This is however not a valid argument. First, since the GDPR states that users must give permission for specific goals of data collecting. Second, if the collection of logs already had privacy or security concerns, this should be fixed instead of ignored.

#### Pseudonymised vs Anonymised

Given the GDPR, there are two scenarios for collecting user behaviour:
*Using pseudonymised data* – In this scenario, the behaviour of individual users can be tracked. The ids are pseudonymised but can be linked to individual users with other data sources. For analytical purposes this gives the best results since behaviour can be tracked over multiple sessions. However, the user should be made aware of the data collection, the goal must be made clear and the user should give their consent. The user also “should have the right to have personal data concerning him or her rectified and ‘a right to be forgotten’.” (European Union, 2016).
*Using anonymised data* - In this scenario, the log files do not contain any data that is relatable to individuals. Not relatable also means that the user cannot be linked to the data with the use of other data sources. For example, the behaviour log files could include timestamps about the time a user logged in. The user might be identified by relating these times with the logs on user logins. Or when instead the user role is stored and there is only one user that has that specific role this could be considered as personal data.

Both scenarios have their pros and cons. Pseudonymised data give a higher level of detail and makes it possible give feedback to individual users. However, the goal must be determined, and the user must give his or her consent before data can be collected. Furthermore, the software company must implement systems to give the user access to their private data and let the user delete their private data. Security breaches should be reported, and the storage of personal data should be limited. For anonymised data these points are not needed. However, the company should make certain that the user cannot be linked directly or indirectly to the data. Table 2.4 gives a summary for this.

**Table 2.4:** Pseudonymised VS anonymised data.

| | Pros | Cons |
|---|---|---|
| Pseudonymised data | Tracking on user level<br>Higher level of detail<br>Give individual users feedback | Goal must be determined on beforehand<br>User must give consent<br>User must get access to private data<br>User has right to be forgotten<br>Privacy of users not guaranteed |
| Anonymised data | No consent needed<br>User privacy guaranteed<br>Goal is not fixed<br>No retention limit | Lower level of detail<br>Only aggregated data<br>Make certain no personal data is collected |

#### Privacy Solutions

The solution proposed by Tillem et al. (S25) is in the grey area. The collected data is processed while partially encrypted to cover private sensitive data. The data can be decrypted when the process miner and software company work together, so this is not anonymous data. But to retrieve the private data of individual users, the proposed user privacy solution is undone for all users, since the software company then must have access to the database.

#### Business Privacy

The literature on privacy in process mining is focusing more on business privacy. Either in the case when different businesses work together, or when process mining is outsourced. When companies work together it is important that no business information is leaked to other companies. For companies that outsource process mining and use process mining as a service, it is also important that the process mining provider does not have access to the actual data.

**Privacy by Design**

The concept of privacy (and security) by design is not widely adopted in the literature on process mining or customer journey. This is important for both scenarios on data collection. Companies should think about how they collect and handle data from a privacy perspective. How can data be pseudonymised and how do we communicate this to user whose data is collected? Or how do we make the data completely anonymous? What are the implications if the data is leaked and how do we prevent this from happening? The GDPR forces companies to consider these scenarios by introducing high fines for non-compliant companies.

**RQ III - What privacy aspects are relevant to collecting user behaviour?**

This literature review shows that privacy in the literature is limited. The concepts discussed above could help to bridge this gap. According to the GDPR, Privacy by design should be used, which is also applicable to collecting user behaviour data. The privacy part is twofold: both the privacy of the users as well as the privacy of the business should be considered. This last concept is especially important is businesses work together or if data analysis is outsourced.

## 2.4 Conclusion

This Literature review shows that the first steps in customer journey process mining have been made. The individual parts are already discussed; however, they do not show real insight in how collaborative software is used. Customer journey maps offer a solution to show the (collaborative) behaviour of users in software. However, further research is needed.

The GDPR forces companies to consider the privacy of users in collecting and processing user behaviour data. Privacy by design should be the standard. The goal of collecting the data must be clear and users must be informed before collecting can start. Business privacy also plays a role in process mining and should therefore not be forgotten.

This literature review answered RQ I, RQ II and RQ III. From the literature review it is now clear what techniques are available on customer journey process mining and what the privacy aspects are relevant. The remainder of this research will use these results to answer the remaining Research Questions.

## Chapter 3

# Solution Design

The solution design discussed in this chapter is an approach track user behaviour in software using Process Mining. Based on the requirements in Section 3.1, three methods are considered for software companies to implement user behaviour tracking. Section 3.2 discusses the different methods and helps software companies to decide which method fits best for their situation. In Section 3.3, the Stakeholders for the different methods are identified. The three different methods are discussed in Section 3.4, Section 3.5 and Section 3.6. For each method, the corresponding tasks are also discussed. Each method is described using Business Process Model, which includes a visualisation that follows the Business Process Model and Notation.

## 3.1 Solution Requirements

Software companies are all different in size, maturity, working method, product and so on. Companies also have different goals for gathering user feedback. This section determines the scope of the solution, by settings the following requirement:

- Privacy by design (Section 3.1.1)
- Suitable for small and large software companies (Section 3.1.2)
- Scalable (Section 3.1.3)
- Suitable for collaborative user behaviour tracking (Section 3.1.4)
- Suitable for giving feedback to users (Section 3.1.5)

These requirements are discussed below.

### 3.1.1 Privacy by Design

Following a privacy by design approach is essential to guarantee the privacy of users. Article 25 of the GDPR explicitly mentions ”Data protection by design and by default” (European Union, 2016, p.48). Based on the literature review, privacy by design is not yet widely adopted in the literature on process mining or customer journey. Nevertheless, the solution should use a privacy by design approach, both to protect the user privacy and the business privacy.

Different frameworks exist to help with privacy by design. Cavoukian described 7 Foundational principles for privacy by design. Although this paper is not written in the context of the GDPR, it still provides principles that are relevant for conforming to the GDPR. Another framework is the *‘data protection by design’* framework of the privacy company (Privacy Company, 2019). The framework proposes to use anonymous data where possible. according to their framework, if the data is completely anonymous (per definition of the GDPR), no extra measures are needed. In all other cases, the schema from the framework should be followed (Privacy Company, 2019). Another framework that could be used is that of NOREA. Their privacy control framework can be used to audit the control objectives regarding privacy and personal data based on key elements of the GDPR (NOREA, 2018).

The solution should be compatible with these frameworks or other privacy frameworks so that checks on privacy are built in. This will make sure that the gathering of user feedback is GDPR compliant.

### 3.1.2 Suitable for Small and Large Software Companies

The solution should be suitable for small and large software companies. Small software companies should be able to start with the solution and get feedback on user behaviour with little effort and from a situation where no data is yet available on user behaviour. Large software companies should be able to get details insight in their application and use more advanced methods such as behaviour tracking in collaborative software (Section 3.1.4) and personalised feedback (Section 3.1.5).

### 3.1.3 Scalable

In addition to the previous requirement in Section 3.1.2, the solution should also be scalable. It should be easy for software companies to extend the solution to gather more and different information. If a company decided to change the strategy on user behaviour tracking to also allow for personalised feedback, this should be possible. Thus, the solution should be flexible enough to handle such changes.

### 3.1.4 Suitable for Collaborative User Behaviour Tracking

User behaviour tracking in Computer Supported Collaborative Work (CSCW) applications should be supported by the solution. Tracking user behaviour in collaborative software could give insight in how users work together. However, there are also extra challenges in terms of user privacy and collecting collaborative data. For collecting data on how users work together, identifiable information on users must also be collected because otherwise it is not possible to differentiate between users. This means that the data cannot be completely anonymous. Therefore, this is only recommended for medium to large software companies.

### 3.1.5 Suitable for Giving Feedback to Users

Information on the use of the application could be relayed back to the user. The solution should provide the steps needed for this. The actions of users can be compared to that of other users so that personalised advice can be given. Due to the complexity, this is primarily meant for medium to large software companies with some experience on other forms of user behaviour tracking.

## 3.2 Solution Methods

Based on the requirement in Section 3.1, three methods were constructed: functionality tracking, customer journey tracking and personalised feedback tracking. The first method focuses on the usage of functionality in applications. The second enables companies to follow customer journeys in the application. Finally, the third method includes feedback to users.

All three methods fulfil the first three solution requirements (privacy by design, suitable for small and large software companies and scalable). The fourth requirement (suitable for collaborative user behaviour tracking) is considered by customer journey tracking method and personalised feedback tracking. The last requirement (suitable for giving feedback to users) is realised in the personalised feedback tracking method. Table 3.1 gives an overview on how the solution requirements are fulfilled by each method.

**Table 3.1:** Overview of solution requirements fulfilled by each method.

| Tracking | Privacy | Size | Scalable | Collaborative | Feedback |
|---|---|---|---|---|---|
| Functionality | Anonymous | S/M | Yes | No | No |
| Anon. customer journey | Anonymous | S/M | Yes | No | No |
| Coll. customer journey | Pseudonymised | M/L | Yes | Yes | No |
| Personalised feedback | Identifiable | M/L | Yes | Yes | Yes |

In Figure 3.1 a decision tree is shown to determine which method is most applicable for a company. The method of choice is the one that supports the needs of the company and is the least privacy evasive for the user. For example, a medium to large company that want to track customer journey in their application but has no need to give feedback to users or track collaborative user behaviour should use anonymous customer journey tracking.

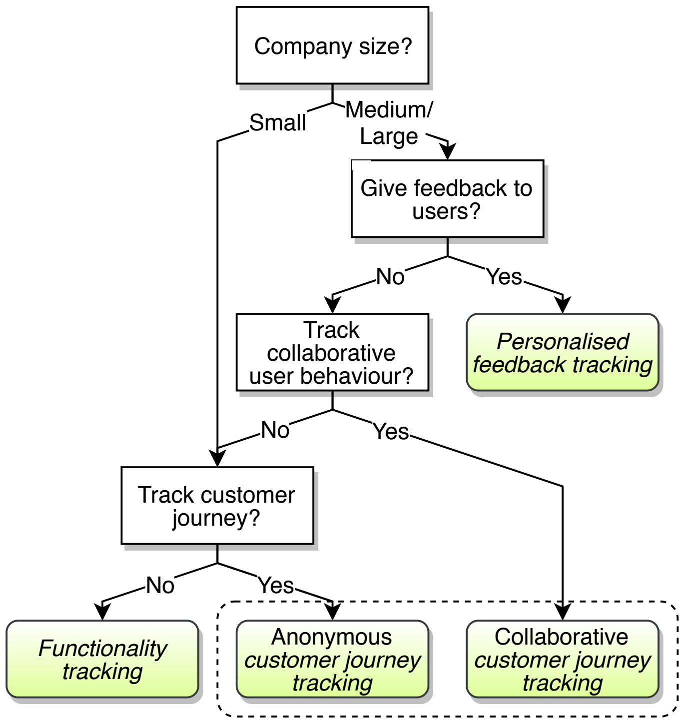



**Figure 3.1:** Decision tree for selecting tracking method.

The scalability of each method is not limited to scaling within the method itself, but also by scaling to another method. A company that implemented functionality tracking can easily extend the implementation to conform with customer journey tracking. In the same way, customer journey tracking can be extended to support personalised feedback tracking.

Each method consists of tasks that are performed by stakeholders. For each method, a process model is given that uses the Business Process Model and Notation (BPMN) (OMG, 2011). All process models consist out of five 'swimlanes', where every swimlane represents a stakeholder role. These roles are *generic* and can differ per company. The *generic* definitions used for each stakeholder can be found in Section 3.3. The models and corresponding tasks for each method can be found in the following sections:

Functionality Tracking: Section 3.4
Customer Journey Tracking: Section 3.5
Personalised Feedback Tracking: Section 3.6

## 3.3 Solution Stakeholders

The following stakeholders that interact directly with the methods were identified. These are *generic* stakeholders and could vary per company.

**Product Management** - Responsible for the planning and production of the product. Makes decisions on the direction of the software product. Information on the usage and behaviour of users can be helpful in making these decisions.

**Legal** - Gives legal advice to the company. Most large companies have their own legal department. Smaller organisation could outsource legal advice. The role for the legal department in this case is to make sure that the tracking is conform the GDPR and other legislation.

**Development** - Create the software based on the input of Product Management. For the solution they are responsible for implementing the tracking in the application.

**Hosting** - Hosting makes sure that the hardware and software is available to the users. The host department is in this case responsible for making sure that databases are available to store and retrieve the logs. Depending on the company this could also be done by the development team.

**Process Miner** - The process miner tries to translate the logs on user behaviour to useful information for Product Management. This role can be outsourced or could be part of the development team.

## 3.4 Functionality Tracking

Functionality tracking is suitable for small to medium companies that have little to no experience with user behaviour tracking. From a privacy perspective this is also the least invasive. User data can be collected anonymously therefore the user does not have to give their consent.

An example of functionality tracking is the number of times each button in the application is used, or how many times people an error message is show to a user. These metrics will give the software company basic insight in how the application is used and can help determine the focus of future development.

An overview of the tasks in this method is shown in Figure 3.2. The sections below discuss the details of each task.

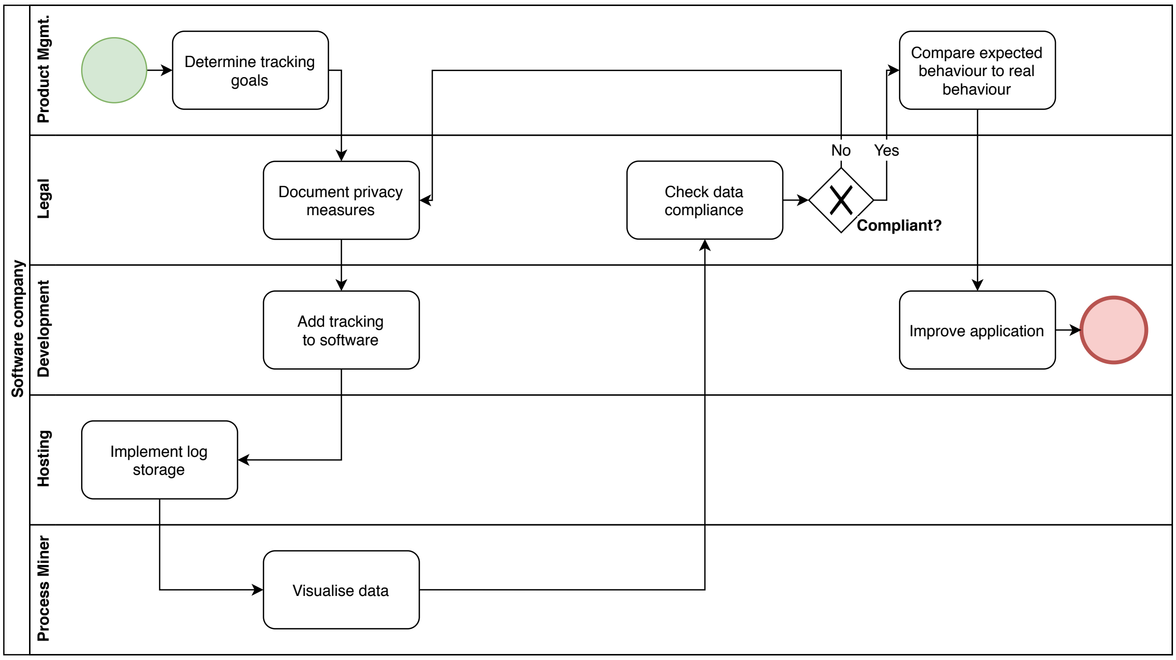



**Figure 3.2:** The model for the functionality tracking method.

### 3.4.1 Determine Tracking Goals

The first task for Product Management is to determine what the goals are of implementing functionality tracking. The characteristics of this task are shown in Table 3.2. Based on budget, time and technical possibilities it is determined if the set goal can be achieved. Information on implementation can be discussed with Development. the Process Miner can tell what is needed for visualisation. Mind mapping tools or a business model canvas can be used to map all information. The outcome is the feasibility of using functionality tracking. If the method is feasible, Product Management determines which functionality should be tracked and on what level of detail.

This can vary from only logging which pages the user visited to exactly track which specific buttons and functions are used.

**Table 3.2:** Characteristics of determine tracking goals.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Product Mgnt. | Development, Process Miner | Budget, Goals, Technical possibilities | Feasibility, Functionality to track | Mind mapping tool, Business Model Canvas |

### 3.4.2 Add Tracking to Software

Development adds the tracking to the software. The characteristics of this task can be found in Table 3.3. Functionality marked as important by Product Management should be tracked on the desired level of detail. Depending on what is available, the developer can choose between using an existing tool that logs data on functionality or build an own implementation. In consultation with Hosting, a storage strategy is worked out. The outcome is working tracking software which delivers the desired data on functionality to track.

For functionality tracking, the developer must only collect minimal data on user interaction. For example, the count on how many times a button is clicked should be collected.

**Table 3.3:** Characteristics of add tracking to software.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt., Hosting, Process Miner | Functionality to track | Tracking software, Tracking data | Development tools, External tracking tools |

### 3.4.3 Implement Log Storage

Hosting is responsible for storing the tracking data. Characteristics of this task are shown in Table 3.4. Hosting should deliver the means to store the data and deliver an interface for both Development and the Process Miner to store and retrieve the data respectively. The data could either be stored within the company or at a cloud provider. The outcome of this task is the log storage and the retention period of the data.

**Table 3.4:** Characteristics of implement log storage.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Hosting | Development, Process Miner | Tracking data | Log storage, Retention period | Database, Cloud storage |

### 3.4.4 Visualise Data

The task of the Process Miner is to analyse the data from the log storage and create a visual representation of the data. Table 3.5 shows the characteristics of this task. The visualisation could be done with for example graphs on the use of functionality over time or the total use of some functionality.

**Table 3.5:** Characteristics of visualise data.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Process Miner | Product Mgnt. | Log storage | Tracking information | Process mining tools |

### 3.4.5 Check Data Compliance

The task of legal is to check that the implementation is completely anonymous. Table 3.6 shows that the input is the process and the metadata of stored logs. In consultation with Development, the data is checked for any identifiable user data. If this if found, this should be fixed first before proceeding to the next task.

**Table 3.6:** Characteristics of check data compliance.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | Development | Retention period, Log storage | Privacy issues | GDPR, Privacy guideline |

### 3.4.6 Compare Expected Behaviour to Real Behaviour

After the privacy of users is checked, the tracking can be activated and Product Management can get insight in which functionalities of the application are used to what extent. This can be compared to the expected usage. Heavily used functionality could get more attention whereas functionality that is rarely used could be phased out. Reports on error messages shown to customers could show functionality that is unclear or not working as expected. The results could also be used as input for discussing the application usage with customers.

**Table 3.7:** Characteristics of compare expected behaviour to real behaviour.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Product Mgnt. | User, Process Miner | Tracking information | Improvements | Interviews |

### 3.4.7 Improve Application

Based on the results from the previous task, improvements can be made. Unused functionality can be removed, and error prone functionality can be improved.

**Table 3.8:** Characteristics of improve application.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt. | Improvements | Software improvements | Development tools |

## 3.5 Customer Journey Tracking

Customer journey tracking is suitable for small to medium companies that want to optimise the customer journey in their software. This method also comprises collaborative customer journey tracking for medium to large companies that offer CSCW software. Customer journey tracking extends functionality tracking by also logging the session in which the action was performed. For non-collaborative tracking this can be done anonymously. For collaborative tracking pseudonymised data is needed. The actions that were performed by the user in a session form the journey of a user. This journey can be compared to the expected journey to see if there are any discrepancies. In CSCW software, the journey can also contain interactions with other users. Tracking collaborative journeys can be used to see if there are bottlenecks in the collaboration between users. For example, it could be that a user must wait on actions of other users before they can continue their work. The process model is shown in Figure 3.3.

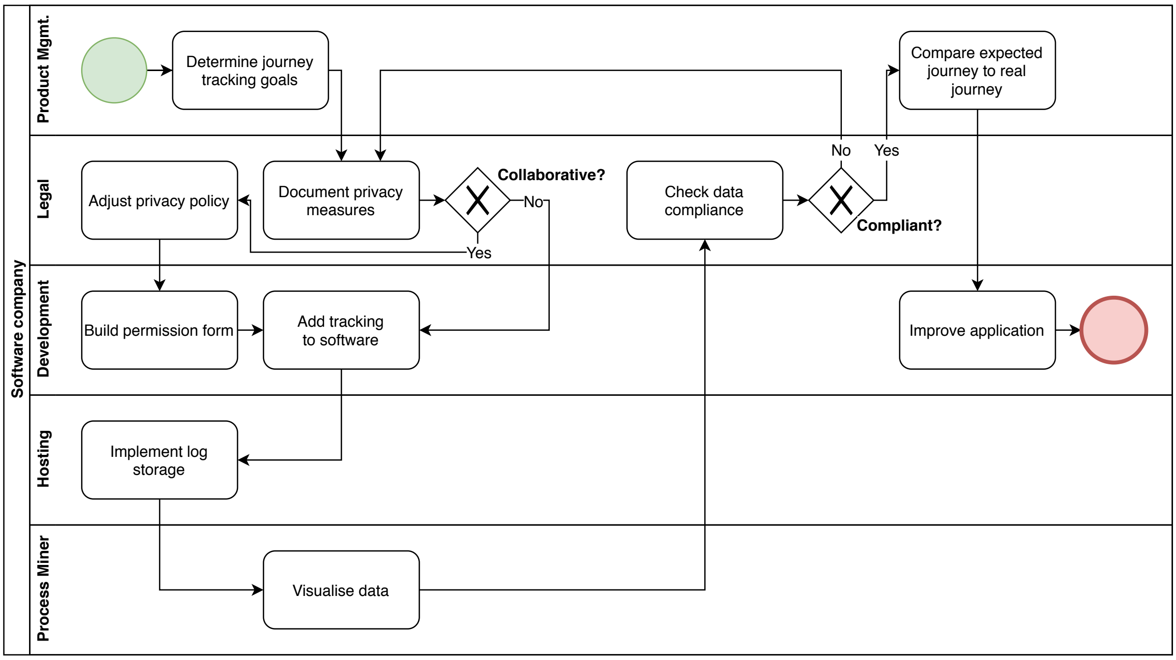



**Figure 3.3:** The model for the customer journey tracking method.

### 3.5.1 Determine Journey Tracking Goals

The first step is customer journey tracking is to determine the tracking goals. Table 3.9 gives an overview of the characteristics of this task. These are comparable to that of Determine tracking goals in Section 3.4.1. The difference is that instead of functionalities to track, Product Management must determine which customer journeys are relevant to track.

For collaborative applications Product Management can choose to either add collaborative tracking or stay with single user behaviour tracking. For collaborative tracking, anonymous tracking is not an option since actions of different users must be linked. Based on the type of application, the maturity of the software and considering the privacy of users, Product Management can decide to support collaborative tracking. This decision determines if the next task is to *document privacy measures* (Section 3.5.2), or that this can be skipped and continue to *add tracking to software* (Section 3.5.5).

**Table 3.9:** Characteristics of determine journey tracking goals.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Product Mgnt. | Development, Process Miner | Budget, Goals, Technical possibilities | Feasibility, Functionality to track, Collaborative | Mind mapping tool, Business Model Canvas |

### 3.5.2 Document Privacy Measures (Collaborative)

For collaborative tracking, identifiable information on users will be collected. It is the task of the legal department to make sure that the data is processed according to the applicable privacy regulations. By applying a privacy by design practice, privacy problems during or after implementation can be avoided. Following a privacy design framework such as described in Section 3.1.1 can be used to ensure that all requirements of the GDPR are met. The characteristics of this step can be found in Table 3.10.

**Table 3.10:** Characteristics of document privacy measures.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | Product Mgnt., Development, Hosting | Functionalities to track, Goals | Privacy documentation | GDPR, Privacy guideline |

### 3.5.3 Adjust Privacy Policy (Collaborative)

After privacy measures are clear, the legal department can start working on adjusting the privacy statement to let the users know what data is collected, how the data is collected, why the data is collected, what measures are taken guarantee the privacy of users and how they can get insight in their data. Table 3.11 shows the characteristics of this step.

**Table 3.11:** Characteristics of document privacy measures.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | User | Privacy documentation | Privacy policy | GDPR, Privacy guideline |

### 3.5.4 Build Permission Form (Collaborative)

If the collected data will not be completely anonymous, the users should give their consent first. A permission form where the users can adjust their privacy is needed. Based on the privacy documentation, the development team can build a permission form that reflects the choices that the user has. This can also be the place where the user can retrieve and delete their personal data. Table 3.12 shows the characteristics of this step.

**Table 3.12:** Characteristics of build permission form.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Legal | Privacy documentation,<br>Privacy policy | Permission form | Development tools |

### 3.5.5 Add Tracking to Software

After building a permission form (in the case of collaborative tracking), the development team can add tracking to the software. Table 3.13 Shows the characteristics of this step. This step is the same as Section 3.4.2, except that more that should be collected so that users can be followed during their session. For collaborative tracking, the privacy measures as described in the privacy documentation should be followed.

**Table 3.13:** Characteristics of add tracking to software.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt.,<br>Hosting,<br>Process Miner | Privacy documentation,<br>Functionality to track | Tracking software,<br>Tracking data | Development tools,<br>External tracking tools |

According to van der Aalst, the bare minimum needed for process mining is a "*Case Id*" and "*Activity*" (van der Aalst, 2016). A case or, in terms of process mining, a *trace* is a sequence of activities. The *Case Id* links different activities and forms a trace. In user behaviour process mining, a trace is defined as the actions a user executes when performing a specific use case. However, unlike normal process mining, the use case can only be determined based on the actions of the user. The intended user action is only known after the user performed that action. To determine what activities belong to a case, it is assumed that a user performs the use case during one session. The session id is thus the *Case Id*. During a session a user can perform multiple use cases. Therefore, the session attribute is not mapped directly to one use case. The order of activities is determined based on A *Timestamp* or *Time since beginning of session*. The following attributes should thus be collected on user actions:

*Action Id*, *Session Id* and *Timestamp*.

In collaborative tracking, the data of different users needs to be linked to determine collaboration. A *Session Id* is too limiting for this since it is not possible to, for example, show that a user waits on an action of another user. Instead, a *User Id* should be used. An *Action Id* is also not enough since it does not provide information on the object the user interacts with. For example, if two users work on the same document, it is not enough to know that both users worked on a document. The name or *Object Id* of the document they both worked on is needed to show that they worked together on a document. For collaborative tracking, the following attributes should therefore also be added:

*User Id* and *Object Id*

### 3.5.6 Implement Log Storage

This step extends the approach described in Section 3.4.3. The characteristics can be found in Table 3.14. For collaborative tracking where identifiable information is collected, the user should be able to retrieve their and delete their data. This should be considered when implementing log storage.

**Table 3.14:** Characteristics of implement log storage.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Hosting | Development, Process Miner | Tracking data | Log storage, Retention period | Database, Cloud storage |

### 3.5.7 Visualise Data

With customer journey tracking, the visualise data step as described in Section 3.4.4 is extended with visualising the journey of users in the application. Table 3.15 describes the characteristics of this step. tasks that the user performs should be grouped so that paths that different users follow become visible. For collaborative journey tracking, the interaction between users can also be visualised.

**Table 3.15:** Characteristics of Visualise Data.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Process Miner | Product Mgnt. | Log storage | Tracking information | Process mining tools |

### 3.5.8 Check Data Compliance

In this step the legal department checks if the privacy of the users is guaranteed. For non-collaborative user behaviour tracking, the data should be anonymous. For collaborative tracking, the privacy documentation should match what is discussed in the privacy documentation from Section 3.5.2. The privacy policy from Section 3.5.3 should also be compared to the actual implementation. Table 3.16 shows the characteristics from this step. If privacy issues are present, then either the privacy policy should be updated, or the tracking should be adjusted.

**Table 3.16:** Characteristics of check data compliance.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | Development | Privacy documentation, Privacy policy, Retention period, Log storage | Privacy issues | GDPR, Privacy guideline |

### 3.5.9 Compare Expected Journey to Real Journey

This step extends Section 3.4.6. The characteristics are also the same as shown in Table 3.17. Product Management can compare the use of functionality and additionally check if the users follow the expected journey. This can be done by comparing the designed workflow to the actual workflow. If users follow different paths, the user flow might not be clear. The user journeys might also show that some tasks are repeatedly executed by the user. In combination with interviews with users, Product Management can use this to decide to make improve the usability of the application.

**Table 3.17:** Characteristics of compare expected journey to real journey.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Product Mgnt. | User, Process Miner | Tracking information | Improvements | Interviews |

### 3.5.10 Improve Application

Just like the improve application step from functionality tracking discussed in Section 3.4.7, the development team can use the input of the previous step to improve the application. The characteristics are shown in Table 3.18. Together with Product Management, the priority of improvements and issues can be determined.

**Table 3.18:** Characteristics of improve application.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt. | Improvements | Software improvements | Development tools |

## 3.6 Personalised Feedback Tracking

Personalised feedback tracking is suitable for medium to large mature software companies that have experience with user behaviour tracking and want to extend this with personalised feedback to the users. In this case, it is not possible to anonymise or pseudonymise the data, since it must be possible to give the user feedback on their own behaviour.

The information from tracking the behaviour of individual users is used to give feedback on their actions in the application. For example, the application could suggest the next action based on what other users did. Personalised feedback tracking extends customer journey tracking, so it is still possible to extract other information on for example the journeys of users or functionalities used. Figure 3.4 shows the process model of this method.

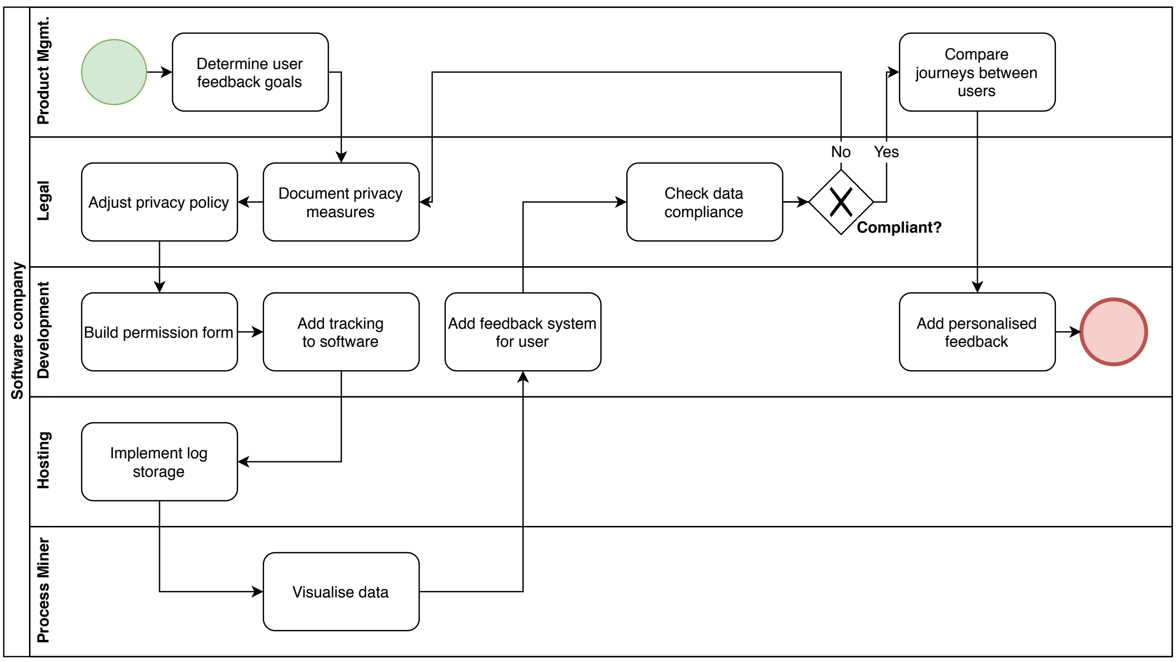



**Figure 3.4:** The model for the Feedback to users method.

### 3.6.1 Determine User Feedback Goals

For feedback to users, Product Management starts with identifying what the goals are of implementing feedback to users. Just like in Section 3.5.1, Product Management uses the budget, goals, and technical possibilities as input for this step. Based on this, Product Management determines the feasibility of giving feedback to users and what kind of feedback should be given to the users. Table 3.19 shows these characteristics and the other characteristics.

The kind of feedback depends on the application. For some applications it makes sense to compare the behaviour of users to other users, whereas with other applications the feedback can better be based on predefined behaviour defined by the software company.

How the feedback is delivered to the user is also application depended. Examples of feedback delivery are personalised emails on useful functions, a help text in the application or customised interfaces.

**Table 3.19:** Characteristics of determine user feedback goals.

| **Owner** | **Stakeholders** | **Input** | **Outcome** | **Tools** |
|---|---|---|---|---|
| Product Mgnt. | Development, Process Miner | Budget, Goals, Technical possibilities | Feasibility, What kind of feedback | Use Cases, Business Model Canvas |

### 3.6.2 Document Privacy Measures

To give feedback to users, identifiable information on the users will be collected. Therefore, the legal department must make sure that the data is processed according to privacy regulations such as the GDPR. Table 3.20 contains the characteristics of this step. For the implementation a privacy by design approach should be followed. To help with this a privacy by design framework such as described in Section 3.1.1 can be used. For business software, not only the privacy of users is important, but also the privacy of the business that have sensitive business information in software tools.

**Table 3.20:** Characteristics of document privacy measures.

| **Owner** | **Stakeholders** | **Input** | **Outcome** | **Tools** |
|---|---|---|---|---|
| Legal | Product Mgnt., Development, Hosting | What kind of feedback | Privacy documentation | GDPR, Privacy guideline |

### 3.6.3 Adjust Privacy Policy

This step is like the adjust privacy policy step of the customer journey tracking (Section 3.5.3). The input is based on the outcome of Section 3.6.2. The characteristics can be found in Table 3.21.

**Table 3.21:** Characteristics of adjust privacy policy.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | User | Privacy documentation | Privacy policy | GDPR, Privacy guideline |

### 3.6.4 Build Permission Form

This step follows the same approach as for the permission form for collaborative user tracking as described in Section 3.5.4. Table 3.22 shows the characteristics.

**Table 3.22:** Characteristics of build permission form.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Legal | Privacy documentation, Privacy policy | Permission form | Development tools |

### 3.6.5 Add Tracking Software

To enable good personalised feedback, the user must be tracked over multiple sessions. Therefore, only tracking the session of a user is not enough. Development should consult with both Product Management and the Process Miner to determine what behaviour should be tracked so that this can be compared to other users. The characteristics can be found in Table 3.23.

**Table 3.23:** Characteristics of add tracking software.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt., Hosting, Process Miner | Privacy documentation, What kind of feedback | Tracking software, Tracking data | Development tools, External tracking tools |

### 3.6.6 Implement Log Storage

The host department. is responsible for storing the data. In consultation with Development and the Process Miner a storing strategy should be developed. The data could be stored on premise or in the cloud. Mechanisms to delete personal data should the user request that must also be in place. The characteristics of this step are shown in Table 3.24.

**Table 3.24:** Characteristics of implement log storage.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Hosting | Development, Process Miner | Tracking data | Log storage, Retention period | Database, Cloud storage |

### 3.6.7 Visualise Data

After the tracking software is added to the application and the log storage is implemented, the Process Miner can start with analysing and visualising the data. The Process Miner can start with showing basic metrics such as with functionality tracking described in Section 3.4.4. However, the focus is on visualising what users do within the application. Different routes through the application can be grouped and compared to each other. The desired situation is discussed with Product Management. Table 3.25 shows the characteristics of this step.

**Table 3.25:** Characteristics of visualise data.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Process Miner | Product Mgnt. | Log storage | Tracking information | Process mining tools |

### 3.6.8 Add Feedback System for Users

A feedback system should be implemented to give the user insight. The precise implementation is dependent on the kind of application and the type of feedback. A possible feedback implementation gives the user insight in what the next step is based on where the user is in the application and what other users did. The characteristics of this step can be found in Table 3.26. The feedback system is implemented in consultation with Product Management and the Process Miner.

**Table 3.26:** Characteristics of add feedback system for users.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Development | Product Mgnt. Process Miner | Tracking information | Feedback system | Development tools |

### 3.6.9 Check Data Compliance

After completing the previous steps, the Legal department can check if the privacy of the user and business is guaranteed. This step is to confirm that the privacy documentation is followed according to plan and that the collected data is concurrent with this. Should there still be any privacy issues or, then the process returns to document privacy measures. From here, either the documentation can change or the implementation in add tracking to software can change. The characteristics of this step can be found in Table 3.27.

**Table 3.27:** Characteristics of check data compliance.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Legal | Development | Retention period, Log storage | Privacy issues | GDPR, Privacy guideline |

### 3.6.10 Compare Journeys Between Users

In this step Product Management determines what feedback is given to the users. This could be based on automated processes or by manually giving feedback to user groups. The Tracking information is used as input and will be processed. after this, the output is personalised information for the user. Table 3.28 gives an overview of the characteristics.

**Table 3.28:** Characteristics of compare journeys between users.

| Owner | Stakeholders | Input | Outcome | Tools |
|---|---|---|---|---|
| Product Mgnt. | User, Process Miner | Tracking information | Personalised feedback | Interviews |

### 3.6.11 Add Personalised Feedback

The last step is to give the user personalised feedback. For this, the information of the previous step is used and given to the feedback system from Section 3.6.8.

**Table 3.29:** Characteristics of add personalised feedback.

| **Owner** | **Stakeholders** | **Input** | **Outcome** | **Tools** |
|---|---|---|---|---|
| Development | Product Mgnt. | Personalised feedback | Feedback to users | Development tools |

# Chapter 4

# Prototype Implementation

For the prototype implementation, a case study at Fortes Solutions was conducted. The method described in Chapter 3 was followed to implement user behaviour tracking in the software application of Fortes. Section 4.1 introduces the case study. Section 4.2 gives an overview of the current situation, whereas Section 4.3 describes the desired situation. in Section 4.4 the selecting of the method is discussed and the actual implementation of this method is given in Section 4.5. At the end of this Chapter, in Section 4.6, the implementation is concluded.

## 4.1 Fortes Change Cloud

FCC is the latest platform of Fortes Solutions. This platform consists of different apps that support (large) organisations in organising their strategy, portfolios and projects. The apps are divided in 5 categories as shown in Figure 4.1 . Each app has its own core functionality. However, the strength of the software is the interconnection between apps and with other application. FCC is a web based SAAS solution and is therefore easy to implement and use for businesses.

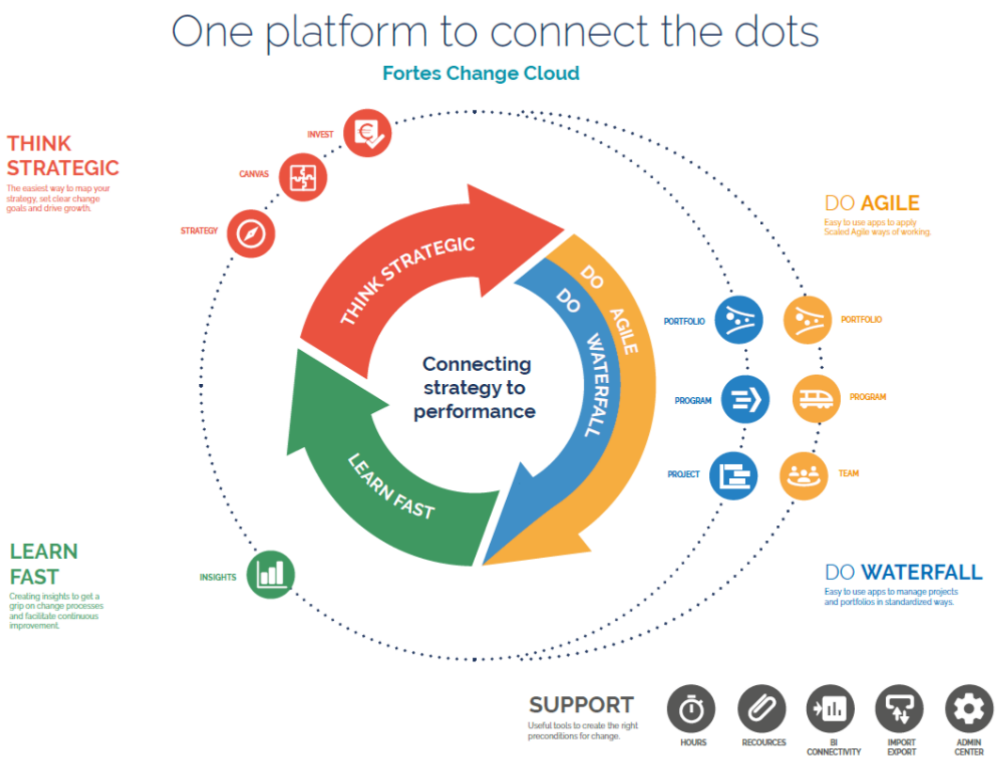



**Figure 4.1:** Fortes Apps.

### 4.1.1 Architecture

This section gives a high-level overview of the FCC architecture. Each app of the FCC is built fit-for-purpose. The apps all have their own scope, lifecycle and loosely coupled code base. The Apps are written in JavaScript and each app has its own frontend and backend. The frontend of each app is built on the React framework and the backend uses the Node.js framework. All apps make use of a shared code base which is extended with app specific functionality.

The React frontend uses components and elements. Components are the smallest parts of the user interface, such as buttons, input fields, panels and icons. These components are used by elements which are pre-build elements that can be reused throughout the application. These elements are built around a functionality and are independent, although they can interact with other elements. Some elements allow for nested elements. For example, a page is an element that can contain other elements. An app in the FCC usually exists out of multiple pages. These pages each contain multiple elements. Developers can create new elements and rearrange elements on pages used the Fortes Software Development Kit (SDK). Figure 4.2 gives an example of the Kanban element. This element has a Title element, a Button element, a Select element, a Menu element and a Card element. The Card element can also contain other elements. Here, it only has a Text element.

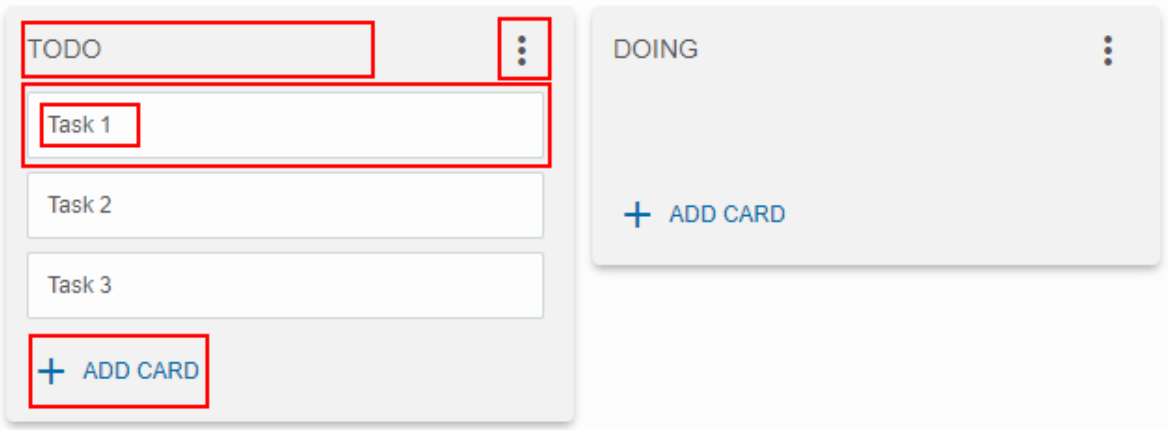



**Figure 4.2:** Example of kanban elements with other elements.

### 4.1.2 Stakeholders

Different stakeholders are involved when it comes to Fortes and FCC. The most important stakeholders relevant to this case study are listed in this section.
**Customer** - The customers are companies that use the software of Fortes. People that work at these companies and have an account in the application are *users*.
**User** - The users in most cases work at the customers and use the software. Users can have different roles, these include: Strategy Developers, Portfolio Managers, Programme managers and Team members.
**Product Management** - The Product Management team within Fortes is responsible for the planning of development of FCC. Product Management consults with the customers to determine new features for the application. The team delivers specifications for new or improved functionality. These specifications are mostly in the form of use cases and an UX design.
**Support** - The support team handles questions from customers. Support tickets from customers are handled by phone or mail. Customer bugs or Requests for Change are relayed to the rest of the organisation.

**Consultancy** - Consultants help large customers with the implementation of the software in their organisation. This includes giving advice and tailoring the software to the needs of the customer.
**Developers** - The developers are Fortes develop the software. New functionality based on the input from Product Management are added and bugs are fixed. Architectural changes to the software are also considered.
**Hosting** - The hosting team makes sure that servers where the applications run function. Servers for the development process are also managed by hosting.
**Others** - Other stakeholders are the Marketing team, the Sales team, and Partners of Fortes.

## 4.2 Current situation

The software development process at Fortes is based on an Agile mindset. Scrum teams work in sprints of 2 weeks. Sprint goals are determined by the product management team and the development team determines what is needed to achieve the goals. The product management team determines the goals based on own insight and the wishes of customers.

Currently, there is no measurement on the usage of FCC. The only information available is that of customers providing feedback. The 'happiness' of customers on existing or new features is not measured but determined based how the enthusiasm of customers during presentations or in feedback sessions with customers. However, these people are not always the end user of the product. Insights in the use of functionality is missing.

## 4.3 Desired situation

By introducing tracking of user behaviour in FCC, the development process can be improved. The product management team will get insight on what functionality customers use, how often they are used and if the software is used in the way intended. This information can then be used to determine what priority functionality should be improved. Development capacity on for example performance improvement could better be applied to functionality that is used often instead of functionality that is rarely used. The information can also be used to include new quick features on tasks that are repeatedly executed by customers. For example, the information shows that users often follow the same pattern of tasks, these tasks could then be combined into one task. In combination with feedback sessions, this information can be used to get a better understanding in what the customer wants to achieve.

## 4.4 Method Selection

Fortes Solutions is a medium sized software company with no to little experience on user behaviour tracking. The software supports collaborative working. However, since the company has no experience on user behaviour tracking and values the privacy of users and businesses, giving feedback to users or track collaborative user behaviour is not needed for now. Fortes is however interested in following the user throughout the application. Based on the decision tree from Figure 3.1, *anonymous customer journey tracking* was chosen.

## 4.5 Method Implementation

In the next section the implementation of the selected method is discussed. From Section 4.4 *anonymous customer journey tracking* was selected. To show how the implementation of the method works, the steps of Section 3.5 are discussed in the remainder of this section.

For this implementation, all roles were performed by the researcher.

### 4.5.1 Determine Tracking Goals

The main goal was to get a better understanding in how users use the application so that the application can be improved.

The prototype implementation covers three apps of the FCC; Strategy, Portfolio and Agile program (Figure 4.3). The focus was on the so called “happy flow”. This is the path the user is expected to follow. For each app, uses cases for three actors (a Director, a Portfolio Manager and a Programme Manager) were written. These use cases can be found in Appendix C. The use cases were selected based on the core functionality of the app. For each app there are more use cases, but these were left out for now.


**Figure 4.3:** The selected apps.

The functionalities to track are the interactions the user has performing the use cases. This can be clicking on a button, visiting a page and other actions.

### 4.5.2 Document Privacy Measures

As discussed in Section 4.4, tracking collaborative user behaviour was of interest. Therefore, the Adjust privacy policy and Build permission form steps were not needed. Moreover, following an approach with anonymous tracking, no other privacy measures are needed if the data stayed completely anonymous. By following a privacy by design approach at the remainder of the steps, it was assured that the data stayed anonymous and could not be traced back to individual users.

### 4.5.3 Add Tracking to Software

Based on the output of tracking goals, the functionalities that should be tracked were determined. First, the attributes that comprise the interaction of a user with the application were selected. After that, tracking was added to the software at strategic points.

#### Select Attributes

For process mining, three attributes were needed. *Session Id* as case identifier, a *Action id* to identify activities and a *Timestamp* to order the activities.

The *Session Id* is generated after a user logs in and is available between apps. If the user logs out or is logged out automatically, the *Session Id* will no longer be used. A hashed string of the *Session Id* was added to each event.

To determine the order of events, a *Timestamp* was used. The ISO8601 standard with milliseconds was used to determine the order even if the user clicks fast.

Determining the *Action Id* was more complex. Many components are shared amongst different apps within the platform. For example, a create button in the Strategy app can be used for creating a strategy, whereas the same button in the Portfolio apps creates a new portfolio. To differentiate between all actions, the following structure for the *Action Id* was used:

app name + page name + tab name + action

To give an example, the Add list button from the portfolio app is used. A screenshot of this button is shown in Figure 4.4. The action for this button is called 'add_column'. Would a user click on this button, then the corresponding action is:

Portfolio + Portfolio + funnel + add_column

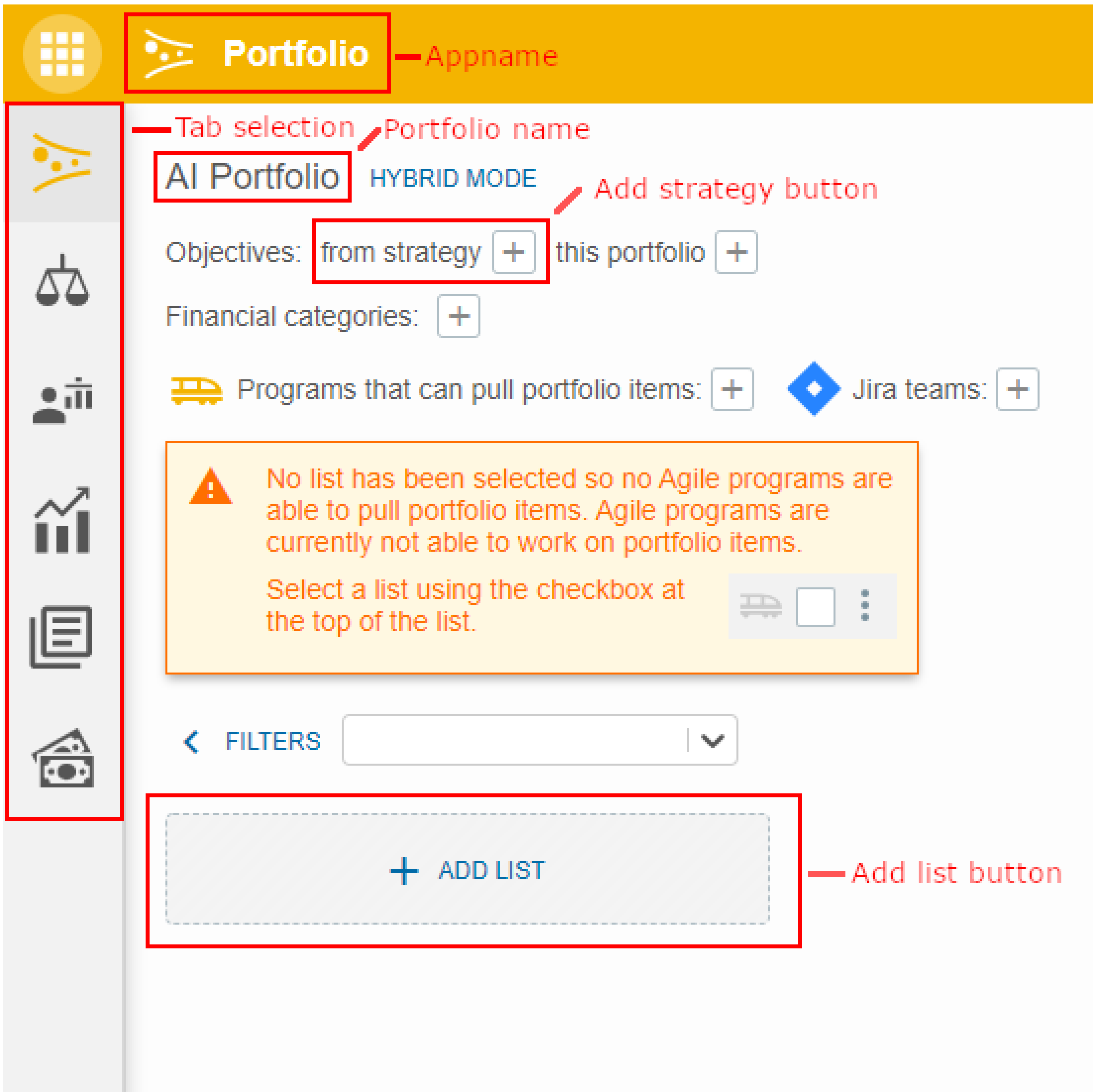



**Figure 4.4:** Example attributes of the portfolio app.

Where Portfolio is the app name, portfolio is the page name, funnel is the tab name and add_column is the action. For easy filtering and aggregation later, these attributes were stored separate columns.

### Tracking

To add tracking to existing the software, a solution that works together with the JavaScript platform (React) was needed. Before building a new solution, the Node Package Manager (NPM) (npm, Inc., 2019) repository was searched for an existing package that was lightweight and suitable for React. The NPM package react-tracking (New York Times, 2018) was chosen because it integrates with React, can be added with little impact on existing code and is flexible in how the data is stored.

React tracking makes use of a decorator function that can be added to classes and methods. If a method is called, the decorator function of that method will trigger and merge with any other tracking decorators that are added to the classes above. The data is written to a window variable on the client from where it can be further processed.

The next sections give an overview on how react tracking was implemented in the FCC.

### App Class

The app class is the entrypoint for each app. it wraps all other components on the page. React tracking was added to this class to dispatch the events from lower components to the tracking factory which is discussed below. Appendix D.1 shows the added tracking decorator function.

### Tab Element Class

Tracking was also added to the tab element, which is responsible for showing the tab selection as seen in Figure 4.4. The code can be found in Appendix D.2. Each time the user switches tab, an event is saved:

```
{element: 'tab', action: `change_tab_${currentNode.get('activeTab')}`}
```

### Other components

To other components, the save principle was applied. Appendix D.3 shows for example the tracking code for the column cardboard. For each user action with this component (create cards, click a card, change the card order, create a column or change the column order) code was added to save events.

for each action that the user can perform (create cards, click a card, change the card order, create a column or change the column order) tracking decorator function was added.

A more advanced future of the React-tracking library is shown for the `onClickCard()` decorator. For the tracking function, it was possible to add the props of the React element, the state React element and the parameters of the called function as parameters. In this case, the id of the clicked card(model) was added. This should always be done carefully, since this also makes it possible to accidentally log data of users. In this case, if instead of `cardModel.getId()` `carModel.getContent()` is called, the content of the cards would have been stored in the logs, which is not desirable since this contains user data.

#### Client-Side Tracking Factory

The code for the client-side Tracking Factory can be found in Appendix D.4. The user data was collected with the custom dispatch function. The `_handlePage()` function keeps track if the user changes tab or page. This is read from the `action: 'change_tab_activetab'` event from the Tab Element Class. After this, the dispatch function adds a hash of the session token, the appname and a timestamp to the data. This data is then pushed to `window.dataLayer`. After a set period (in this case every minute), or if the user closes the window or logs out, the data is sent to the server with the `_uploadData()` function. This prevents that for every action a server call is needed. The `_uploadData()` function uses the `navigator.sendBeacon()` method which is built for the purpose of sending small parts of data such as analytics data to the server. Should this function fail, a REST Call is made instead.

### 4.5.4 Server-Side Tracking Factory

The data from the client-side tracking Factory ends up in the server-side tracking Factory through either a Beacon Call or a REST Call. The code of the server-side tracking factory can be found in Appendix D.5. For this implementation, the data was stored via Log Analytics of Microsoft Azure (Section 4.5.5). It is possible to send the data directly to Azure from the client their browser. For this case, a server-side solution was chosen to give more control over what data is sent to Azure. For example, the data could be validated before sending it to Log Analytics. Also, the client does not have to make a call to an external service and the client cannot send arbitrary data the logs which is more secure.

The `track(data)` method first checks if the need configuration is available. If this is the case the data is processed and send to Azure via a POST call. The data must be signed with the Azure Shared Key. After this, the data is available from Azure.

### 4.5.5 Implement Log Storage

The Hosting department is responsible for making sure that the logs can be stored. In consultation with Development and the Process miner, a fitting storage solution should be found. In this case, a cloud solution was chosen since this aligns with the current cloud-based architecture. Moreover, this was the fastest way of implementing log storage since most cloud providers provide off the shelf fitting solutions that can be started instantly. It also has an API that and integration with Grafana which will be used in the next step. The Azure Log Analytics application, which is designed for storing and querying log data was used.

In the Azure Portal (https://portal.azure.com) it is possible to directly analyse the logs from the Log Analytics workspace. Figure 4.5 shows a screenshot on the workspace.

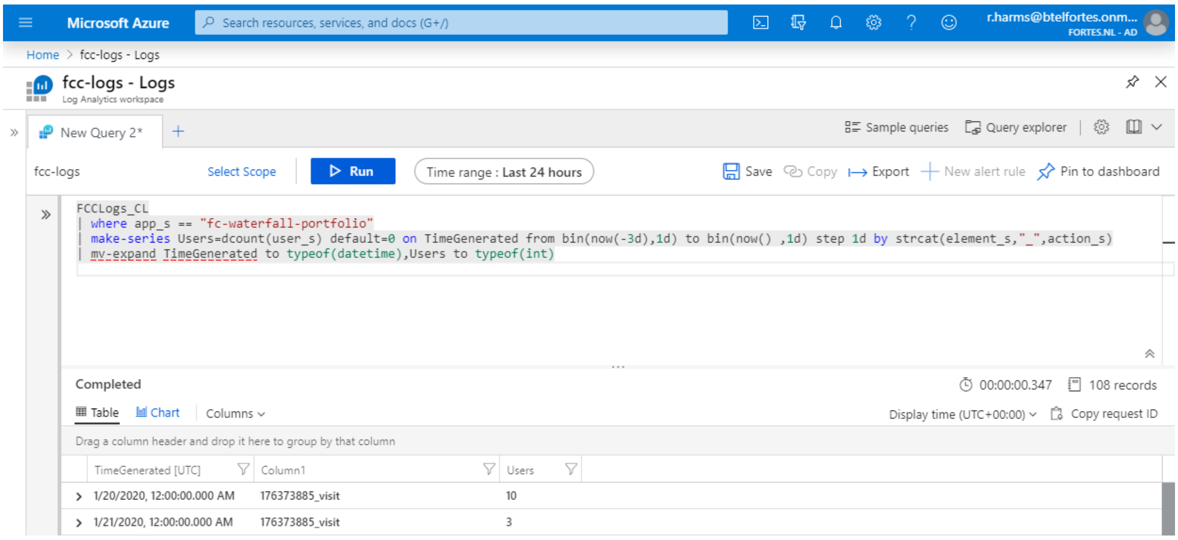



**Figure 4.5:** The Log Analytics workspace.

### Pricing

Pricing for Log Analytics can be found at https://azure.microsoft.com/en-us/pricing/details/monitor/. Currently, sending log data to Log Analytics is free until 5GB per month. If more data is sent, the costs are €2.522 per GB per month (West Europe). The default retention period for Log Analytics is 31 days which is included in the price for data ingestion. To extend the data retention period, a fee of €0.102 per GB per month is currently charged.

### API and Integration

To collect the data in Azure Log Analytics, the Log Analytics Data collector API was used ( https://docs.microsoft.com/en-gb/azure/azure-monitor/platform/data-collector-api). The implementation of this API for this case can be found in Section 4.5.5. For retrieving the data, it is possible to use the Log Analytics Query API (https://docs.microsoft.com/en-gb/rest/api/loganalytics/query).

Azure also has integration Microsoft PowerBi (https://docs.microsoft.com/en-gb/azure/azure-monitor/platform/powerbi) and Grafana (https://docs.microsoft.com/en-gb/azure/azure-monitor/platform/grafana-plugin). It is also possible to export the data as a csv file directly from the Log Analytics Workspace.

### 4.5.6 Visualise Data

For visualisation of the data, two approaches were used. The first approach uses Grafana which is a dashboard utility. The second approach, used for the visualisation of the customer journey, was first done with Rapidminer, but after some unsatisfying results, Celonis was used.

#### Grafana (Dashboard)

As discussed in Section 4.5.5, Grafana allows to import data live from Log Analytics. Grafana was installed on a local server and an API key was generated in Azure to provide access to the data. This API key was entered in Grafana and provided direct access to the data.

The goal was to get insight in the use cases of the apps. Therefore, two kind of dashboards were created as an example. Screenshots of both dashboards can be found in Appendix F.

The main dashboard gives on overview of all apps, the active users in each app over time and the browser usage. This last metric was a suggestion from developers that would like to know what browsers are used. For each app a column is made that shows with the metrics of that app. On the bottom of the page, a link to the app specific dashboard is shown. Figure F.1 shows a screenshot of this dashboard.

The app specific dashboard has information on active sessions, the duration of the session and on the usage of functionality. Data is shown based on the selected period. Figure F.2 shows a screenshot of this dashboard.

To generate the graphs of the dashboards, the relevant data first needs to be queried. Microsoft Azure uses an own language called Kusto query language for retrieving log data. In Appendix F.1 three examples can be found on queries that were used to generate different graphs.

#### Rapid Miner and Celonis (Process Mining)

To visualise the customer journey, a dashboard tool like Grafana does not suffice. Two tools that support Process Mining were therefore used: RapidMiner and Celonis.

RapidMiner is a platform dedicated to data science. To use RapidMiner for process mining, the plugin RapidProM can be used. This plugin contains different tools for process mining, such as converting csv files to event logs, tools for analysis of processes, process mining algorithms for discovery and conformance checking operators.

The data flow used for RapidMiner can be seen in Figure 4.6. Unfortunately, it was not possible to easily access the live data from Log Analytics. Instead, the CSV export function was used, which was then imported in RapidMiner. To identify events, the app name, element id and action id were combined. The user id was set as trace identifier. A start and end event were added to each trace since this gave better results. Moreover, events were sorted chronologically otherwise RapidMiner did not recognise the order of events correctly. Both Inductive Miner and Fuzzy Miner operators were used. The output of both miners can be found in Appendix F, Figure F.3 and Figure F.4 respectively. The petrinet that came out of the inductive miner was then used in the analyse performance operator to make it possible to do some performance analysis. The performance analyser is shown in Appendix F, Figure F.5.

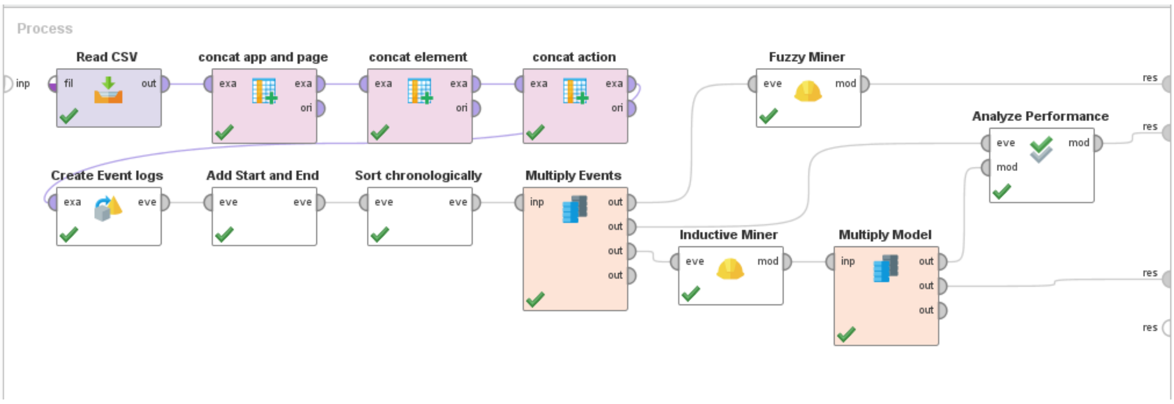



**Figure 4.6:** RapidMiner process.

Unfortunately, the results from RapidMiner were not as expected. The readability of the process in the end results was poor. The layout of the final processes was also not well structured on the test data set. Although the processes could be recognised, it was hard to see which steps the user performed. Therefore, it was decided to try out another (commercial) tool.

Celonis is an alternative to RapidMiner with RapidProM. The tool is cloud based. After registration, a free perk demo environment was made available. The same CSV file used earlier with RapidMiner was used. After attaching the right labels to the columns of the CSV, it a model was generated automatically. In Appendix F, Figure F.6 the process overview generated from the CSV is given.

### 4.5.7 Check Data Compliance

Before proceeding to the final steps, which in practice could also contain activating tracking in production, the data compliance was checked. In the initial tracking code, some mistakes were made related to tracking event. For example, ids of objects were collected and used as identifier for pages. These objects ids relate to for example a specific strategy, portfolio or programme that a customer created. On first sight is was thought that this could give extra insight in which page each user specifically visited. However, afterwards this created a lot of different cases which were hard to aggregate. From a privacy perspective, collecting these ids is also not a good idea, since cases could be linked to users based on the programmes or portfolios they visited. Since this data did not provide useful, the tracking code was adopted to not store these ids and the test data was sanitised from these ids.

Other privacy issues, both business and user, were not found. After the code was corrected, the author proceeded to the next step.

### 4.5.8 Compare Expected Journey to Real Journey

After adding the tracking to the software, implementing storage, visualising the data and check data compliance, the compare step could be performed. However, data was still missing. Instead of enabling the tracking in the production environment, an experiment was performed. This experiment was also used for the validation of this model. Chapter 5 contains more details on the execution and results of this experiment.

The data generated from the experiment was generated using scripted scenarios for the use cases discussed earlier. Therefore, deviations, if any, cannot be generalised to the real-world application. The next example should therefore only be interpreted with this in mind.

After the experiments the data was added to Celonis. The values on the connections between activities was set from case frequency to activity frequency. This showed the total number of times an activity was followed by the activity the connection pointed to. One interesting connection that stood out is shown in Appendix F, Figure F.7. The process shows users clicking on a card on the funnel tab kanban board (the portfolio_funnel_Board_click_card action). Next they perform different actions: adding a label to the card (the path shown to the left), adding an attachment (the path in the centre) or starting an agile program (the path shown to the right). After each path the user closes the card (the Portfolio_funnel_modal_click_Cancel action). The closing action is however followed 7 times by the click card action. This loop suggests that users are doing a task repetitively, probably adding labels to multiple cards. After further analysis a possible outcome could be that users would like to add the same label to multiple cards and the only option now is to open, add a label and close each card individually. A possible solution would in this case be to add a multi select option so that labels can be added to cards in bulk.

### 4.5.9 Improve Application

Based on the results from the previous step, a multi-select could have been added to the software by developers. Also, information on browser and functionality usage could have been used to determine what to develop next.

## 4.6 Prototype Implementation Conclusion

During the prototype implementation all steps of the method were followed. determining what data to collect and more importantly what data not to collect was hard in the beginning. After some trial and error, the current format proved to be best. Also, the visualisation part gave some difficulties. In RapidMiner it was relatively easy to show the first results. However, the visualisation was not intuitive or readable. Celonis offered a better alternative in terms of visualisation. The free edition has however its limitations such as the flexibility to transform the data before visualisation.

The prototype implementation is now at a state that it can almost be used in production. By completing the method, it is now possible to get information on the usage of the FCC. The code can be activated at customers via a ‘feature toggle’, after which the log data will automatically be collected. Not all functionality is tracked yet but it can be added when needed. The Grafana dashboard can be deployed on an internal server at Fortes, after which it can be further customised. For the Celonis application, it is still needed to download a CVS of the data from Azure. A live connection or an intermediate server could be developed to solve this.

The final steps of the method in the case study are still limited. This is because no real data was collected. The prototype implementation was finally used for the validation of the model. Chapter 5 has more details on the validation and discusses the data that was collected.

## Chapter 5

# Validation

In Chapter 4, the method was applied to the problem context of Fortes Solutions. This Chapter discusses the validation of the method. The validation assesses if the expected are produced in the intended problem context (Wieringa, 2014). The approach used for the validation is discussed in Section 5.1. The case study was used as validation model, where the artifact was the implementation of the method and Fortes was the Model of Context. To study the validation model, A single-case mechanism experiment was used which is discussed in Section 5.2. To study the usefulness of the method for software companies to improve their software, a workshop was held at Fortes which is discussed in Section 5.3.

## 5.1 Validation Approach

The validation approach was two-fold. First, a single-case mechanism experiment was conducted to validate that the method is suitable to track user behaviour in a privacy preserving way. Figure 5.1 shows the validation model adapted from Wieringa (2014). The model of artifact is the application of the method described in Chapter 4. Fortes acts as the model of context.

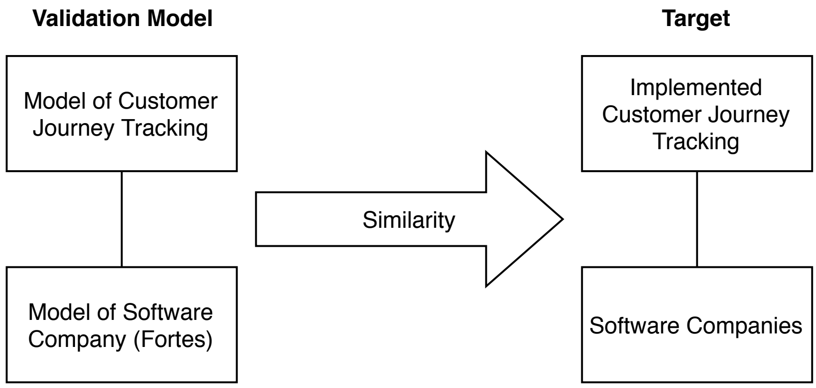



**Figure 5.1:** Validation model for Customer Journey Tracking, adapted from Wieringa (2014, p.61).

A single-case mechanism experiment was used to validate that the method can be used for collecting user behaviour data. In a single-case mechanism experiment, stimuli are applied to the validation model by the researcher to explain the response mechanisms in the model (Wieringa, 2014). This concept is further explained in Section 5.2.

In addition, expert opinion was used to validate that the method can be implemented within a software company and contributes to improving their software. The method was presented to a panel of experts in a workshop. The experts had the opportunity to share their opinion about the effectiveness of the method in a real world situation. Afterwards, the experts were asked to fill in a questionnaire. Section 5.3 continues on the workshop approach.

## 5.2 Single-Case Mechanism Experiment

In Chapter 4, the case of Fortes was introduced. The method for customer journey tracking was applied in their software by the author of this paper. Since all steps of the method were followed, including the adding tracking in the software of FCC, the implementation was used as a basis for validating that the method can be used for user behaviour tracking. The following approach was used to setup the single-case mechanism experiment:

- Prepare
  - Write scenarios for different use cases
  - Prepare experiments with content
  - Verify that tracking works
- Execute
  - Instruct participants
  - Participants follow the scenarios
- Analyse Results
  - Collect tracking data by researcher
  - Check tracking data for usefulness

### 5.2.1 Prepare

For the experiments, the tracking software has been added to three apps of the FCC; Strategy, Portfolio and Agile program (Figure 4.3). These apps were selected because they have dependencies in the workflow and are used by different roles in the organisation (SEO, Programme Manager and Portfolio Manager).

For each role, different goals within the application were selected that fit with the role. These are listed in Table 5.1. For each role, a realistic scenario was written down. These scenarios can be found in Appendix E.

**Table 5.1:** Roles and Goals.

| User | Role | Goal |
|---|---|---|
| Alice | Director | Create a strategy |
| | | Create a portfolio |
| Bob | Portfolio Manager | Fill portfolio |
| | | Create agile team |
| | | Add members |
| | | Add agile programme |
| Carol | Programme Manager | Add portfolio items to programme |
| | | Fill programme |
| | | Add team |
| Admin | administrator | Setup |

After creating the scenarios, the application was prepared for the experiment. First, accounts for the different roles were added. Then settings such as access rights and configuration settings needed for the application to work as intended were applied. After that, some demo content was added to make the scenarios more realistic.

The scenarios were then tested by the researcher to check if the tracking worked as expected. Where needed, tracking was added or the scenario was adjusted.

### 5.2.2 Execution

Three students working at Fortes Solutions were asked to participate in the experiment. Each participant was asked to follow the scenarios for the three roles as described in Appendix E. The students all had experience with the software, however they did not have a background in portfolio or project management and have therefore no experience in using the software as real use users. However, the scenarios were made in such a way that the outcome would capture a realistic use case. An alternative method would be to collect data of real users to get a more realistic scenario on how users behaviour could be tracked. however, the scenarios would be more unpredictable and it would be harder to confirm if users actually followed the expected scenario. The scenarios used in this case offer a more controlled environment.

The three were asked to follow the scenarios using their own computer. The researcher did not actively monitor the session but was still around in the case something was unclear. Where possible, the participants were asked to come up with names or add a random number of items. This was done to see if the tracking could give insight in what users generally did without collecting any personal or business data.

After each participant executed the scenarios, the data from that session was downloaded from Azure. After that, the application was restored to the original state so that the next participant could perform the experiment. Each participant completed the scenario without any problems.

### 5.2.3 Analysing Results

After the experiments, a dataset was available in Azure log monitoring. The set contained data from the three participants, that all completed three scenarios. This in total nine journeys were available. Each of the nine cases has a distinctive id to identify the session. The data was scanned for any inconsistencies. in total, 307 datapoint were collected. One datapoint was removed since this was a session with only one datapoint that possibly occurred by a participant logging in twice. The data was scanned for any personal data and none were found. In this case this could be done by hand, but with larger datasets this could be automated by scanning for irregularities in test sets.

As discussed in Chapter 4, RapidMiner and Celonis were used for the analysis of the data. These are discussed below.

### RapidMiner

For RapidMiner, the inductive miner and the fuzzy miner gave the best results. These results can be found in Appendix F: Figure F.3 and Figure F.4. With the default settings, the paths the participants followed were distinctively visible. However, further analysis was hard since the model was too large and the activity names were not fully readable since the boxes were too small. The inductive miner has a petrinet as outcome, which was used as the model in the analyse performance step. Figure 5.2 shows a cutout of the performance analysis process, in this case for clicking through the portfolio app. The colour indicates this case the average throughput time. As can be seen, there are a few steps that took longer (indicated by the orange/red colour). A larger overview of this analysis can be found in Appendix F, Figure F.5.However, the model has a lot of transitions and even after adding a Reduce Petrinet operator in the model, this was slightly improved but not much. Based on these results, it was decided to give Celonis a try.

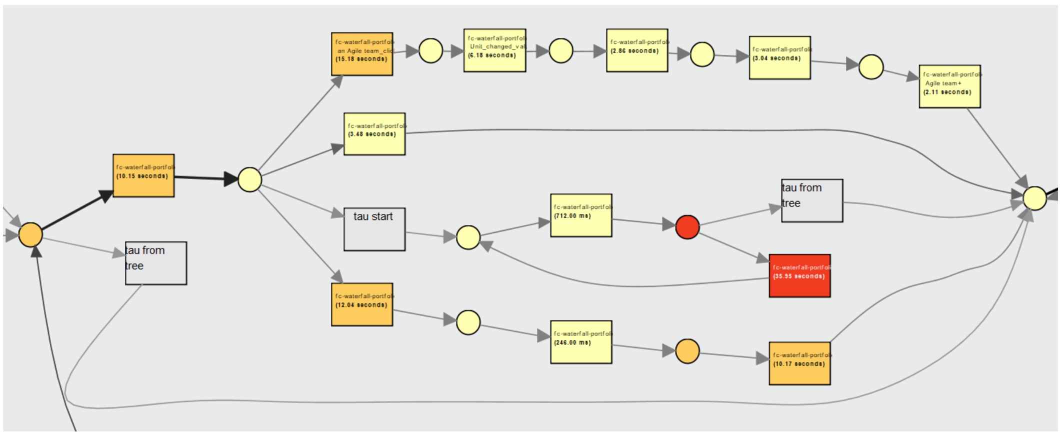



**Figure 5.2:** RapidMiner - part of the portfolio process.

### Celonis

Celonis is an SAAS tool dedicated to process mining. After creating an account, an online environment is set up in which different tools are available. As said in Section 4.5.6, a CSV file was uploaded to this online environment. For the process mining part, the process analytics tool was used. As can be seen in Appendix F, Figure F.6, The initial model shows three distinctive paths. Figure F.7 zooms in on a part of the portfolio process. It shows that different actions are performed after clicking on a card such as adding attachments or adding labels.

In total, Celonis showed data of 10 cases. This did not match with the expected 9 cases (3 students logging in as 3 different users). After exploring the cases, one case was found that only had 1 activity, which was added by mistake. This case was omitted from the data.

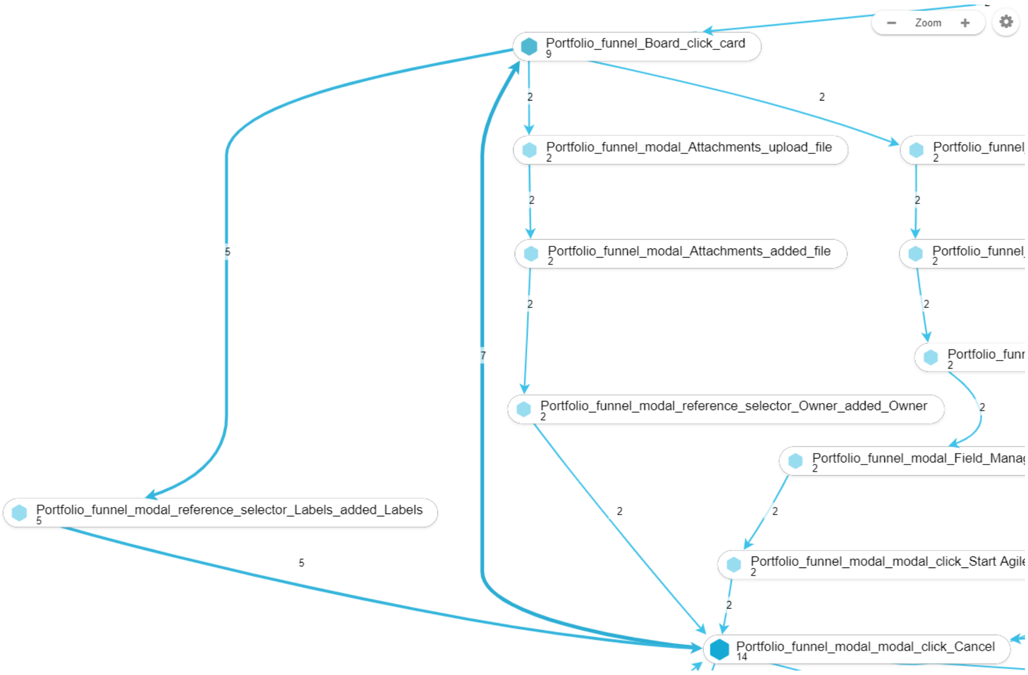



**Figure 5.3:** Celonis - Portfolio process.

Celonis also allows to set the throughput time on the connectors to give an indication on how long certain path take on average. The smallest unit was however 1 minute. Therefore, most paths showed 0 minutes between activities as the users proceeded to the next activity within 1 minute. Celonis however clearly showed the paths the user followed during the sessions. In the used dataset this data was limited to what the scenarios told the user to do. If actual data was used, the path would be less predictable. Celonis however does have good filtering options that allow for grouping activities and reducing the number of connections. As shown in Section 4.5.8, it is possible to identify deviations from the expected behaviour the Celonis process, which could contribute to improving the software.

### 5.2.4 Grafana

Section 4.5.6 also discussed the use of a Grafana dashboard for a more general overview on the data. This dashboard was not validated though an experiment a with the user behaviour tracking. However, for completion it is added here. For this, no specific scenario was used. Instead, the author and another developer at Fortes activated the tracking in their application to generate some general log information. Figure 5.4 shows one of graphs in the dashboard: actions performed in the strategy app. One thing to note is that the data also contains an id on which strategy the user visited. this should be avoided in a future version since this data could compromise business or user privacy. Appendix F contains more screenshots on the Grafana dashboard.


**Figure 5.4:** Grafana - Portfolio process.

## 5.3 Workshop

To validate the method in practice, a workshop on tracking user behaviour was given at Fortes. For this workshop, Developers and Product Management were introduced to the method and asked to give their feedback on the method. A presentation was given on the method, the implementation of the prototype and the results from the singe-case mechanism experiment. After the workshop, an anonymous questionnaire was handed out to the participants to gain feedback on the method. This was used to show if the method is useful for Fortes. Privacy and scalability were also discussed.

### 5.3.1 UTAUT

The questionnaire was be based on the UTAUT of Venkatesh, Morris, Davis, and Davis (2003). The UTAUT model is based on eight use acceptance models and can be used to determine the likelihood of success of new technology. The model is shown in Figure 5.5.There are four key constructs (performance expectancy, effort expectancy, social influences and facilitating conditions) and four moderating variables (gender, age, experience and voluntariness of use) used in the model.

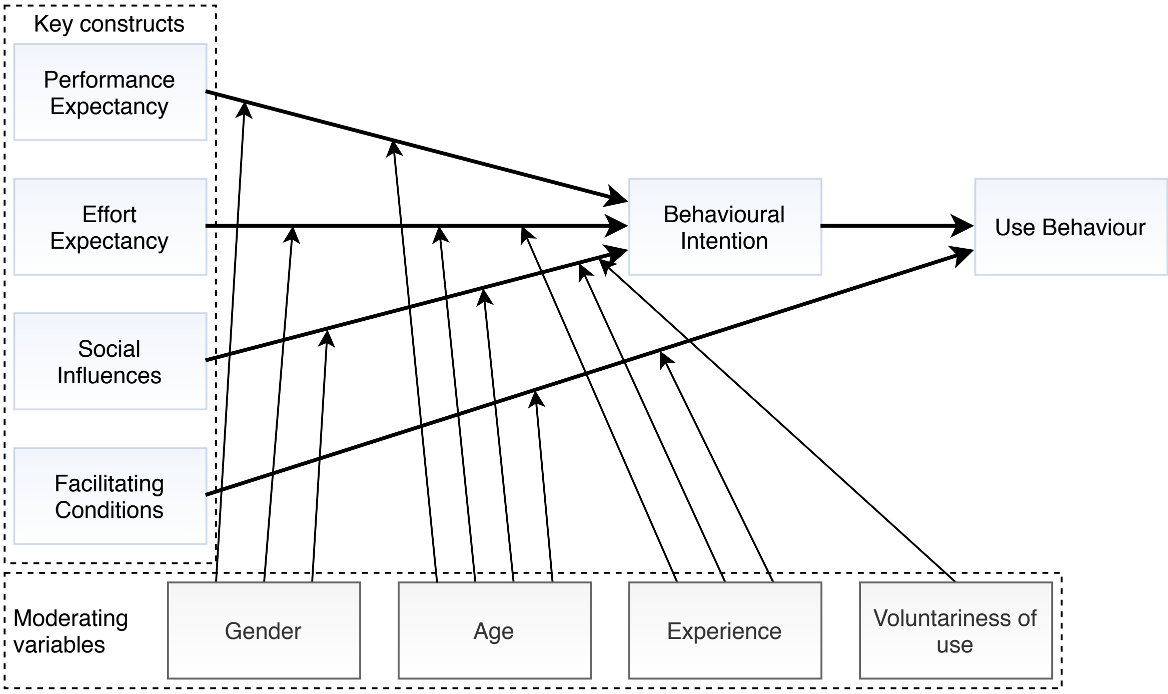



**Figure 5.5:** The UTAUT model. Adapted from Venkatesh, Morris, Davis, and Davis (2003).

The questionnaire can be found in Appendix G. The questions were adopted from Venkatesh et al. and where needed extra questions were added (2003). The questions are grouped by the key constructs of the UTAUT model. For the effort expectancy, question five was added to measure if the employees of Fortes would find implementing the method worth the effort. Furthermore, two additional categories were added: Privacy and Scalable. These were used to find out if the employees agree that the method ensures the privacy of the user/business and if they think the method is scalable. Finally, a final remarks question was added at the end for anything else the employees might want to share. Google Forms was used to send the questionnaire to the employees.

### 5.3.2 Workshop Results

In total, thirteen employees of Fortes attended the workshop, seven of which are developers, two are testers, one is product management, one is hosting and two are students (testing and development).

The presentation introduced the concepts of process mining and customer journey. The three methods from Chapter 3 were explained and the method of choice for Fortes was motivated. After that, all steps from Chapter 4 were discussed. For the developers an example was given on how to add tracking to the software. For hosting and development link to Azure was briefly explored. The actual result from the single-case mechanism experiments (Section 5.2) were shown with a live demo of the Grafana dashboard and the Celonis application. During and after the presentation questions and suggestions were answered and noted. The presentation ended with a request to fill in the survey.

Besides validating the method, the workshop was also meant to transfer knowledge on how tracking can be implemented at Fortes. The prototype implementation in combination with the workshop should give enough understanding and motivation to further implement user behaviour tracking.

During the workshop, a discussion on the privacy of users started. One developer suggested to also track the role of the user (admin, portfolio manager, member, etc.). However, someone else noted that this could form a privacy issues, especially when it is possible to link the pseudonymised data with other data sources. The author shares this opinion. The general notion was that the goal was not to collect any private data but that this could be challenging.

Another developer noted in relation to the privacy topic that it might be possible to instead of logging timestamps to log session time in combination with a login date. This prevents that sessions can be related to users based on the login timestamp but still makes it possible to track the behaviour of users over time. According to the author of this paper, this could be a good solution for tracking user data in a privacy preserving way without losing much relevant data.

In general, the attendees were positive using the method for adding tracking to the software. For the developers, getting feedback on user behaviour is something that is currently not possible.

### 5.3.3 Questionnaire Results

In total, seven employees that attended the workshop filled in the questionnaire. Table 5.2 shows the participants for each group. Because of the limited number of participants in each group, all results were grouped together. The same applies to the age of the participants, which were all between 20 and 39 (Table 5.3).

**Table 5.2:** Number of participants per group.

| Product Management | Developer | Hosting | Data Analyst | Tester | **Total** |
|---|---|---|---|---|---|
| 0 | 4 | 1 | 0 | 2 | **7** |

**Table 5.3:** Age of participants.

| Under 19 | 20-29 | 30-39 | 40-49 | 50-59 | 60 or older | **Total** |
|---|---|---|---|---|---|---|
| 0 | 1 | 6 | 0 | 0 | 0 | **7** |

Considering the moderating variables, the gender was omitted as moderating variable. Age influences all four key constructs. The relatively young age of the employees possibly has a positive influence on the model, which should be noted when generalising the data. The experience with the use of the technology is limited to the workshop for all participants. This limited experience means that the effects of effort expectancy and social influence should be stronger. The voluntariness of use variable .

The results of the questionnaire can be found in Appendix H. The results are grouped per key construct of the UTAUT model. A boxplot diagram shows the results, where the number above each diagram represents the average value. In the Questionnaire, Q16 was asked negatively, which followed the original UTAUT questions. In the results, this question was turned around from “The method is not compatible with other systems I use.” to “The method is compatible with other systems I use.” and the results were inverted as well (disagree became agree and so on). This makes the result easier to compare.

### Use Behaviour

For the performance expectancy, the results are slightly positive. The exception is Q4: “If I use the method, I will increase my chances of getting a raise.”, which was answered mostly negative. This could be explained with the cultural environment, in which is less performance driven.

The effort expectancy questions also show mostly optimistic. This indicates that most employees feel good about using the method. The added question five shows that the employees are positive about implementing the method. Moreover, the interaction with the method is clear for all employees (Q5).

The social influence is mostly neutral. This might indicate to for the method to be successful, more support from higher management could be improved.

The same applies to the facilitating conditions that also shows an average to above average positive result. It should be noted that the method was already implemented by the researcher. This could possibly explain why employees answered mostly positive to the compatibility of the method (Q16), since the method was implemented with that in mind.

The behavioural intention, which are influenced by performance expectancy, effort expectancy and social influences, tend to follow these constructs. The opinions on usage prediction (Q18) are divided. Prediction (Q19) and planning (Q20) are slightly negative. Combining this with the key constructs, a neutral to slightly positive attitude to using the method.

By combining the Behavioural intention with the Facilitating conditions the actual use can be explained. In this case, the facilitating conditions are positively influenced by the case study which already implemented a large part of the method. Since the behavioural intention is mostly neutral, it might be needed to improve the social influence from higher up or the experience with the method so that people are more motivated. Since most employees agree that the method is useful for their job (Q1) and that implementing the method is worth the time (Q9). This indicates that the method is easy to implement and that the results of the tracking are useful for the company.

#### Privacy and Scalable

Questions Q21 and Q22 were added to get insight in if the employees think the privacy aspect is thought of enough. The answers varied between strongly disagree and agree, and both averaged around neutral. This indicates that the employees are not completely confident about the privacy aspect.

Questions Q23, Q24 and Q25 about scalability were answered positive. Employees feel confident that the method can be scaled easily and could be used at both small and large software companies.

### 5.3.4 Discussion on Workshop

From the workshop discussion and questionnaire results, the employees of Fortes are generally positive on using the method. Using the method could contribute to getting some insight in the application. however, some employees still question the privacy aspect. The discussion on privacy during the workshop however indicates that there are viable solutions for gathering user feedback in a privacy preserving way.

# Chapter 6

# Conclusion and Discussion

This chapter concludes the research results and discusses the contributions and validity of the research. The research results are discussed per research questions in Section 6.1. Section 6.2 discusses contributions to both software companies and the literature. In Section 6.3, the validity of the research is presented. Last, in Section 6.4, the recommendations to Fortes are given and future work is discussed.

## 6.1 Research Results

**RQ I - What techniques can be used for process mining of user behaviour?**

According to this research, four distinctive categories of process mining were identified in Section 2.2.1: business process mining, service mining, mining software process and mining user behaviour. Al four use the general techniques of process mining, but also have their own approaches and challenges. For process mining of user behaviour, the literature suggests using the XES standard or tools such as ProM or Disco. Both are heavily dependent on the scientific research done with those tools and less on actual business using these tools. This literature introduced a method that can be used for process mining of user behaviour in a real business context. From the case study (Chapter 4), the tool Celonis looked most promising for showing the user behaviour in software applications.

**RQ II - What techniques can be used for mapping the customer journey in collaborative software?**

Customer journey mapping in collaborative software starts with regular customer journey mapping. in Section 2.2.2 different papers were identified that explored customer journey mapping in combination with process mining. However, literature on customer journey mapping in collaborative software was not found. On the other hand, the usage of process mining for collaborative software is touched by the literature. The customer journey tracking method as discussed in Section 3.5 introduces the steps needed for collaborative user behaviour tracking. However, an example to show if and how this could be applied in practise is not currently present. From this research it is clear that the collaborative part requires more effort from software companies in term of privacy, implementation of the tracking and analysis of the data. Therefore, the author recommends software companies that deliver CSCW software to start with customer journey tracking without the collaborative part. In the case that it is needed to get insight in how users work together, the tracking can be extended.

**RQ III - What privacy aspects are relevant to collecting user behaviour data?**

User behaviour data inherently concerns the privacy of users since their data is processed. This research identified literature on two different forms of privacy: user privacy and business privacy. For user privacy, the GDPR was taken as a starting point. From here, two scenarios on collecting user behaviour were given: using pseudonymised data or using anonymised data. Which one to use depends on what the goal of collecting the data is. The ground rule is privacy by design, which inherently means that using anonymised data is preferred unless this is not possible. Besides user privacy, business privacy should not be forgotten. Especially in software used by companies, sensitive data could be stored that should not leave the application. The literature has some suggestions on how process mining can be outsourced without sharing sensitive information.

**RQ IV - How can the tracking of user behaviour be added to existing software?**

This research suggests three methods for software companies that want to start with customer journey process mining. A decision tree is given in Figure 3.1 that companies can use to decide which method they should choose. The functionality tracking method is the most basic method but therefore also the easiest to implement and with the least privacy concerns, but it is not possible to analyse the actual journey of customers. With the customer journey tracking method, this can be done. For non-collaborative tracking, this can be done anonymously. Collaborative user behaviour tracking can only be done with pseudonymised data. Finally, Personalised feedback tracking is the most advanced form of customer journey process mining and therefore the hardest to implement. Since the methods have some overlap, a company can start with functionality tracking and easily extend the method to customer journey tracking and then personalised feedback tracking.

**RQ V - How can tracking of user behaviour be applied in a privacy preserving way?**

Based on the literature, a privacy by design approach was applied to the three methods introduced in Chapter 3. According to the author of this paper, anonymous data should be used when possible. In all other cases, pseudonymised should be used. To guarantee business privacy, no customer data should be logged, only metadata on the actions the user performs. All three methods have a step to check data compliance to make sure no privacy sensitive data is collected. This all contributes to the goal of implementing user behaviour tracking in a privacy preserving way.

## 6.2 Contributions

This research contributes to both software companies that want to get insight in user behaviour and the scientific field of customer journey process mining.

### 6.2.1 Contribution to Software Companies

For software companies, three methods were introduced that software companies can use to apply user behaviour process mining to their software. This will give the company more insight in how users use their application. The case study can be used as an example for implementation. All steps needed for user behaviour process mining are given, which means that no prior experience with user behaviour tracking is needed for the company or that the company already needs a form of data collection. The method is made with the privacy of users and businesses in mind, which is an important topic when dealing with user data.

### 6.2.2 Contribution to the Literature

Regarding the literature on customer journey process mining, this research further explored the sub-field of process mining, mining user behaviour. Process mining normally implies using *existing logs*. In this research this was not the case. For many companies there is no existing data or the data is not suitable for this kind of user behaviour tracking. Additionally, not using existing logs made it possible to apply privacy by design for collecting user behaviour. This research also explored a real business case for user behaviour process mining. To the knowledge of the author, this was not done before. This practice contributes to aligning the literature output with actual business demands.

The research also introduces concepts of collaborative user behaviour process mining and personalised feedback process mining. These concepts were not deeply explored in this research but can still be useful in future literature in which the concepts can be further explored.

## 6.3 Validity

The validity of the research is discussed in terms of internal and external validity. Several measures were taken for internal validity. For the literature review, a SLR protocol was followed, which ensures that the results are consistent and the research is repeatable. For the solution design, the design science methodology was used to investigate the problem, design a treatment and validate this treatment. The design cycle was followed only once. Extra iterations could have improved the final design based on the feedback of the users. Due to the set of solution requirements, the final solution contained more methods than that could be validated with a case study in time span of the research. However, the methods have some overlap: the customer journey tracking is an extension of functionality tracking and during the case study, the functionality tracking was also explored. The personalised feedback tracking method still needs to be further explored and validated.

For the questionnaire, the validated UTAUT model was used. Due to the limited resources, only 7 employees answered the questionnaire. The employees also did not interact directly with the method, but only had a workshop on the method. All steps of the method were performed by the researcher. A case study in which the method is actually used by the actual stakeholders could further improve the validity of this research.

Regarding the external validity, A limitation of the single case mechanism experiment is that only one case was explored. The case was applied to a real software company. The validation to check if the tracking method worked was done with an experiment based on scenarios which delivered the expected data. No real data was used and in practise it will be harder to collect data based on use cases. Regarding the privacy aspect of the research, only the GDPR was discussed. Although most local privacy laws in Europe are based on the GDPR, however, differences exist. Also, laws in other parts of the world are not discussed which means that the privacy aspect of this research cannot be generalised to the rest of the world. The underlying concepts such as privacy by design are however still valid.

## 6.4 Recommendations and Future work

This section is twofold. First are the recommendations to Fortes and other software companies. After that future work is discussed.

### 6.4.1 Recommendations for Fortes

The next step for Fortes is to implement the tracking software further into their application. For this, it is important to follow the method again and make sure that the prototype implementation of the method still fits their needs.

The prototype implementation can be used as a starting point for adding tracking to software. Based on what functionality Fortes wants to track besides what is already possible, this can be extended.

For creating a baseline of use cases, the automated test servers can be used. These automated tests click through the application based on different scenarios. By linking the sessions of the tests to the use case that is being tested (by, for example, adding a use case id to the tracking software), a baseline of the use cases can be made, which could be compared to the actual users. Furthermore, the data from the automated tests will show if the tracking works as expected. After this, the software can be implemented at customers to get insight in what functionality customer use and if use cases are followed as expected. For other software companies, the method can be applied in the same way as the case study in Chapter 4.

### 6.4.2 Future Work

For the scientific future work, this research can be extended to further explore the collaborative feedback tracking and feedback to users method. These two methods were not fully validated during this research. Furthermore, the implementation could be applied to other software companies to validate that the same approach works for them. Future research could also check if other privacy laws might apply to adding tracking to software. However, since privacy laws can differ per country, a country specific approach is probably needed.

# References


Anderl, E., Schumann, J. H., & Kunz, W. (2016). Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys. *Journal of Retailing*, *92*(2), 185–203. doi:10.1016/j.jretai.2015.10.001

Astromskis, S., Andrea, J., & Mairegger, M. (2015). A process mining approach to measure how users interact with software: An industrial case study. In *International conference on software and systems process, icssp 2015* (Vol. 24-26-Augu, pp. 137–141). Association for Computing Machinery. doi:10.1145/2785592.2785612

Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, *88*(December 2017), 407–414. doi:10.1016/j.jbusres.2017.11.047

Bernard, G. & Andritsos, P. (2017a). A Process Mining Based Model for Customer Journey Mapping, 49–56.

Bernard, G. & Andritsos, P. (2017b). CJM-ex: Goal-oriented exploration of customer journey maps using event logs and data analytics. In *Ceur workshop proceedings* (Vol. 1920). CEUR-WS.

Bolt, A., de Leoni, M., & van der Aalst, W. (2018, May). Process variant comparison: Using event logs to detect differences in behavior and business rules. *Information Systems*, *74*(1), 53–66.

Burattin, A., Conti, M., & Turato, D. (2015). Toward an Anonymous Process Mining. In *2015 3rd international conference on future internet of things and cloud* (pp. 58–63). IEEE Comp Soc, Tech Comm Internet; IEEE Comp Soc. 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, CA 90720-1264 USA: IEEE COMPUTER SOC. doi:10.1109/FiCloud.2015.9

Cavoukian, A. (2009). Privacy by Design - The 7 foundational principles - Implementation and mapping of fair information practices. *Information and Privacy Commissioner of Ontario, Canada*, 5.

Diamantini, C., Genga, L., Potena, D., & Ribighini, G. (2014, May). A methodology for building log of collaboration processes. In W. Smari, G. Fox, & M. Nygard (Eds.), *2014 international conference on collaboration technologies and systems (cts)* (pp. 337–344). Intelligence Adv Res Projects Activ, Off Director Natl

Intelligence; Honeywell Int Inc; Adventium Labs; Intelligent Automat Inc; Knowledge Based Syst Inc; LexisNexis Corp; Ball Aerosp & Technologies Corp; Intel Corp; MEI Res Ltd; Microsoft Res; Springer V. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE. doi:10.1109/CTS.2014.6867586

Dijkman, R., Turetken, O., van Ijzendoorn, G. R., & de Vries, M. (2018). Business processes exceptions in relation to operational performance. *Business Process Management Journal*. doi:10.1108/BPMJ-07-2017-0184

Dixit, P., S. Garcia Caballero, H., Corvò, A., F. A. Hompes, B., C. A. M. Buijs, J., & van der Aalst, W. (2017). Enabling Interactive Process Analysis with Process Mining and Visual Analytics. In E. VanDenBroek, A. Fred, H. Gamboa, & M. Vaz (Eds.), *Proceedings of the 10th international joint conference on biomedical engineering systems and technologies* (pp. 573–584). AV D MANUELL, 27A 2 ESQ, SETUBAL, 2910-595, PORTUGAL: SCITEPRESS - Science and Technology Publications. doi:10.5220/0006272605730584

Dong, L., Liu, B., Li, Z., Wu, O., Babar, M. A., & Xue, B. (2018). A Mapping Study on Mining Software Process. In J. Lv, H. Zhang, X. Liu, & M. Hinchey (Eds.), *24th asia-pacific software engineering conference, apsec 2017* (Vol. 2017-Decem, pp. 51–60). IEEE Computer Society. doi:10.1109/APSEC.2017.11

Earley, S. (2014). Usability for internal systems: What's the payoff? *IT Professional*, *16*(5), 66–69. doi:10.1109/MITP.2014.80

European Union. (2016). Regulation 2016/679. *Official Journal of the European Communities*, (4.5.2016), 1–88. doi:http://eur-lex.europa.eu/pri/en/oj/dat/2003/l{_}285/l{_}28520031101en00330037.pdf

Følstad, A. & Kvale, K. (2018). Customer journeys: a systematic literature review. *Journal of Service Theory and Practice*, *28*(2), 196–227. doi:10.1108/JSTP-11-2014-0261

Gadler, D., Mairegger, M., Janes, A., & Russo, B. (2017). Mining Logs to Model the Use of a System. In *11th acm/ieee international symposium on empirical software engineering and measurement (esem 2017)* (pp. 334–343). International Symposium on Empirical Software Engineering and Measurement. ACM; IEEE; IEEE Comp Soc; IBM. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE. doi:10.1109/ESEM.2017.47

Harbich, M., Bernard, G., Berkes, P., Garbinato, B., & Andritsos, P. (2017). Discovering Customer Journey Maps using a Mixture of Markov Models. (December), 3–7.

Heuchert, M., Barann, B., Cordes, A.-K., & Becker, J. (2018). An IS perspective on omni-channel management along the customer journey: Development of an entity-relationship-model and a linkage concept. In N. P. X. L. Drews P. Funk

B. (Ed.), *Mkwi 2018 - multikonferenz wirtschaftsinformatik* (Vol. 2018-March, pp. 435–446). Leuphana Universitat Luneburg.

IEEE CIS Task Force on Process Mining. (2019). Xes wg certificates. https://www.win.tue.nl/ieeetfpm/doku.php?id=shared:xeswg:certificates Accessed: 17 February 2019.

Irshad, H., Shafiq, B., Vaidya, J., Bashir, M. A., Shamail, S., Adam, N., ... Workflow Management, C. (2015). Preserving privacy in collaborative business process composition. In P. Lorenz, M. S. Obaidat, & P. Samarati (Eds.), *12th international conference on security and cryptography, secrypt 2015* (pp. 112–123). SciTePress.

Jorbina, K., Rozumnyi, A., Verenich, I., Di Francescomarino, C., Dumas, M., Ghidini, C., ... Raboczi, S. (2017). Nirdizati: A web-based tool for predictive process monitoring. In J. Mendling, M. Weske, R. Clariso, B. Pentland, W. van der Aalst, H. Leopold, & A. Kumar (Eds.), *2017 bpm demo track and bpm dissertation award, bpm-d and da 2017, co-located with 15th international conference on business process management, bpm 2017* (Vol. 1920). CEUR-WS.

Kitchenham, B. & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. *Engineering*, *2*, 1051. doi:10.1145/1134285.1134500

Krumeich, J., Werth, D., Loos, P., Atos, Benteler, Sap, ... et al. (2014). Interactive process discovery and improvement in people-driven business processes. In D. Kundisch, L. Beckmann, & L. Suhl (Eds.), *Multikonferenz wirtschaftsinformatik, mkwi 2014 - multi-conference on business informatics, mkwi 2014* (pp. 1185–1198). University of Paderborn.

Liu, C., Van Dongen, B., Assy, N., van der Aalst, W., & Society, I. C. I. (2017). Component behavior discovery from software execution data. In *2016 ieee symposium series on computational intelligence, ssci 2016*. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/SSCI.2016.7849947

Liu, C., Wang, S., Gao, S., Zhang, F., & Cheng, J. (2018, November). User Behavior Discovery from Low-Level Software Execution Log. *IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING*, *13*(11), 1624–1632. doi:10.1002/tee.22727

Liu, C., Zhang, J., Li, G., Gao, S., & Zeng, Q. (2018, August). A Two-Layered Framework for the Discovery of Software Behavior: A Case Study. *IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS*, *E101D*(8), 2005–2014. doi:10.1587/transinf.2017EDP7027

Maita, A. R. C., Martins, L. C., López Paz, C. R., Rafferty, L., Hung, P. C. K., Peres, S. M., & Fantinato, M. (2018). A systematic mapping study of process mining.

*Enterprise Information Systems*, *12*(5), 505–549. doi:10.1080/17517575.2017.1402371

Mannhardt, F., Petersen, S. A., & Oliveira, M. F. (2018). Privacy Challenges for Process Mining in Human-Centered Industrial Environments. *2018 14th International Conference on Intelligent Environments (IE)*, 64–71. doi:10.1109/ie.2018.00017

New York Times. (2018). React tracking. https://github.com/nytimes/react-tracking. https://github.com/nytimes/react-tracking Accessed: 20 December 2019.

NOREA. (2018). Norea guide privacy control framework. https://www.norea.nl/download/?id=4160. https://www.norea.nl/download/?id=4160 Accessed: 10 December 2019.

npm, Inc. (2019). Npmjs.com. https://npmjs.com/. https://npmjs.com/ Accessed: 20 December 2019.

OMG. (2011). Business process model and notation (bpmn). https://www.omg.org/spec/BPMN/2.0/PDF. https://www.omg.org/spec/BPMN/2.0/PDF Accessed: 15 December 2019.

Padidem, D. K. & Nalini, C. (2017). A study on classification of users shopping behavior process model using click stream data. *Journal of Engineering and Applied Sciences*, *12*(Specialissue12), 9548–9553. doi:10.3923/jeasci.2017.9548.9553

Pourmirza, S., Dijkman, R., & Grefen, P. (2017, June). Correlation Miner: Mining Business Process Models and Event Correlations Without Case Identifiers. *INTERNATIONAL JOURNAL OF COOPERATIVE INFORMATION SYSTEMS*, *26*(2, SI). doi:10.1142/S0218843017420023

Privacy Company. (2019). Privacy by design framework. https://www.privacycompany.eu/knowledge-base-nl/privacy-by-design-framework. https://www.privacycompany.eu/knowledge-base-nl/privacy-by-design-framework Accessed: 30 July 2019.

Rubin, V. A., Lomazova, I., Van Der Aalst, W. M. P., Huawei Technologies Co, L., International Software Process, A., & Nanjing, U. (2014). Agile development with software process mining. In *2014 international conference on software and systems process, icssp 2014* (pp. 70–74). Nanjing: Association for Computing Machinery. doi:10.1145/2600821.2600842

Rubin, V. A., Mitsyuk, A. A., Lomazova, I. A., & van der Aalst, W. (2014). Process mining can be applied to software too! In *8th acm/ieee international symposium on empirical software engineering and measurement, esem 2014*. IEEE Computer Society. doi:10.1145/2652524.2652583

Schönig, S., Cabanillas, C., Di Ciccio, C., Jablonski, S., & Mendling, J. (2018). Mining team compositions for collaborative work in business processes. *Software and Systems Modeling*, *17*(2), 675–693. doi:10.1007/s10270-016-0567-4

Setiawan, F. & Yahya, B. N. (2018, July). Improved behavior model based on sequential rule mining. *APPLIED SOFT COMPUTING*, *68*, 944–960. doi:10.1016/j.asoc.2018.01.035

Sirijaitham, P., Porouhan, P., Palangsantikul, P., & Premchaiswadi, W. (2018). Improving efficiency of OTT systems using fuzzy mining technique. In *15th international conference on ict and knowledge engineering, ict and ke 2017* (pp. 1–5). IEEE Computer Society. doi:10.1109/ICTKE.2017.8259632

Thiede, M., Fuerstenau, D., & Barquet, A. P. B. (2018). How is process mining technology used by organizations? A systematic literature review of empirical studies. *BUSINESS PROCESS MANAGEMENT JOURNAL*, *24*(4), 900–922. doi:10.1108/BPMJ-06-2017-0148

Tillem, G., Erkin, Z., & Lagendijk, R. L. (2017). Mining Encrypted Software Logs using Alpha Algorithm. *4*(Icete), 267–274. doi:10.5220/0006408602670274

Trischler, J. & Zehrer, A. (2012, January). Service design: Suggesting a qualitative multistep approach for analyzing and examining theme park experiences. *Journal of Vacation Marketing*, *18*(1), 57–71. doi:10.1177/1356766711430944

van der Aalst, W. (2013, October). Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior. *IEEE Transactions on Services Computing*, *6*(4), 525–535. doi:10.1109/TSC.2012.25

van der Aalst, W. (2016). *Process Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-49851-4

van der Aalst, W., Adriansyah, A., Alves De Medeiros, A., Arcieri, F., Baier, T., Blickle, T., & Bose, J. (2012). *Process mining manifesto* (F. Daniel, K. Barkaoui, & S. Dustdar, Eds.). Lecture Notes in Business Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-28115-0

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. doi:10.2307/30036540

Verbeek, E. & van der Aalst, W. (2018). Ieee 1849-2016 xes standard. http://www.xes-standard.org/ Accessed: 17 February 2019.

Wang, Y., Zacharewicz, G., Traore, M. K., & Chen, D. (2018). An integrative approach to simulation model discovery: Combining system theory, process mining and fuzzy logic. *JOURNAL OF INTELLIGENT & FUZZY SYSTEMS*, *34*(1), 477–490. doi:10.3233/JIFS-17403

Wieringa, R. (2014). *Design science methodology*. doi:10.1007/978-3-662-43839-8

Wolny, J. & Charoensuksai, N. (2014, April). Mapping customer journeys in multichannel decision-making. *Journal of Direct, Data and Digital Marketing Practice*, *15*(4), 317–326. doi:10.1057/dddmp.2014.24

Wooff, D. A. & Anderson, J. M. (2015). Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase. *Journal of Statistical Theory and Practice*, *9*(2), 227–249. doi:10.1080/15598608.2013.862753

# Appendix A

# Data Extraction Form

| Name | Source | Year | Summary | Type | Tool | Privacy | User behaviour | Process Mining | Customer Journey |
|---|---|---|---|---|---|---|---|---|---|
| S1 | Article | 2018 | Customer journey for cashback e-commerce website. With personal retention strategies | Case Study | | no | yes | no | yes |
| S2 | Article | 2018 | Variant comparison | Case Study | ProM | no | yes | yes | no |
| S3 | Article | 2018 | Method for exceptions | Case Study | Disco | no | no | yes | no |
| S4 | Review | 2018 | Overview of software process mining studies, techniques and tools (2004-2016) | Literature Review | ProM, Disco | no | no | yes | no |
| S5 | Review | 2018 | Customer journey literature review (1991-2012) | Literature Review | - | no | yes | no | yes |
| S6 | Conference | 2018 | Omni-channel, technical side | Case Study | - | no | yes | no | yes |
| S7 | Article | 2018 | the abstract user operation activity from the lower level method-call log. Part about software process mining | Case Study | ProM | no | yes | yes | no |
| S8 | Article | 2018 | behavioural models for ProM, Process mining on ProM | Case Study | ProM | no | yes | yes | no |
| S9 | Review | 2018 | Mapping study (2005-2016) | Literature Review | - | no | yes | yes | no |
| S10 | Conference | 2018 | Privacy in process mining for employees | Design | - | yes | yes | yes | no |
| S11 | Article | 2018 | Process Mining framework to discover team attributes and composition of collaborative activities in business processes | Case Study | DPILMiner | no | yes | yes | no |
| S12 | Article | 2018 | Tracks real life user behaviour with smartphones/watches. Find episodes (sets of activities). No regards for privacy. | Case Study | Association Rule Miner, Episode miner, ProM, ProA | no | yes | yes | no |
| S13 | Conference | 2018 | Some basic principles, short, but clear basic steps | Design | Disco | no | yes | yes | no |
| S14 | Review | 2018 | Process mining usage from a service perspective (1992-2016) | Literature Review | - | no | no | yes | no |
| S15 | Article | 2018 | System inference D2FD method | Design | ProM, SimStudio | no | yes | yes | no |
| S16 | Conference | 2017 | Customer journey process mining | Design | - | no | yes | yes | yes |
| S17 | Conference | 2017 | Online tool for Customer journey based on event logs | Design | CJM-ex | no | yes | yes | yes |
| S18 | Conference | 2017 | simple tool for data analysis | Design | interpretA | no | no | yes | no |
| S19 | Conference | 2017 | system use tracking | Case Study | Disco | no | yes | yes | no |
| S20 | Conference | 2017 | Customer journey based on marlov models from log files | Case Study | - | no | yes | yes | yes |
| S21 | Conference | 2017 | case study web application predictive process monitoring | Case Study | nirdizati | no | no | yes | no |
| S22 | Conference | 2017 | Software process mining, kieker framework | Design | ProM | no | no | yes | no |
| S23 | Article | 2017 | clickstream user behaviour | Case Study | ProM | no | yes | yes | no |
| S24 | Article | 2017 | Mining without identifier | Case Study | Disco | no | no | yes | no |
| S25 | Conference | 2017 | Encrypting for business and user privacy | Design | - | yes | yes | yes | no |
| S26 | Article | 2016 | Clickstream dataset analysed | Case Study | - | no | yes | no | yes |
| S27 | Book | 2016 | Book about process mining | Book | - | yes | yes | yes | yes |
| S28 | Conference | 2015 | User behaviour analysis | Case Study | Disco | no | yes | yes | no |
| S29 | Conference | 2015 | privacy, encryption of events | Case Study | ProM | yes | no | yes | no |
| S30 | Conference | 2015 | Focus on business privacy | Case Study | ProM | yes | no | yes | no |
| S31 | Article | 2015 | Data analysis of customer journey | Design | - | no | yes | no | yes |
| S32 | Conference | 2014 | Collaborative activities | Case Study | ProM | no | yes | yes | no |
| S33 | Short Paper | 2014 | short paper on customer journey | Short Paper | - | no | yes | no | yes |
| S34 | Conference | 2014 | User based recommendations | Design | - | no | yes | yes | no |
| S35 | Conference | 2014 | interesting about software tracking | Case Study | ProM, Disco | no | yes | yes | no |
| S36 | Conference | 2014 | | Case Study | Disco | no | yes | yes | no |
| S37 | Article | 2014 | data from interviews | Design | - | no | yes | no | yes |
| S38 | Article | 2013 | Service behaviour, resources | Design | - | no | yes | yes | no |
| S39 | Article | 2012 | Service design | Case Study | - | no | yes | no | yes |
| S40 | Article | 2012 | Manifesto | Design | - | yes | yes | yes | no |

# Appendix B

# List of Papers

| # | *Paper* |
|---|---|
| S1 | Ballestar et al. 'Customer segmentation in e-commerce: Applications to the cashback business model', 2018 |
| S2 | Bolt et al. 'Process variant comparison: Using event logs to detect differences in behavior and business rules', 2018 |
| S3 | Dijkman et al. 'Business processes exceptions in relation to operational performance', 2018 |
| S4 | Dong et al. 'A Mapping Study on Mining Software Process', 2018 |
| S5 | Følstad and Kvale 'Customer journeys: a systematic literature review', 2018 |
| S6 | Heuchert et al. 'An IS perspective on omni-channel management along the customer journey: Development of an entity-relationship-model and a linkage concept', 2018 |
| S7 | Liu, Zhang, et al. 'A Two-Layered Framework for the Discovery of Software Behavior: A Case Study', 2018 |
| S8 | Liu, Wang, et al. 'User Behavior Discovery from Low-Level Software Execution Log', 2018 |
| S9 | Maita et al. 'A systematic mapping study of process mining', 2018 |
| S10 | Mannhardt et al. 'Privacy Challenges for Process Mining in Human-Centered Industrial Environments', 2018 |
| S11 | Schönig et al. 'Mining team compositions for collaborative work in business processes', 2018 |
| S12 | Setiawan and Yahya 'Improved behavior model based on sequential rule mining', 2018 |
| S13 | Sirijaitham, Porouhan, Palangsantikul, and Premchaiswadi 'Improving efficiency of OTT systems using fuzzy mining technique', 2018 |
| S14 | Thiede et al. 'How is process mining technology used by organizations? A systematic literature review of empirical studies', 2018 |
| S15 | Wang, Zacharewicz, Traore, and Chen 'An integrative approach to simulation model discovery: Combining system theory, process mining and fuzzy logic', 2018 |

| # | *Paper* |
|---|---|
| S16 | Bernard and Andritsos 'A Process Mining Based Model for Customer Journey Mapping', 2017a |
| S17 | Bernard and Andritsos 'CJM-ex: Goal-oriented exploration of customer journey maps using event logs and data analytics', 2017b |
| S18 | Dixit et al. 'Enabling Interactive Process Analysis with Process Mining and Visual Analytics', 2017 |
| S19 | Gadler et al. 'Mining Logs to Model the Use of a System', 2017 |
| S20 | Harbich et al. 'Discovering Customer Journey Maps using a Mixture of Markov Models', 2017 |
| S21 | Jorbina et al. 'Nirdizati: A web-based tool for predictive process monitoring', 2017 |
| S22 | Liu et al. 'Component behavior discovery from software execution data', 2017 |
| S23 | Padidem and Nalini 'A study on classification of users shopping behavior process model using click stream data', 2017 |
| S24 | Pourmirza, Dijkman, and Grefen 'Correlation Miner: Mining Business Process Models and Event Correlations Without Case Identifiers', 2017 |
| S25 | Tillem et al. 'Mining Encrypted Software Logs using Alpha Algorithm', 2017 |
| S26 | Anderl et al. 'Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys', 2016 |
| S27 | van der Aalst *Process Mining*, 2016 |
| S28 | Astromskis, Andrea, and Mairegger 'A process mining approach to measure how users interact with software: An industrial case study', 2015 |
| S29 | Burattin et al. 'Toward an Anonymous Process Mining', 2015 |
| S30 | Irshad et al. 'Preserving privacy in collaborative business process composition', 2015 |
| S31 | Wooff and Anderson 'Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase', 2015 |
| S32 | Diamantini et al. 'A methodology for building log of collaboration processes', 2014 |
| S33 | Earley 'Usability for internal systems: What's the payoff?', 2014 |
| S34 | Krumeich et al. 'Interactive process discovery and improvement in people-driven business processes', 2014 |
| S35 | Rubin, Mitsyuk, et al. 'Process mining can be applied to software too!', 2014 |
| S36 | Rubin, Lomazova, et al. 'Agile development with software process mining', 2014 |
| S37 | Wolny and Charoensuksai 'Mapping customer journeys in multichannel decision-making', 2014 |
| S38 | van der Aalst 'Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior', 2013 |
| S39 | Trischler and Zehrer 'Service design: Suggesting a qualitative multistep approach for analyzing and examining theme park experiences', 2012 |
| S40 | van der Aalst et al. *Process mining manifesto*, 2012 |

# Appendix C

# Use Cases

## Strategy app

As a Director, I want to add a new strategy
As a Director, I want to add objectives to a strategy

## Portfolio app

As a Director, I want to add a new portfolio
As a Director, I want to create a funnel with lists
As a Director, I want to make a list in the funnel pullable for programmes
As a Director, I want to add objectives from my strategy to my portfolio

As a Portfolio Manager, I want to add programme items to the funnel
As a Portfolio Manager, I want to add labels to portfolio items
As a Portfolio Manager, I want to add attachments to portfolio items
As a Portfolio Manager, I want to add an owner to a portfolio item
As a Portfolio Manager, I want to move a portfolio item to another list

As a Portfolio Manager, I want to start an Agile team from a portfolio item
As a Portfolio Manager, I want to add members to the portfolio
As a Portfolio Manager, I want to add a programme to a portfolio that can pull portfolio items

## Programme app

As a Programme Manager, I want to pull portfolio items to my programme
As a Programme Manager, I want to add lists to my programme
As a Programme Manager, I want to add programme items to my programme
As a Programme Manager, I want to link programme items to a pulled portfolio item
As a Programme Manager, I want to add labels to my programme items
As a Programme Manager, I want to make a list in the programme pullable for teams
As a Programme Manager, I want to add teams to a programme that can pull programme items

Appendix D

# Tracking Code

## D.1 App Class

**Listing D.1:** App Class (App.js)

```
import React, { Fragment, Component } from 'react';
import { dispatch as trackDispatch, track } from 'factories/tracking';

@track({}, { dispatch: data => trackDispatch(data) })
class App extends Component {
    // App Code
}

export default App;
```

## D.2 Tab Element

**Listing D.2:** Tab Element (elements/Tab.js)

```
import React, { Fragment } from 'react';
import track from 'react-tracking';

@track({ element: 'tab' })
class TabElement extends Element {

  @track(() => ({ tab: true, active: currentNode.get('activeTab') }))
  componentDidMount() {
      // Code
  }

  @track(() => ({ action: `change_tab_${currentNode.get('activeTab')}` }))
  handleTabChange = () => {
        // Code
  }
}
export default TabElement;
```

## D.3 Column Cardboard Element

**Listing D.3:** Column Cardboard Element (elements/ColumnCardBoard.js)

```
import React from 'react';
import track from 'react-tracking';
import Element from 'mvc/react/Element';

@track({ element: 'Board' })
class ColumnCardBoard extends Element {

  @track({ action: 'add_multiple_cards' })
  createCards = (columnObject, titles, lastCard) => {
      // Code
  }

  @track((props, state, [cardModel]) => (
              { action: `click_card_${cardModel.getId()}` }
              ))
  onClickCard = cardModel => {
      // Code
  }

  @track({ action: 'change_card_order' })
  onCardOrderChange = (order, sortable, event) => {
      // Code
  }

  @track({ action: 'add_column' })
  createColumn = columnName => {
      // Code
  }

  @track({ action: 'change_column_order' })
  onColumnOrderChange = (order, sortable, evt) => {
      // Code
  }
}

export default ColumnCardBoard;
```

## D.4 Client-Side Tracking Factory

**Listing D.4:** Client-side React tracking factory (factories/tracking/index.js)

```
import track from 'react-tracking';
import sha256 from 'tiny-hashes/sha256';
import currentUser from 'mvc/models/currentUser';

const INTERVAL = 60 * 1000; // interval in ms.

let activeTab;
let activePage;

/**
 * Custom REST call function. It is not needed to handle the response here.
 * @param url the tracking url to the PTB
 * @param String Stringified JSON in the form of {data: data};
 * @private
 */
function _restCall(url, data) {
  const headers = new Headers();
  headers.set('Content-Type', 'application/json');
  let fetchOptions = {
    method: 'POST',
    headers: headers,
    credentials: 'same-origin',
    body: data
  };
  //Just sending data, response is not of interest.
  fetch(url, fetchOptions);
}

/**
 * Upload function. Uses beacons by preference,
 * but has a fallback to REST call.
 * @private
 */
function _uploadData() {
  // Tracking data is stored on the window.dataLayer.
  // Check if tracking data is actually available.
  let success = false;
  if (window.dataLayer && window.dataLayer.length > 0) {
```

```
    // Tenant Id and token are added to url (beacons do not support headers).
    let url = `/app/tracking/track?tenantId=${currentUser.getTenantId()}` +
              `&token=${currentUser.getToken()}`;
    // Data is written to data.data for easy use server-side
    let data = JSON.stringify({
      data: window.dataLayer.splice(0, window.dataLayer.length)
                            .filter(obj => obj.timestamp !== undefined)
    });
    // IE11 has no sendBeacon functionality
    if (navigator.sendBeacon) {
      try {
        // Chrome temporary does not send beacons with application/json.
        // See http://crbug.com/490015
        success = navigator.sendBeacon(url,
                        new Blob([data], { type: 'application/json' }));
      } catch (err) {
        // Fall back to a REST call.
      }
    }
    if (!success) {
      _restCall(url, data);
    }
  }
}

/**
 * Handles the data.page attribute.
 * If there are tabs, The active tab is added.In modals, '_modal' is added.
 * Checks if 'change_tab' event is fired multiple times and filters this.
 * On tab mount, active tab is set and this event is filtered.
 * @param data the data object
 * @returns {{page}|null}
 * @private
 */
function _handlePage(data) {
  // Keep track of the active Page, if page changes, reset the active tab.
  if (data.page && data.page !== activePage) {
    activePage = data.page;
    activeTab = undefined;
  }
  // On tab mount, the active tab is set.
  if (data.tab) {
```

```
    activeTab = data.active;
    return null;
  }
  // Tab name is added to the page.
  data.page = activeTab ? activePage + '_' + activeTab : activePage;
  // If in a modal, take the original page and active tab.
  if (data.modal) {
    data.page += '_modal';
    return data;
  }
  // When changing tab, set the tab to the active tab.
  if (data.action && data.action.match(/^change_tab/g)) {
    let newActive = data.action.split('_').pop();

    // The handleTabChange is sometimes fired multiple times.
    // If the user changes to the same tab it should not be logged.
    if (newActive === activeTab) {
      return null;
    }
    activeTab = newActive;
  }
  return data;
}

function init() {
  // Tracking data is send when the user: closes the browser,
  // refreshes the page or navigates away from the page.
  window.addEventListener('beforeunload', () => {
    _uploadData();
  });
  // Tracking data is send in a set interval;
  setInterval(_uploadData, INTERVAL);
}

function dispatch(data) {
  if (Object.keys(data).length > 0) {
    data = _handlePage(data);
    if (data) {
      data.user = sha256(currentUser.getToken());
      data.app = currentUser.getAppName();
      data.timestamp = new Date().toISOString();
      (window.dataLayer = window.dataLayer || []).push(data);
```

```
    }
  }
}

/**
 * The logout action is executed before unload event is triggered.'
 * Therefore, the token expires and the upload function will fail.
 * Therefore, data is uploaded right before the user logs out.
 */
function logout() {
  _uploadData();
  window.removeEventListener('beforeunload', () => {
    _uploadData();
  });
}

export { init, dispatch, logout, track };
```

## D.5 Server-Side Tracking Handler

**Listing D.5:** Server-side tracking handler (handlers/Tracking.js)

```
const serverConfig = require('../../config');

let Promise = require('bluebird');
let request = require('request');
let CryptoJS = require('crypto-js');

let logger = require('log4js').getLogger('Tracking');

let configCorrect;

// Name of the Custom log workspace.
const LOG_TYPE = 'FCCLogs';
// Indicates which field in the data is used as timestamp.
// If not set, Azure will use the time the request was send.
const TIMESTAMP_FIELD = 'timestamp';
const METHOD = 'POST';
const RESOURCE = '/api/logs';
const API_VERSION = '2016-04-01';
const CONTENT_TYPE = 'application/json';
const ERROR_NO_DATA = 'Error, no tracking data provided. ' +
                      'make sure tracking data is in body.data';
const ERROR_AZURE = 'Sending logging data to Azure failed: ';
const USAGE =
    'usage: \n' +
    '{\n' +
    '\t"tracking": {\n' +
    '\t\t"workspaceId": "workspace or id_from_azure",\n' +
    '\t\t"sharedKey": "shared_key_from_azure"\n' +
    '\t}\n' +
    '}';

/**
 *
 * @type {module.tracking}
 */
module.exports = class tracking {
  track(data) {
    if (configCorrect == undefined) {
```

```
      if (!serverConfig.tracking || !serverConfig.tracking.workspaceId ||
          !serverConfig.tracking.sharedKey) {
        logger.warn(
             'Tracking is enabled client-side but not server-side. ' +
               'Data will not be saved. Add tracking setting.',
              USAGE
        );
        configCorrect = false;
      } else {
        configCorrect = true;
      }
    }
    if (!configCorrect) {
      return Promise.resolve();
    }
    // If there is no data, do not send data to azure and show a warning,
    // because this indicates that the request was or wrong or not needed.
    if (data && data.data) {
      // Date in UTC format. Set on the custom header x-ms-data.
      let processingDate = new Date().toUTCString();
      // Azure accepts JSON format.
      let body = JSON.stringify(data.data);
      let contentLength = Buffer.byteLength(body, 'utf8');
      // Azure needs a signed authorization key as header.
      let stringToSign = `${METHOD}\n${contentLength}\n${CONTENT_TYPE}` +
                               `\nx-ms-date:${processingDate}\n${RESOURCE}`;
        let hash = CryptoJS.HmacSHA256(stringToSign,
              CryptoJS.enc.Base64.parse(serverConfig.tracking.sharedKey));
      let signature = CryptoJS.enc.Base64.stringify(hash);
      request.post(
          {
            url: `https://${
                serverConfig.tracking.workspaceId
                }.ods.opinsights.azure.com${RESOURCE}` +
                `?api-version=${API_VERSION}`,
            headers: {
              'Content-Type': CONTENT_TYPE,
              Authorization: 'SharedKey ' +
                      serverConfig.tracking.workspaceId +
                      ':' + signature,
              'Log-Type': LOG_TYPE,
              'x-ms-date': processingDate,
```

```
              'time-generated-field': TIMESTAMP_FIELD
            },
            body: body
          },
          error => {
            if (error) {
              logger.error(ERROR_AZURE, error);
            }
          }
      );
    } else {
      logger.warn('No tracking data', ERROR_NO_DATA);
    }
    // Always resolve. We do not bother the client if something went wrong.
    return Promise.resolve({});
  }
};
```

# Appendix E

# Scenarios

## Tracking user behaviour

You are invited to try out the new functionality of Fortes Solutions. This script contains three different scenarios played by three different roles: The SEO, a portfolio manager and a programme manager. The scripts below follow a simplified use case of the application. Try to follow all the steps and make sure that you are logged in with the right role. You are free to click around in the application as you like, for example, to look at the examples for strategies, portfolios or programs that are already available.

You are free to use your own computer. Make sure that you use either **Firefox** or **Chrome** and that no add blocker addon or other script block addon is enabled.

During this experiment, your actions in the application will be logged and stored. These will be used for analysis of the behaviour of the software. No personal data or content data will be collected.

## Scenario

Software solutions is a company that creates a wide range of software products. The company has multiple departments that focus on different products. **Alice**, the SEO, wants to work out a new product that creates software based on Artificial Intelligence (AI). **Bob**, the portfolio manager, is asked to manage the portfolio. **Carol** oversees the Agile Development Program and is responsible for the execution of this program.

You will play each role (Alice, Bob and Carol) one by one. Each role is dependent on the previous role so make sure you don't miss a step. The scripts below describe the minimal steps needed. The content you add is not of importance. However, you are encouraged to come up with meaningful items.

If something is unclear, try first to find a solution for yourself. In the case you get stuck or if something is not working correctly, don't hesitate to ask for help.

# SEO (Alice)

**Step 1 - Login**
Go to https://trackingdemo.fortes.nl/strategy.
login with username: **alice**, password: ****.

**Step 2 - Add a new strategy**
Name: **AI software**.
Manager: **Alice A**.

**Step 3 - Add objectives in the newly created strategy**
Add objectives to financial and to customer.
Add more objective to your liking.
Make all items in the financial category blue.
Drag some objectives around and change the size of the cards.

**Step 4 - Portfolio app**
Go to https://trackingdemo.fortes.nl/portfolio.

**Step 5 - Add a new portfolio**
Name: **AI Portfolio**.
OU: **Portfolio Management**.
Portfolio model: **Portfoliomodel**.
Portfolio type: **Hybrid**.
Managers: **Alice A** and **Bob b**.

**Step 6 - Create the funnel tab**
Go to the newly created portfolio.
On the funnel tab, add three lists: **TODO**, **DOING**, **DONE**.
Make the DOING list pullable.
Add some objective from strategy at your liking.

**Step 7 - Log out**
Log out.

# Portfolio Manager (Bob)

**Step 1 - Login**
Go to https://trackingdemo.fortes.nl/porfolio.
login with username: **bob**, password: ****.

**Step 2 - Portfolio app**
Open AI Portfolio.
On the funnel tab, add multiple cards in TODO.
Click on Manage Labels and create a new label.
Add labels to cards.
Add an attachment to a card, either a link or a file.
Set yourself as an owner of a card.
Move some cards to DOING.

**Step 3 - Create an Agile Team**
Stay on the funnel tab from the previous step.
Create a new card named **Agile AI Team** and start an agile team for that card.
Name: **Agile AI Team**.
Organisational Unit: **Project Management**.
Folder: **Futuristic Projects**.
Managers: **Carol C**.
Check that the agile team is linked and go back to the funnel tab.

**Step 4 - Add members**
In the AI Portfolio, click on the members button.
Add **Carol** as a **Reader**.

**Step 5 - Add agile programme**
On the funnel tab, add **Agile Developers** to programs that can pull portfolio items.

**Step 6 - Log out**
Log out.

# Programme Manager (Carol)

**Step 1 - Login**
Go to https://trackingdemo.fortes.nl/programme.
login with username: **carol**, password: ****.

**Step 2 - Programme app**
Open Agile Developers.
Pull one of the portfolio Items.
Add some lists (e.g. **TODO**, **DOING**, **DONE**) on the splitting portfolio items tab.

**Step 3 - KANBAN**
Go to the kanban tab.
Add some program items and move them around.
Link one card to a portfolio item.
Create some labels to your liking.
Add the created labels to cards.

**Step 4 - Teams**
On the kanban tab, make one list pullable for teams.
Add the artificial intelligence team to teams that can pull program items.

**Step 5 - Log out**
Log out.

# Appendix F

# Grafana, RapidMiner and Celonis

## F.1 Dashboard Kusto queries

**Listing F.1:** Example of average session time query

```
let Result = materialize(FCCLogs_CL| where app_s in ($app_name) and
    $__timeFilter() | summarize min(TimeGenerated), datetime_diff('minute',
    max(TimeGenerated),min(TimeGenerated)) by app_s, user_s);
Result | make-series average=avg(Column1)  default=0 on min_TimeGenerated
    from bin($__timeFrom(),1d) to bin($__timeTo(),1d) step 1d by app_s
| mv-expand min_TimeGenerated to typeof(datetime),average to typeof(double)
```

**Listing F.2:** Example of actions performed per app query

```
FCCLogs_CL
| where app_s in (${app_name}) and page_s in (${page}) and element_s !in (""
    ,"browser") and  $__timeFilter()
| summarize count() by page_s, strcat(element_s,"_",action_s)
| sort by count_ desc
```

**Listing F.3:** Example of sessions that use functions query

```
FCCLogs_CL
| where app_s == "fc-waterfall-portfolio" and strcat(element_s,"_",action_s)
     in ($action) and $__timeFilter()
| make-series Users=dcount(user_s) default=0 on TimeGenerated from bin(
    $__timeFrom(),1d) to bin($__timeTo() +1d ,1d) step 1d by strcat(element_s
    ,"_",action_s)
| mv-expand TimeGenerated to typeof(datetime),Users to typeof(int)
```

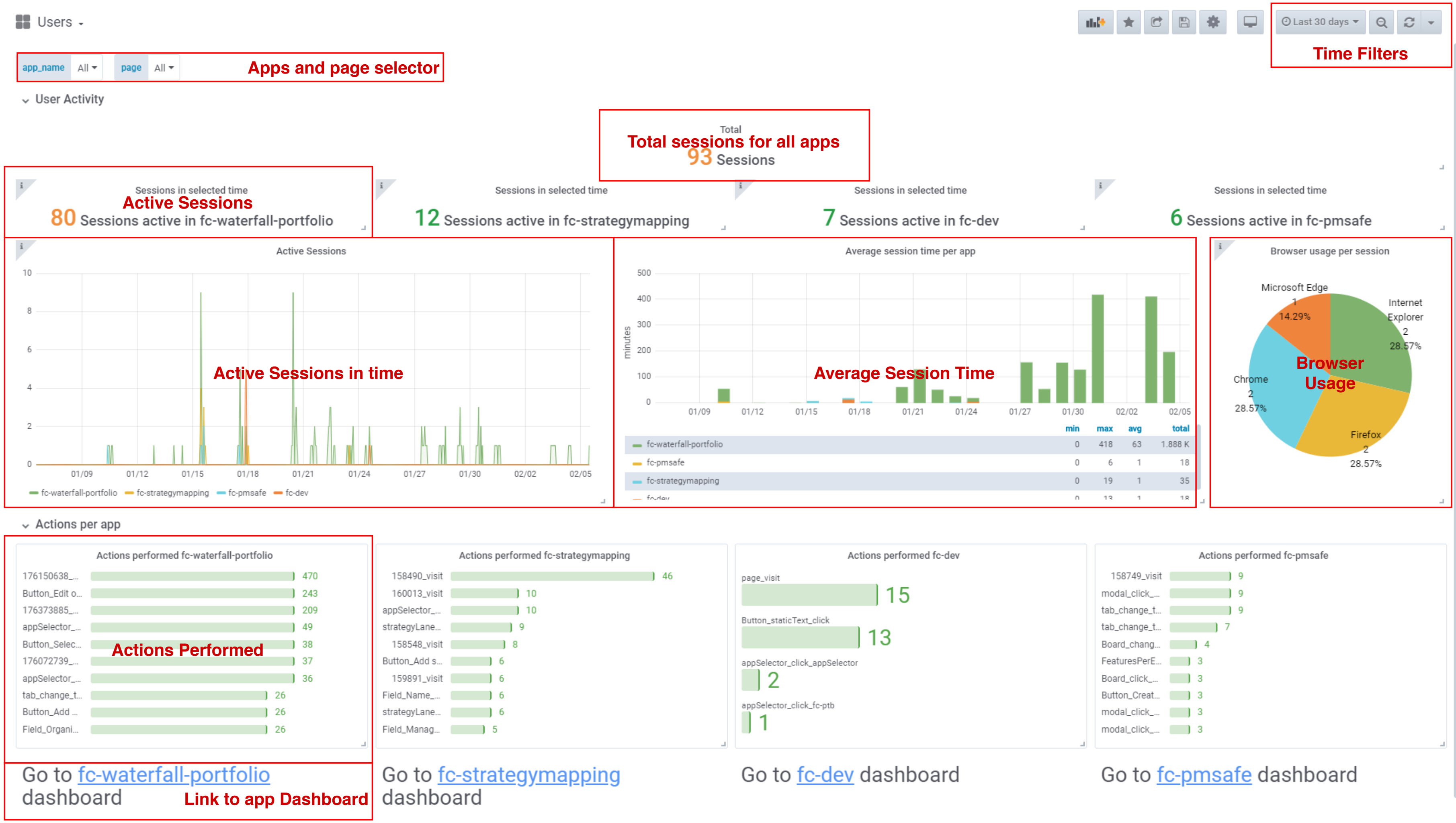



**Figure F.1:** Grafana - Overview Dashboard.


**Figure F.2:** Grafana - App Dashboard.

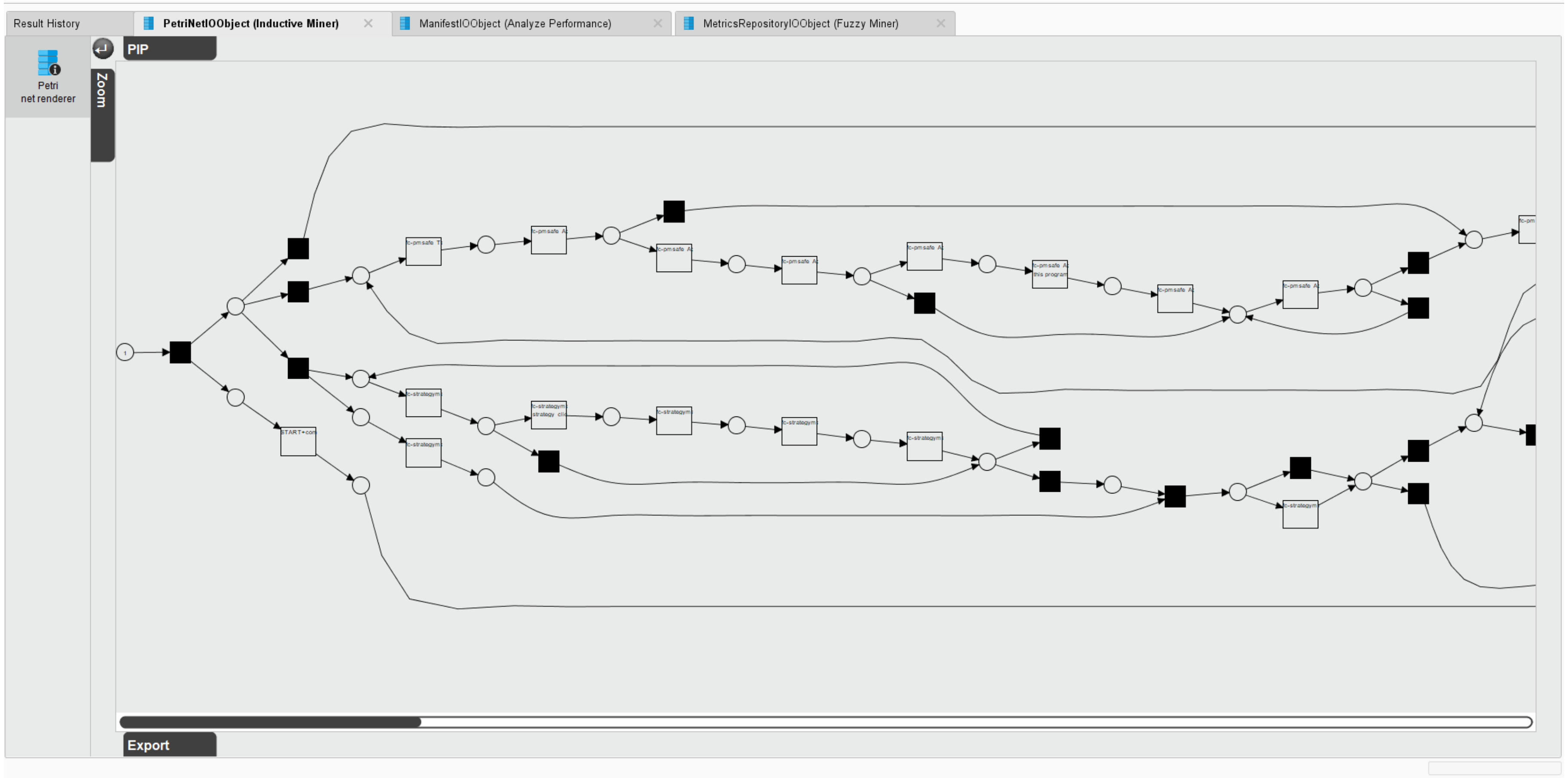



**Figure F.3:** Rapidminer - Inductive Miner results.

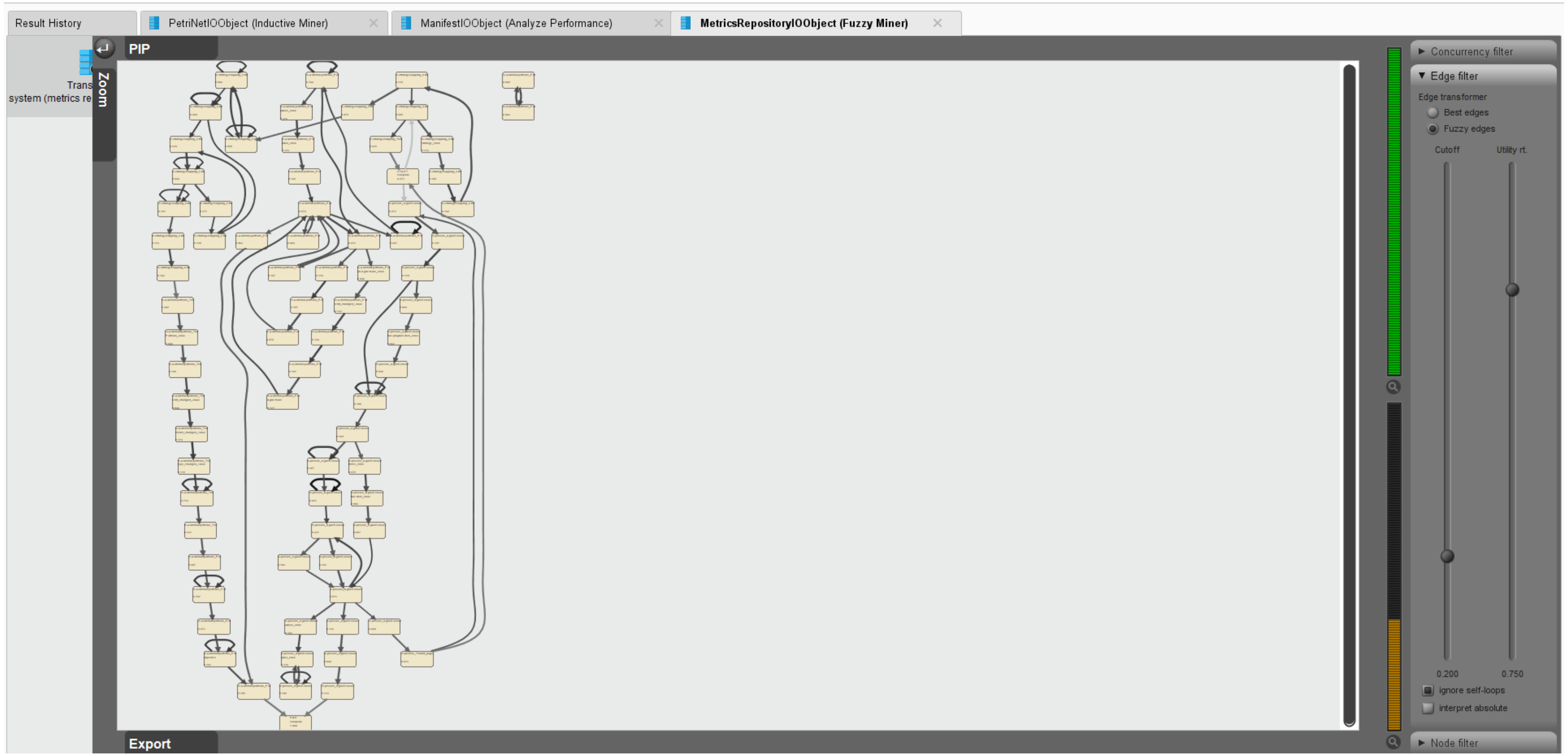



**Figure F.4:** Rapidminer - Fuzzy Miner Results.

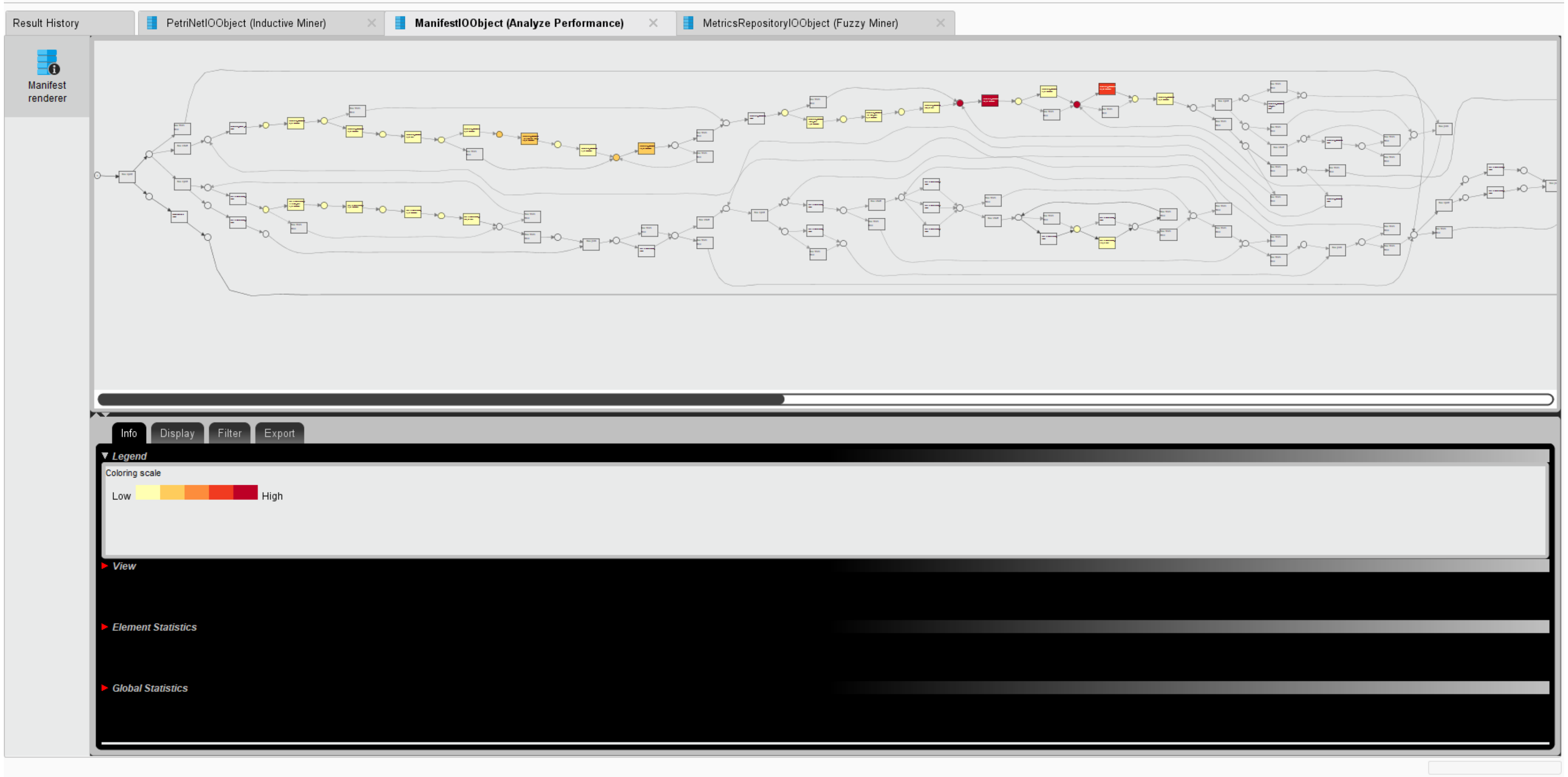



**Figure F.5:** Rapidminer - Performance Analysis.

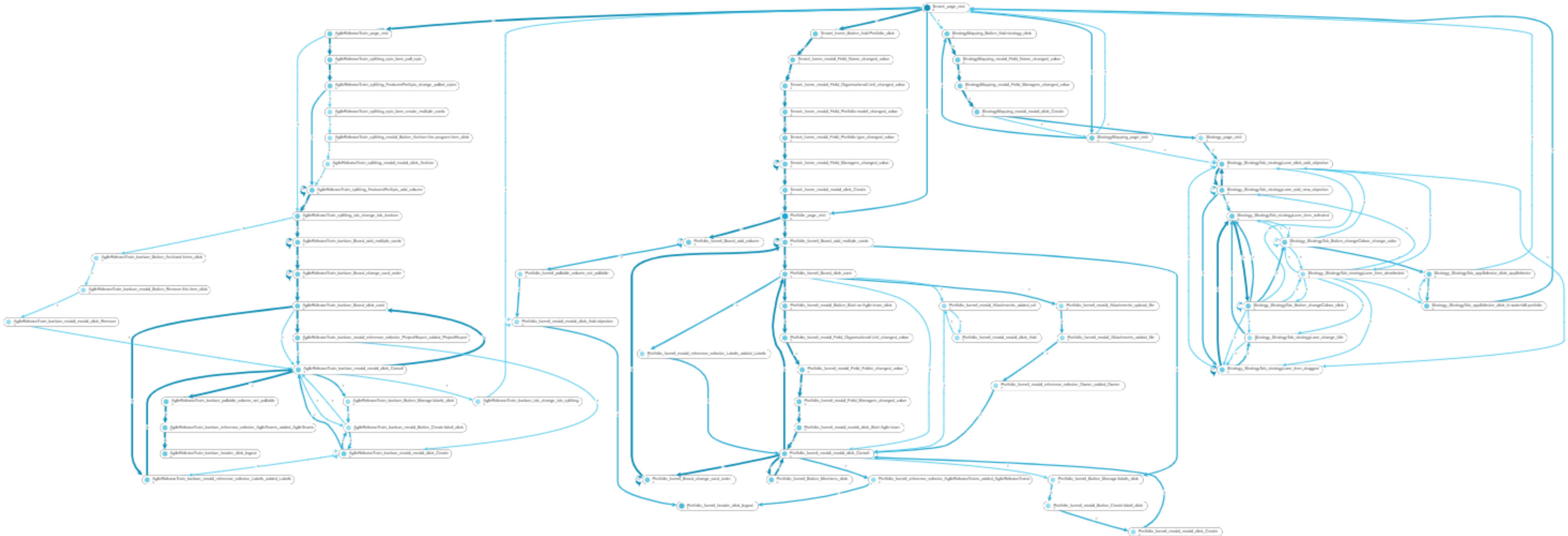


**Figure F.6:** Celonis - process overview.


**Figure F.7:** Celonis - Portfolio process.

# Appendix G

# Questionnaire

The following questions were used to validate the method. The questions are based on the items used in estimating UTAUT (Venkatesh et al., 2003) and adapted to validate the method.

## Demographic questions

*Which option fits your job description best?*

- Product Management
- Developer
- Hosting
- Data Analyst
- Tester
- Other...

*What is your age?*

- Under 19 years old
- 20-29 years old
- 30-39 years old
- 40-49 years old
- 50-59 years old
- 60 years or older

# UTAUT questions

The following questions are grouped per determinant of the UTAUT. A five point Likert scale from strongly disagree to strongly agree.

*Performance expectancy*

**Q1** I would find the method useful in my job.
**Q2** Using the method enables me to accomplish tasks more quickly.
**Q3** Using the method increases my productivity.
**Q4** If I use the method, I will increase my chances of getting a raise.

*Effort expectancy*

**Q5** My interaction with the method would be clear and understandable.
**Q6** It would be easy for me to become skilful at using the method.
**Q7** I would find the method easy to use.
**Q8** Learning to operate the method is easy for me.
**Q9** The rewards for implementing the method is worth the time.

*Social influence*

**Q10** People who influence my behaviour think that I should use the method.
**Q11** People who are important to me think that I should use the method.
**Q12** The senior management of this business has been helpful in the use of the method.
**Q13** In general, the organisation has supported the use of the method.

*Facilitating conditions*

**Q14** I have the resources necessary to use the method.
**Q15** I have the knowledge necessary to use the method.
**Q16** The method is compatible with other systems I use.
**Q17** A specific person (or group) is available for assistance with method difficulties.

*Behavioural intention to use the method*

**Q18** I intend to use the method in the next 12 months.
**Q19** I predict I would use the method in the next 12 months.
**Q20** I plan to use the method in the next 12 months.

*Privacy*

**Q21** The privacy of users of the software is guaranteed.
**Q22** The privacy for businesses of the software is guaranteed

*Scalable*

**Q23** I think that the method can be easily extended.
**Q24** The method could be used at small software companies.
**Q25** The method could be used at large software companies.

*Final Remarks*

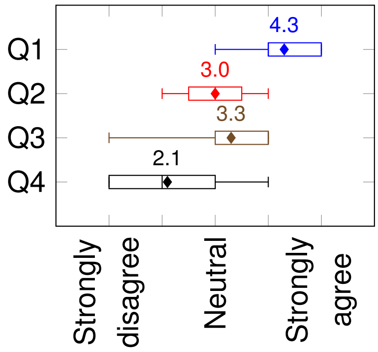
- Open Question

# Appendix H

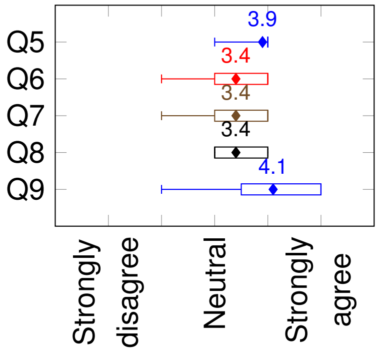
# Questionnaire Results

Questionnaire results grouped per determinant of the UTAUT. The corresponding questions can be found in Appendix G. The number above each boxplot represents the average value.

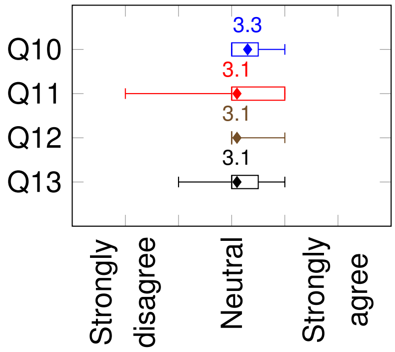
Performance expectancy



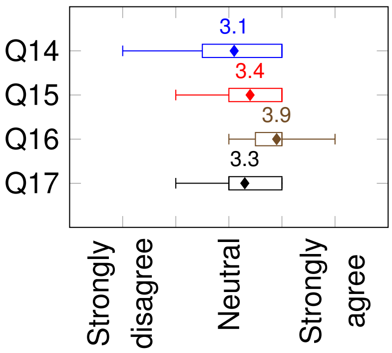

Effort expectancy



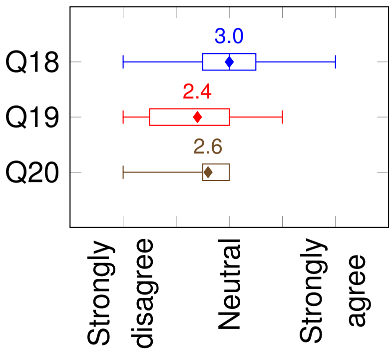

Social influence



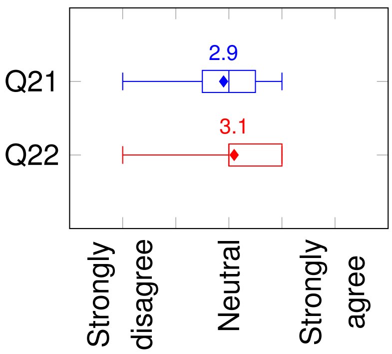

Facilitating conditions




Behavioural intention to use the system




Privacy




Scalable