MINING E-MAIL CONVERSATIONS TO ENRICH EVENT LOGS: AN EXPLORATORY CASE STUDY OF A HIRING PROCESS IN A NORWEGIAN MUNICIPALITY

Goday-Verdaguer, Aina, Department of Technology Management, SINTEF Digital, Trondheim, Norway, aina.goday@sintef.no

Mannhardt, Felix, Department of Technology Management, SINTEF Digital, Trondheim, Norway, felix.mannhardt@sintef.no

Torvatn, Hans, Department of Technology Management, SINTEF Digital, Trondheim, Norway, hans.torvatn@sintef.no

Abstract

Process improvement is an important challenge for the public sector, which struggles with reduced budgets and raised expectations on service quality. Process mining is a set of techniques that use data from information systems to understand how business processes were actually performed. It has been used widely to help with the process improvement challenge by identifying improvement and automation potential based on facts. We analysed the hiring process at Røros Kommune, a Norwegian municipality which hires about 150 new employees each year, by using interview-based process mapping as well as process mining. Our goal was to identify whether there is potential for automation as well as to explore the use of recorded event data for process analysis, which was not yet done beforehand. We used both structured data extracted from two existing information systems as well as unstructured data from e-mail communication between employees. The hiring process is comprehensive consisting of many steps, involving many responsibilities, and making use of several information systems. Many manual activities give raise to a large potential for mistakes. We found that data from the information systems was not always reliable and, therefore, developed a novel data correlation tool for connecting relevant e-mails to process activities. The results show the complexity of the hiring process as well as lacking support for several parts of the process through the used information systems. Including the e-mail communication showed promising results to improve understanding; however, many challenges remain given the complex and ambiguous relation between e-mails and process activities. Our case is unlikely to be unique, for instance a lot of Norwegian municipalities use the same information systems for similar tasks, and the low data quality can likely be replicated there. Thus, any automation or artificial intelligence project likely requires much work on improving data quality.

Keywords: Process Mining, Public Sector, Case Study, E-mail Mining.

1 INTRODUCTION

Organisations are under constant pressure to provide a better and more efficient service by improving their business processes. Public administration organisations such as municipalities are not an exception to this. Taxpayers demand better services while municipalities should spend less of their money. Automation of routine steps without added value or reorganisation of processes to avoid bottlenecks and duplicate work are examples for process improvement that can improve service while saving resources (Reijers & Mansar, 2005). Recently, the paradigm of Robotic Process Automation (RPA) has been promising the automation of routine steps without large upfront investments into redesign of information systems or processes. RPA works by means of scripted or learned agents that interact with the existing systems (Syed et al., 2020). However, successful automation requires knowledge about which processes to automate and continuous monitoring to safeguard the benefits (Geyer-Klingeberg et al., 2018).

As described in (Geyer-Klingeberg et al., 2018) process mining can be well integrated into several stages of an RPA initiative: from discovery of automatable processes parts to the holistic monitoring of RPA agents deployed into a process. Process mining is a family of techniques that analyse the actual process execution by leveraging timestamped events that give evidence when certain activities were performed for an instance of the process (process case). Each event needs to be associated with a timestamp, a process case, and an activity label to be useful for process mining methods (Van Der Aalst, 2016). The main goal of process mining is to discover a truthful as-is model of the process under consideration. This as-is model can then be analysed for issues or automation potential and conformance with regulations or the desired to-be process model can be checked (Carmona et al., 2018). However, process mining initiatives often encounter various data quality issues that affect the truthful analysis of a process due to, e.g., unreliable timestamps that do not reflect the actual time of performing an activity

In this paper, we describe an analysis of the hiring process at Røros municipality in preparation for process automation by using a combination of process mining and traditional process mapping through interviews. The analysis was conducted as part of a small project with the goal to identify whether there is potential for automation in the hiring process. The data used for the analysis consisted of an event log generated with structured data extracted from the information systems supporting the process. However, we encountered the already mentioned data quality problem when only considering an initial event log with data from the information systems. To bridge this gap, we enriched the initial event log with events that were automatically discovered from e-mail conversations between the employees responsible for certain activities in the hiring process. The main contribution of this paper is the enrichment of event logs with information from e-mails, a discussion of the used e-mail matching method, and a detailed description of the process mining application, which is likely to be repeatable in other public administration organisations.

The paper is structured as follows. In Section 2 we give some background on the hiring process and the municipality. In Section 3 we describe the method used in our analysis. We then describe and discuss the results in Section 4, review related work in Section 5 and conclude the paper in Section 6.

2 BACKGROUND

Røros municipality has about 550-600 employees, which corresponds to ca. 430 full-time equivalent positions. Services offered by the municipality include the operation of schools and child-care facilities (Barnehage), care facilities, culture facilities, fire services and many more. A large number of employees also generate a need for vacation substitutes (ferievikar) and temporary replacements in case of illness, which need to be employed in addition to the regular hiring. In total, the municipality hires about 150 new employees each year.

The hiring process is comprehensive, consists of many steps and involves many responsibilities. Different information systems are used in the process and a lot of information needs to be collected, shared, and coordinated between the different units. Today, the process is carried out with many manual operations and a large potential for mistakes. Many systems (about 10-15) need to be coordinated when giving a new employee access to systems, order equipment, and communications. There is a great need for looking at this process and see potential for development and improvement to reduce the number of

manual operations. This would reduce the number of mistakes and increase the efficiency of the process considerably.

The three main information systems that support the hiring process are: a) the recruitment, b) the archive and c) the e-mail systems. The recruitment system is used by Human Resources (HR) internally to register all the actions that only they perform related to a hiring case. These actions are: job announcement, application deadline, sending a preliminary answer to the applicant, receiving or sending rejections, sending offers and employment contracts, and arranging interviews. A new case is created in the recruitment system when the work position is publicly announced. The second support system, the archive, is where all the actions performed by the several stakeholders -candidates, unit leaders, HR and IT- are registered. Among others, some of the actions registered in the archive system are: job analysis, job announcement, candidate application, candidate answer, make the candidates list public, etc. There is no specific time when the case is created in the archive system, but it happens sometime after creating the case in the recruitment system. Last, e-mails are used to communicate among stakeholders. Not only files containing job descriptions, job offers and signed contracts are sent over e-mail between HR, unit leaders and candidates, but also doubts and requests regarding user credentials, access and equipment are sent to the IT department through e-mail.

3 METHODS

To get a comprehensive vision of a process, there are three points of view that need to be taken into account: 1) rules and procedures, 2) direct observation of the process by people involved in it, and 3) what does the data recorded in the information systems tell about the process. Taking into account these perspectives, the starting point for this project with Røros municipality concerned the rules & procedures and direct observation. We began by carrying out a planning workshop with manual process mapping to identify relevant data sources and the extent of the existing process documentation. Thereafter, we proceeded with the analysis of data extracted from the systems by performing two analysis iterations of a typical process mining project as suggested by the PM^2 methodology (van Eck et al., 2015) each one consisting of the stages data extraction, data processing, data analysis and evaluation. We describe the methods followed for each of the steps in this section.

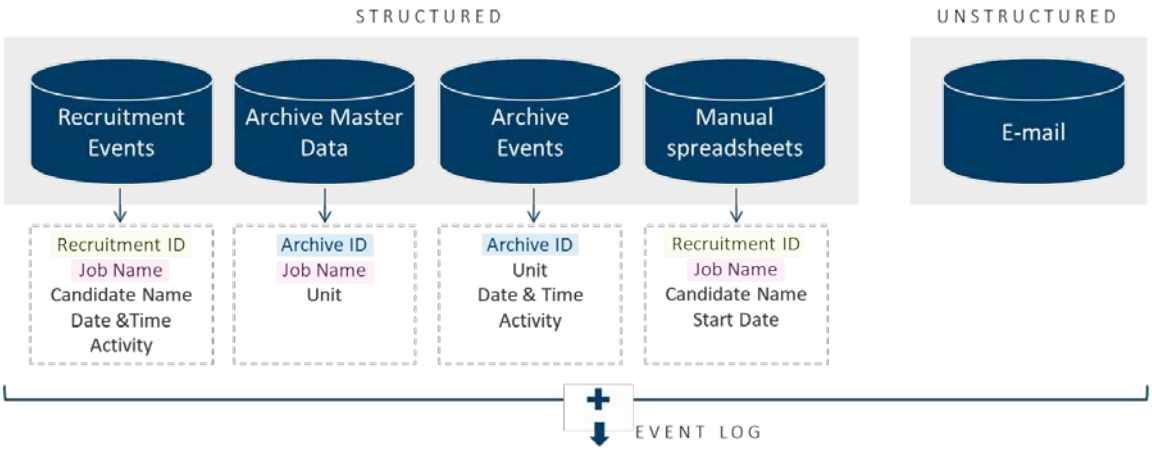
3.1 Planning and Process Mapping

The first task of the project was to carry out an initial manual process mapping and identification of data sources with the HR and IT departments through interviews. The process stakeholders provided a comprehensive list of all the activities involved in the process together with the respective resources, IT systems, as well as their inputs and outputs. Based on this initial list, a process model that reflected the process from the viewpoint of both the HR and IT department was collaboratively designed using BPMN as notation. This process model was then used to identify the sources for collecting the necessary data as well as engaging stakeholders in discussions.

3.2 Data Extraction

The data used to analyse the hiring process at Røros municipality is extracted from the information systems mentioned in Section 2. Despite Røros municipality being the owner of the data in the information systems, it was time-consuming and not straight-forward for them to gain access and extract it. The systems did not allow to export event data in a suitable machine-readable format. This mean that the necessary data needed to be manually requested through the provider of the software.

The information systems we used comprise structured and unstructured data. While the recruitment and archive systems, together with manually collected data in spreadsheets contain structured data, communication through e-mail consists of unstructured information. It is interesting to not only examine structured data from IT systems, but also unstructured corporate data like e-mails, as these may contain significant valuable business information (Jlailaty et al., 2016). Therefore, by bringing together all this data, it is possible to create an event log to comprehensively analyse the hiring process (Figure 1). However, merging the data from the different sources is a laborious job.



Case ID	Activity	Source	Date & Time	Unit	Job Name	Employee	Job Keyword
123	Announcement	Structured	...	Health	Physiotherapist 80% substitute	-	Physiotherapist
123	Position analysis	Unstructured	...	...	...	-	
355	Announcement	...	...	Culture	...	-	
...	...	...	...	...	...	-	
...	...	...	...	...	...	-	
123	Job offer	...	...	...		John Doe	
123	Access	...	...	...	...	John Doe	

Figure 1. Illustration of the sources systems from which data was collected and how the initial pre-processing pipeline matches the events to process cases.

3.3 Data Processing

3.3.1 Structured Data from Information Systems

To begin with, we join the sources of structured data (see Figure 1). On the one hand, there is the Recruitment Events data set and, on the other hand, the archive data which is composed of two data sets: a) the Archive Master Data and b) the Archive Events. While the Archive Events data set contains all the actions performed by the different stakeholders for each case, the Archive Master Data is a list of all the offered jobs. It is therefore convenient to join these two data sets by the archive ID in order to create one archive data set containing all the activities and their job names. Once the archive data is merged, the resulting set needs to be joined with the recruitment data. However, this integration is not so straightforward. Due to different providers offering different information systems, it was not possible to get a mapping of the cases specifying what archive ID belongs to what recruitment ID. Thus, joining these two data sets is done with a regular expression join, matching the job name of the archive data with the job name of the Recruitment Events. Afterwards, a time check is done since several work positions with the same job name can be offered at different points in time. This time check ensures that, for each case, there is at most a 15-day difference between the first event in the recruitment data and the first event in the archive data. The exact parameter value was determined by manually checking the accuracy of the matching until a satisfactory result was obtained. Finally, the manually collected data in spreadsheets is integrated to the resulting event log by matching both through the recruitment ID.

3.3.2 Unstructured Data from E-mail Conversations

Next, we enrich the obtained event log with activities discovered in e-mails by means of e-mail mining. The objective of mining HR e-mails is to extract the timestamps of the activities when the job analysis (stillingsanalyse) was sent by the unit leaders to HR and when HR approved the analysis (godkjent). We now describe the method followed for this data extraction.

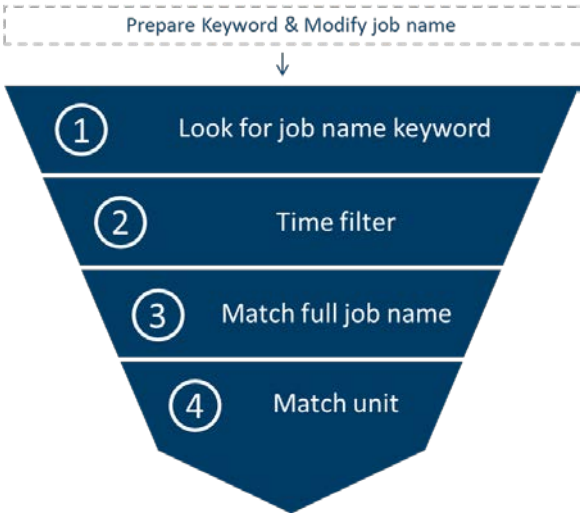


Figure 2. Main funnel of the employed e-mail matching method.

We first correlate process cases with e-mails and then discover the timestamps of activities of our interest for each case. However, before starting the e-mail matching, we prepare the grounds for such data joins by a) determining a job name keyword (e.g., "physiotherapist" for case 123 in Figure 1) and b) by modifying the full job name (remove stop words and signs, except for %). Both will be used in the matching method discussed below.

The matching method follows a funnel approach (Figure 2) where the number of cases matched to an e-mail is reduced in each step. This is done based on the assumption that one e-mail should only belong to one case, but one case can have more than one e-mail. The steps of the method are as follows. We begin by performing a first match between e-mails and cases by means of finding the job name keyword in the e-mails' content. Then, we apply a time filter to ensure time consistence. This filter only allows for matches where the e-mail was received at most 30 days earlier or 14 days later than the first timestamp of the respective case recorded in the event log. Next, we compute (as a percentage) how much of the modified full job name is found in the e-mail content (considering heading, body and foot). The case with the best matching percentage is kept and the others are removed. However, it can still be that two cases have the same matching percentage for one e-mail. Lastly, we check the organisational unit name. The case(s) for which the unit name is found within the e-mail content is(are) matched to the e-mail, regardless of the assumption of 1-e-mail-1-case. Thus, this can result in one e-mail paired with one or more cases. Cases for which the organisational unit is not found in the e-mail content are dismissed and the process is finished.

Figure 3 shows an example of the first three steps of the matching method. Cases 270 and 225 are matched to e-mail 7769 because the job keywords of both cases (highlighted in blue) are found in the e-mail and within the time limits specified above. Next, the job matching percentage is computed, resulting in a 100% match for case 270 since all the words of the modified job name are found in the e-mail (underlined in pink). Therefore, e-mail 7769 is correlated to case 270.

Once the matching process is finished, the e-mail timestamps corresponding to the two activities of interest: job analysis (stillingsanalyse) and its approval (godkjent) are to be found. We used a basic word matching procedure based on whether the e-mail contained the words “stillingsnalayse” and/or “godkjent” (or derivatives of it).The e-mail with the earliest date is taken in both cases. The two activities together with the identified timestamps are then added to the event log that was generated with the data obtained from the HR and archival information systems.

E-mail ID	Received Date	Case ID	First Event Case	Job Name (modified)	Job Keyword	Match
7769	2018-04-12 15:32	270	2018-04-12 16:30	Barnevernkonsulent 100% vikariat	barnevernkonsulent	100 %
7769	2018-04-12 15:32	225	2018-03-30 09:20	Barne ungdomsarbeider Kvitsanden barnehage	barne	25 %

From : John Smith
Sent: Thursday 12 April 2018 15:32
To: John Doe
Subject: Stillingsanalyse og utlysningstekst
Attachment: Ekstern utlysning apr 18.docx; Stillinganalyse april 2018 100% vikariat.docx

Hei,

Oversender stillingsanalyse og utlysningstekst på vikariat som barnevernkonsulent.

Med vennlig hilsen

John Smith
Virksomhets leder barnevern
Tlf.: xx-xx-xx-xx

Figure 3. E-mail matching example. The matched parts of the e-mail are highlighted in the respective colour of the process case as indicated in the table.

3.4 Process Analysis

The main two process mining methods that we employed were process map discovery (Janssenswillen et al., 2019) and dotted chart visualisation (Song & van der Aalst, 2007). Process maps discover the directly-follows relations of the process, i.e., how often a certain process activity was followed by another activity, based on the data recorded in the event log and visualise them along with frequency and performance information. Dotted charts provide an overview of the time distribution of activities across multiple process cases both on a relative and an absolute time axis. Furthermore, we also investigated several of the core Key Performance Indicators (KPI) that were identified during the planning workshop. In particular, we looked at the time spent in the several process stages that were identified.

To facilitate rapid analysis and evaluation of the results with the stakeholders at Røros municipality we developed a prototype process mining dashboard based on the open source Process Mining framework bupaR[1]. Figure 4 shows a screenshot of the developed dashboard showing parts of the extracted data that was used during the final analysis iteration. Next to a process map and a dotted chart visualisation, the dashboard shows a box plots visualisation of relevant KPIs as well as a table with general properties of each process case. We used this dashboard directly in the interaction with process stakeholders during the evaluation workshops. A more detailed discussion of the individual results is provided in Section 4.

3.5 Evaluation

In total, we conducted two evaluation workshops in which we discussed the data processing and the obtained analysis results with the process stakeholder group at Røros municipality.

The first evaluation workshop took place after the initial process mining analysis of the event data extracted from the HR and archival systems (cf. Figure 1). The focus of this workshop was on validating the data quality and using the process mining results to engage the participants in discussion about the process. Next to IT and HR managers, two HR employees that work with the process took part in this workshop.

[1] Available from: https://github.com/bupaverse

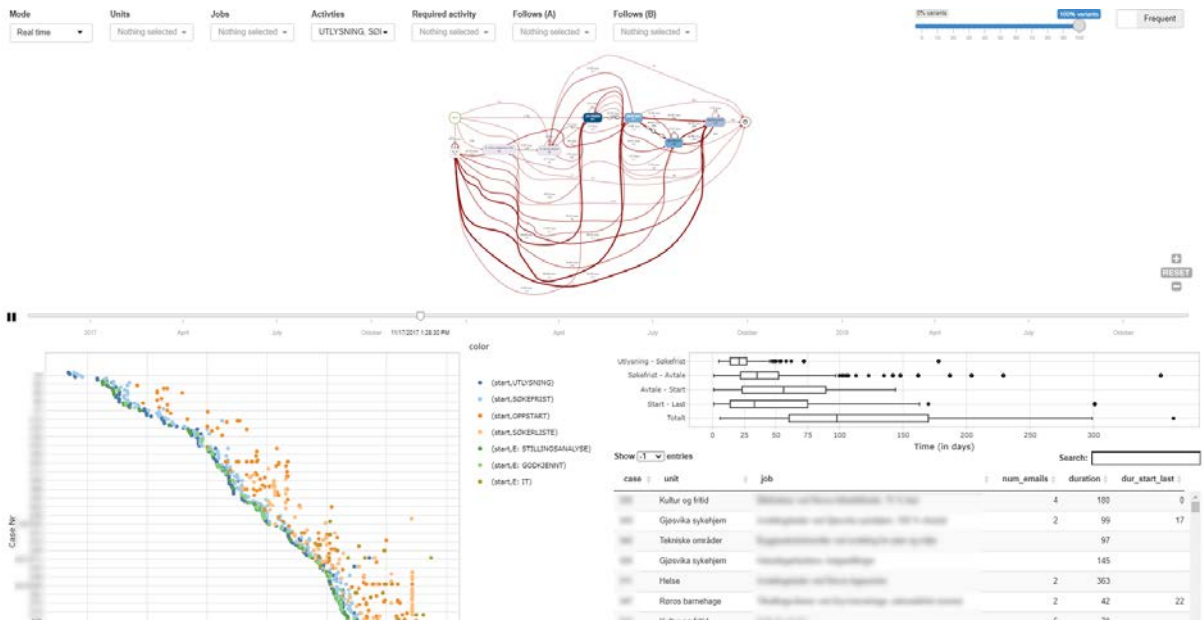


Figure 4. Screenshot of the developed Process Mining dashboard: the top part shows the process map visualisation, the lower left part shows a dotted chart vsualisation, and the lower right part shows an interactive table of process cases.

In the second evaluation workshop, we validated the results of the process mining analysis applied on the refined data set, which includes events based on e-mail conversations. Finally, as part of the second workshop, we conducted a preliminary qualitative evaluation of the accuracy of our e-mail matching method based on a sample of several e-mails. We discussed the accuracy of the computed matching to process instances with the HR and IT manager at Røros municipality. The samples were drawn from three different groups: cases with a very short overall duration, cases with a very long overall duration, and randomly selected cases. The first two categories were added to avoid bias from using only random sampling with a small number of samples.

4 RESULTS AND DISCUSSION

This section presents the results of the analysis that was conducted according to the methods described in Section 3 and discusses their implications.

The main result of the initial process mapping is shown in Figure 5. The underlying BPMN model is a simplified version of the model that was initially created at the first planning workshop. The original model was more complex containing more than 50 activities that need to be performed during the hiring process. However, we identified during the first evaluation workshop that for many of these activities it was not possible to obtain event data. We also identified five distinct process phases and corresponding six milestones that are the main elements in Figure 5. The process starts with a Preparation phase in which the job analysis (stillingsanalyse) needs to be conducted by the hiring unit. The analysis is an important document for the remainder of the process. After approval of the analysis, the job advertisement text is created, a case is opened in the HR system, and the job advertisement (Milestone Job Ad) is published. This step concludes the initial preparation phase of the process. Thereafter, except for creating the list of applications no activities occur until the application deadline has expired (Milestone Deadline). The following process phase is about selecting the most suitable candidates. Since this is mainly outside of scope for the HR department, we excluded this phase from our analysis and only determine the next milestone Selection based on when the HR department stores the final decision consisting of a ranking of candidates. Then, the HR department provides an offer for the most suitable candidate and negotiates the contract. This phase, which we denote Closing phase, is concluded with the milestone Contract when the candidate accepted the position and signed the contract. Finally, the process concludes with the Onboarding phase in which access to several systems and ordering of equipment is performed. This phase should conclude before the actual start of employment. It was found that measuring the duration of each of the process phases would be useful for an initial analysis of the

process. This means that we only need high-quality event for the six activities that represent the milestones to obtain useful insights.

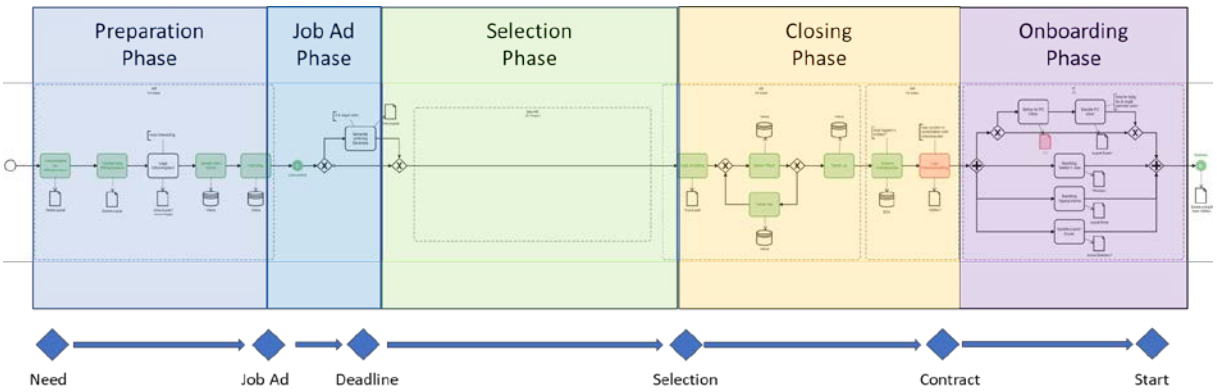


Figure 5. Simplified BPMN diagram in the background with an overlay of the main phase and milestones of the process. The diagram was created during the first workshop using manual process mapping and is shown only to provide a rough overview of the process.

The identification of the five process phases was based on discussing the results of the initial process mining analysis with the process stakeholders during the first evaluation workshop. The initial results that were discussed with the process stakeholders were the dotted chart visualisation in Figure 6 and the process map in Figure 7. Figure 6 shows a dotted chart visualisation. Here each activity observed in the event log is assigned a distinct colour and for every event a small dot is plotted on a timeline (x-axis) that is ordered by the individual process cases (y-axis). For example, the plot shows that the process usually, but not always, starts with a green dot, which indicates the activity UTLYSNING (job advertisement). Figure 7 shows the ordering with which the various activities of the process are performed. Each activity is represented by a node in the graph (e.g., UTLYSNING and OPPSTART) and a directed edge between two nodes represents the temporal order between those two activities as observed in the event log. The start and end of the process are represented by two additional nodes with a circle shape. Each edge is annotated with the frequency with which the ordering was observed, e.g., 13 cases of the process started with STILLINGSANALYSE and 88 cases of the process started with UTLYSNING. Additionally, we show the average waiting time between activities on the edges.

To obtain these results, we considered solely the events extracted from the structured data provided by the two information systems. Unfortunately, the evaluation revealed data quality issues for events about some activities. Some of the data quality issues are already visible in the dotted chart overview. There are several clusters of events that are recognisable as vertical lines in Figure 6. This could indicate batch processing of process cases where some activities are delayed and performed only when multiple cases can be processed simultaneously. However, based on feedback from the process stakeholders, this effect was more likely to be caused by using events obtained from the archival system. Whereas the archival of certain process documentation (e.g., the job analysis) should be performed in a timely manner, it would sometimes be only done in batches. The problematic data quality is also clearly visible in the unexpectedly complex process map shown in Figure 7, which is filtered to show a process map that covers 70% of the most common process traces. A process trace here refers to the exact sequence of activities performed for a single process instance. The complexity in Figure 7 originates from the activity *STILLINGSANALYSE* that seems to be performed at any phase of the process. However, this activity is supposed to be performed at the start of the process before the position is announced (UTLYSNING). Thus, the first evaluation concluded that additional data sources needed to be consulted to get a reliable picture of the process behaviour.

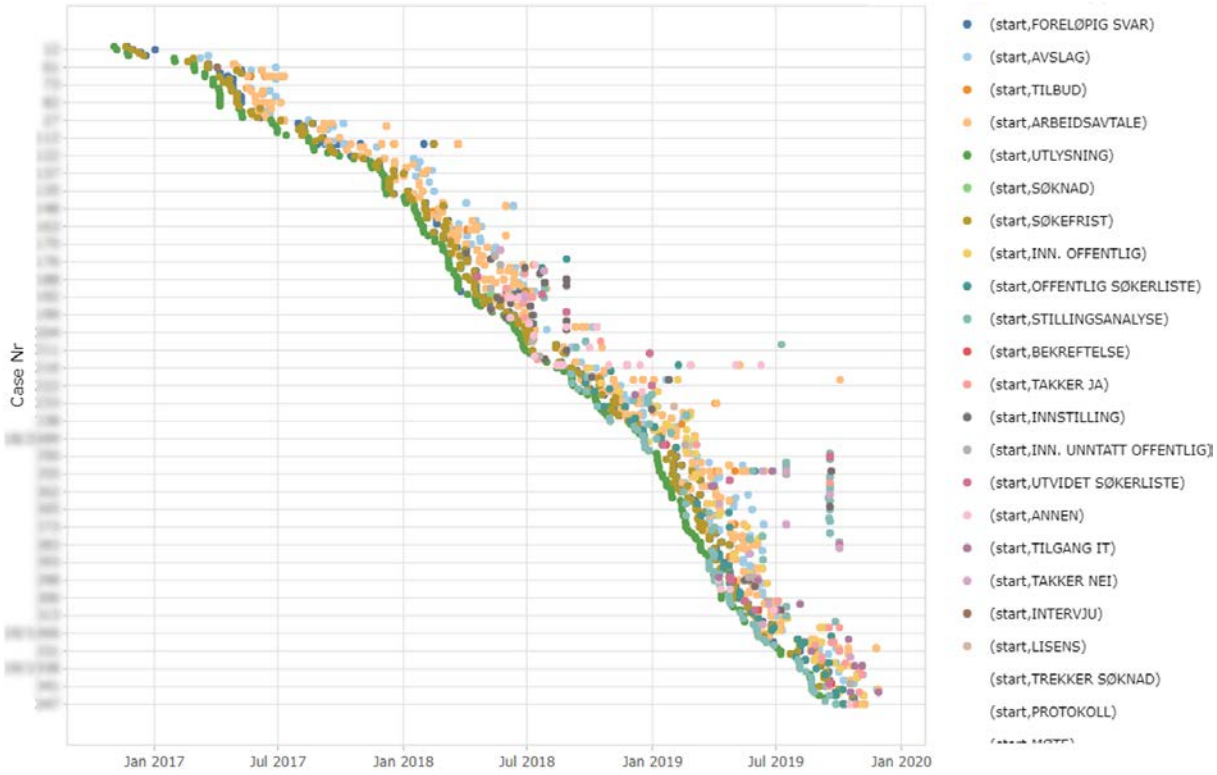


Figure 6. Overview of the event distribution in the hiring process over calendar time. Several vertical bars are visible, which indicate batching or data quality problems.

In the second evaluation round, we considered the additional events that were obtained by applying our e-mail matching method as described in Section 3 to mitigate the data quality issues related to the activity *STILLINGSANALYSE*. We added events based on e-mail conversations related to two activities to the event log: *E: STILLINGSANALYSE* and *E: GODKJENT*. Note that events related to the approval were not present in the initial event log since this part of the process is only handled by e-mail. Furthermore, we removed the problematic events from the archival system. Based on the revised event log, we again discover the process map showing 70% of the most frequent traces that is shown in Figure 8. The result indicates that our e-mail matching method was successful in identifying more accurate events for the activity *STILLINGSANALYSE* in some cases. For these cases, it is clear in Figure 8 that STILLINGSANALYSE is the first activity to be performed, contrarily to what is shown in Figure 7, where it takes place at any point in time. If the Preparation Phase (see Figure 5) -which is the one containing STILLINGSANALYSE- were to be automated, such automation would be hindered by the data used to produce Figure 7. This reveals the need for having high-quality data if any automation project is to be done.

Despite the enriched event log showing promising results for analysis with a more streamlined process map, Figure 8 also shows that we were only able to correlate events for 25 process cases out of the 72 that are visualised in the process map. Even though these 25 cases enable us to improve the trust in the analysis, the e-mail matching method still needs to be further improved to obtain more cases with E: STILLINGSNALYSE. In addition, we also investigate the individual correlation results for several of the e-mails that were matched. Out of the 14 process cases validated, 9 were randomly selected and 5 were selected on purpose due to their deviating process duration (either very long or very short). Out of the total, six process cases had a fully correct match (i.e., stillingsanalyse and godkjent), only one process case had a partially correct match (only stillingsanalyse), and for the remaining ones, wrong e-mails were matched. It is notable that most of the wrong matches occurred for the not randomly selected outlier cases. From these ones (n = 5), only one case was matched properly, whereas from the randomly sampled process cases (n = 9), only three matchings were incorrect. This shows that e-mail correlation to process instance is possible, but further work is required to reduce the number of false positive matches.

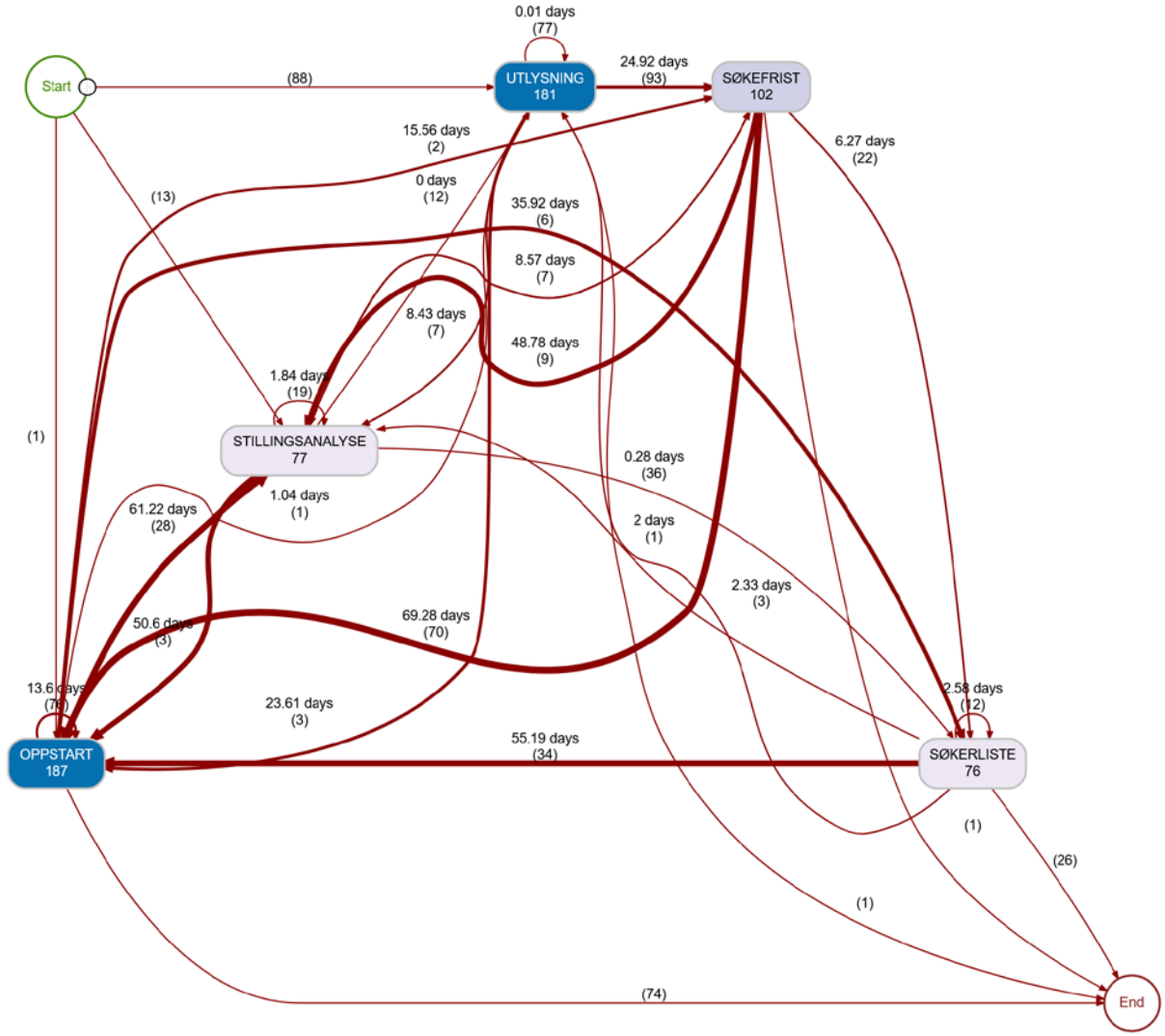


Figure 7. *Process map covering 70% of the most frequent trace variants of the initially prepared event log.*

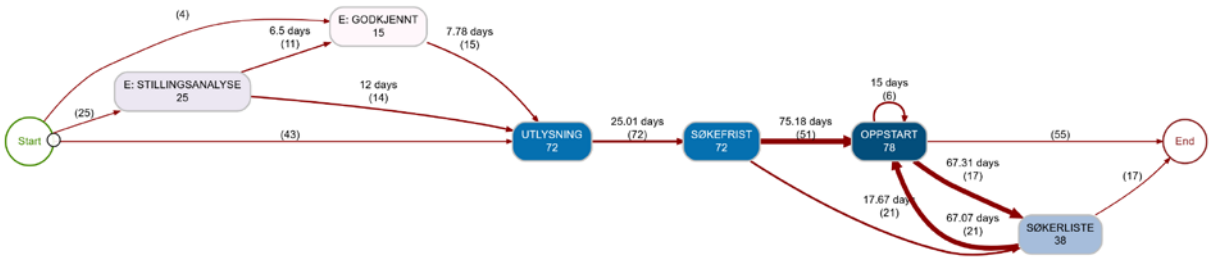


Figure 8. *Main process flow of the 70% most frequent trace variants after enhancing the event log with two activities obtain from e-mail mining: E: STILLINGSANALYSE and E: GODKJENT.*

Finally, we recognise some limitations of the used e-mail matching method. When looking for the modified full job name in the e-mail content, not only full words are matched but also partial words. This results in higher matching percentages. Looking at the example in Figure 3, case 225 gets a 25% matching percentage because one out of the four words that compose the job name is found within the e-mail: "barne". However, this is just a partial match of the words "barnevernkonsulent" and "barnevern". Thus, if only full words were to be matched, the matching percentage for case 225 would be 0. Another limitation is the fact that we searched for words in the full content of the e-mail and, therefore, did not make a distinction between heading, body and foot. This results in job names being matched to position names mentioned in the signature of the e-mail, which is not correct.

5 RELATED WORK

Several work on e-mail mining has been done addressing e-mail summarisation, e-mail organisation, activity extraction or e-mail task management among others. However, we found no approach that directly matched our requirements of matching the e-mails to the process cases based on identifying several domain concepts to be present in the e-mails.

Jlailaty et al. (2016) create a framework for mining process models from e-mail logs using unsupervised machine learning techniques. They use hierarchical clustering to group e-mails according to process and instance, and k-means clustering to group them according to activity. Activity names are then given by users. Ismaili (2019) adopts a supervised machine learning approach with minimized human effort to discover business processes from e-mails. First, e-mails in a training set are manually labelled. Then, a predictive model is trained, which later will automatically label the remaining e-mails. Process instances are then identified using a clustering algorithm. Jlailaty et al. (2017) implement unsupervised mining techniques together with semantic similarity measurement methods to mine activities from e-mail logs. A hierarchical clustering algorithm is used to cluster e-mails according to process and activity. For the similarity function of the algorithm they test two methods: Word2vec and Latent Semantic Indexing (LSA). Finally, a user chooses the most suitable label among the ones recommended by the proposed labelling technique. They do not identify instances nor discover the process model, but this is suggested as future work. Finally, following our same idea of reducing non-value adding activities with automation Jlailaty et al. (2018) extract business activities and their associated metadata from e-mail logs. Their vision is to reduce the number of e-mail exchanges to improve communication between personnel and to reduce waste. For example, the extracted information can be used to make suggestions when writing new e-mails, such as attaching a receipt for confirming a payment activity. This is done by extracting process-oriented sentences from e-mails, clustering them according to activity types and bringing together the metadata associated to the instances of each activity type.

While the presented studies focus on discovering business process models only with information extracted from e-mails logs, we use such information to enrich the event logs generated with structured data from the information systems. Moreover, we are not the only ones facing the challenge of matching e-mails with processes, activities and instances. In practice one e-mail can be associated with zero or multiple activities, instances and processes, based on which Ismaili (2019) makes a similar assumption to ours: each e-mail is related to one activity, one process and one instance.

6 CONCLUSIONS

In this paper we describe the analysis of the hiring process at Røros municipality to identify whether there is potential for process automation. We have observed a discrepancy between the observed process by stakeholders & rules/procedures and the process discovered with data from information systems, which has revealed data quality issues. To overcome some of these issues, we have enriched e-mail logs with activities extracted from e-mail conversations, which has shown to be successful by enabling the analysis of a process that is closer to reality. Nevertheless, we advise Røros municipality to assess what effect would automation have on the process and to address any data quality issues before identifying opportunities for automation. Ensuring high-quality data and a more standardized process will be important pre-requisites for preparing successful process automation. The results obtained in this study

are likely to be transferable to other municipalities who use similar information systems and, hence, might also have data quality issues.

The contributions of this paper are several. First, the enrichment of event logs with information from e-mail conversations to get a better view of the real process. Second, the e-mail matching method that correlates cases and e-mails. Last, a detailed discussion of a real process mining application in a public sector organization. In addition, Røros municipality has gained insight into the hiring process from a data perspective, giving them an extra viewpoint that together with their previous knowledge offers a comprehensive view of the hiring process.

We have identified several ways how the e-mail matching method can be improved. One was is to provide a list with the different units and their leaders, as these are the possible e-mails senders. Since not all e-mails have a specified unit, identifying the senders will enable the identification of the unit and, thus, a better e-mail-case match. Another way is to parse the e-mails dividing them into heading, body and foot. Occasionally, the job name keyword was matched to the job tittle of the sender, leading to a wrong match between case and e-mail. Looking for the job keyword and the job name in the heading and body, whereas looking for the unit name only in the foot of the e-mail will result in better matches.

References

Carmona, J., van Dongen, B., Solti, A., & Weidlich, M. (2018). *Conformance Checking*: Springer.

Geyer-Klingeberg, J., Nakladal, J., Baldauf, F., & Veit, F. (2018). *Process Mining and Robotic Process Automation: A Perfect Match.* Paper presented at the BPM (Dissertation/Demos/Industry).

Ismaili, A. (2019). Emails Analysis for Business Process Discovery. *Algorithms & Theories for the Analysis of Event Data (ATAED'2019)*, 54.

Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., & Vanhoof, K. (2019). bupaR: Enabling reproducible business process analysis. *Knowledge-Based Systems, 163*, 927-930.

Jlailaty, D., Grigori, D., & Belhajjame, K. (2016). A framework for mining process models from emails logs. *arXiv preprint arXiv:1609.06127*.

Jlailaty, D., Grigori, D., & Belhajjame, K. (2017). *Mining business process activities from email logs.* Paper presented at the 2017 IEEE International Conference on Cognitive Computing (ICCC).

Jlailaty, D., Grigori, D., & Belhajjame, K. (2018). *Email Business Activities Extraction and Annotation.* Paper presented at the International Workshop on Information Search, Integration, and Personalization.

Reijers, H. A., & Mansar, S. L. (2005). Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega, 33*(4), 283-306.

Song, M., & van der Aalst, W. M. (2007). *Supporting process mining by showing events at a glance.* Paper presented at the Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS).

Syed, R., et al. (2020). Robotic Process Automation: Contemporary themes and challenges. *Computers in Industry, 115*, 103162.

Van Der Aalst, W. (2016). Data science in action. In *Process mining* (pp. 3-23): Springer.

van Eck, M. L., Lu, X., Leemans, S. J., & van der Aalst, W. M. (2015). *PM2: A Process Mining Project Methodology.* Paper presented at the International Conference on Advanced Information Systems Engineering.