The Dark Side of Process Mining. How Identifiable Are Users Despite Technologically Anonymized Data? A Case Study from the Health Sector

Friederike Maria Bade(✉), Carolin Vollenberg, Jannis Koch, Julian Koch, and Andre Coners

South Westphalia University of Applied Sciences, Haldener Str. 182, 58095 Hagen, Germany
bade.friederikemaria@fh-swf.de

Abstract. Over the past decade, process mining has emerged as a new area of research focused on analyzing end-to-end processes through the use of event data and novel techniques for process discovery and conformance testing. While the benefits of process mining are widely recognized scientifically, research has increasingly addressed privacy concerns regarding the use of personal data and sensitive information that requires protection and compliance with data protection regulations. However, the privacy debate is currently answered exclusively by technical safeguards that lead to the anonymization of process data. This research analyzes the real-world utility of these process data anonymization techniques and evaluates their suitability for privacy protection. To this end, we use process mining in a case study to investigate how responsible users and specific user groups can be identified despite the technical anonymization of process mining data.

Keywords: Process mining · Privacy measures · Healthcare sector · Hospital information system

1 Introduction

Healthcare providers, especially hospitals, are under increasing pressure from policymakers and patient advocates to manage rising healthcare costs while improving the quality of care. The lack of efficiency due to poorly coordinated processes is considered a fundamental problem in achieving cost and quality goals [1, 2]. Since most of the information flow is mapped through the Hospital Information System or occurs “through informal communication, unsystematic processes, and uncontrolled access to information” [3], information deficits can often occur at interfaces. Therefore, a patient’s data must always be recorded and updated in concrete terms so that efforts and risks can be reduced on the one hand and quality can be increased on the other hand. In addition, the involvement of various disciplines and departments in the care process of a patient hospital journey aggravates a continuously guaranteed information flow of patient data. The different departments involved in a patient’s journey often have only little insight

C. Di Ciccio et al. (Eds.): BPM 2022, LNCS 13420, pp. 219–233, 2022.
https://doi.org/10.1007/978-3-031-16103-2_16

into what is happening in other disciplines or departments [1, 4]. The coordination of the involved stakeholders is often "hampered by informal communication, unsystematic processes, and uncontrolled access to information" [5]. This often leads to errors, mistakes, and confusion in information flows.

Process mining is an increasingly used technique in the field of information systems and is used to analyze and improve processes by using event logs. Different organizations from various industries have gained interest and use process mining in other cases and applications [3, 6]. Process mining has already also been successfully applied in healthcare and has helped to provide various insights to improve healthcare processes [1, 7]. However, the purely administrative processes in healthcare areas that require exclusively IT-based processing, such as care documentation, appointment scheduling, or billing processes, have so far only been addressed peripherally in the case of process mining [8, 9]. Therefore, process mining is also an emerging research area within the healthcare sector. Process mining is a growing research topic, which has not only increased in research interest in the last decade but also more recently [10, 11]. However, the main focus of research is on the technical implementation of process mining. At this point, research focuses on developing new algorithms and improving existing process mining techniques [12]. Research on practical aspects of process mining, as well as the adoption of the technology and the influences of process mining on the organization and employees, is less available [12]. While the benefits of process mining are widely recognized, the scientific community also expresses concerns about the irresponsible use of personal data within process mining and the aspects of anonymization of the data used by process mining. Thus, the ethical and legal issues are also of great interest and importance to researchers acutely. Therefore, it is becoming increasingly crucial for scientific efforts in this area to address privacy and confidentiality issues in process mining.

According to Pika et al. [13], Grishold et al. [12], Mannhardt et al. [14], or vom Brocke et al. [15], research should also specifically address these issues in the future and should put more effort into the examination of the privacy issues and risks, especially in sensitive areas like healthcare.

However, the debate on how far anonymized data of process mining can be identifiable has so far been answered and addressed by the use of technical data transformation techniques to anonymize process analysis data [4, 9, 13, 16, 17]. Therefore, considerable research and practical efforts have been made in recent years to develop, implement, and integrate appropriate privacy and confidentiality protection techniques in process mining. Though, this consideration of privacy aspects has been treated here still too inadequately and one-sidedly, because on the one hand it only considers the personal data of, e.g., patients, but not those of the process executors, and on the other hand, it only refers to the measures of data protection and privacy of obvious personal data via usernames [18].

Our research starts here and shows that privacy debates beyond the usual personal data are important and need to be part of ethical and technical discourse by showing that process mining can provide a way to provide role-based and personal information, e.g., when data has been changed and despite technically anonymized data. Here, we want to clarify to what extent process mining can attribute the execution of process steps (deviation from the target process) or the change of data to a specific originator. Especially we want to examine if the user's identification in the existing system (here: Hospital Information System) is explicitly technically prevented by the system. Therefore our research question is the following:

RQ: Can process mining provide identification of users in explicitly technically anonymized systems and assign errors directly to them?

We have been able to explore an approach that allows, using process mining data, to identify deviations from the standard process and user routines to associate users or groups of users to data changes in the system. The case of the hospital in West Germany with its hospital information system presented in this study gave us the unique opportunity to investigate how technically anonymized process mining data can be used, despite anonymization, to identify sources of error among assigned users to investigate whether the technical security measures are sufficient to protect the privacy of the users. The data basis used for this purpose was worked out with the responsible persons and users and the hospital's ethics committee in an experimental setup so that no ethical or legal concerns could arise for our research itself.

2 Background

Process mining is an emerging and essential technology in business process management (BPM). Particularly in information systems, the interest lies in how technologies can change and optimize processes. In the last decade, there has been a significantly increased interest in process mining, both in research and in practice [6]. Due to the increasing amount of available event data in organizations' easily accessible information systems, a variety of opportunities arose to analyze and optimize processes using information from this event data [2]. Process mining aims to gain a traceable overview of a process, to provide insights into a process and the actual process flow, and to support improvements [2, 3]. However, the awareness of privacy issues, and thus the ethical issue of using process mining and with this the use of personal data, has increased significantly [10, 19].

In the context of the application, process mining enables the discovery, verification of conformities, and improvement of processes [2]. In the context of discovery, the corresponding technique of process mining helps in the discovery of the real process by creating a process model using event logs of the process. The conformance checking type of process mining compares the evaluated real process model with the predefined process model and reveals whether the reality matches the predefined model [2, 6]. The enhancement type of process mining "aims to modify or extend the a priori model. For example, by using timestamps in the event logs, the model can be extended to show bottlenecks, service levels, lead times, and frequencies" (van der Aalst, 2012) [2, 20].

Process mining encompasses techniques used to analyze and optimize processes. These techniques provide data-driven methods of process analysis that focus on the evaluation and extraction of information from event logs - information stored in IT systems about individual and actual process steps. Event logs store information such as the entity, e.g. a person or device, that performs or triggers an activity. In addition, event logs store timestamps of an event or data elements recorded with an event. These event logs may contain direct and indirect identifiers of personal data and may disclose the user's personal data or groups of users. The discussion on revealing this data increases due to the new General Data Protection Regulation. In this regard, it can be mentioned that data protection in hospitals is exceptionally high, as sensitive data is involved. There is a lively discussion and debate in the scientific discourse on anonymization techniques of datasets collected through data mining, such as perturbation, anonymization and cryptography. There are various privacy preserving techniques such as Generalisation, Suppression, Distortion, Swapping or Masking. Each technique pursues the goal of transforming personal data. Thus, it is expected to reduce the original information in the dataset by a certain amount. In particular, there are leading research efforts; as per Murhy [21], the consensus is that the anonymization technique of Suppression is the most efficient and resource-saving solution of all [21].

For the underlying process mining technologies, several privacy-preserving techniques on the research side attempt to anonymize the data collected and processed in the event log [16].

Pika et al. [13] already analyzed and evaluated existing privacy approaches to anonymize process data for process mining. They tested the suitability of three different approaches, confidentiality framework, PRESTA, and differential privacy model for event logs. The analysis showed a trade-off between privacy and utility. The methods that maintain higher data utility for process mining purposes (e.g., encryption) do not provide strong privacy protection.

Process mining has also gained a lot of interest in recent years, especially in the healthcare sector [4, 17, 22–24]. Hospitals, in particular, face the challenge of streamlining their processes and the documentation of patient data. The processes in hospitals are characterized by the fact that several departments may be involved in the process of patient care. These different organizational departments often have their own specific IT applications, add different information to the hospital information system, or need different information about a patient. As a result, problems often arise in obtaining data related to healthcare and hospital processes in particular.

Hospital processes are characterized by a high degree of complexity, and the extremely flexible implementation of patient care, which always depends on the needs and condition of the individual patient and their individual treatment. In addition, as already mentioned, hospitals involve various actors, staff from different disciplines, and diverse departments. This setting, therefore, offers great potential for the use of process mining due to the involvement of different actors, the various existing systems, including the hospital information system, and the huge amount of available data and event logs [1, 4]. However, data security and anonymization of these data, especially in healthcare, is a high priority and therefore regulated by high standards, laws, and guidelines. In particular, personal data must be anonymized to ensure data security and to protect personal rights.

Process mining can be used to identify and quantify activity patterns that reflect how users act in processes [13]. Despite the use of privacy transformation techniques to anonymize data, the use of process mining offers the possibility to provide both anonymized and non-anonymized data about user information and thus personal data. However, this also depends on the given conditions of the existing system. It is only possible if the user data is available in the existing information system, only then can process mining provide transparency about which process steps are performed by whom and at what time [19]. Patient and staff identities could be revealed with the help of background knowledge and the event log.

If process mining is used within an information system that does not allow for a clear user assignment because only anonymized data is available or no user assignment is made, process mining with the existing event logs cannot contribute to the identification of users, e.g. to the identification of vulnerabilities. Thus, despite technical security measures to anonymize the data, the question should be asked whether the use of process mining is ethically correct, as conclusions can still be drawn about patients and staff through the background information on the process. Especially when the process flows in practice often do not correspond to the predefined process models, it is possible to get the link to the particular employee. As process flows in practice often do not correspond to the predefined process models, but differ in their actual execution, which leads to a "significant gap between what is prescribed or supposed to happen and what happens in reality" (Mans et al. 2008) [24] the assignment of identification of personal resources could be possible with particular background information. In this way, the weak points can also be identified, poor quality, loss of time, and higher process costs can be avoided. Only an accurate assessment of reality can help in reviewing process models to optimize business processes [25]. Still, this accurate assessment of reality process steps by process mining can also be critical in terms of anonymization of personal data.

3 Methodology

To achieve sustainable results in the rather explorative nature of our study, we use a case study approach. Since we want to gain an understanding of the anonymization and identifiability of users or groups of users in anonymized event logs, as per our formulated RQ, a case study approach, according to Yin, is most appropriate [26]. As it allows us to study a phenomenon in a firmly grounded context by using triangulation of different data sources – Process Mining data, field observations, interviews, and documentation - to gain insights, this approach is well suited for our RQ [26].

The case study method is suitable for gaining a thorough and detailed understanding of factors (such as anonymization and identifiability in our case) [27], and it involves the use of case organizations to prove existing theories of process mining anonymization systems based on empirical evidence. In addition to the primary process mining data collected, we use data triangulation from event logs, database schemas, and development and execution logs generated by the respective PM application systems to provide scientific rigor [20].

Figure 1 shows an overview of our case study approach, which consists of four main phases: case selection, data collection, data analysis, and case conclusion.

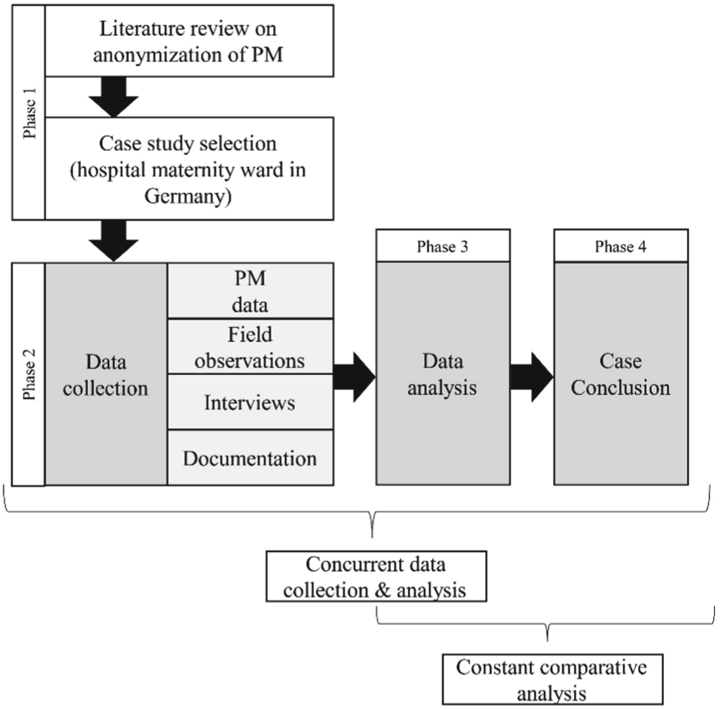
Phase 1
Literature review on anonymization of PM
Case study selection (hospital maternity ward in Germany)
Phase 2
Data collection
PM data
Field observations
Interviews
Documentation
Phase 3
Data analysis
Phase 4
Case Conclusion
Concurrent data collection & analysis
Constant comparative analysis

Fig. 1. Research method.

3.1 Case Selection

To arrive at a representative selection of objects of study, we deliberately chose a hospital maternity ward in Germany to cover a broad spectrum of particularly sensitive and especially anonymous process data and event logs. In this context, it was key that the users and data originators were informed in advance about the data use and analysis and gave their consent to our research. The organization investigated in this study is a large hospital group in Germany with more than 20,000 employees. This case study organization comprises more than 100 facilities, including more than 20 hospital sites with over 6,000 beds, above 30 residential and nursing care facilities with more than 3,000 nursing places, more than 30 medical care centers, and about 10 outpatient care services. Over one million patients are cared for annually in our case study organization.

In our case study presented here, we focus on the data collection process at birth, where the first data of a new life is collected and entered into the Hospital Information System as core data for each newborn. During a hospital stay, however, there are always changes in the master data recognized, which lead to unintentional errors and consequential (negative) effects. These irregularly executed processes of master data manipulation ultimately have an extremely negative impact on the performance, conformance, and compliance of process fulfillment within the whole treatment of the newborn. Further, this affects the work of the medical staff in the different departments a newborn is transferred to during hospital stay. But it also affects the newborn and can somehow disturb or affect the newborn's mother. For example, if not all information about the newborn is available after the release from the hospital, examinations must be repeated in the hospital; this means additional work for medical staff, higher financial expenses for the hospital, and an unnecessary burden and stressful process for the young mother and the newborn as well.

The treatment of a newborn in the studied maternity ward and the regarded master data process takes into account, among other things, the course of a premature baby from birth in the delivery room to the time of transfer to the regular pediatric ward and discharge from the hospital as demonstrated in Fig. 2. After the birth of a premature child, the newborn is initially physically located in the delivery room. The child's first master data is entered into the Hospital Information System by the doctor or the participating obstetrician. Following that, the child gets to the premature care department of the neonatology unit. There, the premature child gets special care from a specialized team of doctors and nurses as it may need extra help, e.g., breathing or eating. All data (e.g., weight or length of the newborn) and information about the treatment of the newborn are entered into the Hospital Information System by the different specialized team members involved. Later it is transferred to the regular perinatal department of the center before the child leaves the center and gets to the pediatric department of the hospital, is being discharged home, or transferred to another hospital.

Because the child and the patient master data are entered, used, and passed through several different departments, the medical staff in the various departments of the hospital, considered in the case study, frequently recognized changes, missing master data, or wrong adjustments of the master data and had problems assuming the relevant data, e.g., assigning the premature child, or often had to do extra administrative work by asking other staff to obtain information. In addition, often, the staff had to search for patient data because the system does not list the data or incorrect data are recorded, e.g., the name of the premature child or the accurate date of birth. In concrete terms, data loss and changes between the individual departments occurred over time again and again. Accordingly, these documentation processes of premature child's master data (e.g., the name of the newborn) and the additional according to data (e.g., the treatments, the parameters of treatment, and continuously raised parameters of the newborn) were selected for our research, by using process mining to proof the anonymization of the user or user groups that are responsible for changes in master data.

Along the described process of treating a premature child (cf. Fig. 2), the mentioned documentation processes were selected as the focus of our research on using process mining to examine and identify the user or user groups responsible for changes in master data. Our application of the process mining methodology is summarized in three steps according to van Dongen et al. (2005).

In step 1, we defined the scope of the extraction by screening out the granularity of the data and the most significant possible observation period as well as associated attributes from the Hospital Information System. This dataset went back a total of 20 months. In step 2, the event logs were analyzed by applying process discovery and conformance checking methods with the process mining solution *ProM* [28]. In step 3, the discovered process model was evaluated using the fitness, precision, and generalizability measures proposed by van Dongen et al. (2005). Finally, in step 4, these data were anonymized using a common and established technical protection measure within the *ProM* program package, specifically including all resource names (proper names, usernames, user abbreviations, logins), case IDs, and roughly detailed timestamps (months and years).

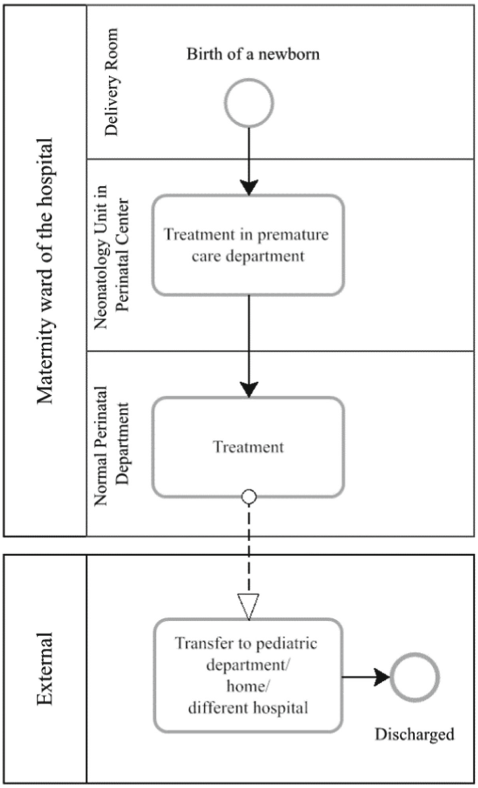
Maternity ward of the hospital
Delivery Room
Birth of a newborn
Neonatology Unit in Perinatal Center
Treatment in premature care department
Normal Perinatal Department
Treatment
External
Transfer to pediatric department/ home/ different hospital
Discharged

Fig. 2. The delivery process of a premature child.

3.2 Data Collection and Data Analysis

In our case study, we followed the established approach of Yin and enrich our research approach with more flexible data collection (field observations, 18 interviews, 5 programmer documentations, and 2 data documentations of the Hospital Information System) in addition to the primary process mining data [26]. To gain a sound and low-threshold insight into the existing process landscape, we used pre-built connectors, in particular XESame [29, 30], which contained a readymade data mapping template and allowed us to generate the standard XES event log formats from the HIS underlying SQL database. To gain insights into the process of the master data documentation during the birth and treatment of a premature child, we have collected data implementing process mining in the hospital studied. We have used process mining to review this process and get depth insights into the as-is process. Further, the aim was to avoid errors and extra work and deliver master data correctly. As semi-structured interviews allow us to gain a depth understanding of the whole documentation process and the role of process mining in this context, we chose this method as our primary data collection. For this purpose, we developed an interview protocol that was largely based on the collection of process mining data and included, for example, the single-cell process steps and possible deviations or inconsistencies. The collected and analyzed process mining data thus semantically specified the interview structure used. The interviews allowed us to observe the participants and gain insights on their non-verbal behavior as well. One author observed the typical input process of master data and characteristic interaction processes within the HIS and observed the system-based recording of planned and delivered care as well as the software-based documentation of the individual steps of care planning. In addition, interaction with the documentation file in HIS was observed to understand the context of the documentation process, including the documentation folder and templates.We collected a total of 340 min of interview material, which was transcribed verbatim, and computer coded by two authors according to the methodology of Flick's depth analysis [31]. The used interview pool consisted of 8 respondents: 3 Nurses, 2 Assistant Doctors, the Head of Patient Management, a staff member of the clinical IT, and an external Application Manager of the Hospital Information System as demonstrated in Table 1.

Table 1. Overview of data collection.

Participant	Data scope	Participant's role
Respondent 1	(2 × semi-structured interview in total 70 min)	Nurse
Respondent 2	(4 × semi-structured interview in total 60 min)	Nurse
Respondent 3	(1 × semi-structured interview in total 20 min)	Nurse
Respondent 4	(1 × semi-structured interview in total 20 min)	Assistant doctor
Respondent 5	(2 × semi-structured interview in total 30 min)	Assistant doctor
Respondent 6	(6 × semi-structured interview in total 110 min)	Staff member clinical IT systems/interface manager
Respondent 7	(1 × semi-structured interview in total 15 min)	Application manager hospital information system (extern)
Respondent 8	(1 × semi-structured interview in total 15 min)	Head of patient management

4 Results

The processes of the Hospital Information System extracted by process mining contained 3,913 historized complete executions of the examined master data. After reviewing the correctness of the studied process, we found that in about 17% (n = 661), the master data did not comply with the compliant-process flow and were permanently changed later than the initial entry. Concerning the research question already introduced, the following observation was made: In 71% (n = 469) of these error cases, the cause of the error could be assigned to a specific identifiable user or user role by the other users running the process, even though the system and process mining had anonymized the data through technical protection measures. Since the other users involved in the process knew each other's work methods and routines, the users involved were able to identify different user roles and assign errors to specific users. For example, only a few qualified users have access to particular program parameters when entering legal regulations according to the Narcotics Act as well as release processes, a certain sequence of operations that are required for the specific work assignment of the persons, or the assignment of defined shift sequences that are linked to individual qualification profiles of persons.

Based on the recorded traces, we have so far been able to determine that the occurrence of these irregular process executions or irregular changes to the master data has certain activity patterns. In doing so, we could establish a tangible link between certain execution routines and the erroneous master data manipulations and changes. This enabled it to identify a user group or, in many cases (n = 188), even the specific process executer via the corresponding process knowledge of the process owners. This became possible based on the detailed sequence of process steps with the help of process mining. Within that, the errors in the process were made possible to be assigned by certain individual process sequences or execution sequences to the individually unique user assignments, although only anonymized data were available.

As an example, we present here the process variants identified so far. Each of which led to the most frequent manipulations and changes of master data in the case study data. This exemplary identified process flow is shown in Fig. 3 and consists of the defined, collected process steps above.

As can be seen in Fig. 3, certain process steps are executed with different frequencies. This allowed us to directly assign certain sequence combinations and variations to specific users or user roles via certain combinations in the process variance or execution, taking into account process knowledge and contextualizing the execution variations. For example, internal staff performs the steps in a different sequence than external temporary staff. Intensive and Aesthesia nurses work through the processes in a different sequence depending on the duty roster and assigned nurse manager. We were able to assign deviations from the standard process found with process mining as these execution routines to a certain user group (in this case the ward doctors) based on their type and sequence of process steps. By looking at the detailed sequences of the process steps, process mining was thus able to clearly show which user role was responsible for the errors in the process, although only anonymized data was available.

This shows how security measures can be circumvented without much effort. Despite anonymizing advanced identifiers such as decoupling usernames and timestamps or generally making usernames unidentifiable to ensure user anonymity, we were able to demonstrate that individual users could be identified. Because process mining is becoming increasingly important and thus widespread, especially in healthcare, our research results are of great importance for the upcoming process mining use in practice

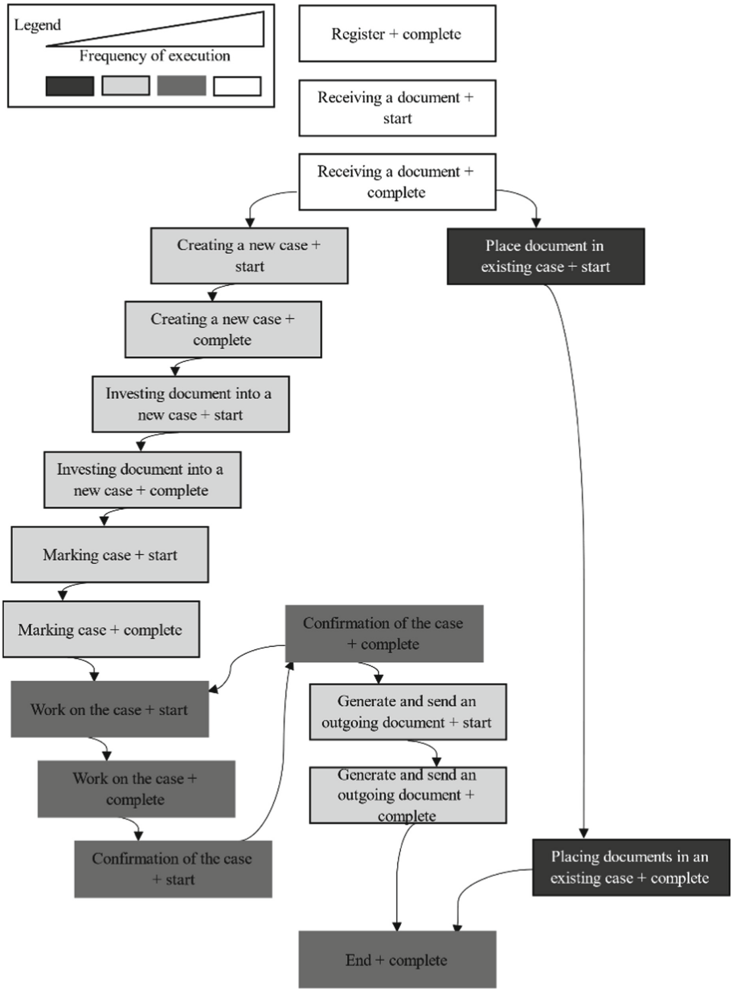
Legend
Frequency of execution
Register + complete
Receiving a document + start
Receiving a document + complete
Creating a new case + start
Place document in existing case + start
Creating a new case + complete
Investing document into a new case + start
Investing document into a new case + complete
Marking case + start
Marking case + complete
Confirmation of the case + complete
Work on the case + start
Generate and send an outgoing document + start
Work on the case + complete
Generate and send an outgoing document + complete
Confirmation of the case + start
Placing documents in an existing case + complete
End + complete

Fig. 3. Master data process in process mining.

and research. They believe in the sufficiency of technical safety measures previously assumed in the literature in research.

We suggest, therefore, a solution approach to overcome this lack of anonymization by only examining process sections instead of the overall processes. If the countermeasures considered safe in the literature were taken and supplemented by making the complete processes unrecognizable, the desired anonymization could be achieved. As part of the authors' proposed solution, an attempt was made to achieve some degree of decontextualization. This decontextualization of the process flow was achieved, among other things, by breaking the process flow down into smaller substeps and isolating it instead of viewing it as a whole. On the one hand, this led to the desired anonymization of the underlying process executors, but on the other hand, it also significantly limited the informative value regarding the improvement possibilities and the recognition gain of the PM application. Since, in this case study, the master data in the dataset may contain changes from different users and much of the deanonymization is due to activity pattern recognition, partitioning the processes into sub-processes proved to be extremely effective. However, for the majority of the maternity ward processes studied, it also had significant drawbacks, as the context of the overall process was lost for process optimization, which is ultimately the overall goal of process mining. Still, in the case of keeping the personal data and anonymization in process mining up, this approach seems suitable.

5 Discussion

From its basic idea, process mining technology enables the reconciliation of governance, risk, or compliance issues. At the user level, this article has shown that under certain conditions, it can also be used to identify processes in d. In particular, these conditions are known user routines and processing steps. We have shown that administrative-like tasks characterize downstream business processes in healthcare. These administrative processes are usually performed by nurses rather than physicians (e.g., obstetricians or general practitioner doctors).

The work presented here shows that, contrary to popular belief that process mining does not allow inferences to be made about user data if they are anonymized in the existing system, this is nevertheless possible.

This is possible if the process flows identified with process mining also reveal user routines that can be assigned to a specific user group or even an individual user with the help of process expertise. This evaluation of user routines and process deviations using process mining in anonymized systems leads to identifying unintentional and user-induced changes in core data. The study showed that these routines allow inferences to be made about the user or user group that made the changes without mapping user data and personal information in systems. To counteract the problem, processes were broken down into process snippets to prevent the activity pattern recognition enabled by process mining. Since this solution can only be a temporary interim solution, because the contextuality that is lost in the process can have disadvantages, as already mentioned, this naturally raises new questions for the theory as an implication. For example, to design security measures that demonstrably do not have the shortcomings we have pointed out, it would have to be possible to exclude activity pattern recognition. On the other hand, it must be ensured that there is no decontextualization as a consequence not to impair the effectiveness of process mining.

We will address these questions in the future, and they need to be addressed by other research teams and researchers as well.

Our research also has considerable added value for practical applications. We have demonstrated that the technical protection measures, which are mostly found to be sufficient in the literature, have serious weaknesses. As already mentioned, we were able to take most processes out of context and make deanonymization impossible. However, there were still processes that offered such a high identification potential that even these measures were insufficient. Department heads can only handle release processes in particular. If errors were to occur here, technical anonymization would not be possible by nature. This raises the question of the extent to which process mining may be used in companies that deal with very sensitive data. Although the measures basically protect the data of uninvolved parties (usually customers and/or clients), the activities of the employees are accessible to everyone via process mining, which means that the measures are not sufficient to protect the privacy of the users.

Our findings are, of course, principally limited, as they are initially restricted in the work presented here to a single case study and the data collected and available there. However, we estimate the transferability of the results outside the case study context used here to be very high since medical procedures, as well as information systems, are largely used internationally and must comply with international standards throughout. We also have to assume that other process mining technologies and used procedures may provide different results in anonymization quality. Nevertheless, the identifiability of process participants via mere variations and combinations of process flows and steps is, of course, technically not solvable without including further steps. We are also aware that our interviews varied in scope depending on the interviewee. We did include this in the coding, but care should still be taken to make the interviews more consistent in subsequent research.

References

1. Mans, R.S., Schonenberg, M.H., Song, M., van der Aalst, W.M.P., Bakker, P.J.M.: Application of process mining in healthcare – a case study in a dutch hospital. In: Fred, A., Filipe, J., Gamboa, H. (eds.) BIOSTEC 2008. CCIS, vol. 25, pp. 425–438. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-92219-3_32
2. van der Aalst, W., et al.: Process mining manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011. LNBIP, vol. 99, pp. 169–194. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28108-2_19
3. van der Aalst, W.M.P.: Process Mining. Data Science in Action. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
4. Rovani, M., Maggi, F.M., de Leoni, M., van der Aalst, W.M.: Declarative process mining in healthcare. Expert Syst. Appl. **42**, 9236–9251 (2015). https://doi.org/10.1016/j.eswa.2015.07.040
5. Fredriksen, E., Martinez, S., Moe, C.E., Thygesen, E.: Communication and information exchange between primary healthcare employees and volunteers - challenges, needs and possibilities for technology support. Health Soc. Care Community **28**, 1252–1260 (2020). https://doi.org/10.1111/hsc.12958
6. Ghasemi, M., Amyot, D.: Process mining in healthcare: a systematised literature review. Int. J. Electron. Healthc. **72** (2016). https://doi.org/10.1504/IJEH.2016.078745
7. Erdogan, T.G., Tarhan, A.: A goal-driven evaluation method based on process mining for healthcare processes. Appl. Sci. **8**, 894 (2018). https://doi.org/10.3390/app8060894
8. Mans, R., Reijers, H., Wismeijer, D., van Genuchten, M.: A process-oriented methodology for evaluating the impact of IT: a proposal and an application in healthcare. Inf. Syst. **38**, 1097–1115 (2013). https://doi.org/10.1016/j.is.2013.06.005
9. Martin, N., et al.: Recommendations for enhancing the usability and understandability of process mining in healthcare. Artif. Intell. Med. **109**, 101962 (2020)
10. Munoz-Gama, J., et al.: Process mining for healthcare: characteristics and challenges. J. Biomed. Inform. **127**, 103994 (2022). https://doi.org/10.1016/j.jbi.2022.103994

11. Eggers, J., Hein, A., Böhm, M., Krcmar, H.: No longer out of sight, no longer out of mind? How organizations engage with process mining-induced transparency to achieve increased process awareness. Bus. Inf. Syst. Eng. **63**(5), 491–510 (2021). https://doi.org/10.1007/s12599-021-00715-x
12. Grisold, T., Mendling, J., Otto, M., vom Brocke, J.: Adoption, use and management of process mining in practice. BPMJ **27**, 369–387 (2020)
13. Pika, A., Wynn, M.T., Budiono, S., ter Hofstede, A.H.M., van der Aalst, W.M.P., Reijers, H.A.: Privacy-preserving process mining in healthcare. Int. J. Environ. Res. Public Health **17**, 1612 (2020). https://doi.org/10.3390/ijerph17051612
14. Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., Michael, J.: Privacy-preserving process mining. Bus. Inf. Syst. Eng. **61**, 595–614 (2019). https://doi.org/10.1007/s12599-019-00613-3
15. vom Brocke, J., Jans, M., Mendling, J., Reijers, H.A.: A five-level framework for research on process mining. Bus. Inf. Syst. Eng. **63**(5), 483–490 (2021). https://doi.org/10.1007/s12599-021-00718-8
16. Rafiei, M., van der Aalst, W.M.P.: Privacy-preserving data publishing in process mining. In: Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds.) BPM 2020. LNBIP, vol. 392, pp. 122–138. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58638-6_8
17. Loxton, M.: Process Mining in Healthcare. (2016, Unpublished)
18. Weise, M., Kovacevic, F., Popper, N., Rauber, A.: OSSDIP: open source secure data infrastructure and processes supporting data visiting. Data Sci. J. **21** (2022). https://doi.org/10.5334/dsj-2022-004
19. Nuñez von Voigt, S., et al.: Quantifying the re-identification risk of event logs for process mining. In: Dustdar, S., Yu, E., Salinesi, C., Rieu, D., Pant, V. (eds.) CAiSE 2020. LNCS, vol. 12127, pp. 252–267. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49435-3_16
20. Houghton, C., Casey, D., Shaw, D., Murphy, K.: Rigour in qualitative case-study research. Nurse Res. (2013)
21. Murthy, S., et al.: A comparative study of data anonymization techniques. In: 5th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance (2019)
22. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. J. Biomed. Inform. **61**, 224–236 (2016). https://doi.org/10.1016/j.jbi.2016.04.007
23. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B.: Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16071-9
24. Mans, R.S., Schonenberg, M.H., Leonardi, G., Panzarasa, S., Quaglini, S., van der Aalst, W.M.P.: Process mining techniques: an application to stroke care. Stud. Helath Technol. Inform. **136**, 573–578 (2008)
25. van der Aalst, W.M.P., Günther, C.W.: Finding structure in unstructured processes: the case for process mining. In: Basten, T., Juhás, G., Shukla, S.K. (eds.) Seventh International Conference on Application of Concurrency to System Design, ACSD 2007, Proceedings, Bratislava, Slovak Republic, 10–13 July 2007. IEEE Computer Society, Los Alamitos (2007). https://doi.org/10.1109/acsd.2007.50
26. Yin, R.K.: Case Study Research and Applications: Design and Methods. SAGE Publications, Thousand Oaks (2017)
27. Yin, R.K.: Qualitative Research from Start to Finish. Guilford Publications, New York (2015)

28. Dongen, B.F., Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., Aalst, W.M.P.: The ProM framework: a new era in process mining tool support. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005). https://doi.org/10.1007/11494744_25
29. Verbeek, H.M.W., Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: Xes, xesame, and prom 6. In: Soffer, P., Proper, E. (eds.) CAiSE Forum 2010. LNBIP, vol. 72, pp. 60–75. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-17722-4_5
30. Günther, C.W., van der Aalst, W.M.P.: A generic import framework for process event logs. In: Eder, J., Dustdar, S. (eds.) Business Process Management Workshops, pp. 81–92. Springer Berlin Heidelberg, Berlin, Heidelberg (2006). https://doi.org/10.1007/11837862_10
31. Flick, U.: Triangulation. In: Mey, G., Mruck, K. (eds.) Handbuch Qualitative Forschung in der Psychologie, pp. 185–199. Springer, Wiesbaden (2020). https://doi.org/10.1007/978-3-658-26887-9_23