Process Mining and Perceived Privacy Violations: A Pilot-Study

Evelyn Zuidema-Tempel[1], Faiza Allah Bukhsh[2], Robin Effing[3] and Jos van Hillegersberg[3]
[1]*Research Group Digital Intelligence & Business, Saxion University of Applied Sciences, M. H. Tromplaan 28, Enschede, The Netherlands*
[2]*Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Drienerlolaan 5, Enschede, The Netherlands*
[3]*Faculty of Behavioural, Management and Social Sciences, University of Twente, Drienerlolaan 5, Enschede, The Netherlands*

Keywords: Process Mining, Privacy, Perceived Privacy, Process Model Abstraction, BPM, GDPR.

Abstract: Despite the existence of various methods and abstraction techniques to reduce the privacy risk of process models generated by process mining algorithms, it is unclear how process mining stakeholders *perceive* privacy violations. In this pilot-study various process model visualisations were shown to 6 stakeholders of a travel expense claim process. While changing the abstraction levels of these visualisations, the stakeholders were asked whether they perceived a violation of their privacy. The results show that there are differences in how individual stakeholders perceive privacy violations of process models generated via process mining algorithms. Results differ per type of visualization, type of privacy risk reducing methods, changes of abstraction level and stakeholder role. To reduce the privacy risk, the interviewees suggested to include an authorization table in the process mining tool, communicate the goal of the analysis with all stakeholders, and validate the analysis with a privacy officer. It is suggested that future research focuses on discussing and validating process visualisations and privacy risk reducing methods and techniques with various process mining stakeholders in organisations. This is expected to reduce perceived violations and prevents developing techniques that are aimed at reducing privacy risk but are not considered as such by stakeholders.

1 INTRODUCTION

Process mining is a technique that is designed to discover, monitor and improve actual processes by extracting knowledge from event logs readily available in today's information systems (van der Aalst, 2016). Process mining is increasingly being adopted in industry as a data-driven innovation (Grisold et al., 2021). With a forecasted growth of 40% to 50% for the coming years and passing $1 billion in revenues in 2022, the market is expected to continue to rise (Gartner, 2020).

Like process mining, privacy is also gaining attention in industry and society due to some recent scandals. High profile cases such as the unwilling exposure of user accounts at Yahoo and Facebook and recent data breaches such as the case at the Marriot hotel (Hill & Swinhoe, 2021) put the spotlights on privacy. Also, data privacy is included in many national and international legislations, such as in the General Data Protection Regulation (GDPR) in Europe (GDPR, 2022). The GDPR includes the principle of privacy by design, meaning that companies are encouraged to implement technical and organizational measures necessary at the earliest stage of the design of processing operations to ensure the principles of privacy and data protection (GDPR, 2022).

Because event logs which are needed to mine processes often contain personal information of process stakeholders, the principles of privacy and data protection also apply to process mining. In academics, privacy and process mining as separate research topics are gaining more attention, but privacy preserving process mining is still in its infancy (Mannhardt et al., 2019, Pika et al., 2020). To the best of our knowledge, only a few studies have been conducted to process mining and privacy (Elkoumy et al., 2022, Sohail et al., 2021, Pika et al., 2020, Rafiei & van der Aalst, 2020). These studies mainly focus on privacy preserving methods and abstraction techniques of process mining tools.

Abstraction is about simplifying process models by removing edges, clustering nodes, and removing

Zuidema-Tempel, E., Bukhsh, F., Effing, R. and van Hillegersberg, J.
Process Mining and Perceived Privacy Violations: A Pilot-Study.
DOI: 10.5220/0011745200003467
In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023) - Volume 2*, pages 289-296
ISBN: 978-989-758-648-4; ISSN: 2184-4992

nodes to make the process model more comprehensible for the person looking at it (Maneschijn et al., 2022). Through abstraction, irrelevant details in a process model can be reduced (Polyvyanyy et al., 2015).

Remarkably, privacy preserving methods and abstraction techniques to reduce the privacy risk have hardly been evaluated from a stakeholder's perspective. We have not been able to identify research that addresses perceived privacy violations of stakeholders when abstraction techniques are being applied to process models. As of yet, it is unclear how stakeholders perceive a violation of privacy when these methods and abstraction techniques are being applied. And given the increased focus on privacy and the fast growth of the process mining market, it is vital that more research is conducted to process mining privacy from a stakeholder's perspective. To this end, the objective of this research is to identify and evaluate perceived privacy with respect to different abstraction levels from a stakeholder's perspective.

This paper is one of the first studies that investigates perceived privacy in process mining from a stakeholder's perspective. Given the novelty of the topic and thus the lack of theories and experiences, we conducted an explorative pilot-study. In this study, perceived privacy is identified and evaluated by showing 6 process mining stakeholders 3 different process model visualisations of a travel expense claim process. The process mining software tool Disco is used to generate the process model visualisations. When changing the abstraction level of the visualisations, the 6 stakeholders are asked whether they perceive a violation of privacy. Also, the stakeholders are asked to give recommendation on how to make the process mining visualisations more privacy proof. The remainder of this research paper is structured as follows; Section 2 highlights the background research, and section 3 will discuss the methodology used. Section 4 contains results and section 5 contains the discussion and future work.

2 THEORETICAL BACKGROUND

In this section we define perceived privacy and survey which privacy preserving techniques for process mining generated by process mining algorithms are currently available in scientific literature.

2.1 Process Models and Perceived Privacy

Westin (1967) defines privacy as "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (as cited by Könings et al., 2016). The definition of privacy differs from the definition of perceived privacy, which is the individual's perception that their personal information is safe from potential compromise (Johnson et al., 2020). The risk of violating perceived privacy is measured by identifying participant's willingness to share their data (Bhatia & Breaux, 2018). In process models that are generated using process mining algorithms, various abstraction techniques can be applied to reduce the privacy risk (Elkoumy et al., 2022, Pika et al., 2020). The majority of process mining abstraction techniques focus on the data preparation phase and the data visualisation phase (Maneschijn et al., 2022). It is important to note that these techniques aim at reducing risk of privacy, but not on how reduced privacy risk is being *perceived* by stakeholders.

2.2 Process Models Abstraction at the Data Preparation Phase

Several abstraction techniques for process models exist to reduce the privacy risk. First, process models are abstracted by using external domain knowledge taken from existing process documentation to semi-automatically match events and process activities (Baier et al., 2014). Second, clustering of data, supervised learning techniques and behavioural pattern modelling are used to reduce the privacy risk in the data preparation phase (Diba et al., 2020, Rafiei & van der Aalst, 2020, Bose & van der Aalst, 2009). Third, Rafiei & van der Aalst (2021) describe various group-based privacy preservation techniques to improve privacy in the data extraction and preparation phase. These are k-anonymity (remove all the trace variants occurring less than k times), and l-diversity (reducing granularity of data representation) and t-closeness (further refinement of l-diversity).

Also, various algorithms and web-based tools exist to anonymize data for process mining, hence reducing the privacy risk. These are among others PC4PM (Rafiei et al., 2021) and PRETSA (Fahrenkrog-Petersen, 2019). Sohail et al. (2021) mention that higher preservation of privacy is realized by applying various noise-adding plugins in the academic process mining tool PRoM. These plug-ins

add noise to log filters, add k-frequent randomly positioned activities, filter out high-frequency activities and reverse process traces. Furthermore, Van Zelst et al. (2021) identify seven different dimensions of event abstraction techniques in the data preparation phase of process mining. These are based on supervision strategy, fine-granular event interleaving, probabilistic nature of the outcome, data nature, use of alternative perspectives, event class/ activity class relation and event instance/activity instance relations. Finally, Pika et al. (2020) state that case identifiers and activity labels can be encrypted in data to increase privacy.

The abovementioned abstraction techniques come with several challenges. When leaving out details in a process mining analysis or process data, the utility of identifying outliers and variances and determining on concrete measures to address these outliers and variances may be affected. (Mannhardt et al., 2018, Rafiei & van der Aalst, 2020, Elkoumy et al., 2022). Also, the identity of a data subject might be at risk and can be recognized or identified when singling out individuals from a supposedly anonymized or pseudonymized event log or when applying encryption (Elkoumy et al., 2022, Mannhardt et al., 2018). Remarkably, all the mentioned abstraction techniques have hardly been validated by process stakeholders in organisations. Hence, they do not focus on how these techniques are *perceived* by process mining stakeholders as techniques of reducing privacy risk.

2.3 Process Models Abstraction at the Data Visualization Phase

Process model abstraction in the analysis and data visualisation phase of a process mining project can be performed using various options of the academic process mining tool PRoM. The fuzzy miner abstraction pattern plug-in uses conflict- and edge resolution to remove insignificant lower-level information in the process map visualization (van der Aalst and Günther, 2007). Using the fuzzy miner, Maneschijn et al. (2022) defined four threshold abstraction levels (A, B, C and D) based on the utility ratio, edge cut-off score and node cut-off score in process mining models. Mannhardt et al. (2018) have created guidelines for using process mining in human-centered industrial environments. These guidelines focus on informed consent of data subjects of among others which data is stored, withdrawal of data, but also of having the option to change the aggregation settings when using or analysing the data by primary users. Finally, Rozinat (2017) provides practitioner experience reports on privacy of process mining in an organizational setting. These are clarification of goals to various stakeholders, have external parties sign an NDA, and use encrypted analysis to ensure responsible process mining and reducing privacy issues at the data analysis and visualization phase. However, advice given in the consultancy papers of Rozinat (2017) lack empirical foundation and scientific validation.

The available literature about reducing the privacy risk when conducting process mining in the data visualisation phase is much scarcer than the available literature applied to the data preparation phase. And the technical possibilities that do exist to reduce the privacy risk focus mainly on PRoM. This tool is mostly used for research purposes and is not aimed at organisations and business. Abstraction techniques that go beyond the technical possibilities presented are scarce and lack scientific validation. In addition, even when applying the abovementioned techniques, it remains unclear how stakeholders of a process mining project *perceive* privacy and potential privacy violations. And given the fast growth of the process mining market and the high focus on privacy, it is vital to understand and act on how various process mining stakeholders perceive privacy.

3 METHODOLOGY

In this explorative pilot-study which is executed at Saxion University of Applied Sciences (UAS) in the Netherlands, perceived privacy in the data visualization phase of a process mining project is identified and evaluated. In order to follow a structured research approach, the PM2 process mining methodology of van Eck et al. (2015) is used. This methodology includes a complete overview of the steps to be executed when conducting a process mining analysis. These steps are planning, data extraction, data processing, analysis, evaluation and improvement.

3.1 Planning

The planning phase of this research includes a literature study, process selection, and stakeholder identification and selection. The results of the literature study to process mining privacy and abstraction described in section 2 of this research paper. The travel expense claim process is selected as the process to visualize. This process has been selected because of data availability, and it is assumed that this process is relevant, understandable and

recognizable for the interviewees.

Next, relevant stakeholders of the travel expense claim process are identified and selected. These stakeholders are firstly identified in the PM2 methodology of van Eck et al. (2015). These stakeholders are a business expert having knowledge of the process and a data analyst who prepares the data, creates the visualizations and analyses the results. In addition, stakeholder identification took place by identifying the stakeholders available in the travel expense process dataset used in this research. This dataset contains an employee, supervisor, administrator, budget owner and a director. An elaboration on the dataset can be found in section 3.2 and 3.3. At Saxion UAS the supervisor is the budget owner. Therefore, the budget owner as a separate stakeholder will not be included in this research. In addition, the director at Saxion UAS is not directly involved in the travel expense claims process. Therefore the director will also not be considered as a stakeholder in this research. The stakeholders that are identified and taken into account in this research are the employee, supervisor, administrator and the analyst. Purposeful sampling is used to select the interviewees from the direct network of the researchers. The participants are known to be involved in the travel expense process at Saxion UAS. Although the sampling method has generalizability issues, purposeful sampling is not costly and not time-consuming, hence fits with the explorative nature of this research (Stratton, 2021).

3.2 Data Extraction

The second step is data extraction. Due to confidentiality reasons, no permission is given to use data of travel expense claims from Saxion UAS. Therefore, the publicly available travel expense process data of the BPI challenge of 2020 (van Dongen, 2020) is used to visualise the travel expense claim process. The five available CSV datafiles of this process contain events pertaining to two years of travel expense claims of Eindhoven University of Technology. In 2017 events are collected for two departments, in 2018 for the entire university. The datafiles are explored by converting the CSV datafiles to an Excel file and compared with each other on similarities, differences, and relevance of included variables. Because the datafiles show overlap in included variables, only the datafile Permit log has been selected. This datafile includes most variables with personal information from stakeholders, such as variables of total declared amount, requested expense, and overspent amount. These variables are not always completely included in the other datafiles.

3.3 Data Processing

Data processing consists of data cleaning. By placing a filter on the variables of the Permit log datafile, the quality of the process data is investigated. The data is cleaned by removing variables without attributes, empty rows, and columns. Variable formats are adjusted to be able to fit the process mining tool. The result of the data cleaning process is a datafile containing 86,582 rows of process data. In this research the process mining software tool Disco is used for creating the process visualisations. Disco focuses on enterprises and we are investigating perceived privacy from an organisational stakeholder perspective. The cleaned data is imported to Disco.

3.4 Analysis

In the analysis phase the process model visualisations were created using Disco and shown to the 6 interviewees. The interviews took place in June 2022 at Saxion UAS. In total 4 employees and 2 supervisors are interviewed. The administrator and analyst are excluded from this research due to their unwillingness to participate. The interviewees all work at the same academy at Saxion UAS. Because of the confidentiality and traceability of the interviewees, more information on the interviewees will not be provided.

To identify whether the interviewees perceive privacy violations, 3 main process visualizations were shown to the interviewees. Following the interface design of Disco, these visualizations contain a process map with frequency statistics and performance statistics, a statistics overview with variable frequency, and a statistics overview showing the process data variables per variant and per individual case. When the interface settings allowed it, the abstraction levels of the visualisations were modified. Table 1 contains the methods used in this research to change the abstraction level of process mining visualisations in Disco. These alterations in abstraction level are described in the literature review as l-diversity (Rafiei & van der Aalst, 2021), conflict and edge resolution (van der Aalst & Günther, 2007) and changing the aggregation settings of a process visualization (Mannhardt et al., 2018). The numbers in the table correspond with the available abstraction techniques in Disco as shown in figure 1 in section 4.1.

Following the measurement of perceived privacy by Bhatia & Breaux (2018), the interviewees were asked whether they perceived a violation of their privacy when the visualizations would be shown to their supervisor, fellow employees, administrator or process analysts. The same question was asked after the abstraction level of the process visualization changed. Next, the interviewees were asked for suggestions on how to reduce the privacy risk of the visualizations when they perceived such a risk.

3.5 Evaluation and Improvement

Per visualization, answers given by the interviewees on perceived privacy, perceived privacy violations and suggestions to reduce the privacy risk were described and compared. The results of this evaluation can be found in section 4.

4 RESULTS

In this section the interview results per process visualization which follows the Disco interface design are described and compared.

4.1 Process Map Visualisation

Figure 1 shows the visualization of the travel expense claim process that was shown to the interviewees, hence executing the analysis phase (step 4) of the research methodology. The numbers 1 until 4 in figure 1 are added manually and correspond to the methods to change the abstraction level of the process model as described in table 1.

Table 1: Methods to change the abstraction level of a visualisation in Disco.

Nr.	Methods
1	Activity slider to reduce the number of shown activities in a process map visualisation
2	Path slider to reduce the number of shown paths in a process map visualisation
3	Showing frequency statistics (yes/no)
4	Showing performance statistics (yes/no)

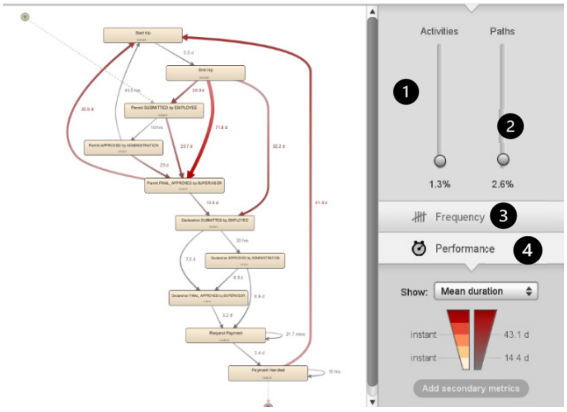

Figure 1: Process map of a travel expense claim process.

Results show that the 6 interviewees differ in how they perceive privacy and whether they perceive a violation of privacy when the abstraction level of the process map changed. Two employees perceived no violation of privacy when sharing the process map visualization with employees, supervisors, administrators and analysts. These employees claim that sharing such a visualization that includes their own data is beneficial for the organization as it helps to detect fraud. In addition, these interviewees mentioned that sharing the process map with all process stakeholders increases transparency in the organization. Changing the level of abstraction in the process map via the activity slider and path slider, or showing the frequency- or performance statistics did not yield different results. Also, these interviewees did not perceive differences in perceived privacy or any violations of privacy if cases in the process map could be linked to individuals.

The other two employees perceived a violation of their privacy when it was possible to identify individual cases. This privacy violation was only perceived if the process map would be shown to employees. Showing the process map including traceability to individual cases to supervisors, administrators or analysts did not lead to a perceived privacy violation. According to these interviewees, working with individual process data is considered part of the job of these stakeholders. The activity slider and path slider were considered suitable methods to make the process map more privacy proof as long as individual cases could not be logically identified and traced back to individuals. Showing the frequency statistics and performance statistics to all stakeholders was not perceived as a violation of privacy. The two supervisors agreed by stating that privacy is not being violated when cases cannot be traced back to individuals by employees.

4.2 Statistics Interface

The statistics interface of Disco shows frequency statistics and relative frequency statistics per variable in the dataset. The two interviewees that did not perceive any violations of their privacy at the process map visualization also did not perceive privacy violation when the statistics interface was shown to stakeholders. Instead, perceived organizational benefits such as fraud detection and increased transparency were perceived. One of these two employees did mention that when names of people are included in the resource stream while this was not pre-discussed in advance, this employee would perceive a violation of privacy when the interface was shown to employees.

The remaining two employees perceived a privacy violation when the statistics interface was shown to employees and contained an employee number, declaration number, organizational entities and declaration amounts. Especially in a small team the interviewees expected that declaration amounts could be traced back to individuals. According to the interviewees this is undesirable, because it is unclear where the information ends up and what the intention is of people that share this information. The interviewees expect that fellow colleagues are going to form an opinion on the declarations, resulting in gossip and rumor that the employees feel uncomfortable with. The employees mentioned that when managers, analysts and administrators see the statistics interface with all its variables, privacy was not being violated. Working with the data is considered part of the job.

The two interviewed managers mentioned that privacy of employees is being violated when sharing declarations and amounts with employees if the data can be traced back to an individual employee or user ID. On the one hand it was mentioned that employees could use this information for purposes other than the process optimization, but on the other hand it was mentioned that when working towards a more transparent organization where teams want to improve their own processes, then this information must all be shared.

4.3 Cases Interface

The Cases interface of Disco shows all individual attributes of the variables per process variant and per case. At the Cases interface, more interviewees perceived a violation of their privacy. Only one respondent mentioned that showing individual cases to all stakeholders was solely beneficial for transparency reasons and detecting fraud. Also, showing the cases visualization to all stakeholders was perceived to accomplish a damping effect when submitting a travel expense claim. The interviewee considered this to be positive for the financial situation of the organization. In addition, it was expected that when people know that everyone could see the data, then people would behave better and more fairly.

The remaining employees felt that their privacy was being violated when the variables and attributes of individual cases was being shared with direct colleagues (employees) and analysts. This mainly goes for variables that are related to identify (resource and organizational entity) and variables related to declaration amount. The main arguments mentioned for privacy violation were that having this information is not part of their jobs and uncertainty on the ethical behavior of colleagues.

The interviewed supervisors agreed with the majority of the employees and mentioned that the cases interface should only be accessible to managers and administration. The main reason mentioned was that insight on case level is needed to help to detect fraud. Privacy is considered being violated when analysts and employees see the individual cases and their attributes. Especially when these stakeholders see the resources and amount, individual privacy is being violated.

The findings on whether each individual interviewee perceived privacy violations (yes or no) are summarized in table 2. In this table a distinction has been made between the visualizations that were shown to the interviewees.

Table 2: Perceived privacy violations per interviewee.

Interviewee	Process map	Statistics	Cases
Employee 1	No	No	No
Employee 2	No	No	Yes
Employee 3	Yes	Yes	Yes
Employee 4	Yes	Yes	Yes
Supervisor 1	Yes	Yes	Yes
Supervisor 2	Yes	Yes	Yes

4.4 Reducing the Privacy Risk

Various methods were suggested by the interviewees to reduce the privacy risk of the process visualizations. First, various employees and supervisors suggested to include an authorization table in the process mining tool. This is expected to

prevents undesired curiosity by stakeholders. In addition, an authorization table is expected to help in only showing travel expense claim information to stakeholders that need information on case level to execute their jobs. To improve transparency in the organization and eliminate perceived privacy violations, it was suggested by the supervisors to communicate and discuss the goal and output characteristics of the process mining visualization with all process stakeholders. Objections to the analyses and visualizations can be discussed in advance. Validating the analysis with a privacy officer to verify whether the analysis adheres to the GDPR regulations and communicating about the validation with all involved stakeholders is also mentioned as an option to eliminate perceived privacy violations. Pseudonymizing or anonymizing the data in the statistics interface or cases interface was not considered a suitable option by all interviewees to eliminate perceived privacy violations. As the team in which the employees and supervisors operate is relatively small, the interviewees mentioned that it is highly likely that pseudonymized or anonymized data can be traced back to individuals. In addition, pseudonymization and anonymization is expected to make it harder for supervisors and administrators to do their job and detect fraud.

5 DISCUSSION AND FUTURE RESEARCH

The objective of this research was to identify and evaluate perceived privacy with respect to different abstraction levels from a stakeholder's perspective. The results show that there are differences in whether individual stakeholders perceive privacy violations when they are shown process models generated via process mining techniques. Perceived privacy violations vary between type of visualization, type of technical privacy reducing method, change of abstraction level and also between role of the stakeholder.

This research has practical relevance as it shows the importance of addressing perceived privacy with process mining stakeholders in organizations. Identifying difference and similarities in how various process mining stakeholders perceive privacy could help organizations decide which privacy preserving methods and techniques can and cannot be applied. Communicating with these stakeholders on perceived privacy and taking actions to reduce or prevent privacy violations could benefit the adoption of process mining in organizations. This research is scientifically relevant as it is one of the first research conducted on stakeholders perceive process mining in organizations. Not all currently available techniques to reduce the privacy risk are perceived by stakeholders as techniques that actually reduce the privacy risk. Suggestions given by the interviewees to reduce the privacy risk were only a fraction of the identified techniques in literature to reduce this risk. To prevent developing techniques that are aimed at reducing privacy risk but are not considered as such by stakeholders in organizations, it is suggested to keep validating privacy risk reducing techniques with process mining stakeholders in organizations. This validation can contribute to designing techniques to reduce the privacy risk of process mining that are valued by stakeholders in organizations. Also, the PM2 process mining methodology of van Eck et al. (2015) does not include privacy in relation to a process mining projects. This research shows that privacy is of importance when executing a process mining project. Therefore it is suggested that future research on process mining methodologies takes privacy into account.

The main limitation of this research is the sample size of 6 interviewees. Based on the rule of logic it is possible to deduce knowledge on how various process mining stakeholders perceive privacy. But more research on perceived privacy related to process mining is needed to justify and develop claims that are generalizable to other processes, organisations, stakeholders and process mining tools. Furthermore, no validated questionnaire or list of interview questions exist on perceived privacy violations in relation to process mining. As a result, a list of questions was developed based on the available interfaces of Disco. This method has issues related to construct validity. Future research should emphasize on expanding the questionnaire to other tools and contexts to gain more in-depth knowledge on perceived privacy related to process mining.

REFERENCES

Baier, T., Mendling, J., & Weske, M. (2014). Bridging abstraction layers in process mining. *Information Systems*, *46*, 123–139. https://doi.org/10.1016/j.is.2014.04.004

Bhatia, J., & Breaux, T. D. (2018). Empirical measurement of perceived privacy risk. *ACM Transactions on Computer-Human Interaction*, *25*(6). https://doi.org/10.1145/3267808

Bose, R. P. J. C., & Van Der Aalst, W. M. P. (2009). Context aware trace clustering: Towards improving process mining results. *Society for Industrial and Applied Mathematics - 9th SIAM International Conference on Data Mining 2009, Proceedings in Applied Mathematics*, *1*, 397–408. https://doi.org/10.1137/1.9781611972795.35

Diba, K., Batoulis, K., Weidlich, M., & Weske, M. (2020). Extraction, correlation, and abstraction of event data for process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–24. https://doi.org/10.1002/widm.1346

Elkoumy, G., Fahrenkrog-Petersen, S. A., Sani, M. F., Koschmider, A., Mannhardt, F., Von Voigt, S. N., Rafiei, M., & Waldthausen, L. Von. (2022). Privacy and Confidentiality in Process Mining: Threats and Research Challenges. *ACM Transactions on Management Information Systems*, *13*(1), 1–17. https://doi.org/10.1145/3468877

Fahrenkrog-Petersen, S. A. (2019). Providing privacy guarantees in process mining. *CEUR Workshop Proceedings*, *2370*, 23–30.

Gartner. (2020). Market Guide for Process Mining. *Gartner*, *September*, 1–33. https://www.gartner.com/doc/reprints?id=1-SBXXPQO&ct=190625&st=sb

GDPR. (2022). *GDPR*. General Data Protection Regulation. https://gdpr.eu/

Grisold, T., Mendling, J., Otto, M., & vom Brocke, J. (2021). Adoption, use and management of process mining in practice. *Business Process Management Journal*, *27*(2), 369–387. https://doi.org/10.1108/BPMJ-03-2020-0112

Hill, M., & Swinhoe, D. (2021). *The 15 biggest data breaches of the 21st century*. https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html

Johnson, V. L., Woolridge, R. W., Wang, W., & Bell, J. R. (2020). The Impact of Perceived Privacy, Accuracy and Security on the Adoption of Mobile Self-Checkout Systems. *Journal of Innovation Economics & Management*, *n°31*(1), 221. https://doi.org/10.391 7/jie.pr1.0065

Könings, B., Schaub, F., & Weber, M. (2016). Könings, B., Schaub, F., & Weber, M. (2016). Privacy and trust in ambient intelligent environments. *Next Generation Intelligent Environments (Pp. 133-164). Springer, Cham.*

Maneschijn, D. G., Bemthuis, R. H., Bukhsh, F. A., & Iacob, M. E. (2022). A Methodology for Aligning Process Model Abstraction Levels and Stakeholder Needs. *In ICEIS (1) (pp. 137-147).*

Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., & Michael, J. (2019). Privacy-Preserving Process Mining: Differential Privacy for Event Logs. *Business and Information Systems Engineering*, *61*(5), 595–614. https://doi.org/10.1007/s12599-019-00613-3

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

Pika, A., Wynn, M. T., Budiono, S., Hofstede, A. H. M. T., van der Aalst, W. M. P., & Reijers, H. A. (2020). Privacy-preserving process mining in healthcare. *International Journal of Environmental Research and Public Health*, *17*(5), 1–12. https://doi.org/10.3390/ijerph17051612

Polyvyanyy, A., Smirnov, S., & Weske, M. (2015). Business process model abstraction. *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*, 147–165. https://doi.org/10.1007/978-3-642-45100-3_7

Rafiei, M., Schnitzler, A., & van der Aalst, W. M. P. (2021). PC4PM: A tool for privacy/confidentiality preservation in process mining. *CEUR Workshop Proceedings*, *2973*, 106–110.

Rafiei, M., & van der Aalst, W. M. P. (2020). Privacy-preserving data publishing in process mining. *Lecture Notes in Business Information Processing*, *392 LNBIP*, 122–138. https://doi.org/10.1007/978-3-030-58638-6_8

Rafiei, M., & van der Aalst, W. M. P. (2021). Group-based privacy preservation techniques for process mining. *Data and Knowledge Engineering*, *134*(April), 101908. https://doi.org/10.1016/j.datak.2021.101908

Rozinat, A. (2017). *Privacy, security and ethics in process mining*. https://fluxicon.com/blog/2017/11/privacy-security-and-ethics-in-process-mining-part-2-responsible-handling-of-data/

Stratton, S. J. (2021). *Population research: convenience sampling strategies.* Prehospital and disaster Medicine, 36(4), 373-374

Sohail, S. A., Bukhsh, F. A., & van Keulen, M. (2021). Multilevel privacy assurance evaluation of healthcare metadata. *Applied Sciences (Switzerland)*, *11*(22). https://doi.org/10.3390/app112210686

van der Aalst, W. M. P. (2016). *Process Mining Data Science in Action*. Springer.

van der Aalst, W. M. P., & Günther, C. W. (2007). Finding Structure in Unstructured Processes: The Case for Process Mining. *Proceedings - 7th International Conference on Application of Concurrency to System Design, ACSD 2007*, *Acsd*, 3–12. https://doi.org/10.1109/ACSD.2007.50

van Dongen, B. (2020). *BPI Challenge 2020. 4TU.ResearchData. Collection.* https://doi.org/10.4121/uuid:52fb97d4-4588-43c9-9d04-3604d4613b51

van Eck, M., Lu, X., Leemans, S., & Van Der Aalst, W. M. P. (2015). PM2: a process mining project methodology. *International Conference on Advanced Information Systems Engineering*, 297–313.

van Zelst, S. J., Mannhardt, F., de Leoni, M., & Koschmider, A. (2021). Event abstraction in process mining: literature review and taxonomy. *Granular Computing*, *6*(3), 719–736. https://doi.org/10.1007/s41066-020-00226-2

Westin, A. F. (1967). *Special report: legal safeguards to insure privacy in a computer society*. Communications of the ACM, 10(9), 533-537.