Enterprise Modelling and Information Systems Architectures
Vol. 15, No. 8 (2020). DOI:10.18417/emisa.15.8

Trust and Privacy in Process Analytics

Felix Mannhardt[*,a], Agnes Koschmider[b], Lars Biermann[c], Jana Lange[d], Florian Tschorsch[e], Moe Wynn[f]

[a] Eindhoven University of Technology, Netherlands
[b] Kiel University, Germany
[c] Deloitte München, Germany
[d] Celonis, Germany
[e] Technical University of Berlin, Germany
[f] Queensland University of Technology, Australia

Abstract. *This paper summarizes the panel discussion at the 1st Workshop on Trust and Privacy in Process Analytics (TPPA) co-located with the 2nd International Conference on Process Mining. The panel discussed to what extend trust and privacy is embedded in applications of process mining and took place on 5th October 2020. The virtual session was chaired by Felix Mannhardt and Agnes Koschmider and the invited panelists were Moe Wynn, Jana Lange, Lars Biermann and Florian Tschorsch. The major challenges that this panel identified related to privacy-preserving process mining are to include (user-centric) privacy filters, understanding the privacy-utility trade-off and to link privacy-preserving techniques with dataset quality.*

Communicated by Agnes Koschmider.

1 Introduction

Privacy and trust are two concepts that are closely linked to the responsible application of data science and have been less represented in the research on process mining of the last years. Whereas process mining has been successfully applied in analyzing and improving processes based on event logs in all kinds of environments, there was less focus on the possibly negative impact of such analysis on participants of a process. Here, privacy relates to the concern that event logs may contain personal data of both customers and employees and the challenge of protecting the information about individuals while still being useful for process mining. Often, security aspects are closely connected when the processing of personal data cannot be avoided; however, they form a different concern. Trust is required both from the perspective of trust in organizational and technological measures that event logs are not misused (e. g., for worker surveillance) as well as from the perspective of trust that the results of a process mining analysis faithfully reflect reality (e. g., data quality, traceability, auditability).

In this paper, we summarize the panel discussion at the 1st Workshop on Trust and Privacy in Process Analytics (TPPA). The workshop, which was held co-located with the 2nd International Conference on Process Mining, held virtually due to the ongoing COVID-19 pandemic, was concluded with a 1-hour panel discussion. This discussion brought together selected experts on trust and privacy topics from academia and industry. Four panelists contributed experience on the application of process mining from the consulting perspective (Lars Biermann, Deloitte) and from the vendor perspective (Jana Lange, Celonis), as well as experience from research on privacy-preserving methods (Florian Tschorsch, TU Berlin) and research on data quality in process mining (Moe Wynn, QUT).

* Corresponding author.
E-mail. f.mannhardt@tue.nl

The starting point of the panel was the question: "*Is there something special about Process Mining in comparison to other projects from the Date Science domain when it comes to Trust and Privacy, and if so, what is it?*". Among the participants there was agreement that the special ingredient is largely the type of data that is required for process mining along with the focus on end-to-end business processes improvement. This means that the data used contains information on how people in different roles in an organization work, information on how customers interacted with the process, and information on how the business performs. All this information can be sensitive, and it was recognized that it needs to be protected and that some protection measures are in place in current projects. However, there was disagreement on what kind of measures are necessary and to which degree technological measures are required or organizational measures suffice to protect personal data. One argument brought forward was that often the data used for process mining already exists in the organizations and that if the analysis is done on-premise, there are less concerns regarding privacy. There is more consideration on this topic when it comes to storing data in a cloud environment and even larger concerns when it comes to employee data.

In the discussion, the role of work councils in a process mining project was brought up. Specifically, in regions with a strong employee-friendly policy such as Germany distrust regarding the potential misuse of results comes sometimes from the work council. It was noted that this was not the case in other legislations and cultures, which do not offer such employee protection. In practice, such concerns are usually handled by organizational rather than technical measures. So far and at least in regions where work councils are involved, there are no questions raised towards identifying poorly performing individuals. Thus, it is generally trusted that data is not misused. In this context, it was noted that it would be better to have systems that would not rely on trust in organizational measures, i. e., where it is impossible to misuse the data. However, it was acknowledged that some kind of trust relationship is often still required. Overall, there seemed to be a large gap between research on privacy-preserving process mining methods (k-anonymity, differential privacy, etc.) and what is used and deemed sufficient in practice (pseudonymization).

A brief discussion on the actual re-identification risk of individuals that took part in a process from an event log in a practical setting followed. Whereas a recent paper highlighted the large theoretical re-identification risk of individuals that are connected to process traces (Voigt et al. 2020), highlighted the large theoretical re-identification risk of individuals that are connected to process traces, it is unclear how this risk is perceived in practice. Compared to problematic scenarios pictures in research papers, the reality of analyzing the 'Account Payable Process' in a large organization with hundreds of workers, seems less problematic. Picking out individual employees in a large dataset of a large organization was regarded as difficult and, as noted before, generally not the aim of any project.

This gave rise to the question whether there are differences between large and small sized organizations regarding the privacy risks. Whereas in large organizations, employees are naturally less easy to identify from a cursory look at process mining results, this is less so in medium sized organizations. Indeed, the adoption of process mining in smaller organizations is still lagging. It was raised as question from the audience whether providing privacy-preserving methods with certain guarantees (such as differential privacy or k-anonymity) would help selling process mining in these cases. While this possibility was not discarded, it was noted that other concerns, such as the costs of a process mining solution, would likely play a larger role.

Finally, in conclusion of the panel we want to give each of the panelist the opportunity to concluded with commenting the question: *What future research directions should be explored for Trust and Privacy concerns in Process Mining in their opinion?*

Privacy filter
Privacy of each individual has to be safeguarded at all times. Research on establishing data safes that allow for pseudonymization of data to relate a transaction to a certain department while not being able to analyze the individual could add value. Preventing the setting of certain filters when the analyzed population becomes too small (below five individuals) could prevent the involuntary breach of privacy. (Lars Biermann)

Privacy-preserving algorithm
Privacy is something we need to take care of while doing process mining. One of the future research directions explored for trust and privacy concerns in process mining should be related to finding an algorithm on data in order to select and restrict certain data sets to avoid privacy issues which creates more transparency and finally leads to users and all stakeholders involved trusting process mining (more). This does not exclude a final qualitative human intervention and judgment. (Jana Lange)

Understanding privacy-utility trade-off
First and foremost, we should acknowledge the inevitable privacy loss when processing and analyzing personal data. Privacy-enhancing technologies are able to limit the risks but not able to avoid them altogether. That said, I see two general directions to address this issue: From a technical perspective, we need more self-determined solutions, e. g., controlling and processing data locally in the domain of the data subject. From an organizational perspective, data collectors and analysts should be aware of the responsibility and clearly follow a data minimization strategy. In the end, we need a better understanding on the trade off between the expected insights we gain from process analytics and the involved privacy risks to advance this field. (Florian Tschorsch)

Correlate dataset and privacy-preserving process mining
As process mining researchers, we need to understand where privacy concerns could arise from within a data set (e. g., customer data, employee details, date/time stamps) and how to address these concerns during the data pre-processing phase. We should not naively apply a privacy-preserving technique to a dataset without considering how this dataset will be utilized for process mining. We should develop new data transformation approaches for process mining that incorporate different privacy-preserving techniques and quantify the utility loss. Privacy concerns may also be related to the type of process mining analysis being proposed (process discovery vs resource profiling). To build users' trust in process mining insights, we should strive to make the process mining analysis steps transparent and process mining insights explainable and understandable. (Moe Wynn)

2 Biography Panelists

Lars Biermann (Deloitte) is working in consulting for more than 15 years. During this time a significant number of his projects were dedicated to process optimization. Since using Process Mining for the first time in the course of such a project he developed a dedication to Process Mining and consequentially joined the Deloitte Center for Process Bionics shortly after that. Since than he has delivered multiple Process Mining projects across different industries like Banking, Logistics, Manufacturing or Retail.

Jana Lange (Celonis) is IT Risk Manager at Celonis. Based at their headquarters in Munich, she is currently implementing IT Risk Management related aspects into global business processes helping various stakeholders improve their security. Given her background as Data Protection Officer combined with her dedication to process optimization, she always takes a balanced approach with regards to security and practicability aspects. She strives to constantly sharpen people´s view on the topic´s importance in a digitized and increasingly cloud-based environment. Jana is member of the ISACA Germany Chapter.

Florian Tschorsch (Technical University of Berlin) is assistant professor at the Technical University of Berlin (TU Berlin) and the Einstein Center Digital Future (ECDF), where he heads

the Distributed Security Infrastructures group. In his research, he strives for the integration of security and privacy aspects in distributed system architectures and networking protocols. In particular, he is interested in the security-privacy-performance tradeoffs resulting from application-specific constraints. His application areas comprise, among others, privacy-preserving telemetry and blockchain technologies.

Moe Wynn (Queensland University of Technology) leads the Business Process Management (BPM) research group at Queensland University of Technology (QUT). She is a co-leader within QUT's Tier 1 Centre for Data Science (Data for Discovery Theme). She completed her PhD in the area of workflow management in 2007 from QUT. Her ongoing research focuses on process-oriented data mining (process mining), data quality and robotic process automation for the digital transformation of processes. She has over twelve years of experience in engaging with Australian industry partners to improve business practices through data-driven methods.

References

von Voigt S. N., Fahrenkrog-Petersen S. A., Janssen D., Koschmider A., Tschorsch F., Mannhardt F., Landsiedel O., Weidlich M. (2020) Quantifying the Re-identification Risk of Event Logs for Process Mining - Empiricial Evaluation Paper. In: Advanced Information Systems Engineering - 32nd International Conference, CAiSE 2020, Grenoble, France, June 8-12, 2020, Proceedings. Lecture Notes in Computer Science Vol. 12127. Springer, pp. 252–267