# Responsible Process Mining

Felix Mannhardt(✉)

Eindhoven University of Technology, Eindhoven, The Netherlands
f.mannhardt@tue.nl

**Abstract.** The prospect of data misuse negatively affecting our life has lead to the concept of responsible data science. It advocates for responsibility to be built, by design, into data management, data analysis, and algorithmic decision making techniques such that it is made difficult or even impossible to intentionally or unintentionally cause harm. Process mining techniques are no exception to this and may be misused and lead to harm. Decisions based on process mining may lead to *unfair* decisions causing harm to people by amplifying the biases encoded in the data by disregarding infrequently observed or minority cases. Insights obtained may lead to *inaccurate* conclusions due to failing to considering the quality of the input event data. *Confidential* or personal information on process stakeholders may be leaked as the precise work behavior of an employee can be revealed. Process mining models are usually white-box but may still be difficult to interpret correctly without expert knowledge hampering the *transparency* of the analysis. This chapter structures the topic of responsible process mining based on the FACT criteria: Fairness, Accuracy, Confidentiality, and Transparency. For each criteria challenges specific to process mining are provided and the current state of the art is briefly summarized.

**Keywords:** Fairness · Accuracy · Confidentiality · Transparency

## 1 Introduction

Data-based decisions affect our society and our daily life. Organizations leverage data to obtain *objective insights* that are based on *facts* rather than on guesswork. Being data-driven to guide decisions is in itself hardly new and, certainly, decisions should be based on data rather than being based on arbitrary factors. In fact, the scientific method itself is based on meticulously analysing data to derive trustworthy conclusions.

What changed in recent years, and is increasingly changing every aspect of our life, is the abundance of data and compute power available to most people and organizations. The capability of collecting and analysing a large amount of data is now within the reach for most organization. What used to be a costly and time consuming operation involving a great degree of planning what data to be collected and what methods to build, can now be done ad-hoc on large amounts of *stockpiled* data.


W. M. P. van der Aalst and J. Carmona (Eds.): Process Mining Handbook, LNBIP 448, pp. 373–401, 2022.
https://doi.org/10.1007/978-3-031-08848-3_12

This abundance of data together with the emergence of a wide variety of analysis techniques has led to the formation of the *data science* field. Data science technique are not limited to giving decision support to human decision makers but increasingly *Artificial Intelligence* (AI) is used to automate decisions based on predictive models. *Process mining* is a *data science* method that focuses on improving an organization's processes by leveraging event logs. The core of event logs are timestamped data about all kinds of events that occur in the context of work or business processes [1]. Process mining techniques have been very successfully deployed in numerous organizations and have helped to remove inefficiencies and improve the quality of processes [2].

However, this increased use of data leads to an increased risk of creating negative effects from its usage by accidental or intentional *irresponsible usage* of data [3]. Irresponsible usage of data ranges from invading the privacy of individuals over flawed analysis of data with poor quality or inappropriate methods to unfair automated decisions of systems trained on data biased towards majority groups. The potential misuse of this power gives rise to calls for the *responsible* use of data by creating knowledge and awareness about possible negative consequences and researching technical and socio-technical solutions to prevent these negative consequences.

### 1.1 Responsible Data Science and AI

Many initiatives have called for research and development on methods that can be broadly categorized under the umbrella term *responsible data science* under which sub themes such as *responsible AI* [4] are included. Depending on the individual perspective different criteria or principles that are relevant to obtain *responsible* methods have been proposed.

- Aalst et al. and the Responsible Data Science consortium[1] call for methods that follow the FACT criteria, which stands for *Fairness*, *Accuracy*, *Confidentiality*, *Transparency* [5].
- The ACM FAccT Conference[2] calls for research on *Fairness*, *Accountability*, and *Transparency* principles.
- In Information Retrieval, the FACTS-IR critera include Fairness, Accountability, Confidentiality, Transparency, and also Safety [6].
- Dignum advoates that systems should be designed to follow the principles of *Accountability*, *Responsibility*, and *Transparency* (ART) [4].
- The European Commission provided Ethics Guidelines for Trustworthy Artificial Intelligence[3] mentioning principles such as *Human agency*, *Technical Robustness*, *Privacy*, *Transparency*, *Fairness*, and *Accountability*.

Several other organizations developing or using AI technology have published manifestos or best practices also include similar principles such as *fairness*, *privacy* or *confidentiality*, *accountability*, as well as often also *interpretability* and

[1] https://redasci.org.
[2] https://facctconference.org.
[3] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

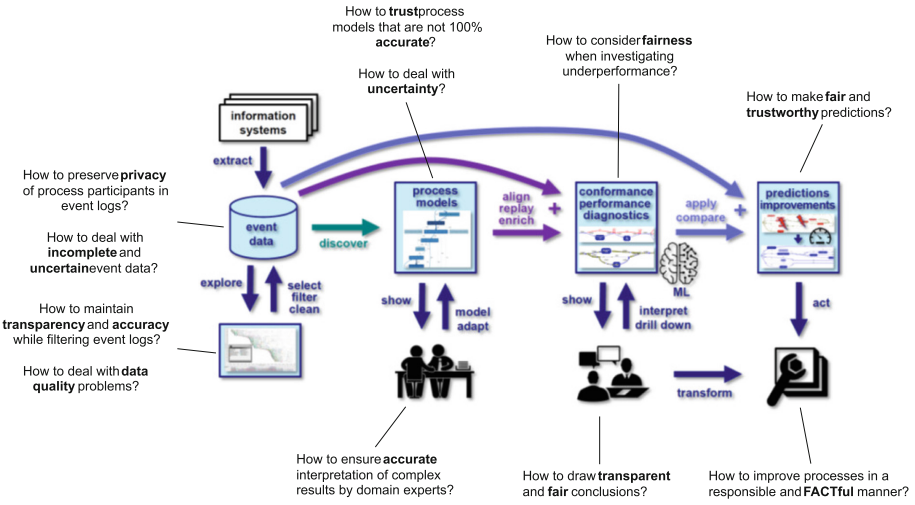


**Fig. 1.** Example challenges for responsible process mining in context of the 360 degree overview on process mining [7]

*safety*. Whereas originating from different perspectives and following slightly different definitions, there is great overlap on the major principles that are deemed relevant for leveraging data in a responsible manner. Naturally, the importance of the criteria differs depending on the application area. Considering *fairness* is crucial when designing AI systems based on machine learning that may possibly discriminate against individuals, whereas *safety* would be important when using such a system for controlling an industrial process. At the core of these "calls for action" is the realization that methods from the standard tool set of data science rarely follow all the desired criteria or principles by themselves. Additional effort is required, either by the analyst or system designer, to ensure their responsible use. This often requires ethical considerations since perfect technological solutions commonly do not exist.

## 1.2 Responsible Process Mining

This chapter instantiates the responsible data science challenges for process mining and summarises the state-of-the-art research on *responsible process mining*. Some of the challenges are specific to process mining and the event log data format whereas others are comparable to any other data science or AI approach. The context in which process mining operates means that many of the responsible data science principles and challenges are highly relevant. Figure 1 provides a non-comprehensive overview on some of the major challenges for responsible process mining in the context of the different process mining tasks. We discuss and, at least partially, answer some of these questions.

The subject of investigation in process mining is a business process, e.g., the handling of loan applications. So, the process mining analysis is not directly

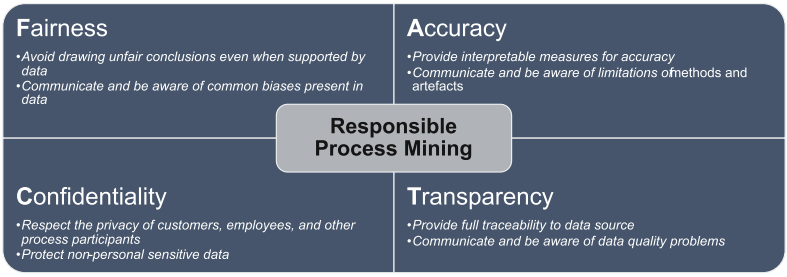


**Fig. 2.** FACT principles for responsible process mining adapted from [5]

focused on individuals. Rather it looks at the manner in which the work is organized and performed. When analysing the loan application process, event logs are commonly not used for deciding the outcome of the loan application but for deciding how to improve the handling of applications to create a better process. Here, better may refer to being more efficient, less costly, more transparent, or any other indicator of process performance. At first glance process mining seems to not have the same impact on individuals as, e.g., deploying face recognition, predictive policing, or automatically scoring applicants for a job using AI methods. However, the manner in which business processes are performed can have an effect on various stakeholders (customers, employees, etc.).

As any other data science method, process mining relies on data to reconstruct how processes were performed and how process can be improved. Thus, the results are highly dependant on the quality of the used event data and the possible biases contained. Some additional quality and *confidentiality* challenges arise from the required sequential ordering of events, grouping of events to a specific process cases, and events being related to activities. In principle, process mining aims to discover *human-interpretable* models that are supposed to be accurate and transparent. However, for *complex* process behaviour process mining techniques often attempt to generalise from incomplete and noisy data. This creates *accuracy* and *transparency* challenges even in the process mining setting.

We follow the definitions of the FACT principles brought forward in [3,5] and illustrated in Fig. 2 to structure the discussion of process mining related challenges. First, we discuss *fairness* and its relevance to process mining in Sect. 2. Then, in Sect. 3, we briefly illustrate aspects of *accuracy* including data quality and model quality. Section 4 is a major part of this chapter and is devoted to *confidentiality*, which is about protecting and respecting sensitive data in event logs including the privacy of individuals. We close the chapter in Sect. 5 with a look at *transparency* focusing on generalization and the interpretability of process mining results.

## 2 Fairness

Algorithmic fairness or fairness of automated systems [8] has been an increasingly prominent topic [9] when it comes to the development and usage of AI systems that are based on black-box machine learning models. Statistical biases embedded in training data may lead to systems making unfair decisions or clearly discriminating against certain groups of people. Prominent examples of such bias are the COMPAS system for predicting the risk of criminals to re-offend, which seem exhibit racial bias by having a higher false positive rate among blacks[4], or gender stereotypes exhibited by automated translation systems such as Google Translate, which applies male gender when translating typically male dominated job names from gender neutral Turkish to English [10]. There are many more examples and we refer to the first chapter of the Fair ML book [10] for a comprehensive introduction.

An important realization regarding bias in data and their usage in any kind of data-based system is that: "Data and data sets are not objective; they are creations of human design" [11]. Data may be incomplete for a certain context leading to *representation bias* that is reflected in the learned model or the data analysis. Even when not being incomplete, data can reinforce existing discrimination that is embodied in the available data (*historical bias*). This cannot be avoided by simply discarding "problematic" attributes from the datasets since bias may be hidden in highly correlated attributes [10]. Many more data biases can be defined depending on the context [9], a notable one being Simpson's Paradox which describe the situation that a statistic may be very different or even opposite for subgroups of a dataset compared to the statistic on the aggregate entire dataset including all those subgroups.

### 2.1 Process Mining Perspective

It seems that the discussion on algorithmic fairness is not directly relevant to process mining. The impact of process mining on individuals is usually indirect, so direct discrimination by a process mining analysis seems unlikely to occur. However, the potential reach of decision made based on process mining may have impacts on individuals. Employees working in an analysed process may be subject to unfair decision, customers may be rejected based on predictive process mining techniques, or processes may be redesigned in a way that is discriminating minorities. These are unfair results that are hidden behind the scenes and may not make headlines in the newspaper, unless discovered. Based on the process illustrated in Fig. 3, we give two examples on how fairness challenges can be part of a process mining project.

Automated decision making can be part of process mining as it may result in redesigned processes with changed decision making. As shown in Fig. 3 additional extensive checks may be added to a loan application process for certain

[4] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

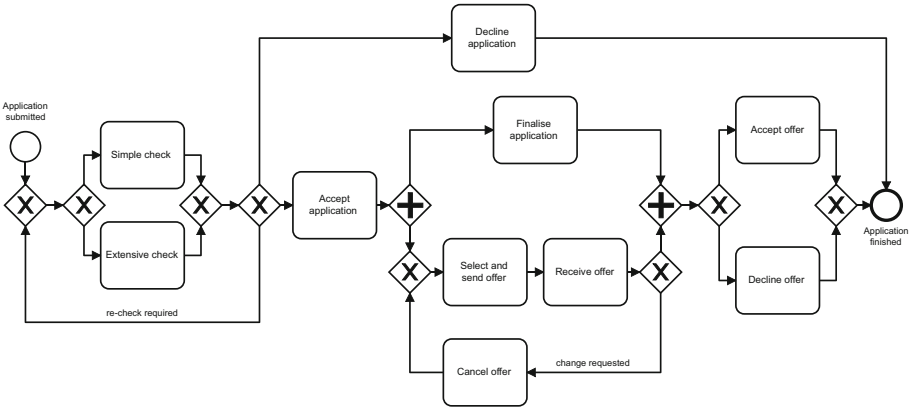


**Fig. 3.** Loan application process in BPMN adapted from the process used in [12]. Additional activities that indicate the kind of checks performed on the loan application before considering it have been added. Based on some criteria either a simple check or a more extensive check of the application is performed and in some cases the check is repeated.

cases leading to *fairness* challenges. This process re-design may be the outcome of a process mining analysis with the goal to minimize the cost of background checks. To further minimize the cost, methods for predictive process mining, action-oriented process mining, or the integration with robotic process automation is used to make the decision whether additional or extensive checks are necessary. Thus, process mining has directly affected the outcome of some process cases. Whereas the final decision is still made by a human, some applicants need to endure much more extensive background checks. This decision is based on machine learning techniques and, thus, inherits all the fairness issue associated with algorithmic decision making.

A second example of a *fairness* challenge that may arise in a process mining context would be affecting the employees working in the process. For example, it may be detected that when certain workers are involved in the processing of the loan application the throughput time is much longer. However, care must be taken not to draw unfair conclusions as those workers may simply handle more difficult cases [3], which leads to biased event data. If the nature of the loan application request is not included in the event log, e.g., due to confidentiality concerns, these confounding factors are difficult to detect and require careful human interpretation.

Besides the obvious ethical concerns that make it relevant to investigate fairness in the context of process mining, there are also upcoming regulations such as the EU Artificial Intelligence Act [13] that may constitute legal threats to consider fairness in any kind of automated data analysis. In the remainder of Sect. 2, we summarise the relevance of fairness for process mining along the main definitions that attempt to formalise fairness for algorithms. For each of the

definitions, we instantiate them in the context of process mining and summarise existing work if available.

### 2.2 Algorithmic Discrimination

In the literature on algorithmic fairness, several types of discrimination that can arise from unfair algorithms have been defined. Similarly, a wide variety of definitions on how fair algorithmic systems can be designed have been researched. We sketch the main fairness definitions and discrimination's in the light of how they are relevant to the different process mining tasks as illustrated in Fig. 1.

Many possible types of discrimination are possible. It is important to realize that discrimination or unfairness does not always need to be caused by direct discrimination [9]. Direct discrimination would be a decision that is solely based on a sensitive or protected attribute a decision is made that negatively affects them. For example, if a predictive process monitoring would be trained on a somehow biased dataset and learn that female applicants for a loan should always received an extra background check causing a worse service quality or an increase rate of rejection. Clearly, this type of discrimination would be easily detected and mitigated. However, often discrimination can be *indirect discrimination* or *statistical discrimination* [9]. In these cases, some negative effect is applied but it is not directly based on a sensitive attribute. Rather the attribute or some statistical distribution is strongly correlated to a sensitive attribute. For example, when analysing the performance of a process with process mining methods one may identify a group of workers as being slower than other as they are assigned more difficult cases [3] or receive less support than others. Similarly, when improving a process design, one may focus on the 80% most frequent variants and, thereby, discriminate against minority groups with special needs that trigger infrequent activities and, thus, are often not visible in the standard process mining visualization.

### 2.3 Algorithmic Fairness

To counter and detect discrimination, there are attempts to formalize the notion of fairness of an algorithmic decision based on data. Again, there are many definitions that formalize different kinds of fairness that can be provided by algorithms [14,15]. It is important to realize that none of them is universally applicable and that it depends on the context which one is suitable. Often fairness definitions are introduced on the example of a simple binary classification task. The four main types of fairness notions based on [15] are: (1) based solely on the predicted outcome, (2) based on the predicted outcome and comparing it with the actual outcome (ground truth), (3) taking additionally into account the probability of the predictions, (4) notions based on similarity of the non-sensitive attributes, and (5) notions based on causal reasoning.

We introduce a few selected of these notions in the context of process mining and assume a simple binary classification model with the protected attribute *gender* for concise presentation as done in [15].

- Group fairness or statistical parity is of type (1) and satisfied when subjects from the protected group, e.g., females, have equal probability of being assigned a positive outcome. However, this notion is only applicable if there is no other, unprotected, attribute that would justify a difference in probability to be assigned a positive outcome.
- Predictive parity is of type (2) and satisfied when the precision or positive predictive value of the classifier is equal for both groups. The fraction of females and males that are predicted to be in the positive group from those that are in the positive group in the ground truth is the same. So, there are equal chances for a positive prediction for those that are in the positive group in the training data. Thus, the definition only works if there are really similar probabilities to be in the positive class.
- Treatment equality is of type (2) and satisfied when both groups have the same ratio of false negative and false positives. This allows to compare if the number of misclassifications (either positive or negative) is different between the groups.
- Fairness through unawareness is of type (4) and satisfied if the sensitive or protected attribute is removed from the dataset. Clearly, this will not resolve the issue of other correlated attributes.
- Fairness through awareness is of type (4) and uses a distance metric between individuals and compares the distance of the outcome with the distance between individuals. However, how to define that distance measure if not always easy.
- Counterfactual fairness is based on causal reasoning, thus, requires the definition of a causal graph instead of any binary classifier. Here the definition is satisfied if the predicted outcome does not depend on the protected attribute in the causal graph [15]. However, building causal graphs generally requires domain knowledge.

Being only a small selection of possible fairness notions, we refer to [15] for a comprehensive overview. None of the provided definitions is universally accepted and provides fairness in every sense of the concept.

The only work so far that directly addresses the challenge of fairness from a process mining viewpoint is written by Qafari et al. [16]. Here the problem of creating a fair classifier for data extracted from an event log that is enriched with process performance information is investigated. The approach firstly advocates to exclude the sensitive attribute or feature from building the classifier and then builds a C4.5 decision tree based on a discrimination-aware decision tree learning method. As fairness decision *predictive parity* is employed. An interesting problem is raised that *relabeling* may not always be desirable, in which case the fairness guarantees cannot be achieved. This is left as future work.

Though not explicitly addressing fairness, several proposals for applying causal machine learning techniques in the context of process mining have been made. For example, Bozorgi et al. [17,18] looked at discovering causal rules from event logs as well as taking some form of cost into account when making suggestions for intervention in running cases as part of a prescriptive process mining

approach. By making the causalities explicit its may be feasible to include fairness constraints into decisions.

### 2.4 Open Challenges

Many open research challenges for considering fairness in process mining exists. So far, there is hardly any research on fairness that is specific to process mining neither from a technological nor from an organizational perspective, with the notable exception of [16]. A clear research challenge is to develop specific notions for fairness in process mining from the more generic fairness definitions. Whereas one could take the stance that the existing definitions from the wider machine learning field are sufficient, we motivated the need to consider fairness explicitly also regarding process mining techniques.

## 3 Accuracy

Models need to be *accurate* to be useful in the real world. An analyst relying on a statistical analysis or an engineer developing a machine learning model for classification needs to have confidence that the analysis or the model captures the real-world phenomenon correctly. Differently to a model based on, e.g., physical laws or logic that can be shown to be correct in any application setting the kind of statistical models often used in data science can rarely be proven to be correct. Thus, the level of accuracy with which a real-world phenomenon is captured or the level of confidence that a user can have when using that model are important aspects of any such model. The accuracy of models depends on many factors and it is often not straightforward to measure it properly. A classification model may, on average, be classifying near perfectly between pictures showing different breeds of dogs on an independent test set but if the relevant breed is highly underrepresented the classifier may still be unusable in the real world due to the class imbalance. It may also be that the classifier provides very good accuracy but makes its decision based on the wrong features picking up on spurious correlations introduced when preparing the training data: a data quality problem.

### 3.1 Process Mining Perspective

Understanding and being able to measure the *accuracy* of a process mining analysis is an integral part of responsible process mining. Whereas it may seem obvious to only use results that accurately reflect the process reality, this is frequently impaired in practice by the need to abstract from that reality.

Process discovery techniques are often unable to create the perfectly accurate model but are forced to balance between several quality dimensions [1] that are competing with each other. For example, to obtain a process model that is understandable by a human analyst, some observed behavior may need to be omitted. In some cases, the process behavior is too complex to be captured by

a single case notion and multi dimensional or multi entity representation are required to avoid drawing inaccurate conclusions [19]. Conformance checking techniques such as alignments [20] often face the challenge that there are multiple possible explanations for a non-conformance between observed and prescribed process behavior. However, it may be infeasible to show all of them due to the large number of possibilities. Finally, the quality of the input data is often a substantial issue when applying process mining in a real-world scenario [12,21].

This brief look at possible challenges for *accuracy* indicates that the topic is very broad and difficult to discuss comprehensively in the scope of this chapter. Thus, we limit ourselves to briefly describe several challenges and selected solution proposals. We categorize them into solutions for *data quality* and *model quality*.

### 3.2 Data Quality

Data quality is known to be often poor [22] and this may lead to non-factual or misleading representation of the real business process. Garbage-in garbage-out is a often used phrase to illustrate this issue. Whereas the data quality issue is not particular to process mining there are some peculiarities of event logs that call for specific solutions.

Often data quality problems in process mining are related to the strict data requirements on timestamps (R1), case identifiers (R2), and event labels (R3) [23]. Wrong or coarse granular timestamps lead to discovering wrong causalities in process models or parallelism where none exists. Inconsistent event labels make it difficult to assign clear semantics to the activities of a discovered process model. These are just two examples of how data quality issue impair process mining. Automated repair approaches to combat some of the data quality problems exist. For example, in [24] autoencoders are used to add missing values. However, any such method may affect *transparency* [25] as it is unclear what part of the data was inferred and what part of the data can be considered truthful beyond doubt. A discovered process model may be perfectly *accurate*, but when it is based on data with poor quality any conclusions become disputable. Notions of data quality and remedies are already introduced and discussed in [12], therefore we go not further into detail on the data quality challenge.

One noteworthy topic connected to data quality is *uncertainty* at the level of the event log data [26], e.g., by adding metadata to express the uncertainty [27]. Pegoraro et al. [26] advocate to explicitly encode the uncertainty about events and traces in order to leverage it in a transparent manner during the analysis. Based on this event log with explicit uncertainty representation conformance checking techniques can be adaped [28] to an obtain more trustworthy diagnostics that also provide more transparency about the possible different scenarios compatible with the (uncertain) observations.

### 3.3 Model Quality

How to decide whether a process model is of good quality? In fact, even when it comes to the question on how to measure *accuracy* there is hardly an agreement in process mining. Classically, process mining quality dimensions consist of fitness, precision, generalization, and simplicity as introduced in [1]. For most of these quality dimensions measures have been proposed that are based on conformance checking, e.g., through alignments as indicated in [20]. However, this common practice of measuring model quality has been challenged at least for precision with Tax et al. [29] proposing several axioms that the prevalent measures do not fulfil.

The issue with initially proposed quality measures led to several new methods and definitions for measuring various model quality dimensions being proposed [30–34]. Main complications for model quality in process mining are that process models commonly exhibit infinite behaviour (through loops) and the absence of negative examples, i.e., behaviour that the model should not contain [1].

Recently, there have been several proposals that aim to extend process discovery and the model quality measures to the stochastic setting in which process models include probabilities and the likelihood of observing a certain trace is taken into account [35,36] allowing to better estimate the relevant subset of the behavior modelled. This may help to truly quantify the confidence that an analyst can have in a model.

A somewhat related issue on the confidence an analyst can put in the performance of a process discovery algorithm was brought up by Van der Werf et al. [37]. They observed that process discovery techniques not always discover better process models when provided with a better sample of the process behavior, i.e., a larger event log with observations of process behavior.

### 3.4 Outlook and Challenges

The extensive discussion around how to measure quality shows that even defining *accuracy* for process discovery is not straightforward. In practice, this creates the challenge to choose which measure should be used in which context and when can a model be considered good for an analysis purpose. Another very relevant perspective for responsible process mining regarding model quality is how the discovered process model representation is understood by the user of such model. We will come back to this issue when considering *transparency*.

## 4 Confidentiality

Confidentiality generally refers to the protection of certain sensitive data or information from disclosure. In the context of an organization many different kind of information is usually confidential. Intellectual property such as the design of machines or software may be confidential to protect it from competitors

but also general information on the business such as the amount of sales in a certain area is usually kept confidential. A subset of the confidential information in the sphere of an organization relates to personal data. Here, the concern is on the right to *privacy* for individuals of which personal data is processed by the organization. Personal data may relate to customers, employees, suppliers or other people that interact an organization's processes. Privacy rights have received a lot of attention with several high-profile data breaches and increased regulation such as Europe's General Data Protection Regulation (GDPR) [38].

### 4.1 Process Mining Perspective

In the context of process mining, the information contained in event logs may be sensitive for several reasons. Event logs contain data providing detailed *information on the operations of an organization*, e.g., the order volume or the production capacity. Uncontrolled disclosure of such information may be undesired as it could negatively affect the organization. Event logs contain *information on individuals*, e.g., customers, which may be subject to the privacy regulations.

Assume a hospital process is analyzed. Case data is related to the individual patient and *confidentiality* challenges to protect sensitive data and *privacy* are obvious [39]. However, the employees that work in processes are often also directly affected by process mining results and may be directly represented in the event logs e.g. via the resource attribute in XES. This can create an additional *confidentiality* challenges to prevent work surveillance [40,41].

Protecting the privacy of individuals in event logs is difficult, as sequential event data is highly vulnerable to re-identification [42]. In fact, when assuming some background information, privacy leakages exists in the vast majority of presumably anonymous event logs that are used in the process mining community [42]. As events are linked together through a case, and often the traces in an event log are highly unique, already very limited background knowledge on some attributes or events can reveal the identity of an individual.

This "privacy problem" creates challenges in the practical application of process mining. Data gathering is more difficult or impossible when privacy concerns are raised. For example, the hospital may fear that privacy regulations (GDPR, HIPAA [43]) are violated when analysing patient trajectories [39,44] or a works council may object to the usage of process mining technology due to fear of worker surveillance [41]. Regulations threaten organizations with high fines when personal data is used without legitimate purpose or consent. The fines in GDPR may be as high as 4% of the organizations worldwide annual revenue [38]. Thus, there is a clear need for privacy-preserving or protecting techniques for process mining. Such approaches aim to retain the utility of the data without the risk of accidental disclosure of personal data. Please note that we use *protection* here in the sense of *anononymity* and *unlinkability* requirements. Next to those, other requirements such as *notice*, *transparency*, and *accountability* are often imposed by regulations [45]. Note that most privacy-preserving techniques differ from the wide variety of best-effort pseudonymization, perturbation, and generalization

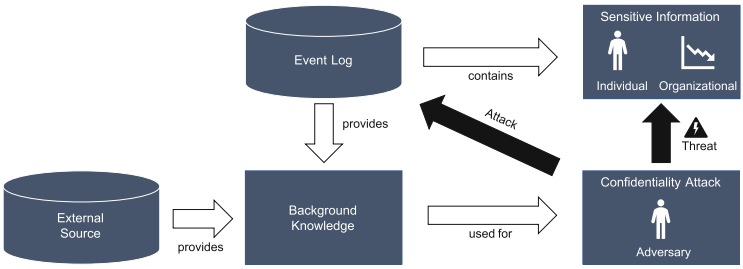


**Fig. 4.** The main aspects of any confidentiality scenario for process mining: What is the sensitive information contained in the event log that needs to be protected? Which background knowledge can be assumed (including provided by external sources)? What are the attacks used by the adversary and which threats are posed?

methods that are used by commercial tools[5]. Unfortunately, it has been shown such naïve replacement of identifiers is often not sufficient to keep information secure in many scenarios.

For each confidentiality scenario we need to characterize at least the *sensitive information* (Sect. 4.2) and the *background knowledge* (Sect. 4.3) of the attacker or adversary as illustrated in Fig. 4. Then, we can identify *confidentiality attacks* (Sect. 4.4) that are assumed to be employed an the resulting *threats* that should be mitigated. Based on the analysis of the available threats, protection techniques have been proposed to mitigate these threats under certain assumptions (Sect. 4.5).

### 4.2 Sensitive Information

Several kinds of sensitive information may be derived from event logs. We consider both the scenario in which an event log contains some business information that needs to be secured as well as the scenario in which personal data of individuals that took part in the process should not be revealed. These individuals could be customers that are the *subject* of the process or workers that perform activities withing the process.

We assume that the **sensitive information** is contained in a given event log as shown in Fig. 4. Sensitive information may be obtained *directly* from the attribute values of individual events of or it may be *derived* by performing some computation over several events in. Often, in the scenario in which personal data of individuals is at risk the sensitive information in the event log is assumed to be connected to the individual through the process cases each of which is about a single individual. We now illustrate several types of sensitive information with

[5] Most commercial tools provide some kind of pseudonymization technique to replace sensitive data by a hashing or replacement. An example is given here: https://fluxicon.com/blog/2017/11/privacy-security-and-ethics-in-process-mining-part-3-anonymization/.

**Table 1.** Example of a loan application event log that contains several types of sensitive information and may be subject to confidentiality attacks revealing this information to an adversary possessing suitable background knowledge.

| SSN | Activity | Time | Resource | Amount | Age | Type | Postcode | Income |
|---|---|---|---|---|---|---|---|---|
| 617-07-5604 | SA: Submit appl. | 09-02-22 23:39 | | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | SC: Simple check | 11-02-22 08:38 | Alice | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | AA: Accept appl. | 12-02-22 11:35 | Joe | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | SO: Send offer | 12-02-22 12:32 | Joe | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | RO: Receive offer | 13-02-22 08:14 | | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | FA: Finalise appl. | 15-02-22 16:30 | Alice | 200k | 30 | Home | 94121 | 50k |
| 617-07-5604 | AO: Accept offer | 19-02-22 23:31 | | 200k | 30 | Home | 94121 | 50k |
| 528-41-8024 | SA: Submit appl. | 01-03-22 12:32 | | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | SC: Simple check | 02-03-22 15:23 | Joe | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | EC: Extensive check | 05-03-22 07:31 | John | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | AA: Accept appl. | 11-03-22 12:21 | Alice | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | SO: Send offer | 11-03-22 15:44 | Joe | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | RO: Receive offer | 12-03-22 12:33 | | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | FA: Finalise appl. | 15-03-22 16:54 | Robert | 60k | 42 | Car | 37287 | 75k |
| 528-41-8024 | AO: Accept offer | 18-03-22 18:23 | | 60k | 42 | Car | 37287 | 75k |
| 330-80-8169 | SA: Submit appl. | 02-03-22 23:30 | | 500k | 22 | Home | 32984 | 45k |
| 330-80-8169 | EC: Extensive check | 05-03-22 08:30 | John | 500k | 22 | Home | 32984 | 45k |
| 330-80-8169 | DA: Decline appl. | 10-03-22 11:30 | John | 500k | 22 | Home | 32984 | 45k |
| 526-34-5246 | SA: Submit appl. | 15-04-22 23:31 | | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | SC: Simple check | 17-04-22 12:47 | Joe | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | AA: Accept appl. | 18-04-22 11:59 | Alice | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | SO: Send offer | 18-04-22 12:29 | Alice | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | RO: Receive offer | 19-04-22 07:52 | | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | CO: Cancel offer | 24-04-22 21:34 | | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | SO: Send offer | 28-04-22 09:21 | John | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | RO: Receive offer | 29-04-22 10:12 | | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | FA: Finalise appl. | 02-05-22 15:43 | Alice | 100k | 30 | Home | 75755 | 30k |
| 526-34-5246 | AA: Accept offer | 05-05-22 05:23 | | 100k | 30 | Home | 75755 | 30k |

the event log in Table 1 that was obtained from the previously introduced loan application process.

An example for sensitive information related to an individual that can be directly obtained is the social security number of the applicant stored in the column *SSN*, which also acts as case identifier here. Obviously, using such direct identifiers of individuals poses a privacy risk as it would allow to directly link all the remaining information contained in the event log to individuals. Analogously, the *Resource* column contains the full name of the employee responsible for handling the process activities. This information would enable direct profiling of the work performance of individual employees, which may be against company policies or forbidden by work regulations. It is easy to remove directly personally identifiable information such as names or identifying numbers of customers or workers as they are not necessary for process mining. For example, it would

be trivial to replace both the *SSN* column and the *Resource* in Table 1 with a surrogate case identifier based on a mapping obtained through one-way hash function or a simple lookup table.

However, it has been shown that obscuring the direct identifiers is not sufficient as also not directly identifying attributes can be problematic [42,46]. *Quasi-identifiers* are values not directly revealing the identity of a person but may be used to do so in combination with other attributes. Common quasi-identifiers are attributes such as gender, birth dates, or postcodes that taken together are often unique for an individual. For example, in Table 1 the combination of columns *Age*, *Type*, and *Postcode* would very likely be uniquely identifying a single customer leading to disclosure of other sensitive information contained in the event log such as the yearly *Income* of the applicant.

So far, we gave examples of sensitive information that is directly stored in the event attributes. However, also the presence of a certain activity in the event log or derived information such as the sequence of events that occurred for a certain case may be considered sensitive. Take for instance the third case in Table 1 in which the loan application is declined ($DA$) after an extensive check ($EC$). Knowledge of such details on how the loan application process was carried out may be used against the individual. Thus, even the sequence of activities performed for an individual case may be regarded as sensitive information. At the same time, the sequence of activities performed may also act as a quasi-identifier as it is often unique and identifies an individual such as an applicant or a patient [42,47].

When it comes to sensitive business information one may think about attributes encoding the cost of a certain activity or information on prices paid by different customer segments (e.g., the interest offered on the loan). Similarly to the case of personal information, the sensitive information may not only reside in the attribute values but also be derived from the sequence of events that occurred or their timestamps, e.g., the throughput times computed for different organizational units may be considered sensitive.

It is important to realize that these computations may also be based on the artefacts that are returned by the classical process mining tasks: intermediate data structures, process models, and conformance checking results. Thus, direct access to the original event log may not be required to gain access to sensitive information. For example, the utilisation of a certain department or group may be determined by considering the number of traces in a certain time period and could be considered sensitive. The cross-organisational process mining scenario is also commonly considered when it comes to motivating the need of protecting sensitive information for process mining. Here, two organizations want to compare their processes to learn from each other or analyse a process that is jointly performed (e.g., supplier and integrator). However, certain sensitive data should not be shared.

### 4.3 Background Knowledge

Apart from the trivial case in which an individual or an organizational entity can be directly identified, attackers often need to possess certain limited background knowledge about the individual, i.e., the process case, the entity, or about remaining parts of the dataset. This is reflected in Fig. 4 by assuming the adversary to use some knowledge to facilitate attacks on sensitive information. Some protection models assume the worst-case scenario in which no restriction on the background knowledge of an attacker is assumed and still some kind of privacy guarantee should be given. However, in many cases it is reasonable to assume only limited background knowledge to be available.

**Background knowledge** may be fully derived from the event log or it may also contain information that is not present in the event log but related to specific cases or events. Thus, it can be any kind of knowledge that gives an attacker information that can be used to identify sensitive information. We keep the definition of the background knowledge deliberately vague as it may be defined in various ways and include arbitrary external data sources. Two more precise definitions for event logs have been introduced in the literature.

Rafei et al. [48] provide several definitions for possible background information in a process mining context. They assume that background knowledge is defined over a simple view of process traces as sequences of event labels, e.g., the third trace in Table 1 would be seen as sequence $\langle SA, EC, DA \rangle$. Three categories are defined: *Set knowledge*, *Multiset knowledge*, and *Sequence knowledge*. The knowledge refers the occurrence of activity labels in the to be attacked process case at one of the three abstraction levels. Thus, an attacker can either know only about the *presence* of activities (set abstraction), their *frequency* (multiset abstraction), or have in-depth knowledge about a certain *ordering* of activities (sequence abstraction). In Table 1, the third trace $\langle SA, EC, DA \rangle$ would already be uniquely identified when having the set background knowledge $\{DA\}$ since that is the only trace in which an application is declined. As another example, the multiset background knowledge of $[SO^2]$ would uniquely identify the fourth case. In many setting such knowledge of process events may be easy to obtain, e.g., one may know that their neighbours received two loan offers in a specific time period.

Von Voigt et al. [42] quantify the re-identification risk of individual cases by assuming different kinds of background knowledge. In addition to knowledge of activity labels as in [48] also case-level attributes are considered to be candidates for background knowledge. For example, in the well-known BPI Challenge 2018 dataset [49] case attributes have been generalized to provide some level of privacy protection. However, still when considering the combinations of all case attributes 84.5% of all cases are unique.

Many other similar abstraction and definitions of background knowledge are possible but have not yet been investigated. For example, partial orders of activities or knowledge about time or resource involved. An adversary may know that two medical diagnostic tests have been performed on the same day and two days later the patient was re-invited for a discussion by the same doctor.

Also knowledge on the absence of a certain activity in the case to be attacked could be informative. As Fig. 4 illustrates also external data source may provide complementary background knowledge. A famous example that involved using external background knowledge is the successful attack on a Netflix dataset by using information from the public IMBD movie ratings [50], which included full names for some users, and compared them to the ratings in the Netflix dataset thereby identifying users in the supposedly anonymized Netflix dataset.

In summary, a precise analysis of background knowledge assumed is important to provide meaningful guarantees against uncovering sensitive information.

### 4.4 Threats and Attacks

Several attacks on confidential data in event logs are possible. We follow Elkoumy et al. [45] and focus on a honest-but-curious attacker scenario. An adversary has access to data or results and tries to identify some sensitive information without trying to break into systems. So, we do not consider scenarios in which access control or similar security measures are broken.

We structure confidentiality attacks structured according to the threat that they pose, i.e., the kind of sensitive information that an attacker or adversary tries to reveal. As already motivated, it is important to consider the kind of background knowledge that is assumed in the analysis of a specific threat or attack to find reliable mitigation strategies. Attacks on confidentiality use this background knowledge to reveal sensitive information that is contained in the event log as shown in Fig. 4.

So, a very general definition of a confidentiality attack on an event log can be given as follows. Given an event log and some sensitive information that is related to that log, a **confidentiality attack** uses some background knowledge, which may be derived from the log or from other available sources, to reveal some subset of sensitive information that is part of the log. We distinguish four general types of threats based on the goal of an attacker and the employed attack method following the categorization in [45].

*Membership Disclosure Threats.* A basic threat is that an adversary could establish that an individual was taking part of the process that is described by the event log. A *membership inference* attack combines background knowledge about the individual to the information released by an event log or a process mining analysis. So, the sensitive information obtained from the event log would consist of the identifiers for a subset of individuals that took part in the process. Whereas this does not reveal the exact case in which an individual took part, it still often allows to draw conclusions about which activities and events an individual was involved in. Let us assume that the event log obtained in our example loan application process scenario only contains loans for starting a business. Already, the information that an individual is part of that event log, i.e., they were applying for such a specific loan type can be sensitive information.

*Re-identification Threats.* Threats that cause the disclosure of the identity of a individual to which some data belongs are called re-identification threats. So, the sensitive information is the subject of a certain case, e.g., the patient identity, or the subject of a certain event, e.g., the identity of the resource or worker that performed the activity recorded by the event. Example attacks are linkage attacks and intersection attacks [45]. *Linkage attacks* use background knowledge to reveal the identity, e.g., a certain combination of attribute values or a certain sequence of events is known to be connected to an individual. In Table 1, knowing that an individual received two offers, i.e., multiset background knowledge of $[SO^2]$, and that their data is part of the event log uniquely re-identifies identity of the applicant in the fourth case. *Intersection attacks* try to establish a mapping between two separately released event logs revealing the identity of an individual. Here the information revealed in a second separately released dataset is assumed to be directly linkable to an individual without containing any sensitive information. However, in combination this information can be used as background knowledge and reveal the sensitive information in the first event log.

*Reconstruction Threats.* In some cases it may be possible to partially or fully infer the original event from seemingly protected data. Here, the sensitive information to be retrieved would be the entire event log. The two main attack methods for reconstruction are *difference attacks* and *model-inversion attacks*. The basic idea for both is to repeatedly consult a model or a statistic with slightly different queries and, thereby, uncovering sensitive data.

*Cryptanalysis Threats.* Data may have been pseudonymized, as often done by commercial tools, or encrypted in an attempt to provide confidentiality. However, naïvely pseudonymized or even fully encrypted event logs are vulnerable to attacks based on the analysis of the frequency [51]. Please note that this may lead in turn to re-identification, membership disclosure, or reconstruction, but may also simply leak sensitive business information such as the number of certain activity executions. The main attack method is a *frequency analysis* based on background knowledge on the activities of the process and their prevalence.

### 4.5 Protection Approaches

Whereas still in an early stage, the research on privacy and confidentiality has received increased attention in the past years and several protection techniques with diverse assumptions and guarantees that protect against the mentioned threats have been proposed. However, none of the proposed methods is generally applicable to any possible confidentiality and privacy problems. Each of them makes certain assumptions regarding the attack scenario including the background knowledge of the assumed attacker. Conversely, depending on the input log the methods result in some loss of utility. Thus, the goal of the process mining analysis (discovery, conformance, etc.), their data requirements, and the characteristics of the process that generated the event log need to be considered.

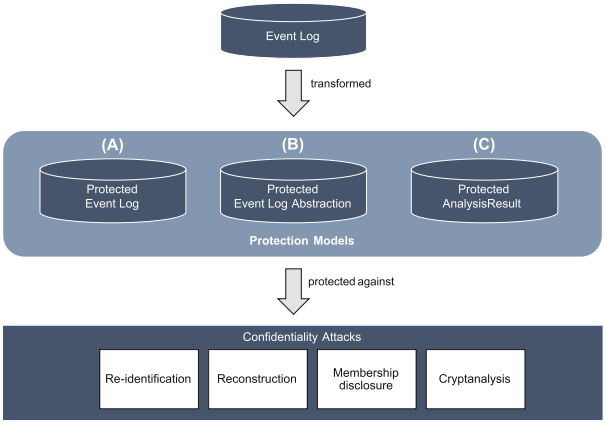


**Fig. 5.** Different protection models have been proposed that protect the data contained in an event log by transforming it into protected representations: a protected event log, a protected abstraction the event log, or a protected analysis result.

Protection models can work at different levels of a process mining analysis as shown in Fig. 5. Following [48], we differentiate between several tasks for protection models. Some models protecting the event log itself and provide a protected copy of the original event log. Other techniques provide protected abstractions over the original event log, e.g., a directly-follows graph representation which can only be used for certain process mining activities. In [48], these two tasks are denoted as Privacy-Preserving Data Publishing (PPDP) and Privacy-Preserving Process Mining (PPPM), respectively. We add a third possible task, which is to protect process mining results, e.g. a process model or a conformance checking result, without an intermediate representation.

Regardless of the task at hand, techniques can also be distinguished into roughly three categories of protection models [45]: *group-based privacy models*, *indistinguishability-based models*, and *confidentiality frameworks* including encryption. We now introduce the main properties of the three different protection model categories and briefly introduce exemplary techniques.

**Group-Based Privacy.** The prototype of a group-based privacy protection model are those that provide *k-anonymity* [52]. The basic idea is that a tabular dataset containing rows with information about individuals is *k-anonymous* when the values for each combination of sensitive attributes or columns (quasi-identifiers) appear at least $k$ times. So, data similarity is used as a criterion here. The intuitive idea is that the individual will have the same sensitive data as the $k-1$ other individuals in the same group and, thus, with sufficiently large $k$ it protects against re-identification. Usually, this is achieved by data *suppression* or *generalization* until the $k-anonymity$ property is achieved. Whereas this model is interpretable and easy to understand, unfortunately, it has been shown to be suspect to certain attacks based on background knowledge [53]. Several

extensions have been proposed that mitigate some of those including: l-diversity [53], and t-closeness [54].

For process mining, two methods are providing group-based privacy protection models. The TLKC model by Rafaei et al. [47] and the PRETSA approach by [55]. Both aim at the release of a proctected event log, option (A) in Fig. 5. PRETSA utilizes generalisation based on a prefix-tree that is build on top of the activity sequence in the event log and provides *k-anonymity* and *t-closeness* guarantee to prevent the disclosure (membership and re-identification) of resources or workers that performed certain activities. The TLKC model protects the identity of cases, e.g., a customer, and provides a relaxed variant of *k-anonymity*. Additional, it supports protecting information in the time and organizational perspective. Both approaches make assumptions on the background knowledge. Maintaining data utility is challenging for both methods when many unique traces exist.

**Indistinguishability-Based Privacy.** Differently to the group-based models providing guarantees such as k-anonymity for a given dataset, indistinguishability-based privacy models give a guarantee that two versions of a dataset are indistinguishable to a certain degree. A central model is *Differential Privacy* (DP). The idea is that there is one datasets $A$ without an individuals information and another one $A'$ including an individuals information. A mechanism provides DP with a parameter $\epsilon$ when the results of a (randomized) query mechanism statistically differ between $A$ and $A'$ only by a small factor that is controlled by the $\epsilon$-parameter [56]. This provides a strong guarantee that is independent of the background knowledge as the guarantee needs to hold for any dataset $A$ and $A'$. There have been many variants of the differential privacy concept [57]. For example, adding a relaxation parameter $\delta$ to better tuning the utility while loosing some of its strengths ($(\epsilon, \delta)$-DP), only requiring the values of the datasets to not differ too much and ignoring addition and removal of items (bounded DP), or extending the guarantees in the case individuals appear multiple times in the dataset (group DP), which is a possible scenario for event logs.

For process mining, several adaptions have been proposed. The first one was given by Mannhardt et al. in [58] who assume a protected event log to be queries through a privacy engine. Laplacian noise is added to the counts returned by each query, thereby guaranteeing DP with regard to the individual cases. Queries are defined for both directly-follows relations [1] and complete activity sequences. The method was later extended in [59] to also protect contextual information that is encoded in the attributes of the event log. Furthermore, the guarantee was extended to *local DP*, which means that a perturbed event log itself can be released. One major issue of these methods is that obviously invalid process behavior may be added. Recently, the approach was improved to consider the semantic of the added noise [60]. Contextual information, in particular process performance indicators, is also protected in the work by Kabierski et al. [61]. Finally, there is a very recent proposal by Elkoumy et al. [62] that provides only

a bounded DP guarantee but improves the utility of the protected data by using an oversampling approach instead of adding noise.

**Confidentiality Frameworks.** The third type that we distinguish are protection models that are not directly targeted at protecting individuals but any kind of sensitive information in event logs. Here, mainly encryption schemes have been proposed. A major family of techniques are those based on on homomorphic encryption [63] schemes. The goal is to enable certain computations on an encrypted version of the data. For process mining, this idea is taken up by Rafei et al. in [51] and embedded in a framework that aims to protect against frequency or background knowledge-based attacks by disassociating events from their respective cases. It could be used to outsource computations on secured data or in a cross-organizational setting. However, it does not protect the resulting analysis results (B) and (C) from an internal process analyst. The cross-organization setting is also targeted by Elkoumy et al. in [64]. A secure multi-party computation [65] method is proposed that avoids to leak sensitive information in the cross-organisational process mining scenario.

The above categorization and list of techniques covers the major share of the work in the process mining field so far.

### 4.6 Outlook and Challenges

Protecting the privacy and confidentiality of data while keeping it useful for analysis is a difficult problem. Information needs to be hidden while the objective is to get as much signal from data as possible. Unsurprisingly, many open challenges exists for *confidentiality* in process mining and, apart from academic prototypes [66,67], none of the proposed techniques has seen uptake in commercial solutions. Seven main challenges for research in the field of privacy and confidentiality in process mining are identified by [45]:

- *Interpretable Quantification of Privacy Disclosure.* Protection mechanisms should be interpretable when it comes to the remaining risk. Guarantees and attacks are often not obvious for non-experts making adoption by industry difficult.
- *Balancing Risk and Utility.* Any protection mechanism may impair the utility of the source data and poses a trade-off that needs to be made upfont. In the exploratory process mining setting this is a challenge for adoption. In [68] it is proposed that mechanisms need to be *utlility aware*.
- *Level of Granularity.* Process mining analyses happen at various levels of granularity and various perspectives. Some tools require only activity sequences and timestamps, which most current protection models focus upon. Others also consider the resource perspective including potential sensitive data on employees. In some cases, access to an event log may not be necessary and privacy guarantees should be given at the level of a released process model as proposed in [69]. A one-fits-all approach to privacy is unlikely to work, which opens opportunities for further research.

- *Distributed Privacy.* In many settings attempts on data sharing between organizations are made which creates the problem of protecting privacy in an inter-organizational setting. This setting is currently less well researched.
- *Computational Challenges.* Some of the approaches proposed are computationally expensive. Thus, research on making those suitable for real-life settings is required.
- *Traceability and Transparency Challenges.* Often personal data still needs to be collected and stored at some point during the analysis. GDPR requires to trace the processing and usage of data to fulfill the different rights (right to consent, right to access, right to be forgotten [38]). This is challenging for process mining where data comes from different distributed data sources. Similarly, GDPR requires organizations to be transparent about the usage of data. Traceability is a pre-requisite but not sufficient for achieving transparency. Investigating how to provide information on the purpose for which data was used in process mining is a research challenge.

Many of these challenges are geared towards the improving technological solutions that provide some form of privacy guarantee in various settings. However, as already reported in [40] many aspects of privacy and confidentiality as well as the compliance to regulations such as GDPR cannot be solved by technological measures alone. However, there is little research from the organizational side apart from anecdotal discussion on the role of privacy in real-life process mining projects [41]. To conclude, it is notable that process mining has also been used to check conformance to privacy regulations [70]. Thus, process mining can also help in uncovering confidentiality issues that are present in an organizations processes.

## 5 Transparency

Transparency has been a widely discussed topic for AI systems that are based on machine learning. Often, a key concern is the explainability of black-box classifiers such as Deep Learning models: Why is a certain classification or prediction made and what features are important in the decision of the model?

The core process mining tasks of process discovery and conformance checking aim to provide *white-box* process models that can be interpreted by process stakeholders. Explainability of the discovered models and, thus, transparency is key objective of process mining. Still, there are several aspects of process mining in which transparency is at risk. In the next two section, we focus on two exemplary transparency challenges for process mining: achieving *generalization* without hampering transparency and the *interpretability* of the discovered process model representations.

Besides these two transparency challenges all the common transparency issues of predictive models are inherited when building predictive process mining models. Therefore, we do not discuss this in detail since many resources on explainable machine learning are available and [71] gives a brief overview of how to obtain explainable predictions in the context of process mining.

### 5.1 Generalization

Process discovery aims to abstract from the exact behaviour observed in the event log and return a *concise* model of the underlying process. This often requires to disregard *infrequent behaviour* to obtain simpler process models. Conversely, process discovery techniques often attempt to generalize beyond the observed behaviour since they cannot be assumed to have observed all possible incarnations of the process, particularly in the presence of parallel process behavior. This aspiration creates a *transparency* challenges.

Disregarding infrequent behaviour may hide important parts of the observed data. In particular, infrequent patterns may be of high interest [44]. Very few techniques have been focusing on retaining infrequent data, e.g., in [72] certain infrequent dependencies are not filtered if they can be reliably predicted from data attributes and in [73] it is explored how to selectively include infrequent behaviour by filtering over multiple ranges of parameter values.

In a orthogonal direction, the frequency and probability with which behaviour is observed gets more attention in approaches that can be labeled as: *stochastic process mining*. In [35], Leemans et al. proposed a new conformance checking method with the goal of taking into account routing probabilities, which improves the accuracy of the diagnostics.

### 5.2 Interpretation of Results

Interpretation of results based on process model notations or visualizations can be difficult for stakeholders leading to *transparency* challenges. For example, the presence of loops together with optional activities may enable non obvious process behaviour and the filtering of edges in a directly-follows graph may lead to invalid statistics as is illustrated for many commercial tools in [74].

However, also for discovery approaches based on clear semantics misinterpretations are possible. As an example, the models discovered by the Inductive Miner often contain silent transitions that allow to skip certain behaviour that in combination with loops allow any behaviour. This may be difficult to spot for a non-expert. Whereas there exists research on the comprehension of process models [75], little work has yet been done in the context of automated process discovery.

Recently the question of interpretability of process mining results has been touched upon by Mendling et al. [76] who raise the issue that the quality of process mining results needs to be judged in light of the tasks of a process analyst using the models. A first technical contribution for process discovery in this direction was provided by Fahland et al. [77] with a new variant of the Inductive Miner that was evaluated in a user study in which an analyst's trust in the model as considered. Overall, there has been surprisingly little research on this topic given the claim of process mining to provide white-box models.

## 6 Conclusion

This chapter defined the concept of Responsible Process Mining under the umbrella of Responsible Data Science. Based on the FACT criteria put forward in [5] (Fairness, Accuracy, Confidentiality, and Transparency), we gave an overview of challenges related to these criteria and introduced state-of-the-art approaches for addressing each of them. Due to the broad scope of the FACT criteria, we can provide only a high-level introduction and discussion for each of them. We refer to the individual work or relevant surveys for further details.

In some areas the research on responsible process mining is already much further developed than in others. Little attention has been devoted to *fairness* in the context of process mining, at least compared to its prominence in the machine learning field. The trend to more automated decision taking in process mining may change this in the future. In contrast, the *confidentiality* challenge has been recognized in the process mining research community and has recently received much attention in research. However, adoption by commercial process mining tools has not yet started even though the problem has also been recognized by industry [41].

Criteria such as *accuracy* and *transparency* are very broad and many approaches touch these issues; however, with the notable exception of the work on data quality [12] they are rarely addressed explicitly under the umbrella of responsible process mining. More work is required to develop and address these criteria more explicitly in future process mining research.

## References

1. Aalst, W.: Foundations of process discovery. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
2. Reinkemeyer, L.: Status and future of process mining: from process discovery to process execution. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
3. van der Aalst, W.M.P.: Responsible data science: using event data in a "People Friendly manner. In: Hammoudi, S., Maciaszek, L.A., Missikoff, M.M., Camp, O., Cordeiro, J. (eds.) ICEIS 2016. LNBIP, vol. 291, pp. 3–28. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62386-3_1
4. Dignum, V.: Responsible Artificial Intelligence. Springer (2019)
5. van der Aalst, W.M.P., Bichler, M., Heinzl, A.: Responsible data science. Bus. Inf. Syst. Eng. **59**(5), 311–313 (2017)
6. Olteanu, A., Garcia-Gathright, J., Rijke, M.d., Ekstrand, M.D.: FACTS-IR: Fairness, accountability, confidentiality, transparency, and safety in information retrieval. ACM SIGIR Forum **53**(2), 20 (2019)
7. Aalst, W.: Process mining: a 360 degrees overview. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
8. Friedman, B., Nissenbaum, H.: Bias in computer systems. ACM Trans. Inf. Syst. **14**(3), 330–347 (1996)

9. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. **54**(6) (2021)
10. Barocas, S., Hardt, M., Narayanan, A.: Fairness and Machine Learning. fairmlbook.org (2019). http://www.fairmlbook.org
11. Crawford, K.: The hidden biases in big data. Harvard Bus. Rev. **1**(4) (2013)
12. De Weerdt, J., Wynn, M.T.: Foundations of process event data. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
13. Commission, E.: Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (2021)
14. Gajane, P., Pechenizkiy, M.: On formalizing fairness in prediction with machine learning. CoRR abs/1710.03184 (2018)
15. Verma, S., Rubin, J.: Fairness definitions explained. In: FairWare@ICSE, pp. 1–7. ACM (2018)
16. Qafari, M.S., van der Aalst, W.: Fairness-aware process mining. In: Panetto, H., Debruyne, C., Hepp, M., Lewis, D., Ardagna, C.A., Meersman, R. (eds.) OTM 2019. LNCS, vol. 11877, pp. 182–192. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33246-4_11
17. Bozorgi, Z.D., Teinemaa, I., Dumas, M., Rosa, M.L., Polyvyanyy, A.: Process mining meets causal machine learning: discovering causal rules from event logs. In: ICPM, pp. 129–136. IEEE (2020)
18. Bozorgi, Z.D., Teinemaa, I., Dumas, M., Rosa, M.L., Polyvyanyy, A.: Prescriptive process monitoring for cost-aware cycle time reduction. In: ICPM, pp. 96–103. IEEE (2021)
19. Fahland, D.: Process mining over multiple behavioral dimensions with event knowledge graphs. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
20. Carmona, J., Dongen, B., Weidlich, M.: Conformance checking: foundations, milestones and challenges. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
21. Accorsi, R., Lebherz, J.: A practitioner's view on process mining adoption, event log engineering and data challenges. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
22. Wynn, M.T., Sadiq, S.: Responsible process mining - a data quality perspective. In: Hildebrandt, T., van Dongen, B.F., Röglinger, M., Mendling, J. (eds.) BPM 2019. LNCS, vol. 11675, pp. 10–15. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26619-6_2
23. Suriadi, S., Andrews, R., ter Hofstede, A.H.M., Wynn, M.T.: Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Inf. Syst. **64**, 132–150 (2017)
24. Nguyen, H.T.C., Lee, S., Kim, J., Ko, J., Comuzzi, M.: Autoencoders for improving quality of process event logs. Expert Syst. Appl. **131**, 132–147 (2019)
25. Martin, N., Martinez-Millana, A., Valdivieso, B., Fernández-Llatas, C.: Interactive data cleaning for process mining: a case study of an outpatient clinic's appointment system. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBIP, vol. 362, pp. 532–544. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_43
26. Pegoraro, M., van der Aalst, W.M.P.: Mining uncertain event data in process mining. In: ICPM, pp. 89–96. IEEE (2019)

27. Pegoraro, M., Uysal, M.S., van der Aalst, W.M.P.: An XES extension for uncertain event data. In: BPM (PhD/Demos). Volume 2973 of CEUR Workshop Proceedings, pp. 116–120. CEUR-WS.org (2021)
28. Pegoraro, M., Uysal, M.S., van der Aalst, W.M.P.: Conformance checking over uncertain event data. Inf. Syst. **102**, 101810 (2021)
29. Tax, N., Lu, X., Sidorova, N., Fahland, D., van der Aalst, W.M.P.: The imprecisions of precision measures in process mining. Inf. Process. Lett. **135**, 1–8 (2018)
30. van Dongen, B.F., Carmona, J., Chatain, T.: A unified approach for measuring precision and generalization based on anti-alignments. In: La Rosa, M., Loos, P., Pastor, O. (eds.) BPM 2016. LNCS, vol. 9850, pp. 39–56. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45348-4_3
31. Kalenkova, A., Polyvyanyy, A., La Rosa, M.: A framework for estimating simplicity of automatically discovered process models based on structural and behavioral characteristics. In: Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds.) BPM 2020. LNCS, vol. 12168, pp. 129–146. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58666-9_8
32. Polyvyanyy, A., Solti, A., Weidlich, M., Ciccio, C.D., Mendling, J.: Monotone precision and recall measures for comparing executions and specifications of dynamic systems. ACM Trans. Softw. Eng. Methodol. **29**(3), 17:1–17:41 (2020)
33. Augusto, A., Armas-Cervantes, A., Conforti, R., Dumas, M., Rosa, M.L.: Measuring fitness and precision of automatically discovered process models: a principled and scalable approach. IEEE Trans. Knowl. Data Eng. **34**(4), 1870–1888 (2022)
34. Polyvyanyy, A., Kalenkova, A.A.: Conformance checking of partially matching processes: an entropy-based approach. Inf. Syst. **106**, 101720 (2022)
35. Leemans, S.J., van der Aalst, W.M., Brockhoff, T., Polyvyanyy, A.: Stochastic process mining: earth movers' stochastic conformance. Inf. Syst. **102**, 101724 (2021)
36. Alkhammash, H., Polyvyanyy, A., Moffat, A., García-Bañuelos, L.: Entropic relevance: a mechanism for measuring stochastic process models discovered from event data. Inf. Syst. **107**, 101922 (2022)
37. van der Werf, J.M.E.M., Polyvyanyy, A., van Wensveen, B.R., Brinkhuis, M., Reijers, H.A.: All that glitters is not gold. In: La Rosa, M., Sadiq, S., Teniente, E. (eds.) CAiSE 2021. LNCS, vol. 12751, pp. 141–157. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79382-1_9
38. Regulation, E.G.D.P.: Regulation (eu) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) 2016. OJ L 119(1) (2016)
39. Pika, A., Wynn, M.T., Budiono, S., ter Hofstede, A.H.M., van der Aalst, W.M.P., Reijers, H.A.: Towards privacy-preserving process mining in healthcare. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBIP, vol. 362, pp. 483–495. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_39
40. Mannhardt, F., Petersen, S.A., Oliveira, M.F.: Privacy challenges for process mining in human-centered industrial environments. In: Intelligent Environments, pp. 64–71. IEEE (2018)
41. Mannhardt, F., Koschmider, A., Biermann, L., Lange, J., Tschorsch, F., Wynn, M.T.: Trust and privacy in process analytics. Enterp. Model. Inf. Syst. Archit. Int. J. Concept. Model. **15**, 8:1–8:4 (2020)
42. 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

43. Centers for Medicare & Medicaid Services: The Health Insurance Portability and Accountability Act of 1996 (HIPAA) (1996). Online at http://www.cms.hhs.gov/hipaa/
44. Martin, N., et al.: Recommendations for enhancing the usability and understandability of process mining in healthcare. Artif. Intell. Med. **109**, 101962 (2020)
45. Elkoumy, G., et al.: Privacy and confidentiality in process mining - threats and research challenges. ACM Trans. Manage. Inf. Syst. (2021) accepted
46. Sweeney, L.: Simple demographics often identify people uniquely. Health (San Francisco) **671**(2000), 1–34 (2000)
47. Rafiei, M., van der Aalst, W.M.P.: Group-based privacy preservation techniques for process mining. Data Knowl. Eng. **134**, 101908 (2021)
48. Rafiei, M., van der Aalst, W.M.P.: Towards quantifying privacy in process mining. In: Leemans, S., Leopold, H. (eds.) ICPM 2020. LNBIP, vol. 406, pp. 385–397. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72693-5_29
49. van Dongen, B., Borchert, F.F.: Bpi challenge 2018 (2018)
50. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets. In: IEEE Symposium on Security and Privacy, pp. 111–125, IEEE Computer Society (2008)
51. Rafiei, M., von Waldthausen, L., van der Aalst, W.M.P.: Supporting confidentiality in process mining using abstraction and encryption. In: Ceravolo, P., van Keulen, M., Gómez-López, M.T. (eds.) SIMPDA 2018-2019. LNBIP, vol. 379, pp. 101–123. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46633-6_6
52. Sweeney, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzz. Knowl. Based Syst. **10**(05), 557–570, 101962 (2002)
53. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: l-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data **1**(1) 3-es (2007)
54. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: privacy beyond k-anonymity and l-diversity. In: 2007 IEEE 23rd International Conference on Data Engineering, pp. 106–115, IEEE (2007)
55. Fahrenkrog-Petersen, S.A., van der Aa, H., Weidlich, M.: Pretsa: event log sanitization for privacy-aware process discovery. In: 2019 International Conference on Process Mining (ICPM), pp. 1–8. IEEE (2019)
56. Dwork, C.: Differential privacy: a survey of results. In: Agrawal, M., Du, D., Duan, Z., Li, A. (eds.) TAMC 2008. LNCS, vol. 4978, pp. 1–19. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79228-4_1
57. Desfontaines, D., Pejó, B.: SOK: differential privacies. Proc. Priv. Enhancing Technol. **2020**(2), 288–313, 101962 (2020)
58. Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., Michael, J.: Privacy-preserving process mining - differential privacy for event logs. Bus. Inf. Syst. Eng. **61**(5), 595–614 (2019)
59. Fahrenkrog-Petersen, S.A., van der Aa, H., Weidlich, M.: PRIPEL: privacy-preserving event log publishing including contextual information. In: Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds.) BPM 2020. LNCS, vol. 12168, pp. 111–128. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58666-9_7
60. Fahrenkrog-Petersen, S.A., Kabierski, M., Rösel, F., van der Aa, H., Weidlich, M.: Sacofa: semantics-aware control-flow anonymization for process mining. In: ICPM, pp. 72–79. IEEE (2021)
61. Kabierski, M., Fahrenkrog-Petersen, S.A., Weidlich, M.: Privacy-aware process performance indicators: framework and release mechanisms. vol. 12751, pp. 19–36 (2021)

62. Elkoumy, G., Pankova, A., Dumas, M.: Mine me but don't single me out: differentially private event logs for process mining. In: ICPM, pp. 80–87. IEEE (2021)
63. Gentry, C.: Computing arbitrary functions of encrypted data. Commun. ACM **53**(3), 97–105 (2010)
64. Elkoumy, G., Fahrenkrog-Petersen, S.A., Dumas, M., Laud, P., Pankova, A., Weidlich, M.: Secure multi-party computation for inter-organizational process mining. In: Nurcan, S., Reinhartz-Berger, I., Soffer, P., Zdravkovic, J. (eds.) BPMDS/EMMSAD -2020. LNBIP, vol. 387, pp. 166–181. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49418-6_11
65. Lindell, Y.: Secure multiparty computation. Commun. ACM **64**(1), 86–96 (2021)
66. Bauer, M., Fahrenkrog-Petersen, S.A., Koschmider, A., Mannhardt, F., van der Aa, H., Weidlich, M.: Elpaas: event log privacy as a service. In: BPM (PhD/Demos). Volume 2420 of CEUR Workshop Proceedings., CEUR-WS.org, pp. 159–163 (2019)
67. Rafiei, M., van der Aalst, W.M.P.: Practical aspect of privacy-preserving data publishing in process mining. In: BPM (PhD/Demos). Volume 2673 of CEUR Workshop Proceedings., CEUR-WS.org, pp. 92–96 (2020)
68. Elkoumy, G., Pankova, A., Dumas, M.: Utility-aware event log anonymization for privacy-preserving process mining. EMISA Forum **41**(1), 37–38 (2021)
69. Maatouk, K., Mannhardt, F.: Quantifying the re-identification risk in published process models. In: ICPM Workshops, vol. 433, pp. 382–394. LNBIP. Springer (2021). https://doi.org/10.1007/978-3-030-98581-3_28
70. Zaman, R., Hassani, M.: On enabling GDPR compliance in business processes through data-driven solutions. SN Comput. Sci. **1**(4), 210 (2020)
71. Di Francescomarino, C., Ghidini, C.: Predictive process monitoring. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. LNBIP, vol. 448, pp. xx–yy. Springer, Cham (2022)
72. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Data-driven process discovery - revealing conditional infrequent behavior from event logs. In: Dubois, E., Pohl, K. (eds.) CAiSE 2017. LNCS, vol. 10253, pp. 545–560. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59536-8_34
73. Vidgof, M., Djurica, D., Bala, S., Mendling, J.: Cherry-picking from spaghetti: multi-range filtering of event logs. In: Nurcan, S., Reinhartz-Berger, I., Soffer, P., Zdravkovic, J. (eds.) BPMDS/EMMSAD -2020. LNBIP, vol. 387, pp. 135–149. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49418-6_9
74. van der Aalst, W.M.: A practitioner's guide to process mining: limitations of the directly-follows graph. Procedia Comput. Sci. 164, 321–328 (2019). (CENTERIS 2019 - International Conference on ENTERprise Information Systems/ProjMAN 2019 - International Conference on Project MANagement/HCist 2019 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2019)
75. Figl, K.: Comprehension of procedural visual business process models - a literature review. Bus. Inf. Syst. Eng. **59**(1), 41–67, 101962 (2017)
76. Mendling, J., Djurica, D., Malinova, M.: Cognitive effectiveness of representations for process mining. In: Polyvyanyy, A., Wynn, M.T., Van Looy, A., Reichert, M. (eds.) BPM 2021. LNCS, vol. 12875, pp. 17–22. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85469-0_2
77. Brons, D., Scheepens, R., Fahland, D.: Striking a new balance in accuracy and simplicity with the probabilistic inductive miner. In: ICPM, pp. 32–39. IEEE (2021)