A Literature Review on Process Mining

Seminar paper

Schilling, Lasse, FH Wedel, Wedel, Germany, imca103071@fh-wedel.de

Abstract

Process mining has long been successfully used in businesses. However, some challenges remain and new use cases for process mining emerge. This literature review highlights the challenges process mining is facing today using state-of-the-art literature. The paper rises predictions for promising topics of research in the future. Along with the topics of data, processes, use of process mining, and people this paper provides insights into process mining today. These categories serve as a possible reference frame for practitioners to update their use of process mining and start new process mining initiatives.

Keywords: process mining, literature review, data mining, business process modelling.

Table of Contents

1 Introduction

Process mining refers to different techniques for analyzing large amounts of event data. These event data are available in contemporary information systems of the business. Through evaluating the digital traces, process mining reveals business processes as they are executed (van der Aalst 2016). Process mining generates process transparency of a firm’s as-is process variations. Formerly firms relied on manual process modeling which was prone to subjective bias and incomplete process knowledge (Dumas et al., 2013). In contrast, process mining uncovers process variations that are less known or less frequent which leads to firms becoming aware of the variety of the process landscape for the first time. Consequently, firms can rapidly adapt to quickly changing business requirements (Eggers et al. 2021)

2 Research Design

2.1 Approach

This research was executed in the following steps: planning the review, conducting the review, reporting the findings, and discussing them. The research design uses a few steps proposed by Tranfield et al. 2003 but also sticks to the approach of Webster & Watson 2002. Planning and conducting the review were inspired by Tranfield et al. 2003. Analyzing the Findings and creating a discussion were executed using the concept matrix proposed by Webster & Watson 2002.

2.2 Planning the review

This literature review on process mining discovers the latest trends (since 1/1/2016) in the literature. Other reviews already summed up the prior literature on process mining to a good extent. Thus, this literature review aims to provide an overview of where the future of process mining is heading to. This paper concentrates on which trends in process mining emerge today and will be the content of future research.
Research question: *What is the state of the art of literature on process mining?*

The literature included in this paper was found in the following databases:
Science Direct, EBSCOHost, AIsel.
This literature review was not confined to geographical region, research methodology, or one journal set (Webster and Watson 2002).

These databases were chosen to build a foundation for searching. After searching on Science Direct and AiSel, EBSCOHost provided some new literature. However, at the end of the search, no new important literature was discovered. Hence the databases listed above tend to be collectively exhaustive. (From the perspective of the conducted review, the literature found on IEEE was neither new nor relevant for this review)

2.3 Conducting the review

The process for conducting the review consists of six steps: 1. Optimizing the search string, 2. Removing papers older than 1/1/2016, 3. Removing duplicates, 4. Clustering according to the quality ranking, 5. Using the backward search to identify new papers and 6. Relevance assessment by reading the full text.
Optimizing the search string for this review was iteratively executed. One iteration consists of searching with the same search string in every database listed above. Following the improvement of the search string in every iteration based on the results. Firstly, the search string was designed to alter a large

number of results. Then it was narrowed down to tackle the topic of process mining more specifically and make the number of results manageable. After ten rounds of search string optimization, the following search string offered a promising amount of literature: “("process mining" AND (business process modeling OR business process management OR data mining))”.

To further limit the results in this research the release date was set to 1/1/2016 to filter only for the latest literature.
After removing the duplicates, a total of 320 papers were identified.

The 320 papers were then categorized along with the vhb Ranking (Gesamtliste - vhbonline.de n.d.). A total of 33 papers were peer-reviewed and of A/ A+ quality. These papers in the highest quality category were taken into consideration.

Among the 33 peer-reviewed papers published in an A or A+ ranked journal the backward search led to some new literature that was not A-ranked. Nevertheless, this work is also included in this review if relevant. Notice that only backward searched literature was included which met the criteria of being published after 1/1/2016. Finally, the considered set of literature was identified consisting of 15 papers.

To assess which papers were included in this literature review the quality of the paper as well as the relevance and the actuality were taken into consideration. The quality of the articles included in this literature review had to be published in A and A+ ranked journals according to the vhb-ranking (Gesamtliste - vhbonline.de n.d.). The publication date of the articles had to be on 1/1/2016 or younger to insure actuality. Mitigating the relevance of the paper for this literature review was difficult. To mitigate the relevance of an article for this literature review, the paper had to fulfill the following criterium: process mining had to be the main topic of the article. Many papers only used process mining as a tool without describing it in detail or focusing on the process mining process.
To ensure the criterium was met, the relevance assessment consisted of two steps: (1) The abstract as well as the conclusion was checked whether the inclusion criterium was met. Finally, (2) The whole study was analyzed.

Filter process step	Number of results
Database results with removed duplicates and publishing date after 1/1/2016	320
Grouping according to the quality ranking: Using A/A+ literature	33
Applying backward search	49
Relevance assessment	15

Figure 1. “Filtering process”

3 Findings

The identified paper tackled different topics on process mining. However, some challenges or topics were similar. The topics of data, processes, use of process mining, and people occur significantly often in the literature.

3.1 Data in Process Mining

Process Mining is based on digital traces which are collected in the software systems of the business. The fit of the data in the event logs is thereby critical for process mining. This can also be seen in (a) the literature claiming that limitations for process mining lay in the data set by generating huge effort in the data pre-processing (Emamjome et al. 2020; Goel et al. 2021; Bernard et al. 2016)and (b) the literature specifically addresses this aspect in 14 of the 15 papers (Appendix 1). Thus, data tends to be of vital importance for process mining. Process mining can only be as good as the data in the event logs (Baier et al. 2018; Emamjome et al. 2020; Goel et al. 2021). Data or event logs face different challenges which can be mainly categorized into a) data quality, b) data availability and c) data granularity.

a) **Data quality** was mentioned most in the literature. Identifying, merging, and rectifying data quality issues is error-prone, time-consuming, and thus very costly. The Pre-processing of the data can take up to 60% of the effort invested in the process mining project (Emamjome et al. 2020). While taking up to 80 % of the process analyst's time (Goel et al. 2021). Data quality issues that can impact process mining have been identified (e.g. (Suriadi et al. 2017)), and techniques to detect (e.g. (Fischer et al. 2020)) and repair (e.g. (Dixit et al. 2018)) data quality issues have been developed (Goel et al. 2021).

The causes for data quality issues in the event logs are diverse. They are categorized in a huge variety of dimensions. They can be clustered along with their complexity: 'Quality issues in event logs arise for a variety of reasons - some simple (e.g. incorrect construction of a format mask for a datetime column during ETL) and some complex (e.g. different task completion behaviors across resources - task-by-task completion during the day vs batch completion at the end of the day)'(Emamjome et al. 2020). The origin of quality issues can also be categorized along with the actor or trigger in the process of generating the data: (a) social, (b) personal, and (c) material triggers (Emamjome et al. 2020). Personal refers to individual mistakes while social quality issues arise from different connotations by different actors used in the creation process. The material triggers refer to information systems, software logic, and transmission mechanisms which also affect the quality of the event log(Emamjome et al. 2020). Note that these categories interact with each other. A Review of process mining studies reveals that many data quality problems in event data are created as a result of poor design of interfaces and inconsistencies between the design, tasks, and users' requirements (Lanzola et al. 2014). For example, (Suriadi et al. 2017) identified data quality issues resulting from the form-based design of user interfaces (Goel et al. 2021).

'The main quality issues can be broadly categorized into noise and incompleteness. Noise refers to behavior that is in the data but should not be represented in a discovered process, while incompleteness refers to event logs being only a finite sample of possible process behaviors – a sample that is not necessarily representative and that may be missing important behavior'(Breuker et al. 2016). While others categorized event log quality issues as missing data, incorrect data, imprecise data as well as irrelevant data in the event logs. (Suriadi et al. 2017).

Data has to be cleaned to gain valuable and correct outcomes out of process mining. Therefore, noise has to be removed to minimize information loss and produce an event log so that the event log is valid for analysis (Suriadi et al. 2017).

To do so the literature provides a few concepts and entire frameworks. To handle data quality issues (a) data quality issues can be prevented from happening and (b) the impact of quality issues can be minimized by designing noise and incompleteness resistant algorithms and techniques.
For (a) assessing quality issue prevention the ImperoPD framework (Goel et al. 2021) delivers a suitable approach. The framework addresses data quality in depth by defining 20 capabilities that are required for sufficient data governance. These capabilities are categorized into five domains: Business Strategy Management, Process Management, Information Technology Management, Organization and Project Management, and People Management (Goel et al. 2021).

However, the literature also provides some content for (b) reducing the impact of quality issues in the mining process. Many process-discovery techniques, like the Heuristics Miner and the AGNEs miner, address noise by analyzing the frequency of patterns of behavior. The pattern is used to deduce the process model only if the pattern is found sufficiently often (Goel et al. 2021; Breuker et al. 2016).
Addressing the incompleteness of the data involves switching to a different process representation. Most discovery algorithms use Petri nets as representation. Petri nets specify the set of all acceptable event sequences for a process. Other discovery techniques include the Fuzzy Miner which does not produce Petri nets leading to a higher resilience to incompleteness (Breuker et al. 2016).
Probabilistic techniques are popular in many fields because of their ability to handle noisy, possibly incomplete data (Breuker et al. 2016). Probabilistic approaches use hidden Markov models. These tend to overfitting process models (Breuker et al. 2016). To avoid overfitting, a new approach uses probabilistic techniques directly over the set of all conceivable event sequences (Breuker et al. 2016).

b) **The availability of data** faces three major challenges: Firstly, relevant data is kept under lock mainly due to personal rights regarding surveillance. Employees tend to feel controlled by another tool to check their performance (Grisold et al. 2021). Data Regulations enforced by law can have significant implications for the use of process mining (Grisold et al. 2021). To obey personal rights, data has to be anonymized or can’t be collected in the first place.
Secondly, the place to look for the data is unclear. Data is scattered over various systems (data warehouses, applications, etc.) and unknowingly collected or flowing through the business without being integrated into the process mining application (Grisold et al. 2021). For evaluation, the Data must be accessible in a physical way: Data can be collected from various systems and then stored in a central place. Many organizations use a centralized data lake for this very purpose (Eggers et al. 2021). Along with the physical availability, the knowledge of where to find the data must be spread among the staff. This democratization of the data can be reached by implementing a cross-functional platform for employees to communicate (Eggers et al. 2021).
Thirdly the data, relevant for the event log, is not yet collected. For instance: in the industry special sensors have to be added to the production cycle to collect relevant data. Often sub-processes tend to be more specific and therefore need precise data to be tracked. On the other hand, for some meaningful data, there is simply no sensor to collect it. For instance, if one wanted to include what people think as a contextual factor. Thus, only observable data can be collected and included in process mining. (B. T. Pentland et al. 2020).

Finally, companies need to develop data collection strategies, beginning with an analysis of the information needs or analytical opportunities. They have to define which data to collect from existing sources as well as for which data new sources have to be established (Grisold et al. 2021).

c) Another debate in the literature assesses **the granularity of the data** (B. T. Pentland et al. 2020).

The conventional wisdom is that large numbers of granular categories need to be reduced into a smaller number of categories with higher abstraction level (B. T. Pentland et al. 2020). This keeps the data at a manageable level. This resonated with the conventional belief that a high number of fine-grained categories of events was considered undesirable (B. T. Pentland et al. 2020).

However, the literature suggests the contrary. Enrich the datasets with contextual information and embrace the complexity of the data set (B. T. Pentland et al. 2020). Adding contextual information can result in a very large number of fine-grained categories of events but larger numbers of categories can make process data more informative for theorizing in general (B. T. Pentland et al. 2020). Increasing granularity (with adding context) of the data leads to more entropy which is more informative (B. T. Pentland et al. 2020); The highly granular data is also beneficial for theorizing about process dynamics. Process dynamics are stressing the question of why processes change over time. This can be revealed by process mining (Pentland et al. 2021). Some approaches state that the data in the event logs have not to be of atomic granularity but if it does it is assumed to be beneficial (Pentland et al. 2021). While other approaches already rely on contextual data (which leads to a high granularity). The contextual data makes these approaches much more feasible for theorizing about process dynamics (B. T. Pentland et al. 2020).

Having big datasets with contextual and granular data affects the real-time use of process mining. Processing time scales up to such an extent that real-time evaluation is only possible through big data (Pentland et al. 2021). Parallel computing technologies from Big Data provide enough processing power for probabilistic models. Without parallel computing, the runtime of the entire mining process can consume up to several months (Breuker et al. 2016). “In practical applications, mining can be executed periodically, not every day, so runtimes of minutes or even hours could be acceptable. The prediction itself—the question of “what event will be next”—can be computed in milliseconds, a satisfactory time even for real-time applications (Breuker et al. 2016)”.

3.2 Processes

Process mining is used in a vast landscape of processes. However, no prior framework can classify which processes could benefit from process mining or which processes process mining should be used (Grisold et al. 2021). Following the question of which processes could benefit from process mining the literature provides only a few observations where process mining is used:

The processes where process mining is used are not directly connected to money. This explains why process mining is found not only in value chain activities but in support processes as well (Dunzer et al. 2021). Rather the amount of data that is processed tends to be relevant as well as the number of people involved in the process (Grisold et al. 2021). However, this observation is not always backed in the industry; Cellular workstations deliver less data than assembly layouts but were the subject to process mining in the business (Dunzer et al., 2021).

This counterintuitive observation above highlights another important aspect: the use of process mining was arranged by humans who wanted to look out for processes that could turn multiple ways and therefore seemed to be of higher complexity. The selection of processes for process mining is thereby affected by the street light effect. The street light effect manifests the tendency to look for errors where there is light which means that errors become easier to assess. The street light effect could play a large role not only in the use of process mining but also in its effectivity as well. Not only the choice of processes for process mining is affected by the street light effect but the pre-processing data cleaning as well (Emamjome et al. 2020).

In production companies, three main areas for using process mining in operative processes were identified. First, companies try to alter information about repair and rework steps in the production process. Process mining can identify repair activities as bottlenecks and root-cause for disadvantageous machine disposition. Second, in the quality assurance process mining was applicable as well. Third, the largest proportion of applications of process mining was found in machine utilization and workstation efficiency. Companies can optimize the performance of these operative processes with process mining (Dunzer et al. 2021).

Process mining was also used in continuous flow processes. However continuous flow processes have no logical termination which results in a long run time of the process as well as a huge period between activities. Thus, process mining is difficult under these circumstances so only process discovery was applicable in such processes (Dunzer et al. 2021).

Following the implementation of process mining, practitioners use either a bottom-up or a top-down approach. Therefore, processes can be categorized into end-to-end processes, sub-processes, or process landscapes. In the top-down approach, an end-to-end process is selected for process mining. In the bottom-up approach, the use of process mining starts at the subprocesses. The choice of the approach (bottom-up, top-down) leads to different results of process mining. Thus, the choice of the process (end-to-end, sub-process) impacts the outcome of process mining projects (Eggers et al. 2021). While a top-down approach fosters further awareness and standardization of sub-process as departments are required to adopt process mining for monitoring specified process KPIs, the firms struggle with establishing the self-governed exploratory use of process mining across functions/ departments to discover unknown process complications. The exploratory use can be enabled from the bottom-up approach. But the bottom-up approach consequently lacks a coordinated approach to aggregate process knowledge on a global level. This can only be accomplished iteratively. However end-to-end as well as sub-process can benefit from process mining (Eggers et al. 2021).

Process Mining is also used in connection with RPA. It is important to note that robots typically do not automate complete end-to-end processes but only sub-processes or certain tasks thereof (Wanner et al. 2019). Process mining is used in these sub-processes or tasks to identify the real processes for automation. Hence, implementing RPA is also a driver for using process mining on a specific process.

3.3 Use of Process Mining

The discovery use of process mining focuses on generating a process model based on the evaluation of digital trace data. The literature highlights that process discovery was found most in the literature (Appendix 1) and businesses tend to use process discovery exclusively (Dunzer et al. 2021)
For conformance checking a previously generated process model is checked via evaluating digital trace data to quantify its correctness or to uncover deviations. Businesses only rarely use conformance checking (Dunzer et al. 2021). One reason for this phenomenon could be that the a priori model is missing or is not up to date. Businesses tend to not maintain their process models (Seeliger et al. 2016; Dunzer et al. 2021). In that regard, process mining often falls into oblivion after being used once (Grisold et al. 2021). Therefore, the literature provides content on how to make conformance checking easier. Conformance checking normally requires an a priori model as a reference point, but the latest trends show that conformance checking can be done without relying on an a priori model by transferring it into a graph reachability problem (Seeliger et al. 2016).

Another hindrance to conformance checking was the mapping from data in event logs on the activities in the a priori model. This mapping commonly relied on prior knowledge in the business, which was often missing in the reality(Baier et al. 2018). The exact mapping was generally done manually due to its combinatorial complexity (Baier et al. 2018). Here the literature undertook an effort to tackle this challenge. (Baier et al. 2018) proposed an approach that works in a semi-automated fashion. Also, it was designed to resist noise which makes it quite practical. Under certain conditions, it can still provide valid outcomes even with 50% of noise in the data set. Thus, conformance checking could be more accessible for businesses in the future.

Enhancement refers to a method of adding newly gained information into a previously generated process model. This type of process mining allows the prediction of bottlenecks or the remaining flow time. According to the richness of the event log, it is also possible to discover roles, work distribution mechanisms, and resource characteristics (Computer and 2011 n.d.). With the increasing popularity of predictive analytics, interest in using process mining for predicting the future has risen (Breuker et al. 2016). Predicting the future behavior of processes can enable managers to act proactively in anticipation of events (Breuker et al. 2016).

The mainstream design principle for such systems is to enhance an a-priori business model (constructed with process mining) with additional information. Therefore, the a-priori model is highly important for the enhancement technique because it characterizes the current state of the process. Predictions made in the enhancement technique rely on the characterization in the a-priori model (Breuker et al. 2016). A new approach for enhancement using a probabilistic technique and grammatical interference impresses through its ability to handle noise and possibly incomplete data. These are major challenges in the data quality as presented above. While traditional enhancement methods relied on the a priori model and incorporated the mistakes in this model for the predictions, this approach does not rely on a discovered process model. This approach is not only resilient to noise and incompleteness but also has a staggering performance(Breuker et al. 2016).

Another dominant finding in the literature is focusing on theorizing about process dynamics. Normally, the goal of process mining is to find a clean model that provides a reference for the execution of a process (discovery). Most process mining algorithms assume that the underlying process is stable such that discovery of the stable process and conformance checking are the primary applications (van der Aalst, 2005, 2011b). To that end, a typical goal of visualization has been to simplify overly cluttered graphs into *comprehensible* models (Breuker et al.,2016; van der Aalst, 2011a) (B. T. Pentland et al. 2020). In contrast, the literature now aims to reveal the extent of the mess and displays processes as they unfold in event time due to process dynamics(B. T. Pentland et al. 2020). Process dynamics or process drift refer to the changes in the structure of a process over time. Reasoning that processes drift, the literature uses process mining to gain new insights into it (B. Pentland et al. 2020). With the help of process mining, it was possible to observe that processes are prone to unanticipated bursts of complexity followed by relative inertia (B. Pentland et al. 2020). With process mining, even small or for humans unnoticeable process drifts are possible to detect (Pentland et al. 2021). Theoretical papers now try to find answers to where process changes originate from and how processes change over time. Accordingly, the counterpart to how processes change, the question of what keeps a process on track is also an important aspect in the literature. They observe the trajectory of the process as well and if process changes are predictable. With the availability of contextual data, theorizing about process dynamics gains prominence (B. T. Pentland et al. 2020). This research area could allow predicting the future behavior of single process instances at an early stage of process execution (Breuker et al. 2016; B. T. Pentland et al. 2020).

3.4 People

This paragraph addresses which skills and capabilities of the staff were found in the literature. Additionally, process mining also has implications for the employees which are presented below. The following skills are not exhaustive but were highlighted in the literature as complicated or overly important.

Process awareness is a common goal for process mining initiatives which impacts the people working in the observed process. Process awareness refers to the notion of employees being aware of how they perform and how their actions are linked or embedded into a bigger process (Eggers et al. 2021). Process awareness consists of a multi-layered construct that requires the development of a shared process language and understanding (Eggers et al. 2021).

Process mining creates an evidence-based and objective reference frame for achieving process awareness (Eggers et al. 2021). Thus, process awareness is no longer depending on an individual perception (Eggers et al. 2021). However, raising process awareness based on process mining has still proven difficult for firms.

In addition to process mining, the literature highlights that the skill of process-oriented thinking is crucial for process mining initiatives to enable process awareness (Eggers et al. 2021; Goel et al. 2021). A recent study states that 80% of the 360 firms surveyed use process mining to achieve process transparency and awareness but they face challenges in realizing the expected benefits due to resistance to insufficient process-oriented thinking in the workforce (Eggers et al. 2021). The goal of achieving process awareness does not automatically follow the use of process mining (Eggers et al. 2021). This reveals that process-oriented thinking in the workforce is of vital importance for process mining initiatives.

Another important capability mentioned is communication among people. Communicating and building a shared language on process mining is highlighted as an important capability. Communication enables organizations to access the right data as well as effective techniques to manage data quality (Goel et al. 2021). In addition, communication contributes to the dissemination of protocols about the secure, confidential, and legitimate use of data for process mining analysis. Exchanging and sharing knowledge through communication increases the effectiveness of process mining. This also contributes to the goal of process awareness.

As the importance of data was highlighted above, a commitment to data quality is also a capability for the staff. The commitment to obtain data of high quality increases the success of process mining initiatives (Goel et al., 2021).

These skills and capabilities should be supported by training (Goel et al. 2021; Eggers et al. 2021). The literature highlights the importance of training for implementing the mindset and skills beneficial for process mining and process mining initiatives.

Process mining has implications for people. The increased transparency of processes in the business leads to a fear of surveillance in the staff (Eggers et al. 2021). This fear is persistent/stubborn and can be continuously found even though the workforce received process mining training (Eggers et al. 2021). To resolve concerns about supervision, personal information can be anonymized (Eggers et al. 2021). Furthermore, the protection of sensitive data can be achieved by restricting access via encryption and authentication (Eggers et al. 2021). Privacy-preserving approaches are covered by a method proposed by Mannhardt et al., 2019. This method allows for safely reuse of collected data; however, a proposed privacy engine introduces noise as a trade-off.

Accordingly, the role of a privacy officer who manages the privacy and ethics related to a process mining project is proposed by the literature (Goel et al. 2021).

4 Discussion

This paper shows the latest knowledge and trends in the field of process mining. This paper raises new questions where further research is required

Practitioners can use this paper to further structure the use and implementation of process mining. This paper can help to sense problems in the use of process mining and how to assess them. Furthermore, practitioners can use this paper as a foundation to improve their use of process mining. Consequently, practitioners can use this paper as a starting point for new process mining projects in their companies.

Finally, this paper provides a foundation for future research on process mining. Process mining itself is not new to the literature. Thus, process mining is already discovered to a good extent. Nevertheless, process mining still needs some further research on the following aspects:

Data is (still) of vital importance to process mining. Quality assessments are still costly and time-consuming. Nonetheless, the literature already provides some help. Regarding granularity of the data, the literature developed a new approach by adding contextual data to the event log. This has implications for the run and processing time. The evaluation of real-time data could be affected by processing contextual data due to the enhanced processing time. Using big data analysis in the combination with contextual enriched data sets on real-time processes is hardly found in the literature but could be of practical importance in the future.

Process mining seems suitable for nearly every kind of process. However, no prior framework addresses which processes could benefit from process mining. In the literature, multiple studies are found which show the use of process mining on a specific process. But a systematic approach to which processes would benefit from process mining and how to prioritize process mining initiatives is still lacking.

Although process mining is used in practice, a systematic approach to how businesses engage with the process transparency through process mining is a question stressed by the literature (Grisold et al. 2021)Which implications does the process transparency have for change management? How to leverage the newly gained knowledge about the process? Accordingly, practitioners highlight that they neither measure the effectiveness nor the outcome of process mining (Grisold et al. 2021). A framework for mitigating the concrete and measurable outcome of process mining in a structured manner would be desirable for the future.

A new trend in the use of process mining is the discovery of process dynamics. Identifying the reason why and how processes change is yet subject to theorizing. The use for practitioners however is left scarce. Knowing why processes change could lead to new predictive analytics in process mining. It might be possible to predict the impact of process changes. Leveraging the theory of process dynamics on real live processes would add new insights to the literature. Predicting change and showing implications on real live processes would be a promising field for further research.

This study is not free of limitations. Even though the search string was iteratively narrowed down, this paper might not be exhaustive due to the lack of using further databases as well as excluding possibly important literature through narrowing down the search string. By clustering the results according to their vhb-ranking (Gesamtliste - vhbonline.de n.d.), papers with a lower quality have not been taken into consideration. This could have excluded important literature as well. Finally, the selection of the important literature after scanning the abstract and reading the full paper is prone to selection bias.

5 Conclusion

This literature review provides an overview of the latest trends in process mining. The literature since 1/1/2016 was scanned and structured along with the topics of data, processes, use of process mining as well as people. The literature highlights that data is highly important for process mining. New trends stress embracing the complexity of granular data sets. Contextual data is added to the event log to uncover process dynamics. Theorizing process dynamics marks a new use of process mining. Process mining is used to uncover reasons for process change. Still subject to theorizing, predicting processual changes is promising for the future. While process mining is used in a wide range of processes, practitioners still have difficulties mitigating the improvement of process mining. Accordingly, a structured framework which processes benefit most from process mining and how to prioritize process mining initiatives is still lacking. Thereby, this paper hopefully aids as an entry point for leveraging and researching the latest trends in process mining.

Appendix 1: Concept Matrix (Webster and Watson 2002)

SOURCE	RANKING	DATE	DATA	PROCESSES	PEOPLE	USE OF PROCESS MINING			
						Discovery	Conformance	Enhancement	Process Dynamics
Goel et al., 2021	A	2021	x	x	x				
Bernard et al., 2016	A	2016	x	x		X	X	X	
B. T. Pentland et al., 2020	A	2020	x	x					X
Seeliger et al., 2016	A	2016	x		x		X		
Wanner et al., 2019	A	2019		x		X			
Breuker et al., 2016	A+	2016	x			X		X	
B. Pentland et al., 2021	A+	2021	x	x					X
B. Pentland et al., 2020	A+	2020	x	x					X
Eggers et al., 2021	B	2021	x	x	x	X		X	
Grisold et al., 2020	C	2020	x	x		x	x		
Grisold et al 2021	C	2021	x	x	X	x			
Suriadi et al., 2017	C	2017	x						
Dunzer et al., 2021		2021	x	x		X	X	X	
Emamjome et al., 2020		2020	x		x				
Baier et al., 2018		2017	x		x		x		

References

van der Aalst, W. (2016). Process mining: Data science in action. *Process Mining: Data Science in Action* [online], 1 January 2016, pp.1–467.

Baier, T., di Ciccio, C., Mendling, J. and Weske, M. (2018). Matching events and activities by integrating behavioral aspects and label analysis. *Software and Systems Modeling* [online], 17(2), pp.573–598. Available from: https://link.springer.com/article/10.1007/s10270-017-0603-z [accessed 22 June 2022].

Bernard, G., Boillat, T., Legner, C. and Andritsos, P. (2016). When Sales Meet Process Mining: A Scientific Approach to Sales Process and Performance Management. *ICIS 2016 Proceedings* [online], 11 December 2016. Available from: https://aisel.aisnet.org/icis2016/ISDesign/Presentations/15 [accessed 22 June 2022].

Breuker, D., Matzner, M., Delfmann, P. and Becker, J. (2016). Comprehensible Predictive Models for Business Processes. *Management Information Systems Quarterly* [online], 40(4). Available from: https://aisel.aisnet.org/misq/vol40/iss4/12 [accessed 22 June 2022].

Computer, W.V. der A.- and 2011, undefined. Using Process Mining to Bridge the Gap between BI and BPM. *padsweb.rwth-aachen.de* [online]. Available from: http://www.padsweb.rwth-aachen.de/wvdaalst/publications/p651.pdf [accessed 22 June 2022].

Dumas, M., la Rosa, M., Mendling, J. and Reijers, H.A. (2013). Fundamentals of Business Process Management. *Fundamentals of Business Process Management* [online], 2013.

Dunzer, S., Zilker, S., Marx, E., Grundler, V. and Matzner, M. (2021). The Status Quo of Process Mining in the Industrial Sector. *Wirtschaftsinformatik 2021 Proceedings* [online], 11 February 2021. Available from: https://aisel.aisnet.org/wi2021/EManagementofdigitalprocesses20/Track20/2 [accessed 21 June 2022].

Eggers, J., Hein, A., Böhm, M. and Krcmar, H. (2021). No Longer Out of Sight, No Longer Out of Mind? How Organizations Engage with Process Mining-Induced Transparency to Achieve Increased Process Awareness. *Business & Information Systems Engineering* [online], 63(5). Available from: https://aisel.aisnet.org/bise/vol63/iss5/2 [accessed 21 June 2022].

Emamjome, F., Andrews, R., Hofstede, A. ter and Reijers, H. (2020). Alohomora: Unlocking Data Quality Causes Through Event Log Context. *ECIS 2020 Research Papers* [online], 15 June 2020. Available from: https://aisel.aisnet.org/ecis2020_rp/80 [accessed 21 June 2022].

Gesamtliste - vhbonline.de. Available from: https://vhbonline.org/vhb4you/vhb-jourqual/vhb-jourqual-3/gesamtliste [accessed 23 June 2022].

Goel, K., Emamjome, F. and Hofstede, A. ter. (2021). Data Governance for Managing Data Quality in Process Mining. *ICIS 2021 Proceedings* [online], 12 December 2021. Available from: https://aisel.aisnet.org/icis2021/governance/governance/9 [accessed 21 June 2022].

Grisold, T., vom Brocke, J., Gross, S., Mendling, J., Röglinger, M. and Stelzl, K. (2021). Digital Innovation and Business Process Management: Opportunities and Challenges as Perceived by Practitioners. *Communications of the Association for Information Systems* [online], 49(1), p.27. Available from: https://aisel.aisnet.org/cais/vol49/iss1/27 [accessed 17 June 2022].

Lanzola, G., Parimbelli, E., Micieli, G., Cavallini, A. and Quaglini, S. (2014). Data quality and completeness in a web stroke registry as the basis for data and process mining. *Journal of healthcare engineering* [online], 5(2), pp.163–184. Available from: https://pubmed.ncbi.nlm.nih.gov/24918182/ [accessed 23 June 2022].

Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M. and Michael, J. (2019). Privacy-Preserving Process Mining. *Business & Information Systems Engineering 2019 61:5* [online], 61(5), pp.595–614. Available from: https://link.springer.com/article/10.1007/s12599-019-00613-3 [accessed 23 June 2022].

Pentland, B., Liu, P., Kremser, W. and Hærem, T. (2020). The Dynamics of Drift in Digitized Processes. *Management Information Systems Quarterly* [online], 44(1). Available from: https://aisel.aisnet.org/misq/vol44/iss1/4 [accessed 22 June 2022].

Pentland, B., Vaast, E. and Wolf, J. (2021). Theorizing Process Dynamics with Directed Graphs: A Diachronic Analysis of Digital Trace Data. *Management Information Systems Quarterly* [online], 45(2). Available from: https://aisel.aisnet.org/misq/vol45/iss2/16 [accessed 22 June 2022].

Pentland, B.T., Recker, J., Wolf, J.R. and Wyner, G. (2020). Bringing Context Inside Process Research with Digital Trace Data. *Journal of the Association for Information Systems* [online], 21(5), p.5. Available from: https://aisel.aisnet.org/jais/vol21/iss5/5 [accessed 22 June 2022].

Seeliger, A., Nolle, T., Schmidt, B. and Mühlhäuser, M. (2016). Process Compliance Checking using Taint Flow Analysis. *ICIS 2016 Proceedings* [online], 11 December 2016. Available from: https://aisel.aisnet.org/icis2016/DataScience/Presentations/6 [accessed 22 June 2022].

Suriadi, S., Andrews, R., ter Hofstede, A.H.M. and Wynn, M.T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems* [online], 64, pp.132–150.

Tranfield, D., Denyer, D. and Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management* [online], 14(3), pp.207–222. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/1467-8551.00375 [accessed 22 June 2022].

Wanner, J., Hofmann, Fischer, M., Imgrund, F., Janiesch, C. and Geyer-Klingeberg, J. (2019). Process Selection in RPA Projects – Towards a Quantifiable Method of Decision Making. *ICIS 2019 Proceedings* [online], 6 November 2019. Available from: https://aisel.aisnet.org/icis2019/business_models/business_models/6 [accessed 22 June 2022].

Webster, J. and Watson, R.T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), pp.xiii–xxiii. Available from: http://www.jstor.org/stable/4132319.