Towards Understanding the Role of the Human in Event Log Extraction

Vinicius Stein Dani[1(✉)], Henrik Leopold[2], Jan Martijn E. M. van der Werf[1], Xixi Lu[1], Iris Beerepoot[1], Jelmer J. Koorn[1], and Hajo A. Reijers[1]

[1] Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
{v.steindani,j.m.e.m.vanderwerf,x.lu,i.m.beerepoot,j.j.koorn, h.a.reijers}@uu.nl

[2] Kühne Logistics University, Großer Grasbrook 17, 20457 Hamburg, Germany
henrik.leopold@the-klu.org

Abstract. Process mining is widely used to visualize, analyze, and improve business processes. However, often its application is hindered by the considerable preparation effort that needs to be conducted by humans. One of the key tasks required in this context is obtaining the input artifact for process mining techniques: the event log. The data that is required for building such an event log typically needs to be collected from several databases and then transformed into a suitable format. While it has become clear to both academics and practitioners that the amount of human work is substantial, there is no deep understanding of the exact activities humans need to perform. Therefore, we use this paper to develop a precise understanding of how humans are involved in event log extraction. Based on a structured literature review and qualitative data coding, we derive a taxonomy of human tasks in event log extraction. This taxonomy can serve as input for both future automation efforts, as well as for process mining methodologies.

Keywords: Process mining · Event log extraction · Human tasks

1 Introduction

Many organizations use process mining techniques to visualize, analyze, and improve their business processes. Among others, process mining has been applied in auditing [20], production planning [15], and in the healthcare domain [32]. While process mining has been shown to come with many benefits [18], its application is still often hindered by the considerable preparation effort that needs to be conducted by humans [12]. One of the key challenges in this context is to obtain the input artifact for process mining techniques, the so-called *event log*.

Depending on the specific IT landscape, the data that is required for building an event log must be collected and extracted from several databases and then transformed into an appropriate format that can be processed by available process mining tools. While a variety of automated techniques have been proposed to support organizations in obtaining event logs (e.g., [7,9,27]), human support is still required at several stages. Unfortunately, there is currently no deep understanding of the *exact activities* humans

A. Marrella and B. Weber (Eds.): BPM 2021 Workshops, LNBIP 436, pp. 86–98, 2022.
https://doi.org/10.1007/978-3-030-94343-1_7

Table 1. Excerpt from an event log of an ordering process

No	Case identifier	Event	Timestamp
1	123	Receive order	2021-05-11 11:35
2	123	Collect order items	2021-05-11 12:41
3	234	Receive order	2021-05-12 09:03
4	123	Send invoice	2021-05-12 13:49
5	234	Collect order items	2021-05-12 17:12
6	123	Ship order	2021-05-13 08:54

are performing in this context. While this may seem surprising, this can be explained by the fact that available techniques for event log extraction often focus on rather isolated technical aspects and not on the full range of tasks that are required in a practical setting.

Recognizing that the human involvement in event log extraction comes with considerable time and cost, we set out to develop a precise understanding of how humans are involved in the context of event log extraction. We believe that such an understanding is vital input for both future automation efforts as well as for process mining methodologies. Therefore, the research question of this paper is "*What are the specific manual tasks that humans perform in the context of event log extraction?*". To answer our research question, we first conduct a structured literature review targeting process mining case studies. Based on qualitative data coding, we then use the identified papers to derive a taxonomy of human tasks in event log extraction.

The remainder of the paper is organized as follows. Section 2 discusses the background of this paper. Section 3 describes our research method. Section 4 presents our taxonomy of manual tasks in even log extraction. Section 5 reflects on the implications and limitations before Sect. 6 concludes the paper.

2 Background

Process mining is a family of data analysis techniques that aims to discover, monitor, and improve organizational processes by analyzing data from so-called event logs [1]. Such event logs can be obtained from various IT systems and provide insights into how organizational processes are executed. Table 1 shows an excerpt of a simple event log of an ordering process. It shows that each entry of an event log must have at least three attributes: a case identifier, an event, and a timestamp. The event column reveals *what* happened. The timestamp column shows *when* the event occurred. The case identifier relates each event to a particular process execution, often referred to as *case*. In the example, we only observe two different cases (123 and 234).

Event logs like the one from Table 1 are a prerequisite for any available process mining technique. The process of obtaining event logs is called *event log extraction*. It is a complex and time-intensive process, which requires human involvement at several stages. The data required for constructing an event log often resides in a variety of sources. One of the main challenges is that many information systems are not process-centric and, therefore, do not record events or case identifiers explicitly. This means that we may need to identify and extract event data from a variety of different databases and transform them to a process-centric event log [12].

To provide automated support for this endeavor, many automated techniques have been developed. Recognizing the large variety of potential data sources and requirements in practice, available techniques differ considerably with respect to required inputs, their output, and also their limitations. Among others, there are extraction techniques that build on ontologies [7], redo logs [9], and database objects [27]. All these techniques require human intervention or input at some stage. Unfortunately, the exact role of the human is not always clear. Since many of these techniques address rather specific problems of event log extraction, the required human involvement can often only be understood when these techniques are applied in practice.

Given that the human involvement in event log extraction is both time and cost-intensive, we use this paper to develop a precise understanding of the respective human tasks. We argue that such an understanding is a prerequisite for further automation efforts, as well as for developing process mining methodologies. Therefore, we define our research question as: "*What are the specific manual tasks that humans perform in the context of event log extraction?*".

3 Research Method

To answer our research question, we followed a two-step approach. We first conducted a structured literature review. Based on the result, we then derived a taxonomy of human tasks via qualitative data coding. In the Sects. 3.1 and 3.2, we describe both steps in detail.

3.1 Literature Study

To identify which manual tasks have been performed in the context of event log extraction, we decided to focus on papers conducting case studies in process mining projects. In Scopus, we used the search string: ("process mining" AND "case stud*"). We included peer-reviewed papers that were published in journals or conferences between 2000 and 2020. Next, we filtered out duplicate papers, papers that were not in English, and papers that merely mentioned process mining. The search resulted in 191 papers. We identified papers that could contain tasks related to event log extraction by making a first read of the papers. This resulted in a set of 120 papers that were assessed more closely to extract the actual tasks related to manual efforts during log extraction.

As a next step in the study, we went through the 120 papers to identify human tasks related to event log extraction. To label a task as manual, we defined the following criteria:

1. It is explicitly mentioned that the task was performed manually.
2. There was no explicit counter statement that the task was performed (semi-) automatically.
3. We assume that conceptual models, such as data models and process models, were created manually unless the authors explicitly stated that they were created in an automated fashion.

With this in mind, the final set of papers from which we identified manual efforts is composed of 46 papers. On these papers we performed the coding as presented in the next section.

3.2 Coding and Taxonomy Derivation

To derive a taxonomy of manual tasks, we coded the 46 papers resulting from our literature review based on the coding of qualitative data [33]. We performed three specific steps: First, we performed hypothesis coding. This means that we devised a set of codes we expected to find in the data without actually conducting any further analysis. Second, we performed a holistic coding to identify possible categories that could emerge from and represent the data. Third, we compared both coding schemes and discussed how they relate to each other to achieve a more concise representation of the identified categories. We identified that we could directly match the holistic coding entries to our hypothesis coding, which led us to the taxonomy presented in Sect. 4.

For example, the following codes emerged from the holistic approach: i) "search into the data to select case perspective", supported by "*[...] It is therefore possible to consider the data from at least three different 'case' perspectives, i.e. an incident may be considered as a case, each patient may be considered as a case, or each response unit may be considered as a case. [...]*" [4]; and, ii) "identify, from discussion with domain expert, which case should be considered", supported by "*[...] in this case we constructed three different event logs according to the different perspectives. [...] After several discussions, it appeared that both the team flow and the document type flow represented the business process best [...]*" [10]. Both codes are related to the selection of a case, which can be summarized by our hypothesis code "Select case notion". As can be seen from these examples, we used lower-level coding in our holistic coding, and high-level coding in our hypothesis coding.

4 A Taxonomy for Manual Tasks in Event Log Extraction

In this section, we present the outcome of our research: a taxonomy for manual tasks in event log extraction. We first provide an overview of our taxonomy in Sect. 4.1. In the remainder of the section, we discuss the details of each of the five categories.

4.1 Overview

A visual representation of our taxonomy is shown in Fig. 1. It consists of five categories, represented as four squares and one rectangle. These five categories resemble the main manual tasks we identified in event log extraction. Starting from the context and scope definition, the data source need to be assessed, and attributes need to be selected. Based on this selection, an event log can be extracted from the data source, and needs to be assessed before it can be used in process mining. Note that there is no strict flow. Some tasks might be executed repeatedly or in an arbitrary order. The numbers in brackets indicate in how many papers these task categories were mentioned.

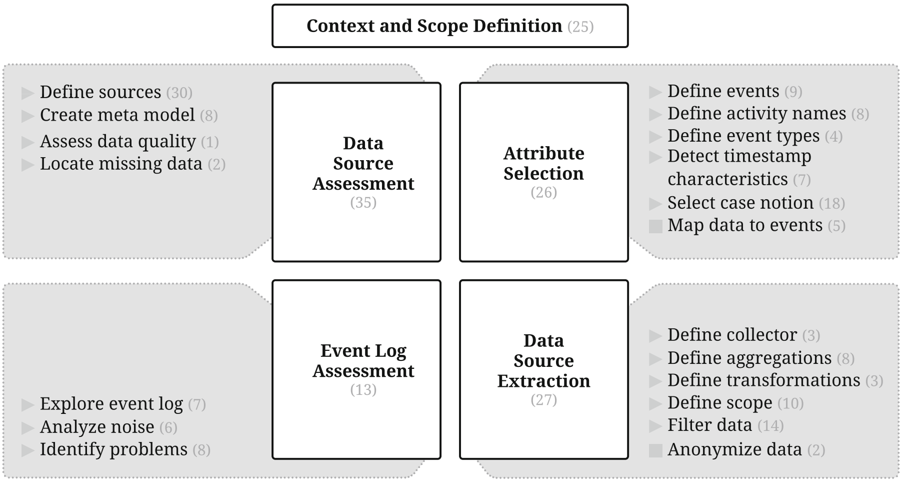

Fig. 1. Taxonomy of manual tasks in event log extraction.

4.2 Context and Scope Definition

The first step as indicated by several process mining methodologies (cf. [13]) is to define the overall context and scope (supported by 25 papers from our literature study). Naturally, this is a completely manual task that mainly consists of discussions between the process analyst and the stakeholders. The scope can vary between very exploratory to very specific. Once there is an agreement on the scope, this defines the subsequent data extraction steps. However, as pointed out in [10], the scope definition often needs to be adjusted or redefined:

> "*[...] it proved very difficult to mark out the boundaries of the process under investigation in the larger DMS [Document Management System]. Therefore, the data scope had to be fine-tuned multiple times, going through the scope adjustment loop after each inspection, until a satisfactory data set was obtained.*" [10]

This excerpt highlights that the definition of context and scope is not a one-off task, but may require repeated manual interventions.

4.3 Data Source Assessment

Event logs are extracted from data sources. As a first step, these sources need to be assessed on their usability and quality. In our literature study, we identified four task categories: reverse engineer a meta model, define sources, audit data, and locate missing data. These categories are supported by 35 papers overall.

Define Sources. Typically, there is a variety of potential data sources available. This means that a human has to understand, analyze, and decide which data sources contain the required data and, therefore, need to be considered for data extraction. That this is a complex manual task is, for example, highlighted in [36]:

> "*The selection of information on business events is an important challenge in event log extraction: while large sets of information may reduce the performance of process mining tools, too little information may affect the quality of the analysis and resulting conclusions.*" [36]

However, currently, there is no clear perspective on how this manual task can be supported or automated.

Create Meta Model. When data is extracted from multiple sources, the process analyst needs to decide on how to properly merge the data. This task can be supported by acquiring (or generating) meta models for each data source. This provides a better understanding of the different data sources [4], and allows the process analyst to already develop an understanding of how the data relates to specific process activities. In our literature study, we identified that authors leverage different meta models for this task, such as entity models [14] and hierarchy models [23].

Assess Data Quality. Prior to data extraction, there should be an assessment of the quality of the available data sources. It may be that data sources turn out to be unsuitable for process mining. An example has been reported in [22]:

> "*Clinical inspection of the data quality of the EHR [Electronic Health Record] revealed data that was considered too unreliable to use for process mining. Issues included unrecorded events and observations, recording on letters and paper records rather than the EHR, mis-diagnosis and inappropriate referrals.*" [22]

This example highlights that some manual effort is not only required for the assessment, but also for fixing the problem. In the case above, the authors decided to manually review large amounts of data, which they then manually transformed into a suitable format.

Locate Missing Data. Suppose the initially available sources do not contain all the necessary data for the process mining project. In that case, there is a need to locate the missing data in other data sources or extract it by other means. In practice, it is often necessary to obtain data managed by departments in the organization that otherwise would not be considered to take part in the process mining initiative. However, suppose the missing data is not available in *any* of the information systems of the organization. In that case, this data may be obtained by other means. In [22], for example, the authors filled the gap of missing data through interviews with domain experts. In any case, the detection and localization of missing data is so far still a manual and time-consuming effort.

4.4 Attribute Selection

Once the data sources have been assessed and selected, the next step is to select the attributes for log extraction. In this category, we identified six task categories, which are supported by 26 papers: define events, define activity names, define event types, detect timestamp characteristics, select the case notion, and map activities to events.

Define Events. One of the key tasks here is to decide which attributes will together represent the events. Attributes can be more generic towards process mining, such as resources, actors, and activities. In some case studies, highly context-specific attributes were selected. An example of such a context-specific attribute is provided in [36]:

> "*Some logistics elements (e.g., the cargo type) can be selected as the instance attributes such that the logisticians could look at the processes from a wider perspective [...] The cargo type of the instance is useful information as it has a big impact on the cargo handling processes*" [36]

The majority of the studies report on their defined events or the relevance of a proper selection of events [17,24,28] without getting much into details on the decisions related to the event definition, nor how it was performed.

Define Activity Names. Defining activity names involves different actions, such as data correlation and abstraction [12]. In [19], the authors use names already available in the data source. While in [36], to perform this task, the authors use available documents. Independently of how this task is performed, it usually still requires manual efforts and more thorough discussions within the case studies in the literature.

Define Event Types. If event types are available (e.g., start, complete, etc.), it might be needed to define which will be necessary for the analysis. In some cases [30], there are so many event types available that there is the need to trim down the data by choosing only event data from a subset of event types. Otherwise, the data would generate spaghetti-like process models.

Detect Timestamp Characteristics. Timestamps are available in data sources in many different formats, sizes, and varieties. The detection of timestamp characteristics before the data extraction can anticipate different problems, such as in [25]:

> "*[...] we do not have any information about the actual timestamps of the start and completion of the service delivered. Consequently, the ordering of events which happen on the same day do not necessarily conform with the order in which events of that day were executed.*" [25]

However, few studies reporting on the event log extraction report on the detection of timestamp characteristics in early stages [5,11]. From our findings, we identified that the studies discuss timestamp characteristics only later on in the process of event log extraction [19].

Select Case Notion. In many settings, there are several options concerning the case notion selection. A typical example could be found in the healthcare domain:

> "*It is [...] possible to consider the data from at least three different 'case' perspectives, i.e. an incident may be considered as a case, each patient may be considered as a case, or each response unit may be considered as a case.*" [5]

As also pointed out by several other authors, this decision is an important one to take and may require several analyses to understand the implications of a potential case identifier [10,17].

Map Data to Events. In this task, process analysts need to identify which tables contain data that relate to events they wish to be part of the event log. What is more, they need to define which mappings provide the intended representation of the events. Often this mapping is guided by activities from a reference model resulting from the previous task category. It can also be directly performed by the process analysts [3].

4.5 Data Source Extraction

Once the attributes have been selected, the actual event logs can be extracted. In this category, we identified six task categories: define collector, define aggregations, define transformations, define scope, filter data, and anonymize data. These categories are supported by 27 papers from our literature study altogether.

Define Collector. In some cases [5,22], the process analyst requests and receives the desired data from a data expert, who represents an involved stakeholder. In other cases, the process analyst has access to the data either via direct access to the data sources [6] or via a dump from the needed data sources [35]. The actual role of the person who collects the data is often left implicit (as e.g. in [8,26]).

Define Aggregations. When data is too fine-grained, there is often the need for the process analyst alongside the domain expert to define which will be the aggregations that need to be manually performed [16,34]. The importance of defining aggregations is reported by [31], which states that the aggregation definition leads the process analyst towards a high-level event log, which reduces the complexity of the resulting process model.

Define Transformations. Different studies report on data transformation supported by domain knowledge [21,29]. In [21], the authors report they transformed data scattered across a set of tables in a database into a unified format. In [29], the authors transformed some data elements using proxy timestamps whenever there were no previously recorded timestamps.

Transformations also refer to the decisions and manipulations performed to transform event data from non process-centric information systems into a process-centric event log. Independently of the transformations performed, it is still necessary to define and report on the data transformations to be consistent and keep documentation of what was performed over the data and why. In this way, it becomes repeatable.

Define Scope. Depending on the available data, it is necessary to define the scope further. For instance, in [17], there is too much data about a process related to many different company branches. In such a case, the authors decided to trim the data by narrowing it down to one branch to "*produce a more focused analysis*" [17].

Filter Data. After the different required definitions to perform the extraction, invariably, data filtering is performed. In many cases, the process analyst can make use of tools to support this task. However, domain knowledge and manual efforts are often driving the data filtering task [10,17,22]. From practice, we know that data filtering is an iterative process, which is often aligned with process discovery and conformance checking.

Anonymize Data. If the process under analysis has data linkable to a specific person, this data should be anonymized (or at least pseudonymized) to preserve privacy [2,26,36]. The detection of the particular attributes that may incur in links to specific persons relies on domain knowledge. Such detection is often performed manually.

4.6 Event Log Assessment

As a final category in the extraction process, the extracted event log is assessed on its quality. In this category, we identified three task categories supported by 13 papers: explore event log, analyze noise, and identify problems.

Explore Event Log. The process analyst needs to explore the extracted event log to reflect back on the previous definitions and to attempt identifying problems that might occur because of remnants of noise, incomplete or imprecise data, or even inadequate definitions. Some studies are explicit about this task, such as in [17]:

> "*The next phase of event log exploration is intended to adjust the data set according to the time and scope of the work, to identify sanitization rules, and, to identify the different analysis dimensions.*" [17]

During this task, the process analyst often leverages process discovery and conformance checking techniques to perform the exploration. However, data and scope adjustments are usually ad-hoc, and performed manually [10].

Analyze Noise. If the definitions, mapping, and filtering previously mentioned are not performed thoroughly, different types of noise in data may still be detected after data is extracted, such as incomplete cases [17], and outliers [6]. And even with thorough definitions, one can still find, for example, infrequent process variants [34]. Although there are several tools to support data filtering, domain knowledge drives the noise analysis and is usually performed by a process analyst alongside a domain expert.

Identify Problems. Once the event log is extracted, it can be assessed on potential problems. Examples of problems identified are: incomplete recordings of events [34], and incomplete cases [17]. This identification task is also driven by domain knowledge. It involves iterations between the process analyst and the domain expert, who will be redefining the scope or the events themselves, and performing data filtering.

5 Discussion

In this section, we discuss the implications and the limitations of our taxonomy for manual tasks in event log extraction.

As for the *implications* of our taxonomy, we argue that it provides an important starting point for reducing the extent of human involvement in event log extraction. We can identify two main use cases in this regard. The first use case concerns *automation*. Without exactly understanding where and how humans are involved, increasing the level of automation is hardly feasible. Our taxonomy, therefore, provides important input on the tasks that can be potentially automated. The second use case concerns *guidance*. Given the nature of some tasks (e.g. defining the scope or defining sources), it is clear that not all human tasks can be automated. However, what can still reduce the time investment and the overall effort is an increased level of guidance for manual tasks. Such guidance may include checklists or other artifacts that provide orientation and help to structure the task execution.

Naturally, our taxonomy is also subject to a number of *limitations*. The first limitation concerns the scope of our literature study. Since we exclusively targeted process mining case studies, we cannot claim to provide a complete picture. We, however, made this choice deliberately since we wanted to learn about the use of process mining in real-world environments. The second limitation concerns the use of a literature study in general. While studying the selected papers, we realized in several places that certain details are missing or not explicitly discussed. Hence, there might exist additional manual tasks that were simply not reported upon or discussed in the analyzed papers. Nonetheless, due to the extent of our literature study, we are confident that our taxonomy provides a rather comprehensive overview.

6 Conclusion

The extraction of event logs comes with substantial human effort. In this paper, we set out to develop a precise understanding of which manual tasks humans perform in the context of event log extraction. We conducted a structured literature review and applied qualitative data coding to systematically derive a taxonomy of manual tasks. Our taxonomy highlights that human work is required in various phases of the event log extraction process ranging from scope and context definition to data source assessment. As such, our taxonomy does not only provide a comprehensive overview but can also serve as input for future automation efforts and for methodological process mining support. In future research, we aim to follow up on the insights provided in this paper and develop techniques that reduce, simplify, or support human work in event log extraction.

Acknowledgments. Part of this research was funded by NWO (Netherlands Organisation for Scientific Research) project number 16672.

References

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
2. van der Aalst, W.M.P., Guo, S., Gorissen, P.: Comparative process mining in education: an approach based on process cubes. In: Ceravolo, P., Accorsi, R., Cudre-Mauroux, P. (eds.) SIMPDA 2013. LNBIP, vol. 203, pp. 110–134. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46436-6_6
3. Andrews, R., van Dun, C.G.J., Wynn, M.T., Kratsch, W., Röglinger, M.K.E., ter Hofstede, A.H.M.: Quality-informed semi-automated event log generation for process mining. Decis. Support Syst. **132**, 113265 (2020)
4. Andrews, R., et al.: Leveraging data quality to better prepare for process mining: an approach illustrated through analysing road trauma pre-hospital retrieval and transport processes in Queensland. Int. J. Environ. Res. Public Health **16**(7), 1138 (2019)
5. Andrews, R., et al.: Pre-hospital retrieval and transport of road trauma patients in Queensland. In: Daniel, F., Sheng, Q.Z., Motahari, H. (eds.) BPM 2018. LNBIP, vol. 342, pp. 199–213. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11641-5_16
6. Benevento, E., Dixit, P.M., Sani, M.F., Aloini, D., van der Aalst, W.M.P.: Evaluating the effectiveness of interactive process discovery in healthcare: a case study. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBIP, vol. 362, pp. 508–519. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_41
7. Calvanese, D., Montali, M., Syamsiyah, A., van der Aalst, W.M.P.: Ontology-driven extraction of event logs from relational databases. In: Reichert, M., Reijers, H.A. (eds.) BPM 2015. LNBIP, vol. 256, pp. 140–153. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42887-1_12
8. Cho, M., Song, M., Yoo, S.: A systematic methodology for outpatient process analysis based on process mining. In: Ouyang, C., Jung, J.-Y. (eds.) AP-BPM 2014. LNBIP, vol. 181, pp. 31–42. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08222-6_3
9. de Murillas, E.G.L., van der Aalst, W.M.P., Reijers, H.A.: Process mining on databases: unearthing historical data from redo logs. In: Motahari-Nezhad, H.R., Recker, J., Weidlich, M. (eds.) BPM 2015. LNCS, vol. 9253, pp. 367–385. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23063-4_25
10. De Weerdt, J., Schupp, A., Vanderloock, A., Baesens, B.: Process mining for the multi-faceted analysis of business processes - a case study in a financial services organization. Comput. Ind. **64**(1), 57–67 (2013)
11. Deeva, G., De Smedt, J., De Koninck, P., De Weerdt, J.: Dropout prediction in MOOCs: a comparison between process and sequence mining. In: Teniente, E., Weidlich, M. (eds.) BPM 2017. LNBIP, vol. 308, pp. 243–255. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74030-0_18
12. Diba, K., Batoulis, K., Weidlich, M., Weske, M.: Extraction, correlation, and abstraction of event data for process mining. WIREs Data Min. Knowl. Discov. **10**(3), e1346 (2020)
13. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P.: PM^2: a process mining project methodology. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds.) CAiSE 2015. LNCS, vol. 9097, pp. 297–313. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19069-3_19
14. Epure, E.V., Hug, C., Deneckére, R., Brinkkemper, S.: What shall I do next? In: Jarke, M., et al. (eds.) CAiSE 2014. LNCS, vol. 8484, pp. 473–487. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07881-6_32
15. Er, M., Arsad, N., Astuti, H.M., Kusumawardani, R.P., Utami, R.A.: Analysis of production planning in a global manufacturing company with process mining. J. Enterp. Inf. Manag. **31**(2), 317–337 (2018)

16. Flath, C.M., Stein, N.: Towards a data science toolbox for industrial analytics applications. Comput. Ind. **94**, 16–25 (2018)
17. dos Santos Garcia, C., Meincheim, A., Garcia Filho, F.C., Santos, E.A.P., Scalabrin, E.E.: Getting insights to improve business processes with agility: a case study using process mining. In: ICSMC, pp. 1336–1343. IEEE (2019)
18. Grisold, T., Mendling, J., Otto, M., vom Brocke, J.: Adoption, use and management of process mining in practice. Bus. Process. Manag. J. **27**(2), 369–387 (2020)
19. Gunnarsson, B.R., vanden Broucke, S.K.L.M., De Weerdt, J.: Predictive process monitoring in operational logistics: a case study in aviation. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBIP, vol. 362, pp. 250–262. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_21
20. Jans, M., Alles, M., Vasarhelyi, M.: The case for process mining in auditing: sources of value added and areas of application. Int. J. Account. Inf. Syst. **14**(1), 1–20 (2013)
21. Jareevongpiboon, W., Janecek, P.: Ontological approach to enhance results of business process mining and analysis. Bus. Process. Manag. J. **19**(3), 459–476 (2013)
22. Johnson, O.A., Ba Dhafari, T., Kurniati, A., Fox, F., Rojas, E.: The clearpath method for care pathway process mining and simulation. In: Daniel, F., Sheng, Q.Z., Motahari, H. (eds.) BPM 2018. LNBIP, vol. 342, pp. 239–250. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11641-5_19
23. Knoll, D., Reinhart, G., Prüglmeier, M.: Enabling value stream mapping for internal logistics using multidimensional process mining. Expert Syst. Appl. **124**, 130–142 (2019)
24. Lemos, A.M., Sabino, C.C., Lima, R.M.F., Oliveira, C.A.L.: Using process mining in software development process management: a case study. In: ICSMC. IEEE (2011)
25. Mans, R.S., Schonenberg, M.H., Song, M., van der Aalst, W.M.P., Bakker, P.J.M.: Application of process mining in healthcare – a case study in a Dutch hospital. In: Fred, A., Filipe, J., Gamboa, H. (eds.) BIOSTEC 2008. CCIS, vol. 25, pp. 425–438. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-92219-3_32
26. Marazza, F., et al.: Comparing process models for patient populations: application in breast cancer care. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBIP, vol. 362, pp. 496–507. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_40
27. Nooijen, E.H.J., van Dongen, B.F., Fahland, D.: Automatic discovery of data-centric and artifact-centric processes. In: La Rosa, M., Soffer, P. (eds.) BPM 2012. LNBIP, vol. 132, pp. 316–327. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36285-9_36
28. Park, M., Song, M., Baek, T.H., Son, S.Y., Ha, S.J., Cho, S.W.: Workload and delay analysis in manufacturing process using process mining. In: Bae, J., Suriadi, S., Wen, L. (eds.) AP-BPM 2015. LNBIP, vol. 219, pp. 138–151. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19509-4_11
29. Partington, A., Wynn, M.T., Suriadi, S., Ouyang, C., Karnon, J.: Process mining for clinical processes: a comparative analysis of four Australian hospitals. ACM TMIS **5**(4), 1–18 (2015)
30. Potavin, J., Jongswat, N., Premchaiswadi, W.: Applying fuzzy-genetic mining in conformance and dependency relations. In: ICT & KE, pp. 228–235. IEEE (2012)
31. Rojas, E., Cifuentes, A., Burattin, A., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Performance analysis of emergency room episodes through process mining. Int. J. Environ. Res. Public **16**(7), 251–263 (2019)
32. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. J. Biomed. Inform. **61**, 224–236 (2016)
33. Saldaña, J.: The Coding Manual for Qualitative Researchers. Sage, Thousand Oaks (2009)

34. Suriadi, S., Wynn, M.T., Ouyang, C., ter Hofstede, A.H.M., van Dijk, N.J.: Understanding process behaviours in a large insurance company in Australia: a case study. In: Salinesi, C., Norrie, M.C., Pastor, Ó. (eds.) CAiSE 2013. LNCS, vol. 7908, pp. 449–464. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38709-8_29
35. Walicki, M., Ferreira, D.R.: Sequence partitioning for process mining with unlabeled event logs. Data Knowl. Eng. **70**(10), 821–841 (2011)
36. Wang, Y., Caron, F., Vanthienen, J., Huang, L., Guo, Y.: Acquiring logistics process intelligence: methodology and an application for a Chinese bulk port. Expert Syst. Appl. **41**(1), 195–209 (2014)