Process Mining Challenges Perceived by Analysts: An Interview Study

Lisa Zimmermann(✉), Francesca Zerbato, and Barbara Weber

Institute of Computer Science, University of St. Gallen, St. Gallen, Switzerland
Lisa.Zimmermann@unisg.ch

Abstract. Process mining analysts need to work with event data to discover (business) processes, interpret results and report meaningful conclusions. Although process mining tools are constantly enhanced and advanced techniques are developed to enrich the functional scope in the field, little is known about the individual needs of analysts and the issues they face while conducting process mining projects. This paper aims to close this gap by uncovering perceived challenges occurring in practice. Based on an interview study with 41 participants, we identify and describe 23 challenges, spanning different project phases and directly affecting the work of process mining analysts. We discuss whether methods and techniques exist that can help to overcome these challenges and where further research is needed to devise new solutions and integrate existing ones better into process mining practice.

Keywords: Process mining · Challenges · Interview study · Process analysis · Work practices

1 Introduction

In the last two decades, the interest of companies to leverage, analyze and monetize their data has massively grown. Therefore, analysts are required to acquire, wrangle and explore data, build a statistical data model and report the obtained results [20]. Especially in the area of process mining [18], where specific algorithms are applied to event data to discover and improve (business) processes, the need for trained analysts familiar with different process mining tools is growing [9]. Although there is an increasing demand to attract analysts to work in process mining, little effort is made to better understand their ways of working [10] and particularly, how they approach the analysis phase [21].

With introducing their research framework, vom Brocke et al. [2] have just recently directed researchers towards the consideration of different levels in analyzing and contributing to the field. Particularly relevant to the context of this paper is the individual level they propose, in which attention is drawn to the "effects of process mining on people's interaction and mode of work".

However, individual entry hurdles and aspects hindering the implementation of process mining projects remain largely unconsidered in the research community so far. In this paper, we try to close this gap by shedding light on these,

A. Augusto et al. (Eds.): BPMDS 2022/EMMSAD 2022, LNBIP 450, pp. 3–17, 2022.
https://doi.org/10.1007/978-3-031-07475-2_1

so far unknown, aspects. In particular, we address the following research question (RQ): **"What are the challenges perceived by individual process analysts during a process mining project?"**. With this research question, we complement the previously published results from [11], who identified challenges in the context of process mining on the organizational level.

Having a comprehensive overview of existing process mining challenges allows us to better understand where there is a need to develop support for the daily work practices of analysts. In this way, the risk of a process mining initiative to fail could be mitigated and analysts would be supported to work efficiently.

To answer the research question, we analyzed data from a semi-structured interview study conducted with 41 process mining analysts from academia and industry. The interviews were conducted in the scope of a broader study during which all participants were asked about the challenges they have already experienced and those they perceive in process mining in general.

As an outcome of this paper, we present a catalog of 23 challenges perceived by individual process mining analysts. Then, we discuss whether approaches exist that can be applied in process mining and reflect on avenues for future research to devise novel solutions or integrate existing ones better into practice to support process analysts.

The rest of the paper is structured as follows. In Sect. 2, we discuss related work. Section 3 presents the research method and in Sect. 4 findings are reported, organized by the project phases they relate to. Then, we discuss the findings and limitations of our work in Sect. 5 and conclude the paper in Sect. 6.

2 Related Work

Our research focuses on challenges reported by individual process mining analysts. Therefore, we build upon existing work on (i) process mining challenges as well as more generally (ii) on challenges reported by data analysts in related fields, such as exploratory data analysis (EDA).

To our knowledge, there is no publication to date that explicitly reports challenges stated by process mining analysts regarding their individual perceptions during an analysis task. However, since general, technical and organizational challenges in process mining have already been reported selectively in other papers, our work is related to them.

One of the first publications in the field explicitly listing challenges is the process mining manifesto [18]. The authors describe 11 rather broad and generic shortcomings across all levels of process mining (e.g. "C2: Dealing with Complex Event Logs with Diverse Characteristics"), motivating researchers to develop and enhance algorithms and methods in different areas. Especially "C11: Improving understandability for Non-Experts" is closely connected to research at the individual level of process mining and remains topical. About six years later [13] reviewed the process mining literature and examined whether the challenges outlined in [18] remain open. Their findings show that despite the wealth of research published in the field over the years, none of the reported challenges have been satisfactorily and exhaustively solved. In [12], the authors report challenges of applying process mining in the healthcare domain, remarking some of the challenges of [18], such as the issues related to concept drift and data quality, but also reflecting on the needs of healthcare organizations, such as the involvement of patients and health stakeholders in process mining projects. Somewhat domain agnostic, [11] published their results of a Delphi Study focusing on opportunities and challenges associated with the use of process mining in enterprises. Based on their survey, the authors identified 32 challenges, three of which are extremely relevant in terms of their support from experts: *"Lack of management support"*, *"Poor data quality"* and *"Complex data preparation"*.

Next to related work on process mining challenges, we considered papers reporting challenges of individual data analysts in general. For example, the authors of [8] conducted semi-structured interviews with 35 data analysts to better understand the enterprise analysts' ecosystem and their challenges. They discuss challenges emerging in 12 different areas during five analysis phases. The authors in [20] follow a comparable method to [8] and extended previous work by the aspect of exploration within the data analysis.

Even though our work focuses on the identification of individual challenges, the boundaries between individual and organizational challenges may not be strictly separable for the individuals interviewed in our study. Therefore, we will compare our results to those of [11] and [8] in Sect. 5 and highlight where our work extends the reported results.

3 Research Method

In this section, we describe the design of our study and outline key aspects of the interview data collection and analysis.

Study Design. To investigate challenges perceived by analysts during a process mining project (cf. research question in Sect. 1), we followed a qualitative approach. Specifically, we designed an interview study as part of a broader observational study during which participants engaged in a realistic process mining task. The task served as an anchor for the interviewees to reflect upon a concrete analysis and challenges emerging in their work practices. To participate in our study, we required participants to: (i) have analyzed at least two real-life event logs in the past two years and (ii) be knowledgeable of at least one of the process mining tools available for the task.

Materials. We designed the process mining task to observe participants as they analyze the road traffic fine management event log [3] guided by a high-level question. The focus of the task was on the mining and analysis phase [5], i.e., we provided participants with a ready-to-use log for their analysis and allowed them to use one or more of the available process mining tools. The interview protocol consisted of semi-structured questions grouped into four parts: (i) activities and artifacts; (ii) goals; (iii) strategies; and (iv) challenges. All the questions were designed to be asked twice: the first time in the context of the process mining task; the second time regarding the participants' general work practices.

Execution. We recruited participants in our professional networks and via snowball sampling [6]. We collected the data in the summer of 2021 via virtual meetings with the participants. A few days before the meeting, we administered a background questionnaire to gather information about the participants' demographics, process mining experience and expertise. On the day of the meeting, we supervised the process mining task and conducted the interviews in a semi-structured way, complementing our interview protocol with questions prompting participants to describe their work experiences within their current organization.

Participants. Overall, 41 people (21 practitioners and 20 academics) from 27 different organizations participated in our study. On average, the participants reported 4.5 years of experience in process mining and most of them indicated experience in related areas, such as data science and business intelligence. 11/20 academics also indicated experience in the process mining industry.

Data Validation and Analysis. Initially, we watched and transcribed the video recordings of the whole session and assessed data quality. On average, each session lasted 83 min, 30.5 min of which were dedicated to the interviews.

For the analysis, we followed a coding approach based on grounded theory [14], coding the whole interview in three rounds, with a focus on the questions asking explicitly about challenges. We considered all statements of the interviewees referring to perceived difficulties or obstacles arising when conducting process mining analyses, similar to the definition provided by [11]. First, we focused on analyzing participants individually and fragmented the text using "in-vivo" and open coding [14] to capture core concepts related to challenges. Then, we used axial coding to refine codes and aggregated them into categories. Finally, we relied on selective coding to focus on the most frequent categories and find relationships among them until we achieved saturation. As a threshold for selecting the final set of challenges, we considered the categories supported by at least 4 participants. Each coding round was conducted by one author and was followed by a check that the other authors conducted independently to ensure consistency. All the authors collaboratively contributed to revising and refining the codes. As a result, we obtained 23 challenges supported by 371 participants' statements. Since the challenges were related to different phases of process mining projects, we organized them along the phases described in [5].

The interested reader may find supplementary material including the interview questions, participants' details and the final coding scheme at https://doi.org/10.5281/zenodo.6422094.

4 Findings

In this section, we present the 23 challenges resulting from our analysis, organized in project phases ranging from "Defining Research Question", "Data Collection" and "Data Pre-Processing" over "Mining & Analysis" to "Stakeholder Evaluation" and "Implementation" [5]. In Fig. 1, we provide an overview of all the challenges. We did not identify any challenge for the "Implementation" phase, in which process improvement measures identified from previous phases are implemented. Four of the 23 identified challenges are considered overarching since the corresponding statements are associated with several project phases.

For each challenge, we report its name and the number of participants mentioning it. We indicate the count of practitioners (P) and academics (A) in the form of #/41 (P = #, A = #). For direct quotes, we note the participant ID (p#).

4.1 Defining Research Question

The first phase of a process mining project is characterized by planning the analysis and defining the research question [5]. 15/41 (P = 8, A = 7) participants reported three different challenges related to this phase.

The first challenge is named **"Question Formulation" (C1)**. 10/41 (P = 6, A = 4) participants stated that having a question is important because otherwise *"you can spend hours and hours doing something that doesn't have an impact"* (p24). However, the identification of a goal for the analysis and related research questions is perceived as difficult: *"it is very often hard to identify the correct question"* (p36). Analysts either struggle with the formulation of the questions, lack specifications from the process owner or report that the prescribed question is too broad or too narrow to enable a meaningful analysis. For example, p24 reported that he *"felt limited"* in one of his analyses because *"in this case [the question] was already specified"*.

The following challenge, **"Access and Use of Process Mining Tools" (C2)** was mentioned by 6/41 (P = 3, A = 3) participants. It includes problems related to the required infrastructure and access to process mining software. Participants reported that organizations *"do not have the tools implemented"* (p22) or that they are *"not sure how these tools can be applied"* (p25). In addition, participants also mentioned that usability *"is always an issue everywhere in our tools"* (p24) and often prevents them from using a certain tool.

Identifying the **"Process Mining Suitability" (C3)** was perceived as challenging by 4/41 (P = 3, A = 1) participants. It was pointed out that for *"a lot of the questions you don't need process mining to answer or you can use process mining as a tool in the toolbox where you have a lot of other tools that you use around"* (p12). It is considered difficult to identify process-mining-specific use cases and convince others about the usefulness of applying process mining. For example, p11 stated that *"it is hard for process mining consultants to convince people that it is something we should have, a new process mining project targeting this and that"*. As a result of these concerns, it is reported that process mining projects are not pursued or stopped in an early stage. Participant p34 stated that he analyzed event logs *"much less than I wanted to do and than it would be useful"* because stakeholders *"are not ready to start process mining studies"*.

Phase	ID	Challenge	Description The code combines statements related to perceived challenges ...	No.
Defining Research Question	C1	Question Formulation	...due to either **performing the project based on an external, prescribed goal**, due to working with a **loosely defined goal** or a **lack thereof** and covers **perceived challenges during the formulation of questions**.	10
	C2	Access and Use of Process Mining Tools	...due to the **unavailability of a process mining tool**, due to a **missing governance to structure the tool usage in the organization** or due to a **general low usability** which results in avoiding the application of a process mining tool.	6
	C3	Process Mining Suitability	...associated with the **application and selection of process mining as a suitable method** for a task. Covers the application of process mining on the individual level as well as its general application and governance in an organization.	4
Data Collection	C4	Data Extraction	...during the **retrieval, i.e., identification and extraction, of event data** from any kind of source.	11
	C5	Data Availability	...of **having enough data available for the analysis project**. This also covers the **problem of process steps being conducted outside of the system** from which data is available.	9
	C6	Data Access	...due to **missing permissions (legal, regulatory, organizational) to access the data.**	6
	C7	Source System & Data Structure Knowledge	...during the data collection **due to missing knowledge** about the **data structure** and the **source system** from which data is extracted **(e.g., relational databases).**	4
Pre-Processing	C8	Data Transformation	...during the **preparation of a process mining conformant event log** based on raw event data. It specifically covers **adding attributes or activities** to the event log and **finding the correct level of aggregation** depending on the case key.	17
	C9	Data Quality	...due to **low data quality** either in the **final event log** or already in the **raw data.**	15
	C10	Data Validation	...during **checking data accuracy** or the **suitability for process mining.**	5
Mining & Analysis	C11	Tool Knowledge	...associated with the **use of a specific process mining tool**. This covers **a lack of familiarity with the tool** or **not having the tool used for a long time.**	18
	C12	Event Log & Data Model Understanding	...associated with **the comprehension of the event log and the data model**. Also, covers **missing insights about the data structure** or problems caused by the **absence of elements they are used to** in a data model.	15
	C13	Process Mining Techniques	...related to **missing or insufficient supported functionalities of specific process mining tools** or in general including **missing support for combining certain functions and methodologies.**	14
	C14	Access to Additional Information	...due to a **lack of information or missing access** of the process analysts to this information.	10
	C15	Process Visualization	...due to a **misleading visualization of the event log in form of a directly-follows graph** (DFG) which can be found e.g., in the data map (Disco) or in the variant explorer (Celonis).	8
	C16	Analysis Experience	...due to a **subjective low level of prior experience as a process analyst.**	7
	C17	Analysis Focus	... of **maintaining the big picture** of the analysis **despite the available details.**	6
Evaluation	C18	Conclusions & Question Answering	... of **answering the research questions** and **drawing conclusions** based on the process discovery.	8
	C19	Recommendations & Next Steps	...of **formulating a concrete recommendation** or **deriving concrete next steps** to foster the improvement of the process.	4
All Phases	C20	Process Domain Understanding	...due to **a lack of/partial domain knowledge** or **unfamiliarity with domain-specific terminology** used in the documentation or activity and attribute descriptions.	22
	C21	Collaboration with Stakeholders	...**during the communication** due to **mismatching expectations**, **lack of understanding**, **different process perspectives** and **different backgrounds** of the process analysts and the business and IT stakeholders.	15
	C22	Business Process Complexity	...due to the **inherent complexity of business processes** and the dependencies among them.	10
	C23	Enablement / Training	...due to a **lack of available training opportunities** or **challenges while conceptualizing trainings.**	9

Fig. 1. Overview of all 23 identified challenges organized by process mining project phases [5]. For each challenge, we report a numeric ID identifying the challenge, its name, its description and the number of participants (No.) reporting the challenge.

4.2 Data Collection

In the “Data Collection” phase, the main goal is to understand and extract the data required for the process mining project [5]. In total, 23/41 (P = 14, A = 9) participants reported challenges related to this phase.

During **“Data Extraction” (C4)**, which was reported to be challenging by 11/41 (P = 6, A = 5) participants, event data is extracted from source systems in which the process is executed. Participants stated that data extraction is time-consuming and that there are issues due to strong dependencies on third parties, such as IT departments. Analysts need to invest into *“explaining to the partners what [they] really need from them and what [they] really need from the data to be able to start”* (p15), which makes *“getting the total data sometimes the biggest challenge”* (p31). Additionally, participants emphasized challenges while identifying the right data (*“how do you find the data that you need in these huge databases?”* [p16]) and while consolidating the data from different sources.

When the data itself can be extracted, 9/41 (P = 6, A = 3) interviewees reported that **“Data Availability” (C5)** is a challenge in their projects. It includes the problem of having sufficiently comprehensive data to enable meaningful analyses. For example, p35 described: *“we had less or not enough event data to check because the process has been changed. And we didn’t have that many cases and it wasn’t enough to say if the process is working or not”*. Participants also reported that process steps are executed outside of the information systems. For example, when *“at the end you have maybe a letter that goes out and, in that case, to have a digital footprint of the whole process, is very difficult”* (p26). Ultimately, process mining *“is limited to what was recorded by the system”* (p8).

“Data Access” (C6) is required to determine what data should be collected but also to be able to understand the data. 6/41 (P = 5, A = 1) of the interviewed analysts described that they experienced challenges in their projects due to missing access to the raw data. They pointed out that legal restrictions or company internal data security/privacy policies limit access to data, making it challenging to get *“permission to get access to the data”* (p11). And indeed, the GDPR[1] and even stricter local regulations of personal data can limit process mining use cases [7].

Furthermore, specifically for practitioners, it is important to understand the functionality of the source system and the underlying database structure. For **“Source System & Data Structure Knowledge” (C7)** 4/41, (P = 4, A = 0), participants reported difficulties in understanding database models and were lacking *“system knowledge if it’s not SAP and the standard process”* (p9). They experienced these difficulties because analysts are often *“not an expert on the system and the settings there”* (p9).

4.3 Data Pre-Processing

The “Data Pre-Processing” phase focuses on the creation of the event log. For this purpose, data quality is assessed and the process events are created [5], which

1 https://www.gdpr.eu.

contain at least the case ID, event description, and event timestamp. Further information can be added as required [18]. 24/41 (P = 12, A = 12), i.e., more than half of the interviewees, reported challenges related to this phase.

The predominant category of challenges in this phase is **"Data Transformation" (C8)**, supported by 17/41 (P = 9, A = 8) participants. It was reported that data transformation is a *"very big part of each process mining initiative"* (p1), although *"it's not straightforward to put them in a process or in an event data or a XES format"* (p18; referring to data retrieved from an ERP System). Besides these general issues, there are more specific, subsidiary challenges related to adding event or attribute information to the event log/data model and finding the right aggregation level for the events. Participants reported that it feels like a *"philosophical question of which activities to add"* (p39) and that not having defined appropriate events is problematic because *"if you don't have those activities, it can be quite hard to yeah, to refine your analysis"* (p14) later on.

The challenge of **"Data Quality" (C9)** was raised by 15/41 (P = 6, A = 9) participants from our study. Interviewees reported that data pre-processing is *"quite challenging because the industrial data, the sensor data... sometimes the quality is very poor"*, and that *"data quality and event log quality are the most important challenge for the further analysis"* (p28). Data quality issues can be manifold, but *"noise in the data"* (p16), problems in the format of the date fields and missing timestamps were particularly prominent in our interviews.

Closely related to poor data quality is also the assessment of data quality, referred to as **"Data Validation" (C10)**, to ensure that data are correct, complete and representative of the process to be analyzed. 5/41 (P = 3, A = 2) participants reported that it is time-consuming *"to check if the data is ok and accurate"* (p35) and that validation is an important step not only in process mining but also in many data-based analysis methods because *"you will always get an answer but the data will not tell you that the answer is invalid"* (p12).

4.4 Mining and Analysis

In the "Mining & Analysis" phase, analysts apply process mining techniques to explore event logs [5]. 38/41 (P = 19, A = 19) of the interviewed participants reported challenges during this phase of the process mining project.

About half of the participants, 18/41 (P = 8, A = 10) reported difficulties connected to their **"Tool Knowledge" (C11)**. They stated that they *"didn't feel very comfortable with the tool"* (p8) or that they *"had to apply filters and [were] not sure where to find it"* (p10). Participant p15 summarized that the tools *"work all in a very similar way and they basically use the same algorithms. But, remembering where those patterns are and how to click in the right sequence, it's not always easy"*. This leads to the assumption that tools require a certain level of expertise and training to perform an efficient and meaningful analysis. However, when looking at the background questionnaires (cf. Sect. 3) only two of the 18 participants reporting this challenge ranked themselves as 'slightly familiar' with process mining tools, while all the others were moderately, very or extremely familiar with process mining software.

Another important aspect during the analysis phase is the **"Event Log & Data Model Understanding" (C12)**, for which 15/41 (P = 8, A = 7) participants described challenges. They include difficulties in understanding attributes of the event log (*"the main challenge was to understand the attributes of this event log because many of them had a similar name"* [p41]) as well as *"understanding the data model"*, which *"is probably the biggest challenge"* (p14). Indeed, different process mining tools support different kinds of data structures. While some participants are used to work with a data model based on several tables, other tools are designed to load only one table representing the event log.

Challenges related to the available analysis techniques and their combination are covered in the category **"Process Mining Techniques" (C13)** which was supported by 14/41 (P = 8, A = 6) of the interviewees. Although techniques continue to evolve and new features are constantly added into tools [9], challenges related to the technical maturity of root cause analysis based on process data, the combination of process mining and robotic process automation (RPA), the *"inability of any algorithm to split labels based on context"* (p39) and shortcomings in the configuration of the dotted chart in ProM are reported to still exist. Of this list, dissatisfaction with results based on the integrated root causes analysis was most frequently mentioned. Participants noted that *"conformance analysis is extremely complex and resource consuming"* (p3) and pose the question: *"How can we bring in and integrate process mining, maybe with other tools or improve its own methodologies in theory to help finding the root causes?"* (p1).

"Access to Additional Information" (C14) was remarked as challenging by 10/41 (P = 7, A = 3) participants. They stated that *"it's often the case that we need some additional knowledge to really get into an event log"* (p8), but at the same time do not have access to *"good documentation"* (p34). Access to stakeholders as a source of information is required, but also perceived as challenging. Participant p17 reported: *"These are all assumptions that we make, so we need to have like a confirmation from the business that's actually a right, attribute that you need to have a look into"*. If access to stakeholders is not available during this phase, analysts are limited to *"check what was obvious"* (p7).

The **"Process Visualization" (C15)** covers challenges reported by 8/41 (P = 2, A = 6) participants. Out of the interviewees, especially academics referred to the directly-follows graph (DFG) representation as an unsolved challenge, but also practitioners reported that they *"don't trust the maps [...] because of this slider, we see paths, which already means you don't see variants. You see paths, the most frequent paths. And that's not the same thing, I mean, that's not something that really happens"* and stated that *"you cannot already tell from seeing the map, ok, that's how it behaves because there's some paths missing"* (p37).

A completely different aspect, namely the prior **"Analysis Experience" (C16)** is reported to be a challenge by 7/41 (P = 3, A = 4) participants. Without further elaboration, participants stated that they *"don't have that much practice, so it [the analysis] was challenging in general"* (p10) and that the *"process mining is very, very easy to learn and I think time consuming to really* *master it"* (p13). Apparently, learning the *"way of thinking as a process mining analysts"* (p22) requires time and experience.

The last challenge we identified for the "Mining & Analysis" phase is **"Analysis Focus" (C17)**. Statements from 6/41 participants (P = 6, A = 0), all practitioners, are related to this challenge. It is reported that it is difficult to stop the analysis at a certain point because there is always the risk of *"diving deep into one specific [aspect] but actually loosing the big picture"* (p24). Analysts can *"lose themselves too quickly into the details"* (p25) partially because it is *"hard to not deviate from your original aim"* (p26).

4.5 Stakeholder Evaluation

During the "Stakeholder Evaluation" phase, the process mining analyst presents and discusses insights from the analysis and answers the research question. Meanwhile, tangible conclusions and next steps are suggested for improving the process regarding identified shortcomings [5]. The challenges encountered in this phase are supported by 11/41 (P = 7, A = 4) participants of our study.

One important aspect of the stakeholder evaluation is to find causality to answer the research question. However 8/41 (P = 4, A = 4) participants reported that they struggle with **"Conclusions & Question Answering" (C18)**. It was stated that there is a danger of jumping to wrong conclusions (*"you have the data loaded and the data are correct, it's fairly easy to do an analysis, so a major pitfall is that you jump to incorrect conclusions"* [p11]), and the majority of the interviewees agreed that it is generally difficult to *"come to, let's say, hard conclusions or to find let's say, OK, this is really what we should change now"* (p20). One of the interviewed analysts additionally links the problem to C13 ("Process Mining Techniques") and stated that *"process mining cannot answer all the questions you have. You need to combine it with all the approaches to identify all the features that affect your process in order to answer the whys, why something is not working"* (p3).

After analysts have derived conclusions, stakeholders are often interested in next steps. 4/41 (P = 3, A = 1) participants stated challenges connected to **"Recommendations & Next Steps" (C19)**. For example, participant p4 stated: *"I think it's challenging to answer this question with recommendation of what to do afterwards"* and indeed, process mining shows *"where your issues are, but it's not helping you to solve them"* (p17). Although it could be argued that this aspect is outside of the defined scope of process discovery [18], four of the interviewed participants considered this a challenge and expect *"recommendations or proposals to change the process"* (p25) to come out of a process mining project.

4.6 Challenges Ranging Across All Phases

In addition to the challenges that could be mapped to one of the process mining phases [5], four additional challenges emerged, spanning across (parts of) the project. They are supported by 34/41 (P = 19, A = 15) participants.

22/41 (P = 13, A = 9) analysts reported challenges around **"Process Domain Understanding" (C20)**. Both, acquiring domain knowledge as well as performing various activities throughout the project in the absence of domain knowledge are perceived as challenging. This seems to not be surprising considering that *"without domain knowledge, you won't achieve much or nothing at all"* (p38). Even though associated with all phases, participants explicitly pointed out the shortcomings during the analysis of the process: *"Business knowledge is something that one really needs to have when analyzing the process"* (p17) because *"if you have more of the domain knowledge, you would know like which path to check first"* (p14). Thus, domain knowledge supports the analysis process and leads analysts to more relevant, business-related findings.

Another group of statements is tied to the **"Collaboration with Stakeholders" (C21)**. 15/41 (P = 9, A = 6) participants reported problems due to stakeholder expectations regarding the process mining technique, different backgrounds of the parties involved (e.g. IT versus business), little or different levels of understanding of process mining, and reluctance of stakeholders to work together with the process mining analyst. Regarding the latter aspect, participants speculated that stakeholders *"don't want that somebody external of their business puts his eyes on it"* (p34) and stated that they are *"hitting a wall"* (p34) within their organizations. Compared to the challenge of not having access to the stakeholder during the analysis (C14) this challenge rather connects to having the stakeholders, but that *"communicating effectively what process mining can and should do to people from businesses is maybe the biggest challenge"* (p36).

The **"Business Process Complexity" (C22)** covers challenges related to the interplay between departments, complex IT landscapes and the resulting intricacy of organizations and processes executed within them. 10/41 (P = 5, A = 5) participants reported that *"in process mining you have the problem that you have this complex behavior"* and that *"real processes, with several process objects are more complex than, let's say, the standard process coming from the vendors"* (p33). Demonstrated use cases of process mining often fail to address business reality. Some of our participants mentioned this challenges in the context of designing the business questions or in the context of the analysis of 'spaghetti-like' processes with many events and endpoints.

Detached from the actual process analysis itself, 9/41 (P = 6, A = 3) participants addressed problems during or due to the **"Enablement/Training" (C23)**. Analysts, who have been involved in training colleagues or academics who teach process mining, reported that it is challenging to provide the correct level of knowledge and to plan the training in a way that it is appropriate for the target group, e.g., for *"students, which are not computer scientists at all"* (p36) or *"for beginners"* where the problem is *"that there are lots of very different fields that you have to have some kind of basic understanding like how process models look like and, um uh, well, basic algorithms and not to misinterpret the process models that you get at the beginning like directly-follows graphs"* (p15). Additionally, one participant reported that it is challenging to find the courses and pointed out that available trainings and lectures should be better promoted.

Concluding, based on interviews with 41 participants, we identified 23 challenges, each one supported by at least four interviewees. We can observe that especially C3, C6, C7, C17 and C19 are mainly perceived by practitioners, whereas all other challenges are reported across the different sectors the participants were working in at the time the interviews were conducted. Based on the explanations given by the interviewees, we related 19 challenges to process mining project phases [5] and identified four cross-cutting challenges occurring in and affecting all project phases. We couldn't identify any significant correlation between single challenges and the self-rated expertise or experience of the participants reporting the challenge.

5 Discussion

In this section, we will review the identified challenges, link them to related work, and discuss whether technical or methodological approaches exist that can help process analysts to overcome them.

Among our findings, one first major cluster of challenges concerns data-related ones. All of these challenges fall into the data collection and pre-processing phase (C4-C6 and C8-C10) and include availability, access, quality, validation, extraction, and transformation. Such challenges have already been discovered in different fields, for example, considering data as a prerequisite for data science projects. The authors of [8] and [20] identified comparable challenges from their interview studies around the acquisition of data, working with different amounts or forms of data and dealing with concrete data quality issues. However, comparable challenges have also been discovered in process mining, where "data" mainly refers to the special format of event logs. The authors in [11] identified specific data-related challenges, referred to as C.4, C.7–C.9, C.12 and C.14, which they mainly captured in the area of governance on the organizational level. Based on our study, we can confirm that these challenges are not only perceived at the organizational level but also affect the work of individuals. Especially C5 and C10 extend challenges reported in [11] by bringing in an individual perspective.

Several approaches have already been proposed to tackle specific aspects of these data-related challenges. For example, Suriadi et al. [15] proposed a structured approach to deal with quality issues in event logs, while Diba et al. [4] provided an overview of existing methods to extract event log data (C4) and create meaningful abstractions (C8). However, for most existing approaches, major shortcomings are reported [4] and their integration into many of the process mining tools is limited. As a result, their broad application in practice is missing [1]. Our work suggests that further research in this direction is required and that existing techniques will need to be better integrated into commonly used data pre-processing tools to support analysts in overcoming data-related challenges and lower the entry hurdle for creating event logs of good quality.

Another group of challenges that we observed is connected to the adoption of process mining by the analysts themselves but also by project stakeholders with whom the analysts interact (C2–C3, C11, C16, C21 and C23). These challenges span across all process mining project phases apart from the data collection and pre-processing phase and include limited access to process mining tools, non-suitability of process mining for the analysis or problems in the collaboration with stakeholders, who might not be willing to share information or do not trust the results of the process discovery. While some of our challenges overlap with organizational challenges reported in [11] (in particular C.23–C.32 on the cultural level), others seem to be more tight to the individual level (especially C2, C11 and C23 from our work) and go beyond what was reported by [11].

To overcome difficulties regarding the adoption of new technologies, different approaches and methods have been proposed over the last decades in various fields, such as manufacturing or information systems [17,19]. However, in the area of process mining only first attempts exist to better understand the transition of stakeholders from old practices to the usage of process mining [7,16], and aspects such as training and enablement have received little attention. Hence, we conclude that access to comprehensive trainings and a deeper understanding of the required skill set of process mining analysts is still missing.

Further challenges emerged, revolving around the individual understanding of the analysts (C7, C12, and C17–C20), such as understanding the process domain and the event log, or having issues with answering the research question and deriving improvement recommendations. These challenges occur mainly in late project phases, i.e., the mining & analysis and the stakeholder evaluation phases. We observed that these challenges are especially related to the individual level since there is limited support for them in the findings of Martin et al. [11]. While the authors in [11] identify challenges like the insufficient domain expertise (C.20) or incomprehensible outcomes (C.16) and insufficient prescriptive capabilities of process mining tools (C.17) on the organizational level, the angle from which these challenges are covered differs from our findings and cannot be connected to the understanding of individuals. Based on our knowledge, there exists little support to help analysts guide their analysis based on the research question and reduce the risk of losing the analysis focus and there is limited guidance for understanding event logs and data model structures [21]. Besides, the authors in [5] even observed that the thoroughness of reports for the stakeholder evaluation phase is decreasing in published case studies in the area of process mining. Thus, we think that research on the factors determining the understanding of analysts needs to be enhanced to enable the implementation of targeted support.

To summarize, we discovered that although approaches exist to tackle some of the discussed challenges, their application in process mining practice is limited. We observed a mismatch between the solutions provided by existing approaches, which are oftentimes targeting technical problems, and the challenges faced by individual process mining analysts in practice. We encourage future research to take the individual perspective into account by proposing new methodologies and evaluating existing ones based on their effectiveness on the work of individuals.

Limitations. Since the data supporting our findings was gathered during semi-structured interviews, our work comes with some limitations typical of interview studies. First, we only present and discuss challenges that were directly derived from the data, meaning that they were explicitly stated by our participants. Therefore, there is a possibility that our findings are not complete, since our participants might have not been able to recall and describe all the challenges they face in their work practices. Additionally, the perception of what constitutes a challenge may be subjective and can vary across participants. Nevertheless, in order to obtain valid and reproducible results, we selected a sample of more than 40 interviewees and only considered challenges that were reported by at least 4 individuals. Moreover, the interviews directly followed an analysis task and thus, reported challenges may be biased by the recent experience of conducting this specific task. Still, the study was designed to be representative of typical process analysis tasks and the interviewees were also asked to consider general challenges and difficulties. Interviews anchored to other types of tasks or triangulated with behavioral data can help to complement and generalize our findings.

6 Conclusion

In this paper, we focused on process mining challenges from an individual perspective. Based on the analysis of 41 interviews conducted with practitioners and academics working in the field, we identified 23 challenges. All of these challenges hinder the work of individual analysts, preventing them from working efficiently and effectively and, in the worst case, discouraging them from conducting further projects in their organizations. While focusing on the individual perspective, we identified that the discovered challenges also affect the organizational, group and technical levels. Although approaches exist to address these challenges, most of them have not been applied to the field of process mining yet and the assessment of their ability to support individual analysts during a process mining project remains open. In the future, we plan to continue our work to better understand the factors that cause process mining challenges and individual support needs. Besides, we aim to investigate potential approaches and solutions that experienced analysts implement to overcome the challenges, paving the way for easier access and improved use of process mining, especially for novices.

Acknowledgment. We thank participants for taking time to participate in the study and for sharing their experience. *Funding.* This work is part of the ProMiSE project, funded by the Swiss National Science Foundation under Grant No.: 200021_197032.

References

1. 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)
2. vom Brocke, J., Jans, M., Mendling, J., Reijers, H.A.: A five-level framework for research on process mining. Bus. Inf. Syst. Eng. **63**(5), 483–490 (2021). https://doi.org/10.1007/s12599-021-00718-8

3. De Leoni, M., Mannhardt, F.: Road Traffic Fine Management Process. Eindhoven University of Technology, Dataset (2015)
4. Diba, K., Batoulis, K., Weidlich, M., Weske, M.: Extraction, correlation, and abstraction of event data for process mining. WIREs Data Mining Knowl. Discov. **10**(3), e1346 (2020)
5. Emamjome, F., Andrews, R., ter Hofstede, A.H.M.: A case study lens on process mining in practice. In: Panetto, H., Debruyne, C., Hepp, M., Lewis, D., Ardagna, C.A., Meersman, R. (eds.) OTM 2019. LNCS, vol. 11877, pp. 127–145. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33246-4_8
6. Goodman, L.A.: Snowball sampling. Ann. math. stat. 148–170 (1961)
7. Grisold, T., Mendling, J., Otto, M., vom Brocke, J.: Adoption, use and management of process mining in practice. Bus. Process Manag. J. (2020)
8. Kandel, S., Paepcke, A., Hellerstein, J.M., Heer, J.: Enterprise data analysis and visualization: an interview study. IEEE Trans. Vis. Comput. Graph. **18**(12), 2917–2926 (2012)
9. Kerremans, M., Searle, S., Srivastava, T., Iijima, K.: Market Guide For Process Mining. Gartner Inc. (2020)
10. Klinkmüller, C., Müller, R., Weber, I.: Mining process mining practices: an exploratory characterization of information needs in process analytics. In: Hildebrandt, T., van Dongen, B.F., Röglinger, M., Mendling, J. (eds.) BPM 2019. LNCS, vol. 11675, pp. 322–337. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26619-6_21
11. Martin, N., et al.: Opportunities and challenges for process mining in organisations: results of a Delphi study. Bus. Inf. Syst. Eng. **63**, 1–7 (2021)
12. Munoz-Gama, J., et al.: Process mining for healthcare: characteristics and challenges. J. Biomed. Inform. **127**, 103994 (2022)
13. R'Bigui, H., Cho, C.: The state-of-the-art of business process mining challenges. Int. J. Bus. Process. Integr. Manag. **8**(4), 285–303 (2017)
14. Saldaña, J.: The Coding Manual For Qualitative Researchers. Sage, Thousand Oaks (2015)
15. Suriadi, S., Andrews, R., ter Hofstede, A.H., Wynn, M.T.: Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Inf. Syst. **64**, 132–150 (2017)
16. Syed, R., Leemans, S.J.J., Eden, R., Buijs, J.A.C.M.: Process mining adoption. In: Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds.) BPM 2020. LNBIP, vol. 392, pp. 229–245. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58638-6_14
17. Taherdoost, H.: A review of technology acceptance and adoption models and theories. Proc. Manuf. **22**, 960–967 (2018)
18. van der Aalst, W., et al.: Process mining manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011. LNBIP, vol. 99, pp. 169–194. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28108-2_19
19. Venkatesh, V., Thong, J.Y., Xu, X.: Unified theory of acceptance and use of technology: a synthesis and the road ahead. J. Assoc. Inf. Syst. **17**(5), 328–376 (2016)
20. Wongsuphasawat, K., Liu, Y., Heer, J.: Goals, Process, and Challenges of Exploratory Data Analysis: An Interview Study. arXiv:1911.00568 (2019)
21. Zerbato, F., Soffer, P., Weber, B.: Initial insights into exploratory process mining practices. In: Polyvyanyy, A., Wynn, M.T., Van Looy, A., Reichert, M. (eds.) BPM 2021. LNBIP, vol. 427, pp. 145–161. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85440-9_9