

What Questions Can I Ask? A Taxonomy and Question catalog for Process Mining Analysis Questions

Lisa Zimmermann¹, Francesca Zerbato², Victoria Gentile³, Manuel Resinas⁴, and Barbara Weber¹

¹University of St. Gallen, Switzerland

²Eindhoven University of Technology, The Netherlands

³École Polytechnique Fédérale de Lausanne, Switzerland

⁴SCORE Lab, I3US, Universidad de Sevilla, Seville, Spain

Analysts play a critical role in examining and interpreting data in process mining projects. However, specific goals for such projects are not always clearly stated nor translated into concrete analysis questions, hindering analysts to derive analysis results effectively. To address this issue, we in this work follow an established method for taxonomy design in information system research and propose a categorization schema for the key components of process mining analysis questions. We collect a diverse set of such questions by conducting a review of analysis reports and a survey with practitioners and by gathering input from a tool vendor. The resulting taxonomy classifies analysis questions along six dimensions: Use Case, Perspective, Primary Goal, Cognitive Step, Context, and Data Level, each with distinct categories and sub-categories. Additionally, the categorized set of the 405 collected analysis questions forms a question catalog. The taxonomy and question catalog are evaluated through interviews, illustrative scenarios, a case study with two organizations that applied them for question design, and a survey with students. The evaluations confirm the applicability of the taxonomy for question categorization and question design. Both contributions empower users to effectively describe, compare, and assess analysis questions and design new questions.

Keywords: *Process Mining; Taxonomy; Question Catalog; Question Design; Project Planning*

Acknowledgements: We sincerely thank all participants for their time and contributions to the questionnaires and evaluations conducted in this study. This work was carried out as part of the ProMiSE project, funded by the Swiss National Science Foundation under Grant No. 200021_197032. The work of M. Resinas was partially funded by projects PID2021-126227NB-C21/ AEI/10.13039/501100011033/ FEDER, UE and TED2021-131023B-C22/ AEI/10.13039/501100011033/ Union Europea NextGenerationEU/PRTR.

1 Introduction

Business processes are at the core of organizational operations and the interest of companies in understanding and improving them as part of their business process management strategy is constantly growing. In their book, Dumas et al. (2018) point out that factors such as globalization, standardization, and innovation has increased the interest of organizations to reflect upon their processes, understand their executions, and optimize them. By providing analysis techniques to fulfill this demand, *process mining* enables organizations to discover, monitor, and improve their real-world processes, helping companies gain fact-based insight into their operations (van der Aalst 2022).

As a result, the field of process mining has expanded significantly, broadening its scope to techniques that go beyond automated process discovery (Augusto et al. 2019), and witnessing a growth in the number of open-source and commercial tools available to organizations. In addition, methodologies such as the Process Diagnostics Method (PDM) (Bozkaya et al. 2009), the PM² Methodology (van Eck et al. 2015), and the L* Life-Cycle Model (Van Der Aalst 2011), or guidelines such as Process Mining for Six Sigma (PMSS) (Graafmans et al. 2021) have been proposed. These methodologies structure projects and guide analysts through various tasks. However, despite these advancements, projects often remain fragmented and organizations face challenges in planning and executing them (Martin et al. 2021).

One critical challenge arises in the early stages of a project when business requirements must be translated into actionable analysis *questions* (Zimmermann et al. 2024). Formulating such questions is an essential step in defining stakeholder needs, scoping data extraction, and guiding the choice of techniques and tools for analysis (van Eck et al. 2015; Gurgen Erdogan and Tarhan 2018). For example, a common question concerning process performance is: “Where are the bottlenecks in the process?” (Mans et al. 2013). This question presumes prior knowledge of the existence of bottlenecks and the availability of specific data elements, such as event timestamps and baseline throughput times. In contrast, broader questions like “What are aspects of the process that impede efficiency?” also point to performance, but encourage exploration of various inefficiencies beyond bottlenecks, such as delays or rework. This example illustrates how different questions, while presumably reflecting similar goals (i.e., investigating process performance), can lead to different analyses and answers and underscores the importance of aligning questions with project objectives.

However, formulating and refining such analysis questions is not straightforward, as analysts report that they are struggling to find “good” analysis questions that are narrow enough to identify relevant insights but, at the same time, broad enough to ensure that such insights are rich (Zimmermann et al. 2024). Analysts frequently work with questions that are “not really thought through” (Kandogan et al. 2014) and need to interact with stakeholders to derive questions for their analysis. During question refinement, the direction of the project is defined, aligning the questions with stakeholder goals and organizational needs. At the same time, this phase is manual and labor-intensive, resembling requirements gathering in software engineering, where stakeholder collaboration and human judgment are vital. Therefore, automation may not be the primary goal; instead, the focus should be on providing tools and resources that support analysts in eliciting and refining questions (Hrach et al. 2023; Wang et al. 2021).

Additional difficulties arise given the varying levels of process mining maturity in organizations

(Stein Dani et al. 2023). Many stakeholders and novice analysts are unfamiliar with the full capabilities of process mining, making it difficult to define feasible and impactful questions that align with organizational goals and correctly interpret results (Badakhshan et al. 2022). This lack of understanding can lead to misaligned expectations and requires analysts to invest additional effort into bridging the gap between business needs and process mining capabilities (Mamudu et al. 2022). Despite existing methodologies provide overarching project structures, they rarely address the nuances of question design. Analysts are left without systematic resources to navigate the interplay between project goals, stakeholder requirements, and data/IT-system constraints. This lack of support risks leading to questions that underutilize the potential of process mining techniques or fail to produce valuable insights.

We argue that to address these challenges, there is a need for resources that help analysts systematically categorize, design, and refine questions. Such resources should not only provide insights into the characteristics of questions but also offer practical examples to inspire and guide question formulation. By supporting analysts in the planning phase of projects, these resources can help ensure that process mining projects are aligned with business objectives and capable of delivering actionable results.

Therefore, in this work, we examine process mining analysis questions in detail, revealing dimensions along which such questions can be comprehensively described. We assess whether such a description and a collection of questions can inspire question formulation. In particular, our contribution is twofold: We present a **taxonomy of process mining questions** and a **question catalog**. The taxonomy allows categorizing analysis questions based on their inherent characteristics, aiming to foster reflection on their purpose and alignment with project goals.

As a design science artifact, the taxonomy was developed following the six phases of Peffers et al. (2007) Design Science Research (DSR) Methodology, ensuring a rigorous development and evaluation process. Specifically, we adopt the taxonomy development method of Nickerson et al. (2013), which has become one of the most widely accepted approaches for taxonomy construction in information systems (IS) research. Oberländer et al. (2019) highlight that while taxonomy development in IS has often followed ad-hoc approaches, Nickerson et al. (2013) provide a systematic and replicable framework, integrating both inductive and deductive elements. Additionally, we follow the evaluation guidelines of Kundisch et al. (2021) to ensure a systematic assessment of the validity and utility of the taxonomy. The taxonomy is refined through multiple design cycles, incorporating a broad range of empirically collected process mining questions and insights from the literature. Complementing the taxonomy, we provide a question catalog containing 405 questions categorized along the taxonomy's dimensions. These questions were collected from analysis reports, case studies, an online survey with process mining practitioners, and a process mining vendor. The question catalog provides information about the distribution of questions across the different sources and taxonomy categories. In addition, it serves as an open-access resource for researchers and practitioners to access questions based on specific characteristics and use them as templates for their process mining projects.

After its design, we evaluated both the taxonomy and the question catalog in several iterations. First, the taxonomy was reviewed by process mining experts, whose feedback informed subsequent refinements. Second, we used illustrative scenarios to demonstrate the applicability of the taxonomy for question classification. Third, its usefulness was tested in real-world scenarios with representatives from two

organizations in Germany and Spain. Fourth, we evaluated whether non-expert process mining analysts (i.e., students) were able to derive meaningful analysis questions with the help of the taxonomy and question catalog and compared the results to a group of students who designed questions without the input of the taxonomy and question catalog. The results of the evaluations confirmed the usefulness of our two artifacts for understanding and describing questions in process mining projects. Moreover, our findings underscore the effectiveness of leveraging existing questions as a starting point for question design and refinement, especially in settings where there is low process mining maturity, analysts are not experienced, or new projects are started. Overall, both artifacts contribute to the research community by providing an overview of the *state space* of process mining analysis questions and a structured foundation for categorizing, formulating, and refining questions, supporting analysts in systematically approaching question design and improving the quality of analysis due to clarified requirements.

2 Related Work

Based on our review, we identified two areas of research related to our study. First, Sect. 2.1 covers methodological studies on process mining analysis questions and their alignment with analysis steps. Second, Sect. 2.2 reviews papers on question catalogs and categorizations, distinguishing between works that focus on use cases or techniques, those that propose collections of questions without a categorization, and those that attempt both.

2.1 Role of Analysis Questions and Methodologies

Questions play a crucial role in guiding data-driven investigations across various analytical domains. Well-formulated questions help ensure that analyses align with business objectives and maximize the value of extracted insights (Nalchigar and Yu 2018). In process mining, defining relevant questions is particularly critical, as it shapes subsequent steps such as data selection, event log preparation, and the choice of analysis techniques (van Eck et al. 2015). Several methodologies, including PDM, the L* Life-Cycle Model, and PM², outline structured approaches for conducting process mining projects (Bozkaya et al. 2009; Van Der Aalst 2011; van Eck et al. 2015). While they emphasize the importance of defining objectives and questions, they do not provide concrete guidance on how to derive them systematically.

To address this gap, Zerbato et al. (2022) investigated how analysts approach process mining in the absence of predefined analysis questions. Their interview-based study revealed that analysts primarily rely on “raw data inspection” and domain knowledge to formulate questions. They proposed six recommendations to support question definition and refinement. However, these recommendations remain abstract, focusing on aligning questions with analysis steps rather than systematically generating new questions in the early stages of a project. In addition, Gorgen Erdogan and Tarhan (2018) developed a goal-driven methodology for healthcare processes, employing the Goal-Question-Metric framework (Basili 1992) to translate goals into concrete questions and map them to process mining features and performance indicators. They demonstrated their approach for a surgery process. Although all of these works emphasize the importance of well-defined and actionable questions, they focus on linking questions to analysis steps rather than deriving them.

2.2 Question Catalogs and Categorizations

A structured categorization of analysis questions is essential for organizing inquiries and enabling users to filter and prioritize relevant questions. Several prior works have attempted to classify process mining analysis, either by structuring use cases, compiling sets of questions, or combining both approaches.

Use-case and technique-driven categorizations. Initial works in this area categorize process mining use cases or techniques rather than focusing explicitly on analysis questions. Ailenei et al. (2012) proposed one of the earliest categorizations, grouping 19 use cases based on available techniques such as process discovery, conformance checking, and various enhancement strategies (e.g., time, organizational and case perspectives). Klinkmüller et al. (2019) later expanded this categorization by incorporating additional categories such as prediction, drift detection, and familiarization. Similarly, van der Aalst (2022) provided a comprehensive classification of process mining techniques, structuring them into six types, including discovery, comparative analysis, and action-oriented process mining. While these categorizations offer valuable insights into analysis objectives, they remain centered on technical methodologies rather than the nature of analysis questions themselves.

Complementary to these technique-driven categorizations, Eggers et al. (2023) and Marcus et al. (2024), took an organizational perspective to systematically classify process mining applications and identify their expected value potential and reveal ideal process mining set-ups, respectively. These contributions focus on categorizing governance and managerial aspects that are often overlooked in technically driven approaches. However, while they provide taxonomies for structuring process mining applications at a strategic level, they do not address the classification of analysis questions. Our work differs because it explicitly focuses on a structured approach to organizing questions in process mining, independent of specific techniques or organizational setups.

Unstructured question catalogs. In other disciplines, question catalogs serve as useful references to structure analysis efforts, such as in visual data analysis (Amar et al. 2005) and machine learning interpretability (Liao et al. 2020). Process mining, however, lacks such a carefully collected and validated question catalog. Some industry solutions, such as the Celonis Business Miner, address this need by providing predefined questions for standard processes (Ullrich and Lata 2023). However, such tools are limited to predefined process templates and do not support analysts in classifying or clarifying questions, or in formulating customized case-specific questions. Our proposed question catalog extends prior efforts by integrating empirically collected questions, ensuring broader applicability beyond standard templates.

Question catalogs with categorizations. Some works attempt to both compile questions and provide a classification schema. In their efforts to develop a framework for the analysis of emergency room processes, Rojas et al. (2017) suggested separating general process mining questions from those that are episode-based, i.e., specific to the process, and provided a set of exemplary questions for both categories. More relevant for our work is, instead, the literature that takes a domain-agnostic angle on use cases and questions. In this line, Milani et al. (2022) linked business-oriented questions to use cases derived from literature, categorizing them into "transparency," "efficiency," "quality," "compliance," and "agility." Their framework distinguishes whether questions are "descriptive," "comparative," "explanatory," or "recommendatory." Although useful, this classification focuses on a subset of question char-

acteristics. Our findings indicate that goals and use cases can be captured in a more nuanced way and across several dimensions.

The work of Barbieri et al. (2023) is the closest to ours. While they focused on developing a natural language querying interface, they also designed a taxonomy to categorize analysis questions. Additionally, they collected analysis questions provided by students (which they categorized using the taxonomy) to test their application. Their taxonomy supports comparisons of the scope of different natural language processing implementations by classifying questions based on the required technique and formulation complexity. However, it was not systematically developed or validated for broader analytical use. While their approach highlights the coverage of their interface, it does not provide a structured framework for practitioners and researchers to systematically classify or formulate process mining questions for analysis planning. Moreover, the collected questions were not assessed for quality (e.g., no exclusion criteria applied), leaving their relevance to real-world use cases uncertain.

3 Method

In this section, we outline our approach to developing and evaluating a classification schema in form of a taxonomy for assessing and categorizing process mining analysis questions. Our work is grounded in DSR, which emphasizes the iterative creation of artifacts designed to solve practical problems and contribute new knowledge (Peppers et al. 2007; Vom Brocke et al. 2020). Specifically, we aim to develop an artifact that provides a systematic framework for understanding process mining analysis questions.

In line with prior research on taxonomy design, various structured approaches exist. For instance, Bailey (1994) or Doty and Glick (1994) advocate different methods emphasizing various types of taxonomies and their theoretical grounding. While these methods provide valuable perspectives, our goal is to capture the nature of actual analysis questions. We aim to consider both, theoretical grounding based on existing classification approaches and empirical evidence from real-world analysis questions. Therefore, we follow the widely adopted methodology for taxonomy design of Nickerson et al. (2013), which aligns with the iterative nature of DSR and ensures methodological rigor in the design process (Oberländer et al. 2019). This methodology has been further refined by Kundisch et al. (2021), who emphasize the importance of problem motivation and rigorous evaluation.

Our methodological approach, illustrated in Fig. 1, followed multiple development and evaluation iterations in line with DSR principles. The employed evaluation options were twofold: first, objective and subjective ending conditions were validated after each development iteration. Second, when they were deemed fulfilled, a broader evaluation, ideally involving external reviewers, was conducted. In our project, the first external evaluation took place after five design iterations. Accordingly, our design, development, demonstration, and evaluation steps can be grouped into two phases. Below, we chronologically outline the complete taxonomy development process, including the collection of analysis questions from literature and surveys at different stages. Further methodological details, as well as the collected questions, are available in the appendix at <https://doi.org/10.5281/zenodo.16743262>.

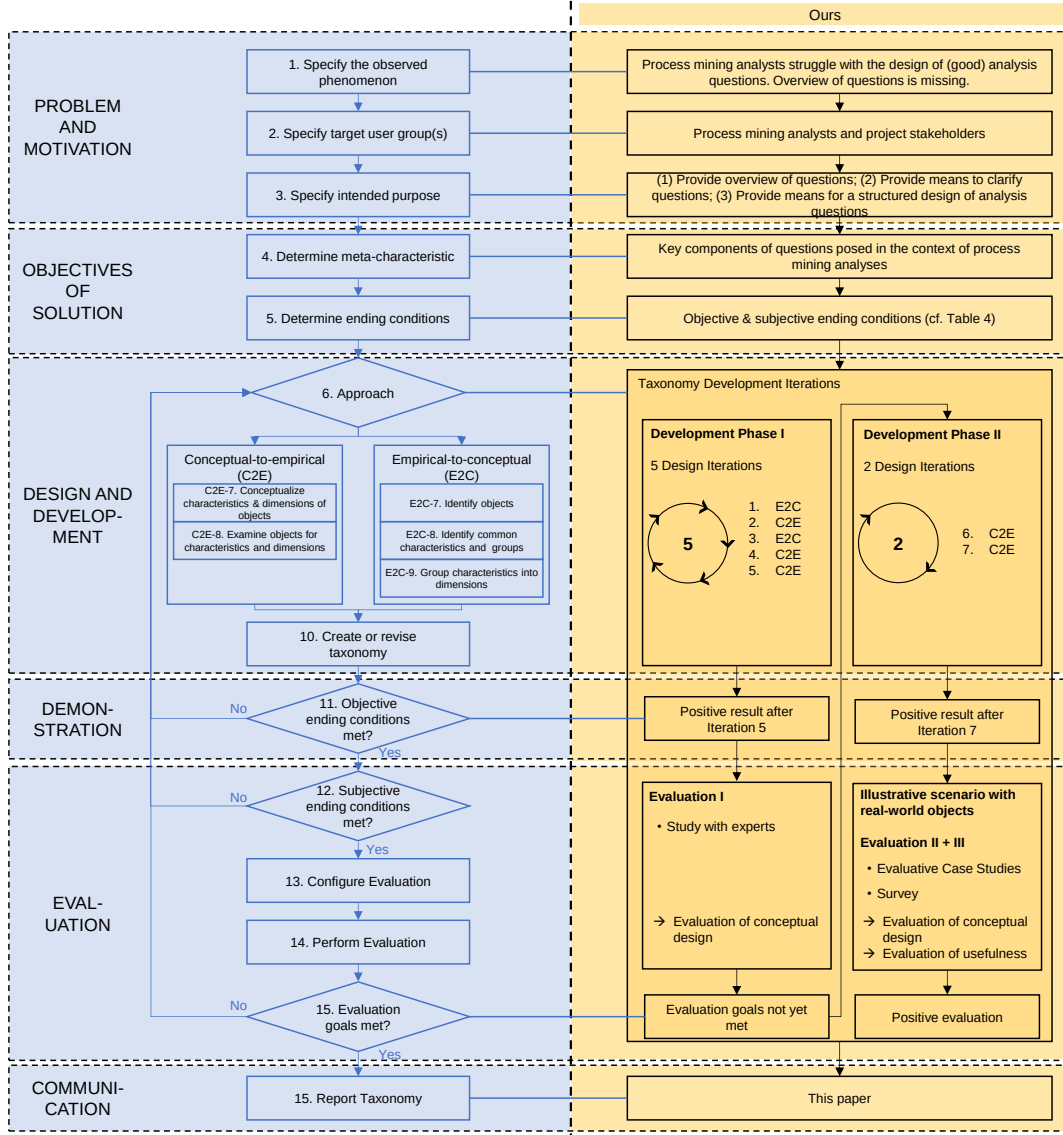


Figure 1: Overview of our taxonomy development approach. The blue connectors “ — ” indicate the correspondence between the steps outlined in this paper (right side, highlighted in yellow) and those in Kundisch et al. (2021).

3.1 Problem and Motivation

The first three steps of the taxonomy design focus on defining the problem and ensuring that the taxonomy encompasses an existing phenomenon. In our case, we build on prior research demonstrating that the design and formulation of process mining analysis questions denote challenges and that currently, little (methodological) guidance exists to identify relevant analysis questions (cf. Sect. 2.1). We derived three concrete intended purposes to guide the taxonomy development: (*Purpose 1*) Provide an overview of components of process mining analysis questions raised in projects in practice; (*Purpose 2*) Provide means for the structured clarification of analysis questions; (*Purpose 3*) Provide means for the structured design of (new) analysis questions.

3.2 Objectives of the Solution

Meta-Characteristic. The meta-characteristic helps to choose specific details for the classification system, based on how it is intended to be used (Nickerson et al. 2013). Given the goals for this study and the intended purposes for the taxonomy (cf. Step 3., Fig. 1), we formulated the following meta-characteristic: *Key components of questions posed in the context of process mining analyses.*

Ending Conditions. As the method for developing the taxonomy is iterative, it is important to set clear ending conditions that can be verified after each development iteration to determine whether the taxonomy development can be terminated. We adopted the full list of ending conditions from Nickerson et al. (2013) and reformulated them slightly to increase clarity. This resulted in the nine criteria in Table 1.

Objective	
A Mutual exclusivity	No question has more than one category in a dimension
B Collective Exhaustiveness	Each question can be assigned to a category in each dimension.
C Uniqueness	Every dimension, and every category within each dimension, are unique and not repeated.
D Saturation	No new dimensions or categories were added in the last iteration.
Subjective	
E Robustness	The taxonomy contains enough dimensions and categories to differentiate process mining analysis questions.
F Conciseness	The number of dimensions and categories provides valuable information without being overwhelming.
G Explanatoriness	The taxonomy is explanatory rather than descriptive.
H Comprehensiveness	The taxonomy includes all dimensions and characteristics to classify all process mining analysis questions.
I Extensibility	New dimensions and categories can easily be added to the taxonomy.

Table 1: Ending conditions for the taxonomy development

After the definition of the ending conditions, the design and development steps can be started. To this end, it should be determined whether a conceptual-to-empirical (C2E) or empirical-to-conceptual (E2C) approach would best advance the phase toward the ending conditions (Nickerson et al. 2013). In the C2E approach, taxonomy updates stem from conceptual considerations and are applied to data items. In the E2C approach, data items are analyzed qualitatively to derive concepts for characterization and grouping, which are then validated against the data (cf. Fig. 1, Steps 6-10). After each iteration, the ending conditions are assessed and their fulfillment is validated (cf. Fig.1, Steps 11.-12.). In case of a positive assessment, an ex-post evaluation should be performed (Kundisch et al. 2021).

In Fig. 2, we provide an overview of the seven development iterations we conducted. The interested reader can find further details for each iteration in the appendix. For the conduct of E2C iterations, empirical data items were required. Therefore, we collected questions from various sources to identify a substantial number of real-world analysis questions (cf. Table 2).

3.3 Design and Development, Demonstration, Evaluation - Phase I

The first five design iterations indicated in Fig. 2 comprised the design and development Phase I. It was concluded by an evaluation with external experts. During this phase, we collected process mining analysis questions from two sources.

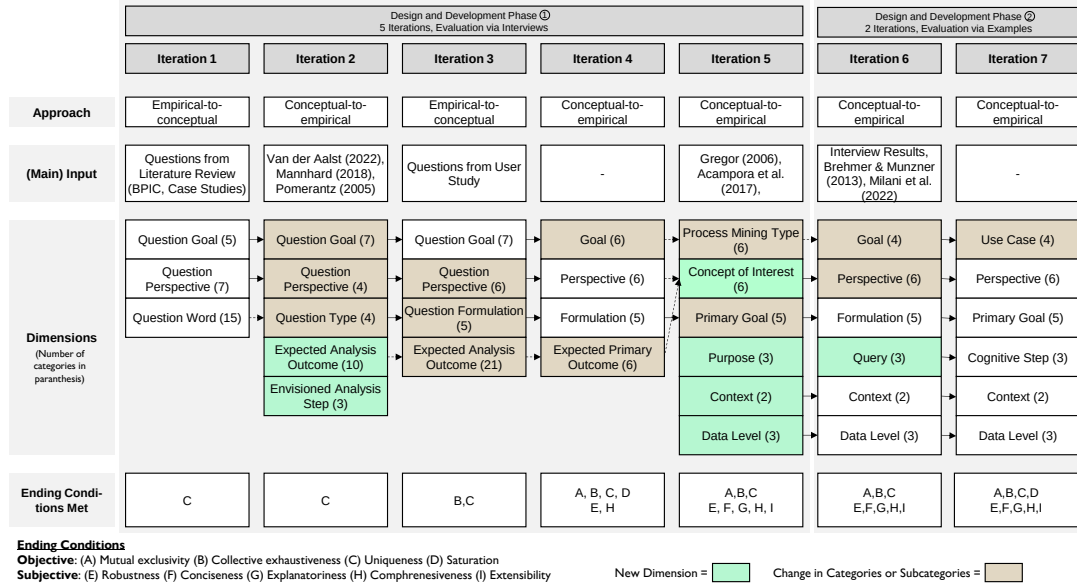


Figure 2: Overview of the taxonomy development iterations.

Development Phase	Question Collection	Source	Questions collected	Questions included
I	Question Collection 1	BPIC (2011-2020)	158	119
		IEEE TFPM Case Studies (2014-2021)	25	21
		Total	183	140
	Question Collection 2	Survey	279	178
II	Question Collection 3	Celonis Business Miner	105	87
Total			567	405

Table 2: Results of the collections: Number of collected and included process mining analysis questions.

3.3.1 Question Collection 1: Literature Review

To gather the first set of questions, we identified relevant literature sources and extracted direct questions from them.

Search Selection of Analysis Reports. Process mining reports, and especially those provided by the IEEE Task Force on Process Mining (TFPM), were previously identified to be a good resource for gaining insights into process mining practices (Capitán-Agudo et al. 2022). They provide insights into practical applications with real-life datasets and cover diverse application fields. Additionally, they are authored by various individuals, enhancing the representativeness of the resulting collection. We collected and reviewed 89 papers from the Business Process Intelligence Challenges (BPICs) spanning from 2011 to 2020, alongside 44 case studies conducted between 2014 and 2021. We carefully screened these resources to identify direct questions and discarded reports without them, resulting in 75 BPIC reports, and 10 case studies that were deemed relevant and included in our research.

Collection of Process Mining Questions. After identifying relevant sources, we reviewed the reports to extract all direct questions. Compound questions linked by conjunctions were divided into separate questions. As summarized in Table 2, this led to the identification of 183 direct questions. We applied exclusion criteria to ensure that the final set of empirical items was relevant to inform the taxonomy design and for the inclusion in the question catalog. In particular, we excluded questions that (EC1) were exact duplicates of previous questions, (EC2) were not written in English, (EC3) fell out of the scope of process mining analysis (e.g., questions that purely affect data pre-processing or the technical feasibility of an analysis), and (EC4) were not understandable (e.g., due to poor or very specific formulation). As a result, we excluded 43 questions, which resulted in a set of 140 included analysis questions.

3.3.2 Question Collection 2: Survey

In addition to the questions gathered from the literature, we aimed to enhance the question catalog with questions encountered by analysts in practice. To this end, we designed an online user survey allowing us to reach process analysts with diverse backgrounds and working in different industries.

Study Design. The user study was designed as a qualitative online survey, following the recommendations of Braun et al. (2021). We targeted practitioners to collect questions they had previously answered (or at least worked on) themselves. During the study, participants were prompted to reflect on up to three recent process mining projects. For each project, they were asked to enter up to five specific questions and link each of these questions to techniques used to answer them. In addition, we asked questions about the type of the analyzed process, the industry, the event log origin, and the analysis goals. Finally, participants were encouraged to reflect on their overall experience and indicate three questions they encountered across all their process mining analyses. Demographic questions and experience-related information concluded the questionnaire. We provide the full questionnaire in the appendix.

Data Collection and Participants. Question Collection 2 took place in August and September 2022 through one-on-one outreach on LinkedIn, targeting academics and practitioners with varying process

mining expertise. Of the 210 people contacted, 42 responded, and 39 fully answered the questionnaire. Those 39 participants came from 14 countries and 17 industries, with expertise levels ranging from basic (4) to average (6), good (13), and advanced (16). A detailed overview of them can be found in the appendix. The 39 participants contributed 279 questions, leading to an average of 7.2 questions per person. We applied the same pre-processing and exclusion criteria as in Question Collection 1 (cf. EC1-EC4, Sect. 3.3.1). 23 questions were duplicates of questions already part of the catalog (EC1), 49 fell out of the scope of process mining analysis (EC3) (e.g., Can we add data from table X to the model?) and 29 questions were poorly formulated or not formulated as a question and thus not understandable (EC4) (e.g., Sales orders first time rate.; Can all the GBS-eligible positions be actually moved?). As a result, we added 178 analysis questions to the catalog (cf. Table 2).

3.3.3 Development Phase I: Iterations 1-5.

In the first development phase (cf. Phase I in Fig. 1), we conducted two E2C iterations in which we leveraged the gathered analysis questions from Question Collection 1 and 2 respectively, and three C2E iterations in which we refined dimensions and categories based on conceptual considerations and literature. At the end of Phase I, we reviewed the taxonomy and assessed the ending conditions. As we perceived them as fulfilled, we conducted the first evaluation (cf. Evaluation I, Sect. 3.3.4).

Figure 2 provides an overview of the taxonomy dimensions that evolved during Development Phase I, as well as the approach and the main input used during each iteration. In the first iteration, we analyzed the questions from Question Collection 1 based on grounded theory coding (Charmaz 2006). We applied in-vivo coding and subsequently merged groups into categories, which we ordered in three dimensions: Question Goal, Question Perspective, and Question Word. In the following iterations, we integrated existing foundations from the literature and added questions from Question Collection 2. The additional questions provided new hints for more fine-grained categories and sub-categories within the existing dimensions. During Phase I, the authors met regularly to review and challenge the current categorization to further improve it. At the end of Iteration 5, all ending conditions but saturation were deemed satisfied. Saturation could not be confirmed as new dimensions were added in the previous iteration. However, as we did not expect the taxonomy to change with another design iteration, we configured and performed the first evaluation.

3.3.4 Evaluation I

We conducted the first evaluation to gather feedback from external experts uninvolved in the design process to refine the taxonomy.

Study Design. We designed Evaluation I as one-on-one sessions combining a practical part with semi-structured interviews and a questionnaire. The study was divided into three parts: (1) an introduction, where the interviewer presented the taxonomy and discussed two example categorizations; (2) an application phase, during which participants had to categorize analysis questions on their own —these questions were either self-formulated by the participant at the beginning of the session or provided samples; and (3) a feedback phase, where participants shared general remarks, assessed the subjective ending

conditions (cf. Table 1), and reflected on the usability of the taxonomy for question classification and formulation. The examples presented during part one and the questionnaire can be found in the appendix.

Participants and Execution. We conducted Evaluation I in October 2023. We reached out to our personal network and involved people with expertise in conceptual design and process mining. This included mainly senior researchers, but we also involved practitioners to receive feedback from people working in practice. We conducted sessions with 14 participants, of which one was excluded due to insufficient experience with process mining. The remaining participants had an average experience of 9.62 years in process mining and 8.62 years in the development of concepts, such as taxonomies, theories, models, or similar. An overview of their demographics can be found in the appendix.

Analysis and Results. During the evaluation, we received qualitative feedback regarding the understandability of the taxonomy and the ease of its application for question classification. This feedback was assessed and summarized by the author conducting the study. While the received feedback was rich, three main concerns were put forward by several participants: (1) The dimensions of the taxonomy are not orthogonal; (2) The taxonomy might not be complete; (3) Process mining (analysis) types are hard to anticipate given a question.

Additionally, we statistically analyzed the participants’ agreement to the subjective ending conditions. The results, presented in Table 3, show an average score of 3.4, indicating overall agreement and supporting the positive direction of our taxonomy development.

Subjective Ending Conditions		Mean	Variance
E	Robustness	3.54	0.27
F	Conciseness	3.54	0.44
G	Explanatoriness	3.31	0.40
H	Comprehensiveness	3.31	0.56
I	Extensibility	3.38	0.76

Table 3: Agreement with ending conditions (Evaluation I): Responses on a 4-point Likert scale ranging from 1 (Fully Disagree) to 4 (Fully Agree).

3.4 Design and Development, Demonstration, Evaluation - Phase II

In the second development phase (cf. Phase II in Fig. 1), we refined the taxonomy based on the results of Evaluation I. While the categorization of the initial set of empirical questions met our ending conditions (cf. Table 1), the expert feedback indicated room for improvement. Therefore, we expanded the question catalog to ensure a comprehensive and representative validation of the ending conditions for further iterations.

3.4.1 Question Collection 3: Celonis Business Miner

In Question Collection 3, we aimed to identify an additional, industry-related set of questions representing the interests of companies. To this end, we identified the work of Ullrich and Lata (2023), which presents the Celonis Business Miner. This tool, integrated into the process mining software of Celonis, the leading vendor in the process mining market (Kerremans et al. 2024), provides access to a set of pre-determined analysis templates based on questions tailored to common processes analyzed by many

Celonis customers. As Celonis has a market share of approximately 50% and the largest cross-industry customer base of any process mining vendor (Kerremans et al. 2024), we considered the information provided by this tool vendor to be representative of the requirements from industry.

Collection of Process Mining Questions. We obtained a document listing all 105 questions embedded in the Celonis Business Miner. Similar to previous data collections, we divided compound questions and applied the exclusion criteria (cf. Sect. 3.3.1). As Table 2 shows, out of 105 questions, 87 were added to the question catalog.

3.4.2 Development Phase II: Iterations 6-7

During development Phase II, we conducted two additional C2E design iterations (cf. Fig. 2) and incorporated questions from Question Collection 3 to assess the ending conditions, ensuring the generalizability of dimensions and categories. The main improvements focused on the dimension *Concept of Interest*, as Evaluation I revealed that several subcategories in this dimension could conceptually span several categories. In particular, we reinstated *Perspective* as a separate dimension. Additionally, as Evaluation I revealed difficulties in classifying questions in the dimension *Process Mining Type*, we replaced this dimension with the *Use Case* dimension. Besides, minor refinements in naming and definitions were made to increase understandability.

The final question categorization confirmed the objective ending conditions, and all authors reviewed and validated the subjective ending conditions. Before concluding the taxonomy development, we evaluated it by demonstrating its applicability based on illustrative scenarios (cf. Sect. 4.7) and performed two evaluations with the target user group. In particular, we designed an evaluative case study (Yin 2018) and an online survey with students. For better readability, we describe them in Sect. 6.

4 Process Mining Question Taxonomy

This section presents the final version of the taxonomy, displayed in Fig. 3. The taxonomy aims to support process mining researchers and practitioners in gaining an overview of the range of analysis questions and in clarifying and designing such questions. It is the result of seven design iterations and captures the key components of questions posed in the context of process mining analyses. Below, we outline the dimensions along with their empirical and theoretical grounding. All sample questions provided are taken from the question catalog (cf. Sect. 5).

4.1 Dimension 1: Use Case

The *Use Case* dimension captures the specific objective for which process mining is employed, highlighting the value it can bring. It is grounded in established classification schemes, such as the six process mining types proposed by van der Aalst (2022) (i.e., process discovery, conformance checking, performance analysis, comparative process mining, predictive process mining, and action-oriented process mining), the process mining use case categorization suggested by Milani et al. (2022), (i.e., transparency, efficiency, quality, compliance, and agility), and the definition of process performance as including time, quality, costs, and flexibility (van Looy and Shafagatova 2016). Based on this foundation, we considered

Dimensions	Categories & Subcategories				
Use Case	Transparency	Performance	Compliance	Automation	
Perspective	Control-Flow	Time	Resources	Data	Costs
Primary Goal	Describe Quantitative Qualitative	Confirm	Explain	Predict Execution Prediction What-If Analysis	Prescribe
Cognitive Step	Identification	Comparison			Summary
Context	Domain-Specific			Domain-Agnostic	
Data Level	Log	Trace	Event	Unspecific	

Figure 3: Process Mining Analysis Question Taxonomy

the applicability and completeness of the categories and their orthogonality to the other dimensions. As a result, we identified the categories **Transparency**, **Performance**, **Compliance**, and **Automation**, as we found them to provide a comprehensive, yet concise set of process mining use cases.

Transparency is concerned with providing visibility into a process and improving the understanding of its execution. It is essential for making informed decisions and might be perceived as a prerequisite for more advanced use cases. Examples for this category include questions that aim at revealing the structure of the process (“How does the process model look like?”), patterns within the execution data (“How many travel declarations are never approved?”), and questions about more specific concepts such as decision rules (“Are there any decision rules among decision points?”), or process drift (“Do the usage patterns of the website by customers change over time?”).

Performance aims at measuring, assessing, or improving processing time, process costs, process quality, or process flexibility (van Looy and Shafagatova 2016). Examples for this category include questions about temporal aspects of the process (“How many invoices are paid late?”, “Where is the bottleneck in the process?”), but also about quality indicators related to the control-flow (e.g., loops, rework, duplicates) and process execution costs (“What are the effort costs in this process?”)

Compliance describes the assessment of a divergence between the behavior captured in the event log and the expected behavior. Examples include questions about deviations from a normative process model, constituting an analysis type commonly denoted as conformance analysis (“How is the conformance rate?”; “What is the root cause for non-conformance?”), but may also pertain to deviations from established business rules, such as temporal or normative constraints (“What actions could be taken to decrease the SLA violation rate?”; “Which customers require credit reviews due to high credit exposure?”).

Automation captures the assessment of the automation level of a process. This includes both the identification of the current automation level (“How automated is your process?”) as well as the assessment of automation potential (“Where can the automation rate be improved?”).

4.2 Dimension 2: Perspective

The *Perspective* dimension describes the way the process is viewed. Independent of its use case, a question can focus on particular attributes and adopt a specific view on the process. In the literature, several perspectives have been considered for process design, process models, and process analysis (Mannhardt 2018; Reichert and Weber 2012). During the development of the taxonomy, we identified the relevancy of this dimension early (Iteration 2) and included the categories described by van der Aalst (2018). Based on the questions, we identified the need to slightly expand this set and added the perspectives **Cost** and **Unspecific** to **Control-Flow**, **Time**, **Resources**, and **Data**.

Control-Flow is defined as the ordering of activities. It describes questions that focus on process activities, their relationships and frequency, including their order and the constraints regulating their execution. Examples include queries such as “What activities are executed after a car loan application is rejected?” or “What are the different process variations over the departments?”.

Time is concerned with the timing of events. It covers questions about the throughput time (or cycle time) of the complete, or parts of the process (“What is the throughput time of a travel declaration from submission to paying?”), questions about bottlenecks, or about delays in the form of mismatches between expected and actual execution times (“How does the current as-is Purchasing Lead Time compare to the planned Purchasing Lead Time in the source system?”).

Resources seeks insight into the organizational entities or resources involved in the process (e.g., actors, organizational units), their attributes (e.g., role), and their interactions (e.g., dependencies, work distribution). It therefore describes a wide range of questions such as “Who is touching the process?”, “What are the possible points for improvement on the organizational structure for each of the municipalities?”, or “Which activities are done manually?”.

Data focuses on the data attributes of cases, their interrelations, and their involvement and influences in the process. Examples for this category include questions such as “How many different purchase order types are used?”, “For what products is the push to front mechanism most used?”, or “How can you reduce short payment terms?”.

Costs describes the financial implications of executing a process. Examples cover hypothetical scenarios (“What are the potential savings if we avoid a specific process flow?”) but might also include questions about actual processing costs (“What are the additional costs that are produced by late payments?”).

Unspecific captures all questions that do not imply a specific perspective. These questions may mention multiple perspectives, such as “By how much would the SLA violation rate decrease if we automate activities X, Y, and Z and if we add more resources for activities V and W?”, where it is not possible to prioritize between time and resource, or no perspective at all (“Is there anything interesting observable regarding the inventory management?”).

4.3 Dimension 3: Primary Goal

The *Primary Goal* dimension emerged from efforts to capture the actual formulation of a question. Earlier studies in linguistics and question-answering emphasized the importance of wording and structure

to classify questions (Antoniou and Bassiliades 2022; Pomerantz 2005). However, in the context of process mining, questions that share the same interrogative term can be substantially different. For example, “How long is the overall process duration?” and “How does the number of calls influence the final outcome?” expect distinctly different types of answers. Additionally, questions with different question words can ask for similar aspects (“What are the reasons for my bottleneck?” and “Why do the bottlenecks occur?”). Therefore, the Primary Goal captures how question formulation guides the expected answer, representing its function in meeting the information need implied by the question. Five categories are observed: **Describe**, **Confirm**, **Explain**, **Predict**, and **Prescribe**. This categorization follows the primary goals of theories in information systems research (Gregor 2006) but is extended by the empirical observation of confirmatory questions that require the existence of a prior hypothesis.

Describe indicates the need for a general description of a phenomenon. Questions within the subcategory **Describe: Quantitative** expect numeric values and oftentimes start with the interrogative terms *How* (“How many corrections have been made for declarations?”) or *What* (“What is the change rate of invoices already entered in the system?”). Instead, the subcategory **Describe: Qualitative** describes questions that seek a qualitative description (e.g., a visualization, attribute names, or textual description). Such questions cover a range of interrogative terms, including *Where* (“Where are differences in throughput times between the municipalities?”), *What* (“What is the relation between the teams that participate in the interaction, incident and change management?”), *Which* (“Which is the location where the “wait user” substatus is most incorrectly used?”), *How* (“How does the usage of the website by customers change over time?”), and *Modal Verbs* (“Are there any other interesting dependencies?”).

Confirm aims at verifying an assumption or hypothesis. Those questions might emerge during the analysis as stakeholders build hypotheses based on initial analysis results (Zerbato et al. 2022), or based on prior domain knowledge. Examples are “Are applicants who are confronted with more requests more likely to not accept the final offer?” or more generic questions such as “Is a specific error in the process correlated with the throughput time?”.

Explain aims at explaining a phenomenon. In her taxonomy, Gregor (2006) describes theories of this kind as “says what is, how, why, when and where”, demonstrating the diversity that explanations can provide. Questions in this category might start with *What* (What causes), *How* (How can it be explained?), or *Why* (“Why do the two IT organizations differ?”).

Predict aims at forecasting future happenings. First, questions such as “Which fund applications will be paid late?” are categorized in the subcategory **Predict: Execution Prediction** and aim at forecasting the future execution of particular cases in the event log. The second type of prediction is labeled **Predict: What-If Analysis** and requires forecasting the effects of (hypothetical) scenarios on the process execution (i.e., implications of potential process changes), such as demonstrated by the question “What would be the efficiency improvements if claims were in-bounded electronically?”

Prescribe captures questions that ask for how to do something. Questions in this category ask for guidance on process changes. They use the interrogative terms *How* (“How can we improve the process?”), *Is/Are*, or *What* (“What actions could be taken to decrease the SLA violation rate?”).

4.4 Dimension 4: Cognitive Step

During taxonomy development, we drew on established categorizations, including the one by van der Aalst (2022), already referred to for Dimension 1 (cf. Sect. 4.1). For the *Cognitive Step* dimension, the classification of comparative process mining, which includes methods for event log comparison, was particularly influential. We observed that questions indeed imply the final operation (e.g., comparisons) required before presenting information as their answer. In line with literature from the visual analytics domain, which recognizes **Identification**, **Comparison**, and **Summary** as integral to analytical tasks (Andrienko and Andrienko 2006; Brehmer and Munzner 2013), we applied the same differentiation for this dimension. To distinguish it from lower-level user actions or system interactions, we named it Cognitive Step to emphasize the higher-level reasoning processes involved in question answering. Notably, in this case, the cognitive step is not limited to human reasoning but can be inherent in analytical techniques.

Identification describes the need to point out single or multiple independent entities or phenomena. Those phenomena may need to be identified based on information in the data (e.g., activity or attribute name), as required for “For what products is the push to front mechanism not used?”, or they may need an assessment or conjecture from the analyst, requiring domain knowledge and industry expertise that is not available from an event log alone (“How can we improve the process?”).

Comparison captures questions that require comparing two or more subsets of (previously identified) elements or phenomena. It includes the comparison of subsets of the data (“How does the lead time compare between different sites?”), the comparison of the data with an external benchmark (“How does the process performance compare to the industry benchmark?”), or the comparison of frequencies across event log elements such as events or data attributes (“Which process steps are reworked most often?”).

Summary involves aggregating multiple data elements or process aspects to provide a comprehensive overview. Examples are “How many travel declarations are booked on projects?” or “Is there a collection of process models which together properly describe the process?”. In contrast to the identification category, summary questions necessitate an aggregation (e.g., sum, mean, correlation) or a descriptive synthesis (e.g., comprehensive process model).

4.5 Dimension 5: Context

The *Context* dimension captures whether the question (as it is formulated) is **domain-specific** or **domain-agnostic**. This characteristic does not influence the analysis but determines in which domain(s) the question is understandable. We integrated it during Iteration 5 (cf. Sect. 3.3.3) after observing that some of the analysis questions referred to specific application areas and processes, while others remained broad, asking for generic concepts. This differentiation aligns with Rojas et al. (2017), who differentiate generic analysis questions from those that are only relevant for their episode-based domain.

Domain-specific describes questions that use specific language or refer to a specific process. Examples range from asking for a generic concept (e.g., throughput time) for a particular process (“How long does each step in the allocation process of a property take?”) to questions on domain-specific phenomena. For

example, “What are the additional costs that are produced by late payments?” is only relevant when the concept of late payment exists.

Domain-agnostic describes questions that are generic and can be raised for any process, given that the data required to answer them is available. An example is “How long is the overall process duration?”.

4.6 Dimension 6: Data Level

Several streams in process mining research deal with aspects concerning the data, such as the event log format, data quality, or event creation. Besides these streams, it is commonly accepted that most process mining techniques require structured data and projects include a data pre-processing phase (Emamjome et al. 2019). One widely accepted event log format is XES, which discriminates the concept of **Log**, **Traces**, and **Events** (Acampora et al. 2017). When analyzing process mining questions, it becomes evident that many inquiries can be categorized based on the data elements they refer to. However, some questions may have **Unspecific** data level requirements.

Log describes questions that address the event log as a whole. Specific traces or events do not need to be particularly considered. Examples are “How does the process model look like?” or “What are the roles of the people involved in the various stages of the process?”.

Trace captures questions that refer to (a set of) specific traces (cases, instances) that satisfy a property (e.g., have a specific value for a data attribute or include a specific pattern). Traces might need to be considered in full (“Why is a specific process variant happening?”) or as subsequences (“How often does Activity X happen before Activity Y?”).

Event captures questions about (a set of) events or event attributes. Such questions may ask for clarification regarding a specific event (“What causes the significantly longer throughput time for specific steps in the process?”, “What is the average number of occurrences for Activity X?”) or require the identification of an event given certain properties (“Where is my bottleneck?”).

Unspecific questions lack a clear focus on a specific data level, often due to ambiguity requiring the consideration of multiple levels or more information about the structure of the event log. For example, the question “Why does the process take that long?” could require the analysis of specific (long) traces, events causing delays, or the entire event log, including all variants and events. In such cases, the analysis depends on the available data, the domain, or an initial data assessment, resulting in an unspecific data level upfront of the analysis.

4.7 Taxonomy - Illustrative Application

In this section, we outline the classification of three analysis questions (cf. Fig. 4) to demonstrate the taxonomy’s applicability to real-world objects, as suggested by Kundisch et al. (2021).

Example 1. The BPIC 2020 provided a real-life event log containing data about the reimbursement process of work-related expenses of employees. In this context, one of the analysis questions raised by the process owner, the Eindhoven University of Technology, was as follows.

Where are the bottlenecks in the process of a travel declaration?

Dimensions	Categories & Subcategories					
Use Case	Transparency	Performance	Compliance		Automation	
Perspective	Control-Flow	Time	Resources	Data	Costs	Unspecific
Primary Goal	Describe	Confirm	Explain	Predict	Prescribe	
	Quantitative Qualitative			Execution Prediction What-If Analysis		
Cognitive Step	Identification		Comparison		Summary	
Context	Domain-Specific			Domain-Agnostic		
Data Level	Log	Trace	Event	Unspecific		

Example 1: Where are the bottlenecks in the process of a travel declaration?

Example 2: Do customers visit different pages when they start using the website versus when they have been using the website for some time?

Example 3: What are the causes of stop in automation?

Figure 4: Categorization of exemplary process mining analysis questions.

Figure 4 shows its classification in the taxonomy. Bottlenecks are recognized as important performance indicators of processes, as they might be root causes for delays or quality defects (Lawrence et al. 1995). Therefore, the question falls into the **performance use case** and the **time perspective**. It starts with the question word *Where* and expects a **qualitative description** (name of the bottleneck events) as an answer. The question does not ask for a comparison or any summary of the bottlenecks, we therefore assume that the **identification** of them is sufficient. As it refers to a specific process (travel declaration) it is **domain-specific**. Lastly, we consider the **data level** and argue that bottlenecks are expected at the **event** level as they are associated with specific events in the process. Note that this classification of the question by an analysts could now serve as a baseline for alignment with the project stakeholders to validate the categorization and ensure common expectations.

Example 2. The second example originates from the BPIC 2016 in which process analysts were asked to provide their insights for a customer journey process of an insurance agency.

Do customers visit different pages when they start using the website versus when they have been using the website for some time?

The summary of its classification can be found in Fig. 4. The question asks to identify changes in the use of a website by customers over time. In process mining, such a question links to a concept referred to as drift detection (Maaradji et al. 2017). We therefore conclude that it falls into the **transparency use case**, as it aims to reveal changing customer behavior, which is represented in the **control-flow perspective**. The question implies the hypothesis that changes exist and therefore seeks **confirmation**. An analyst is required to **compare** the control-flow pattern at different points in time. Similarly to the first example, it is **domain-specific** but instead of mentioning a particular process, it refers to a specific concepts (i.e., website use by customers). Finally, the question implies the differentiation between subgroups of **traces**, namely those that were recorded during early use of the website and those created during later use. Using this classification as a baseline, the analyst could verify whether control-flow (order of visited

webpages) is indeed the main focus or whether changes along other perspectives, such as data attributes or changes of timing, are also of interest.

Example 3. Another example is the following question that had to be analyzed by one of our survey participants (Question Collection 2, cf. Sect. 3.3.2)

What are the causes of stop in automation?

This question seeks an **explanation** as its **primary goal** and falls in the **automation use case**. It does not imply any specific perspective, leaving the potential sources of causes open. Thus, the **perspective** is classified as **unspecific**. While the questioner might not be able to provide further guidance, the available data might impose restrictions on the process perspectives that can be analyzed. Consequently, in pre-analysis planning, any limitations to the perspective could be identified and the remaining ones should be prioritized for the analysis. The question calls for the **identification** of causes. Although specific to process properties, it maintains a **domain-agnostic** formulation and could be raised for any process and domain in which automation is a relevant use case. Similar to its unspecific perspective, also the **data level** is **unspecific**, leaving ambiguity regarding the scope of stops in automation (i.e., whether pertaining to a subset of traces or events). Compared to the previous two examples, this question is more open and without further information or interaction with stakeholders, it leaves room for the analyst to investigate multiple directions to identify an answer. However, especially in settings where an analyst is less experienced, they might not be knowledgeable enough to explore all directions and might miss parts of the answer. A breakdown into smaller, more concrete questions or the formulation of confirmatory questions along different perspectives (e.g., *Is a specific event in the process responsible for the stop in automation / causes re-routing from the workflow engine?*; *Is a specific user responsible for the stop in automation?*) could be facilitated by our taxonomy.

5 Question catalog

As part of this work, we performed three data collections to gather process mining analysis questions (cf. Table 2). As a result, we collected 119 questions from the **Business Process Intelligent Challenges** (BPIC), 21 from **Case Studies** published by the IEEE Task Force of Process Mining (CS), 178 with a **User Study** (US), and 87 questions from the **Celonis Business Miner** (Celonis). The resulting 405 questions cover a diverse set of process mining use cases, perspectives, primary goals, cognitive steps, contexts, and data levels. As this set represents a structured collection of real-world analysis questions, we refer to it as a *question catalog*. In the appendix¹, we provide the full question catalog and encourage readers to utilize this artifact in their projects. Each question within the catalog is classified along the six dimensions of our taxonomy and referenced with its source. For several questions, we also provide information about the process or business context it refers to. Additionally, for each question, we capture its “*Concept of Interest*”. This dimension was part of the taxonomy in Iteration 5 but was later removed due to inconsistencies and overlaps across categories (cf. Sect. 3.3.3). We still report it in the question catalog, as it can help to group questions based on the main process phenomenon they ask about. The most

¹<https://zenodo.org/records/16743262>

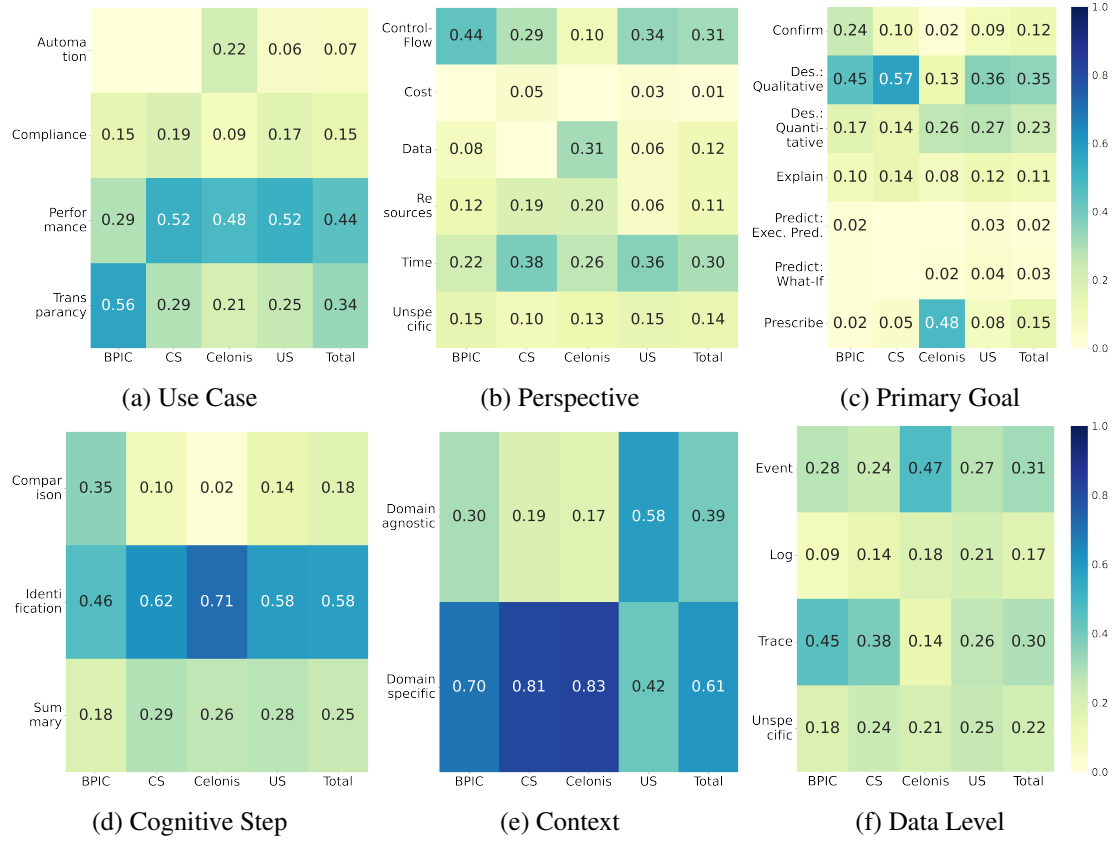


Figure 5: Heatmaps showing the distribution of questions across categories for each taxonomy dimension. **Total** represents the overall distribution of all questions per dimension. Values below 0.01 are omitted.

common concepts in the question catalog are Throughput Time, Process Execution Patterns, Deviations, Process Structure, and Delays. Less common ones are, for example, Fraud, Drift, or Decision Points.

Figure 5 presents the distribution of questions across the taxonomy dimensions, normalized to allow comparisons despite varying question counts per source. The heatmaps reveal that in almost all dimensions, certain categories dominate, highlighting recent process mining focuses. Less frequent categories suggest opportunities for further exploration. This could involve formulating more questions that existing techniques can address or developing new techniques to naturally generate more questions in these areas. Differences between practice-oriented sources (i.e., Celonis, US) and literature-based sources (i.e., BPIC, CS), besides both claiming to cover real-world use cases, may highlight potential gaps between theory and practice and suggest limitations in the representativeness of the literature.

In the **Use Case Dimension** (cf. Fig. 5a) nearly half of the questions pertain to performance, followed by transparency, whereas compliance and automation are less frequent. Questions related to automation are especially common for Celonis, contrasting with their absence in BPIC and CS reports. Regarding the **Perspective Dimension** (cf. Fig. 5b), control-flow and time each account for about a third of the questions, while 14% remain unspecific. When considering the concepts of interests that these questions cover, they are quite diverse, with deviations (“Where do the two IT organisations differ?”), execution patterns (“How can you optimize short lead time orders to improve On-time In-full deliveries?”), and general process performance (“How to improve the process?”) standing out. Celonis questions frequently

explore the data perspective, focusing on data attribute optimization and pattern detection. Cost questions are the least common, appearing only in CS and US, suggesting either less interest in this perspective or a lack of data/tools for cost analysis. When analyzing the **Primary Goal Dimension** (cf. Fig. 5c), one can notice that most questions focus on qualitative or quantitative descriptions, followed by prescriptive questions (mainly found for Celonis), and questions that require confirmations and explanations. Predictive questions are in general rare, constituting only 5% of the question catalog. They can be found in US and BPIC reports. In the **Cognitive Step Dimension** (cf. Fig. 5d) identification is the most common step, followed by summary and comparison. The distribution across sources is generally balanced, except for BPIC, where comparison is the second most prevalent category. The **Context Dimension** (cf. Fig. 5e) shows that questions in retrospective reports (US) tend to be formulated in a more domain-agnostic way, while the overall average suggests a preference for domain-specific questions. Finally, in the **Data Level Dimension** (cf. Fig. 5f) questions targeting logs are less common compared to those addressing events or traces. BPIC questions often concern traces, while Celonis frequently addresses events and in US, both categories are equally distributed. These findings provide insights into current trends and potential areas for further research and development in process mining.

6 Evaluation

Before concluding the taxonomy development, we performed two user studies to evaluate the usefulness of our artifacts (i.e., the taxonomy and question catalog) with end-users representing our target groups of less mature project teams and novice analysts. To facilitate the application, we first developed a structured approach that enables users to access and apply the taxonomy and question catalog in the context of question design (cf. Sect.6.1). We then carried out an evaluative case study with two organizations (cf. Sect.6.2) and a survey with novice analysts (cf. Sect.6.3).

6.1 Process Mining Question Design Approach

The question design approach is based on the idea of leveraging the taxonomy as a filter to the question catalog and allowing resulting questions to be adapted. For the development, we focused on the needs of non-expert users and modified the original artifacts as little as possible to demonstrate their usefulness.

Our approach can be split into three parts: In the **first part**, users reflect on their project goals and the process of interest and select relevant taxonomy categories. This selection then filters the question catalog in the **second part**, where users review the resulting questions, keeping only those that are comprehensible and appear relevant to them. In the **third part**, selected questions can be reformulated or ultimately excluded (e.g., when the reformulation fails). The remaining set of questions is the result of the question design approach and can be utilized as input for a process mining analysis.

To operationalize the question design approach, we implemented an interface to the taxonomy, which we called the Process Mining Question Forge (PMQF), as a web application supporting users through the three parts described above (Zimmermann 2024). PMQF enables users to design questions based on (raw) examples, facilitating interaction with the taxonomy. Examples of the interface of PMQF can be found in the appendix. While aiming for simplicity and usability, the interface itself was not the focus of

our evaluations. Instead, we assessed the taxonomy’s dimensions and categories and its ability to support question design by providing categorized example questions as a resource for practitioners.

6.2 Evaluation II

In Evaluation II, we **aimed to validate whether the taxonomy and question catalog effectively support the design of (new) analysis questions**, particularly in real-world settings and for organizations less mature in process mining. In particular, we wanted to understand if organizations identify benefits from using the taxonomy and question catalog and can apply them without difficulty.

6.2.1 Evaluation Method

Evaluation Design. We designed the evaluative case study so that it could be conducted in the form of a moderated workshop for three reasons. First, workshops let participants work on a concrete and relevant problem they are facing in their work practice. Second, it allowed us to closely interact with the participants and intervene in cases in which they lacked an understanding of commonly known process mining concepts. Additionally, participants also had the chance to clarify their problems which allowed us to understand their situation in detail. Third, workshops are collaborative, providing room for discussions among participants, which allowed us to observe these discussions and understand thoughts that were triggered by the taxonomy and question catalog (Thoring et al. 2020).

For the workshop design, we followed the guidelines of Thoring et al. (2020) and planned it in three parts. First, we introduced the research project, the taxonomy, and two categorized questions, following the approach in Evaluation I (cf. Sect.3.3.4), and discussed participants’ processes and challenges in identifying analysis questions (20-60 min). In the second part, participants explored PMQF independently, following the question design approach (cf. Sect.6.1), while we provided clarifications as needed (60-120 min). The final part consisted of short, semi-structured interviews and a structured questionnaire (30 min) to assess the usefulness of the artifacts and the clarity of taxonomy categories. We included questions on definition clarity, the Technology Acceptance Model (TAM) (Davis 1989), and feedback on the question design approach regarding first impressions, relevance to the organization, and potential improvements. The full evaluation guide and questionnaire can be found in the appendix.

Participants and Procedure. We organized two workshops with different organizations (cf. Table 4). Two authors used their network to identify organizations at an initial process mining maturity level interested in participating. From the identified candidates, we shortlisted two and reached out to their contacts, both of whom accepted the invitation.

We conducted at least one initial, informal meeting with participants of both organizations to clarify that they would fulfill our participation criteria: (a) the participants were required to work in an organization with an initial or repeatable process mining set-up, (b) they needed to be knowledgeable of at least one process within their organization for which they are interested in analysis questions, and (c) they needed to be available for a half-day workshop. In both organizations, we asked our main contact to invite further colleagues who are part of their team and are involved in the process mining initiative as process experts or as analysts. The four participants who accepted the invite are listed in Table 5.

	Organization 1 (Case I)	Organization 2 (Case II)
Country	Germany	Spain
Sector	Manufacturing	Public
Maturity Level	Initial	Initial
Description (Situation and Problem)	Introduced Process Mining and gained experience through the implementation of one of their production processes. In their initial project, process mining did not lead to the expected business value. The results were not “interesting” enough. Therefore they aim to discover more relevant questions to continue the initiative.	Organization is new to process mining and just started to consider business processes in its operations. The IT department drives the business process management initiative. They aim to better understand process mining use cases and gain an overview of what kind of questions are relevant for them.

Table 4: Overview of the two organizations taking part in Evaluation II.

On the day of the sessions, we connected with the organizations online. After a short introduction to the agenda and the PMQF user interface, we followed the evaluation design. At any time, participants were free to jump into discussions or clarify questions. All sessions were recorded.

		Job Role	Expertise	Years	Role in previous project	Difficulties with question design
Case I	C01*	Data Analyst	Average	3	Analyst	Yes
	C02	Production Planner	Basic	0	Evaluated analysis outcomes	No
Case II	C03	Project Manager	Novice	0	—	No
	C04*	Project Manager	Average	3	Analyst	Yes

Table 5: Participants of Evaluation II. Participants highlighted with * were our main contacts.

6.2.2 Evaluation Results

Perceived Usefulness and Ease of Use. Our evaluation goal was to confirm the usefulness of the artifacts. Figure 6 depicts the responses to the TAM questions we asked. All items were rated on a 7-point Likert scale ranging from 1 (extremely likely) to 7 (extremely unlikely). Overall, with an average rating of 2.46, the participants found the approach useful. The mean over the ease of use items equals 2.33. The overall usefulness (cf. Q6) was rated most positively while participants were more reluctant regarding the increase in productivity (cf. Q3) and the time the approach saves (cf. Q1).

Taxonomy Understanding. Participants indicated that they fully or partially understood all the dimensions and categories. C01 remarked that they did not consider conformance analysis as part of compliance in the **Use Case** dimension, but no further concerns were raised.

Question Catalog Understanding. During both evaluation sessions, participants provided valuable feedback on the question catalog. Their main concern was the specificity of some questions that belonged to domains they were not familiar with. They highlighted that it requires substantial effort to relate or adapt these questions so that they fit their own domains. Although for most questions the participants were able to come up with reformulations, they found this step hard and expressed their desire to find more questions specific to their domain of interest based on the initial selection they submitted. As a result, we identify a substantial opportunity to extend the question catalog with (a) a categorization of the

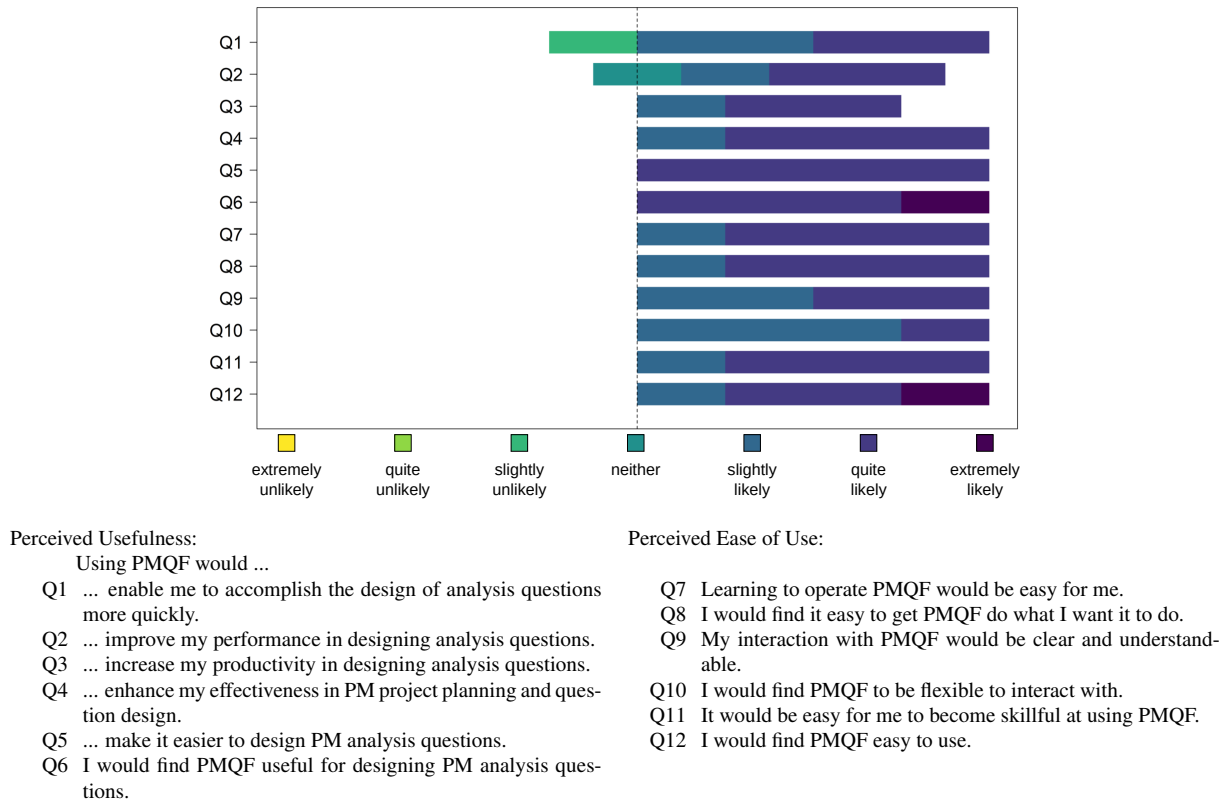


Figure 6: Usefulness and ease of use of our taxonomy and question catalog.

exact domain or process for which the question was originally raised and (b) an option to automatically suggest a translation between domains for domain-specific questions.

Question Design Approach. During the workshops, participants sought more guidance on handling identified questions, including determining data requirements and analysis methods. For example, C01 and C02 pointed out that the question “What are the main root causes of bottlenecks?” hints at an issue they are already aware of but struggle to define, break down, or answer. The feedback highlights the need for further support, particularly for non-experts, to ensure that the designed questions are practically applicable and can be answered by the organizations.

In both workshops, the use of PMQF sparked discussions on at least one new use case. C01 summarized “I believe that when a process is not at a high maturity level, it is a very useful tool. However, for a process with a high maturity level, it falls a bit short. It would be necessary to have more questions specific to our domain”. And finally, “For those processes that we know that we have problems, but we are not sure how to face them, the tool is useful”.

In conclusion, the evaluation confirmed the usefulness of the taxonomy and the question catalog for question design. It further confirmed the understandability of the dimensions and categories of the taxonomy and allowed us to identify concrete pointers for future work.

6.3 Evaluation III

Evaluation III aimed to validate the results of Evaluation II with a broader group of non-expert users outside a professional context, assessing whether the taxonomy and question catalog help them derive

meaningful process mining analysis questions.

6.3.1 Evaluation Method

Evaluation Design. To achieve the objective of Evaluation III, we designed an online survey using a between-subject design (Campbell and Stanley 2015), allowing for a comparison between two groups: one with access to the taxonomy and question catalog (Group A - Test Group) and a control group without such access (Group B - Control Group). The independent variable was the availability of the taxonomy and question catalog, while the dependent variables included question quality, perceived satisfaction with the question design process, and perceived satisfaction with the formulated questions.

During the survey, participants were provided with a project scenario to simulate a realistic use case. The scenario described a Swiss manufacturing company, called MyTruck, seeking to analyze its diesel engine manufacturing process to identify opportunities for improvement. Participants were asked to put themselves in the role of the analyst responsible for the project and formulate three relevant questions. To encourage deeper thinking and discourage the generation of obvious questions, the scenario specified that the process had already undergone an initial analysis, and descriptive information about the control flow and timing was known to the company.

Before being introduced to the scenario, both groups watched an introductory video, summarizing the five common stages of process mining projects (Emamjome et al. 2019). To explore the impact of the taxonomy and question catalog, Group A was given access to PMQF (cf. Sect. 6.1), with an instructional video explaining the approach. This ensured a consistent introduction to the tool while minimizing the risk of discrepancies in participants' understanding. Unlike Evaluation II, participants were not required to use the tool to save and refine relevant questions. Instead, they entered their questions directly into the survey and could use an additional text field for notes and question refinement. Additionally, we included demographic questions and measures to evaluate the experience with the question design process. To assess participants' satisfaction with the outcome, we adapted a scale originally designed for evaluating group decision-making processes (Reinig 2003), focusing on perceived satisfaction with the quality and relevance of the formulated questions. Another adapted scale was used to evaluate satisfaction with the design process itself, measuring whether participants felt they followed a structured and effective approach to question generation (Reinig 2003).

Following this study design, we aimed to balance experimental control and ecological validity while assessing how well the taxonomy and question catalog support non-experts in formulating meaningful process mining analysis questions under realistic conditions. The full questionnaire, including the scenario description and transcripts of the explanatory videos, is available in the appendix.

Participants and Procedure. Participants (i) must have completed at least one introductory lecture in process mining and (ii) should not have practical experience with process mining. These prerequisites ensured that all participants had a foundational understanding of process mining concepts, enabling them to effectively engage with the survey tasks, while maintaining alignment with our target user group of novice, less experienced analysts. By excluding individuals without any understanding of process mining, we aimed to eliminate potential confounding effects caused by unfamiliarity with basic concepts.

During the survey, all participants were asked not to use any external tools or resources to assist them in completing the tasks. To encourage participation, respondents were offered the opportunity to win one of three gift cards. When opening the link to the survey, participants were automatically randomly assigned to the test or control group. The random assignment ensured that observed differences in outcomes could be attributed to the intervention rather than to participant characteristics.

We distributed the survey invitation to student groups who we knew had completed at least one process mining course. In total, 50 responses were recorded, of which 32 were completed. One response had to be excluded due to the use of AI tools during the survey. The remaining 31 participants came from six countries, with eight reporting good, eleven basic, and twelve novice process mining expertise. An overview of the participants of Evaluation III can be found in the Appendix.

Analysis of Question Quality. To assess the quality of the questions generated by participants and identify potential differences between the two groups, two authors independently rated the analysis questions provided. Before the assessment, all questions were extracted from the survey responses and stored in a separate file in a randomized order without attribution to the participants or their group assignment, ensuring that the authors who rated the questions were unable to identify (i) whether a question was provided by a participant from the test or control group, and (ii) which questions originated from the same participant. With this procedure, we aimed to minimize any bias during the rating process. Rating the quality of questions included three steps:

1. Each question was categorized using the taxonomy.
2. The relevancy of each question was rated on a scale from 1 to 4, based on how well a question aligned with the goals of the scenario and key aspects of the diesel engine manufacturing process. The assessment helped to determine whether participants were able to conceptualize questions in line with the challenges outlined in the given scenario.
3. The specificity of each question was rated on a scale from 1 to 4, based on how precise and actionable it was, sanctioning vague or overly broad formulations. High specificity indicated a well-structured question with sufficient detail to guide analysis. By measuring specificity, we aimed to capture the participant's ability to articulate clear and practical analysis goals in a question.

Two authors conducted their rating independently. The inter-rater reliability of steps 2-3 resulted in a Cohen's kappa of 0.75, with a value of 0.89 for relevancy and 0.61 for specificity. To achieve consensus, the authors met to discuss and align their ratings and categorizations for all 93 questions.

6.3.2 Evaluation Results

Question Quality. Among the 16 participants in the test group with access to the taxonomy and the question catalog, the average relevance score was 2.98 (std = 0.95) on a 4-point scale, indicating a generally acceptable alignment with the process improvement objectives given. Specificity scored slightly lower at 2.79 (std = 0.86), suggesting that while most of the questions were actionable, some lacked precision. Frequently raised themes (concept of interest) observable in the questions include throughput time (e.g., "How can you speed up assembly?" - S13), process structure (e.g., "Where is the inspection taking place?" - S3), and energy consumption (e.g., "Can we detect process steps or machine usage

patterns that cause unusually high energy consumption by analyzing event logs?” - S18).

To assess the taxonomy’s impact, we compared these results with the 15 participants in the control group, who completed the task without the taxonomy. Their questions had an average relevance score of 2.73 (std = 0.90), 0.25 points lower than the test group, and a specificity score of 2.44 (std = 0.84), which makes a difference of 0.35 points to the test group. Although these differences are not statistically significant, the effect size for specificity measured by Cohen’s d was $d = 0.41$, indicating a small effect. For relevance, we calculated Cliff’s delta = 0.19, which indicates a small tendency for the test group to generate more relevant questions than the control group. Overall, these results suggest that access to the taxonomy and question catalog may lead to slightly more relevant and precise questions.

To further compare the provided questions, we assessed differences in their categorization according to our taxonomy. However, before the categorization of questions, we had to exclude eight questions from the test group in line with our exclusion criteria for the question catalog (cf. Sect. 3.3). They were not formulated in English (E2; 1 question), were no analysis questions (E3; 5 questions), or were poorly formulated (E4, 2 questions). For the control group, eleven questions had to be excluded, as they were no analysis questions (E3; 10 questions), or poorly formulated (E4, 1 question).

Figure 7 reveals that in the *Primary Goal* dimension, nearly half of the control group’s questions asked for a qualitative description. In contrast, the test group exhibited greater diversity, with 22.5% of questions seeking explanations and 15% aiming for prescriptive insights, which might be arguably more valuable given the scenario of MyTruck. Another interesting difference can be observed in the *Context* dimension. While the test group predominantly formulated domain-specific questions tailored to MyTruck’s diesel engine manufacturing process, the control group produced largely domain-agnostic questions, referencing generic concepts without explicitly mapping them to the scenario’s specific wording and objectives. Minor variations also exist in the *Use Case* and *Perspective* dimensions. The questions of the test group exhibit greater variability in these dimensions, as their distribution is more balanced, whereas the control group’s responses cluster around a few dominant categories. This suggests

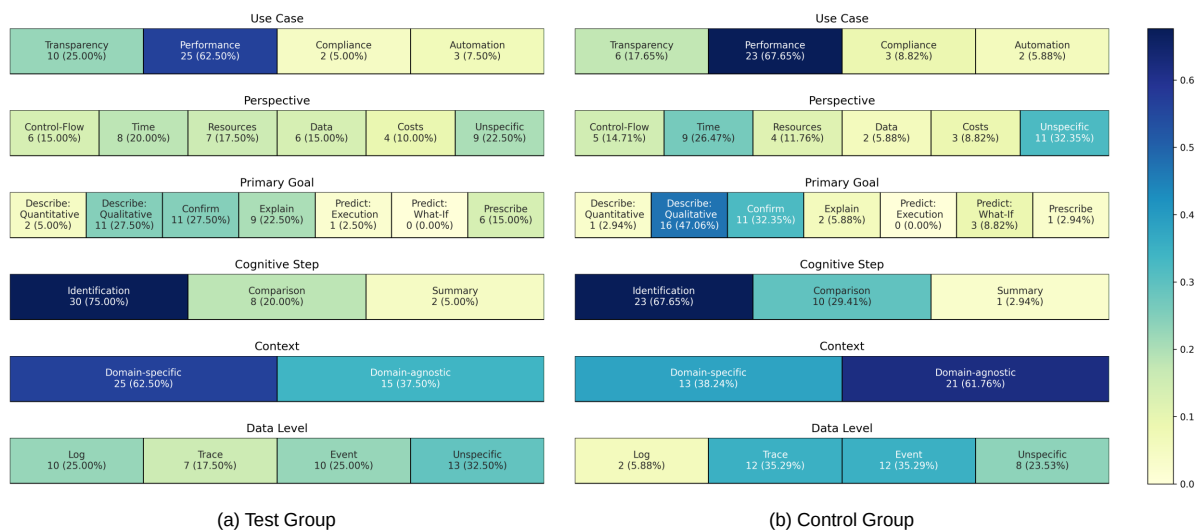


Figure 7: Results of the categorization of questions collected during Evaluation III.

that structured support through the taxonomy and question catalog not only improves question quality but also encourages broader exploration of different analytical perspectives.

Satisfaction Metrics. In addition to the quality of the provided questions, we assessed the participants' experience using the taxonomy and question catalog. Most test group participants were satisfied with both the question-generation process and their outcome, resulting in average ratings of 3.47 and 3.53 (on a 5-Point-Likert Scale), respectively. Interestingly, the control group reported a similar perception, with averages of 3.48 and 3.45, respectively. While both groups found question generation challenging, the slightly higher outcome satisfaction in the test group aligns with their improved question quality.

To contextualize these insights, we reviewed the qualitative feedback from participants regarding their experience with the task. Four students explicitly stated that they had no difficulties using PMQF, with one highlighting that the taxonomy provided a “good structure and overview, simple and understandable” (S2). However, three students reported challenges in selecting appropriate categories, as they found it difficult to map the given scenario to specific terms and concepts within the taxonomy. One participant noted frustration when certain category selections did not yield any analysis questions, leading them to disengage from the tool. However, they later found it useful for refining their questions. Students in the control group without access to PMQF also faced difficulties. Their feedback underscores the general challenge of formulating analysis questions without guidance, particularly due to a lack of domain knowledge (e.g., S21: “I did not know where to start and how possible questions could look like. Moreover there was unclarity about priority and urgency on what kind of questions to ask.”). We did not observe similar statements from students of the test group.

7 Discussion

In this section, we interpret our findings, review our scientific contributions, and discuss practical implications.

7.1 Interpretation of Results

As a result of our design science project, we introduced two artifacts in this paper, namely a taxonomy of process mining analysis question (cf. Sect. 4) and a question catalog (cf. Sect. 5). Both address existing challenges in structuring, categorizing, and designing meaningful process mining questions.

The taxonomy provides a structured framework for classifying questions, and thus helps to clarify them (cf. Sect. 4.7). Categorization and abstraction have long been recognized as tools to foster deeper understanding and reflection (Croft and Cruse 2004). By enabling the systematic classification of questions, the taxonomy supports shared understanding among stakeholders involved in process mining projects and ensures that analysis questions are aligned with organizational goals. Compared to the only other existing classification schema for process mining questions (Barbieri et al. 2023), our taxonomy reflects a distinct focus on meaning and application. Its iterative design process resulted in six dimensions that abstract the underlying goals and focus of questions.

The question catalog complements the taxonomy by providing a repository of real-world questions and offering insights into prevalent and underexplored areas across the taxonomy dimensions (cf. Sect. 5). For example, our analysis revealed that prediction questions (Dimension 3: Primary Goal) are relatively uncommon, which contrasts with the extensive body of literature on predictive process monitoring (Di Francescomarino et al. 2018). The same holds true for compliance and automation use cases, where questions are less common than we expected. Additionally, we found that questions are distributed differently across categories depending on their source, reflecting the field’s evolution. BPIC questions, with the first edition dating back to 2011, focus largely on transparency and control-flow, whereas more recent sources, such as Celonis, emphasize performance and data perspectives.

The catalog also reveals a lack of clear definitions and shared understanding in process mining analyses. Terms like ‘bottlenecks’, ‘rework’, or ‘workload’ lack commonly accepted definitions, posing challenges for categorization, analysis, and automated processing. Moreover, while methodologies suggest starting projects from questions (van Eck et al. 2015; Gurgen Erdogan and Tarhan 2018), there are no guidelines to formulate them in an understandable and unambiguous way. Our taxonomy advances this understanding and can inform question templates for automated analysis, but further research is needed to define what constitutes a *good* question and how to support its formulation.

Finally, we evaluated our artifacts through four studies, each providing distinct insights into its applicability and usefulness. An initial expert evaluation during development (cf. Sect. 3.3.3) identified areas for refinement, leading to updates that improved the clarity and scope of our taxonomy. After these refinements, we applied the taxonomy to real-world examples and demonstrated its ability to systematically categorize and clarify process mining questions (cf. Sect. 4.7), thereby supporting structured analysis planning. We further validated our approach through two empirical evaluations. In Evaluation II, we found that representatives from organizations with low process mining maturity were able to leverage the taxonomy and question catalog as a source of inspiration for their own domains and use cases. This aligns with prior research highlighting the need for structured guidance in early process mining adoption (Martin et al. 2021; Grisold et al. 2021). However, our findings also suggest that additional support is needed to help users effectively adapt and reformulate questions from the catalog to better fit their specific contexts. Future work should explore mechanisms for providing such guidance, for example, through tailored hints or specific adaptation strategies, some of which we discuss below (cf. Sect. 7.3). In Evaluation III, we examined how students with limited process mining experience used the taxonomy and catalog to generate analysis questions for an artificial process mining scenario. Our results indicate that students with access to the artifacts formulated slightly more relevant and specific questions than those without, reinforcing its value as a structured tool for question design. When comparing the categorizations of the generated questions, we observed that the taxonomy encouraged greater diversity and facilitated the formulation of questions aligned with the scenario’s objectives. This suggests that our artifacts can enhance and accelerate novice users’ ability to frame *meaningful* analysis questions.

7.2 Scientific Contribution

Our primary scientific contribution is the taxonomy that offers a comprehensive overview of process mining questions, grounded in literature and empirical data. By organizing questions along multiple

dimensions, it serves as a reference framework for researchers and practitioners to classify and analyze process mining questions systematically. In line with the role of taxonomies as nascent design theories (Gregor and Hevner 2013; Kundisch et al. 2021), our taxonomy provides a structured representation of questions, offering (preliminary and extendable) knowledge that informs both research and practice. As taxonomies help identify and structure key characteristics of a phenomenon, they lay the groundwork for further theoretical advancements and design-oriented research (Goldkuhl 2004). Even though the taxonomy is built on the current knowledge base, it remains adaptable to future developments in process mining. Given the evolving nature of the field, including advancements in analysis techniques and emerging data standards (e.g., OCED) (Fahland et al. 2024), periodic revisions and validations with newly collected questions may be necessary. This ensures that the taxonomy continues to reflect the state space of process mining analysis questions and remains a relevant tool for guiding project planning.

Additionally, we contribute a structured and carefully collected question catalog consisting of 405 relevant analysis questions that have been posed in the context of scientific or practical process mining projects. This catalog not only provides a concrete resource for practitioners but also supports the systematic exploration of the process mining question space.

These resources not only aid in identifying shared patterns across research efforts but also serve as a benchmark for comparing studies addressing similar questions. As such, they contribute to a conceptual foundation that enables the systematic and reproducible study of analysis questions and offer guidance for developing prescriptive approaches that enhance process mining methodologies, akin to how taxonomies support the identification of design characteristics in other domains (e.g., Schöbel et al. (2020)).

Furthermore, the question catalog can help identify research gaps, highlighting areas where fewer questions have been raised. A lack of questions in specific categories may indicate that certain aspects of process mining are underexplored, suggesting a need for further exploration. However, it could also indicate that these aspects are inherently less relevant or challenging in real-world processes and are therefore less asked. Future research could involve empirical validation or expert feedback to assess whether these gaps reflect genuine opportunities for innovation or align with practitioners' needs.

7.3 Practical Implications

In addition to question classification, we demonstrated that the taxonomy and question catalog are useful for supporting analysts and process owners in designing new analysis questions. To this end, we suggest to use the taxonomy as a filter system and allow users to iteratively select questions and adapt them to their needs and application domains (cf. Sect. 6.1). Our results confirm that this is particularly beneficial for process mining projects with low maturity, where stakeholders may need guidance in identifying relevant questions and generating ideas. In a collaborative workshop that brings together the key stakeholders of a process mining project, new questions can be designed that are well understood by all parties and can be actually addressed with process mining analysis. Therefore, our approach extends existing guidelines for designing questions *during* the analysis (Zerbato et al. 2022) with example-based recommendations on how to design questions *before* the analysis.

In educational settings, the taxonomy provides a structured framework for teaching students to formulate and categorize process mining questions. This can help students grasp the breadth of application

areas and equip them with the skills to align specific techniques with the questions posed. Educators can design exercises based on the taxonomy to test the students’ skills in specific areas. Additionally, the question catalog’s real-world examples expose students to practical questions and challenge them to handle ambiguities from unclear definitions and under-specified questions, enhancing their ability to navigate complex scenarios and develop solutions.

To conclude our discussion, we highlight the potential of our work for tool development. Commercial tools already integrate features that support starting analyses from predefined questions and analysis templates (Ullrich and Lata 2023), and similar approaches have been suggested in academic research (Lashkevich et al. 2023). Our work can enrich these efforts by providing a structured and validated framework for question classification and design, ensuring broader coverage and applicability.

Integrating our taxonomy and question catalog with GenAI techniques, such as large language models (LLMs), presents further, promising opportunities. Our artifacts provide a systematic and comprehensive collection of process mining questions, capturing both their diversity and real-world relevance without generalization. In contrast, LLMs tend to prioritize common concepts and lack such a structured coverage of concepts (Chiang 2023; Teubner et al. 2023), which limits their suitability to be relied upon for question classification and design. However, LLMs can complement our approach, especially when integrated into process mining tools, by assisting in adapting questions to specific domains, an area identified as challenging in Evaluation II. As an example, we tested whether ChatGPT-4o could reformulate the question “Is there evidence that cases are pushed to the 2nd and 3rd service line too soon?” after participants C01 and C02 struggled to map it to their manufacturing setting during Evaluation II. The complete prompt and set-up can be found in the appendix. As a result, the model generated two questions: (1) “Is there evidence that engine components are escalated for advanced quality checks before initial assembly checks are fully conducted?” and (2) “Are assembly tasks being moved to later stages of the assembly line without resolving identified issues in earlier stages?”. These two questions may have been easier for an industry expert to assess for relevance and applicability than the original question. While it requires more research to be confirmed, this small example already highlights the potential of providing concrete examples from the question catalog and showcases how LLMs could support such “domain translations” tasks.

Beyond adaptation, recent research suggests that LLMs could also help translate questions into analysis steps to direct analysis operations (Berti et al. 2023; Kampik et al. 2024). This is especially relevant, as we found that once questions were identified as relevant, users missed guidance on how to approach the analysis (cf. Sect. 6.2. Although this use case is beyond the direct scope of this work, it underscores the value of the taxonomy and question catalog as foundational resources for developing AI-enhanced tools that bridge the gap between question formulation and analysis execution.

8 Threats to Validity

Our study introduces results derived from various methods and data sources. We identify three main threats to the validity of our work: completeness, generalizability, and selection bias.

To ensure that the taxonomy is complete, we collected data from various sources and adhered to the

taxonomy design method from Nickerson et al. (2013). However, the completeness of categories within each dimension is limited to the questions we retrieved, as we refrained from adding categories without empirical evidence. To mitigate this threat, we carefully selected the sources and combined a literature review, a survey to reach practitioners, and input from a major process mining tool vendor. In this way, we could cover a broad, representative range of questions up to 2023 (the year in which we conducted the last collection). The taxonomy’s extensibility allows for updates in the future if necessary.

Another threat is generalizability, as our work includes multiple surveys and a case study. In Question Collection 2, we reached a broad spectrum of respondents from various industries and with different levels of expertise, minimizing the threat. In Evaluation I, we involved experts and observed a high overlap in their comments, suggesting minimal risk. Evaluation II focused on organizations with low process mining maturity, making results particularly valid for this group, though applicability may vary for organizations at different maturity levels. Similarly, in our student-based Evaluation III, findings reflect a controlled academic setting, which may not fully translate to industry contexts. Generalizability is also a concern in the taxonomy design. While we followed the established method by Nickerson et al. (2013) and considered extensive validation as suggested by Kundisch et al. (2021), the interpretation and inclusion of dimensions and categories are subjective. Similarly, several of the ending conditions used to finalize the taxonomy are inherently subjective, meaning other researchers might derive different results even when following the same methodology.

Lastly, selection bias is a potential threat due to the subjective nature of question categorization. Clear definitions for each taxonomy category and added assumptions and clarifications in the question catalog address this. Three authors reviewed and discussed the categorizations iteratively to reach an agreement, which reduces the risk of a subjective bias.

9 Conclusion

Concrete questions are perceived as a starting point (Emamjome et al. 2019) and success factor (Mamudu et al. 2022) for process analysis. However, evidence from practice confirms relevant challenges in identifying and formulating such questions (Zimmermann et al. 2024). To address these challenges, we developed a taxonomy and comprehensive catalog of process mining analysis questions. These contributions offer valuable resources for academics and practitioners, enabling them to (a) gain an overview of the state space of process mining analysis questions, (b) systematically categorize questions and (c) inform the design of new questions. Through our evaluations, we confirmed the relevance and usefulness of our contribution. Our findings show that the taxonomy not only aids in the structured categorization of questions but also facilitates a clearer understanding of the different dimensions involved in process mining. This, in turn, can enhance the quality and precision of process analysis. Future work should focus on expanding the question catalog to include a broader range of domain-specific questions and targeting categories that are currently underrepresented. Further work is also needed to guide the formulation of *good* questions and exploring the capabilities of GenAI for translating questions between domains or deriving guidance for their analysis is a promising direction.

References

- van der Aalst W (2022) Process mining: A 360 degree overview. In: van der Aalst W, Carmona J (eds) *Process mining handbook*, Springer, pp 3–34, 10.1007/978-3-031-08848-3_1
- van der Aalst WM (2018) Process discovery from event data: Relating models and logs through abstractions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(3)
- Acampora G, Vitiello A, Di Stefano B, van der Aalst W, Günther C, Verbeek E (2017) Ieee 1849: The xes standard. *IEEE Computational Intelligence Magazine* pp 4–8
- Ailenei I, Rozinat A, Eckert A, van der Aalst W (2012) Definition and validation of process mining use cases. In: *Business process management workshops*, Springer, pp 75–86, 10.1007/978-3-642-28108-2_7
- Amar R, Eagan J, Stasko J (2005) Low-level components of analytic activity in information visualization. In: *Ieee symposium on information visualization*, pp 111–117
- Andrienko N, Andrienko G (2006) *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media
- Antoniou C, Bassiliades N (2022) A survey on semantic question answering systems. *The Knowledge Engineering Review* 37(2):1–42, 10.1017/S0269888921000138
- Augusto A, Conforti R, Dumas M, Rosa ML, Maggi FM, Marrella A, Mecella M, Soo A (2019) Automated discovery of process models from event logs: Review and benchmark. *IEEE Transactions on Knowledge and Data Engineering* 31(4):686–705, 10.1109/TKDE.2018.2841877
- Badakhshan P, Wurm B, Grisold T, Geyer-Klingenberg J, Mendling J, Vom Brocke J (2022) Creating business value with process mining. *The Journal of Strategic Information Systems* 31(4):101,745
- Bailey KD (1994) *Typologies and taxonomies: An introduction to classification techniques*. Sage
- Barbieri L, Madeira E, Stroeh K, van der Aalst W (2023) A natural language querying interface for process mining. *Journal of Intelligent Information Systems* 61(1):113–142
- Basili VR (1992) *Software modeling and measurement: the Goal/Question/Metric paradigm*. University of Maryland at College Park
- Berti A, Schuster D, van der Aalst WM (2023) Abstractions, scenarios, and prompt definitions for process mining with llms: a case study. In: *Int. conf. on business process management*, Springer, pp 427–439
- Bozkaya M, Gabriels J, Van Der Werf J (2009) Process diagnostics: A method based on process mining. In: *Int. conf. on information, process, and knowledge management*, 10.1109/eKNOW.2009.29
- Braun V, Clarke V, Boulton E, Davey L, McEvoy C (2021) The online survey as a qualitative research tool. *International Journal of Social Research Methodology* 24(6):641–654, 10.1080/13645579.2020.1805550
- Brehmer M, Munzner T (2013) A multi-level typology of abstract visualization tasks. *IEEE Trans Visual Comput Graphics* 19(12):2376–2385
- Campbell DT, Stanley JC (2015) *Experimental and quasi-experimental designs for research*. Ravenio books
- Capitán-Agudo C, Salas-Urbano M, Cabanillas C, Resinas M (2022) Analyzing how process mining reports answer time performance questions. In: *Int. conf. on business process management*, Springer
- Charmaz K (2006) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage

- Chiang T (2023) Chatgpt is a blurry jpeg of the web. <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>, accessed: 12 July 2024
- Croft W, Cruse DA (2004) Categories, concepts and meanings. In: Cognitive linguistics, Cambridge University Press, pp 74–106
- Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* pp 319–340
- Di Francescomarino C, Ghidini C, Maggi FM, Milani F (2018) Predictive process monitoring methods: Which one suits me best? In: *Int. conf. on business process management*, Springer, pp 462–479
- Doty DH, Glick WH (1994) Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of management review* 19(2):230–251
- Dumas M, La Rosa M, Mendling J, Reijers H (2018) *Fundamentals of Business Process Management*. Springer, 10.1007/978-3-662-56509-4
- van Eck M, Lu X, Leemans S, van der Aalst W (2015) Pm2: A process mining project methodology. In: *Int. conf. on advanced information systems engineering*, Springer, 10.1007/978-3-319-19069-3_19
- Eggers J, Häge MC, Zimmermann SK, Gewald H (2023) Assessing process mining use cases: A taxonomy of antecedents and value potentials of process mining. In: *Americas conference on information systems*
- Emamjome F, Andrews R, ter Hofstede A (2019) A case study lens on process mining in practice. In: *Otm confederated int. conf. 'on the move to meaningful internet systems'*, Springer, 10.1007/978-3-030-33246-4_8
- Fahland D, Montali M, Leberherz J, van der Aalst WM, van Asseldonk M, Blank P, Bosmans L, Brenscheidt M, di Ciccio C, Delgado A, et al. (2024) Towards a simple and extensible standard for object-centric event data (oced)–core model, design space, and lessons learned. *arXiv preprint arXiv:241014495*
- Goldkuhl G (2004) Design theories in information systems - a need for multi-grounding. *Journal of Information Technology Theory and Application (JITTA)* 6(2)
- Graafmans T, Turetken O, Poppelaars H, Fahland D (2021) Process mining for six sigma. *Bus Inf Syst Eng* 63(3):277–300, 10.1007/s12599-020-00649-w
- Gregor S (2006) The nature of theory in information systems. *MIS quarterly* pp 611–642
- Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS quarterly* pp 337–355
- Grisold T, Mendling J, Otto M, vom Brocke J (2021) Adoption, use and management of process mining in practice. *Business Process Management Journal* 27(2):369–387
- Gurgen Erdogan T, Tarhan A (2018) A goal-driven evaluation method based on process mining for healthcare processes. *Applied Sciences* 8(6):894, 10.3390/app8060894
- Hrach C, Alt R, Sackmann S (2023) Configuration approach of user requirements for analytical applications - challenges, state of the art and evaluation. In: *Information technology for management: Approaches to improving business and society*, Springer, pp 3–22, 10.1007/978-3-031-29570-6_1
- Kampik T, Warmuth C, Rebmann A, Agam R, et al. (2024) Large process models: A vision for business process management in the age of generative ai. *Künstliche Intelligenz* 10.1007/s13218-024-00863-8

- Kandogan E, Balakrishnan A, Haber EM, Pierce JS (2014) From data to insight: work practices of analysts in the enterprise. *IEEE computer graphics and applications* 34(5):42–50
- Kerremans M, Sugden D, Duffy N (2024) Magic quadrant for process mining platforms. Tech. rep., Gartner, Inc., <https://www.gartner.com/doc/reprints?id=1-2HGG0P7J&ct=240502&st=sb>, accessed: 09 July 2024
- Klinkmüller C, Müller R, Weber I (2019) Mining process mining practices: An exploratory characterization of information needs in process analytics. In: *Int. conf. on business process management*, Springer, 10.1007/978-3-030-26619-6_21
- Kundisch D, Muntermann J, Oberländer AM, Rau D, Röglinger M, Schoormann T, Szopinski D (2021) An update for taxonomy designers: methodological guidance from information systems research. *Bus Inf Syst Eng* 64(4):421–439
- Lashkevich K, Milani F, Danylyshyn N (2023) Analysis templates for identifying improvement opportunities with process mining. In: *European conf. on information systems*
- Lawrence SR, Buss AH, et al. (1995) Economic analysis of production bottlenecks. *Mathematical Problems in Engineering* 1:341–363
- Liao QV, Gruen D, Miller S (2020) Questioning the ai: Informing design practices for explainable ai user experiences. In: *Chi conf. on human factors in computing systems*, 10.1145/3313831.3376590
- van Looy A, Shafagatova A (2016) Business process performance measurement: a structured literature review of indicators, measures and metrics. *SpringerPlus* 5(1):1797, 10.1186/s40064-016-3498-1
- Maaradji A, Dumas M, La Rosa M, Ostovar A (2017) Detecting sudden and gradual drifts in business processes from execution traces. *IEEE Trans Knowl Data Eng* 29(10):2140–2154
- Mamudu A, Bandara W, Wynn MT, Leemans SJ (2022) A process mining success factors model. In: *Int. conf. on business process management*, Springer, pp 143–160
- Mannhardt F (2018) Multi-perspective process mining. PhD thesis, Mathematics and Computer Science
- Mans RS, van der Aalst WMP, Vanwersch RJB, Moleman AJ (2013) Process mining in healthcare: Data challenges when answering frequently posed questions. In: *Process support and knowledge representation in health care*, Springer, pp 140–153, 10.1007/978-3-642-36438-9_10
- Marcus L, Schmid SJ, Friedrich F, Röglinger M, Grindemann P (2024) Navigating the landscape of organizational process mining setups: A taxonomy approach. *Bus Inf Syst Eng* 10.1007/s12599-024-00908-0
- Martin N, Fischer D, Kerpedzhiev G, Goel K, Leemans S, Roglinger M, van der Aalst W, Dumas M, La Rosa M, Wynn M (2021) Opportunities and challenges for process mining in organizations: Results of a delphi study. *Bus Inf Syst Eng* 63(5):511–527
- Milani F, Lashkevich K, Maggi F, Di Francescomarino C (2022) Process Mining: A Guide for Practitioners. In: *Research challenges in information science*, Springer, 10.1007/978-3-031-05760-1_16
- Nalchigar S, Yu E (2018) Business-driven data analytics: A conceptual modeling framework. *Data & Knowledge Engineering* 117:359–372, <https://doi.org/10.1016/j.datak.2018.04.006>
- Nickerson R, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. *Eur J Inf Syst* 22(3):336–359, 10.1057/ejis.2012.26
- Oberländer AM, Lösser B, Rau D (2019) Taxonomy research in information systems: A systematic assessment. In: *European conf. on information systems*

- Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. *J Manag Inf Syst* 24(3):45–77
- Pomerantz J (2005) A linguistic analysis of question taxonomies. *Journal of the American Society for Information Science and Technology* 56(7):715–728, 10.1002/asi.20162
- Reichert M, Weber B (2012) Process-aware information systems. In: *Enabling flexibility in process-aware information systems*, Springer, pp 9–42
- Reinig BA (2003) Toward an understanding of satisfaction with the process and outcomes of teamwork. *J Manag Inf Syst* 19(4):65–83
- Rojas E, Sepulveda M, Munoz-Gama J, Capurro D, Traver V, Fernandez-Llatas C (2017) Question-driven methodology for analyzing emergency room processes using process mining. *Applied Sciences* 7(3):302, 10.3390/app7030302
- Schöbel SM, Janson A, Söllner M (2020) Capturing the complexity of gamification elements: a holistic approach for analysing existing and deriving novel gamification designs. *Eur J Inf Syst* 29(6):641–668
- Stein Dani V, Leopold H, van der Werf JME, Reijers HA (2023) Progressing from process mining insights to process improvement: challenges and recommendations. In: *Int. conf. on enterprise design, operations, and computing*, Springer, pp 152–168
- Teubner T, Flath CM, Weinhardt C, van der Aalst W, Hinz O (2023) Welcome to the era of chatgpt et al. the prospects of large language models. *Bus Inf Syst Eng* 65(2):95–101
- Thoring K, Mueller R, Badke-Schaub P (2020) Workshops as a research method: Guidelines for designing and evaluating artifacts through workshops. In: *Hawaii int. conf. on system sciences*, 10.24251/HICSS.2020.620
- Ullrich C, Lata T (2023) Business Miner: Process Mining Insights for Business Users. In: *ICPM Doctoral Consortium and Demo Track, CEUR-WS.org, CEUR Workshop Proceedings*, vol 3648
- Van Der Aalst WM (2011) Process mining: discovering and improving spaghetti and lasagna processes. In: *2011 ieee symposium on computational intelligence and data mining*, pp 1–7
- Vom Brocke J, Hevner A, Maedche A (2020) Introduction to design science research. *Design science research Cases* pp 1–13
- Wang D, Liao QV, Zhang Y, Khurana U, Samulowitz H, Park S, Muller M, Amini L (2021) How much automation does a data scientist want? *arXiv preprint arXiv:210103970*
- Yin RK (2018) *Case study research and applications*, vol 6. Sage Thousand Oaks, CA
- Zerbato F, Koorn J, Beerepoot I, Weber B, Reijers H (2022) On the origin of questions in process mining projects. In: *Int. conf. on enterprise design, operations, and computing*, Springer, 10.1007/978-3-031-17604-3_10
- Zimmermann L (2024) The process mining question forge. In: *Proc of the best dissertation award, doctoral consortium, and demonstration and resources forum at bpm 2024*
- Zimmermann L, Zerbato F, Weber B (2024) What makes life for process mining analysts difficult? a reflection of challenges. *Softw Syst Model* 23(6):1345–1373, 10.1007/s10270-023-01134-0