Structuring Business Process Context Information for Process Monitoring and Prediction

Jens Brunk
University of Muenster - ERCIS
Leonardo-Campus 3, 48149 Münster, Germany
jens.brunk@ercis.uni-muenster.de

***Abstract*—The advance of Big Data, the Internet of Things (IoT) and with it the integration of various systems - generally referred to as digitalization - provides huge amounts of data that can be leveraged by modern Business Process Management (BPM) methods. Predictive Process Monitoring (PPM) represents a novel branch of process mining, which deals with real-time analysis of currently running process instances and also with the prediction of its future behavior. Most of the early PPM techniques base their analyzes and predictions solely on the control-flow characteristic of a business process, i.e. the process events. Recently, researchers are attempting to incorporate additional process-related information, also known as the process context, into their predictive models. To use the available context information to full capacity, we require an understanding of the concept of business process context information. Based on both empirical and conceptual sources, we develop a taxonomy that provides a comprehensive overview of the characteristics of business process context information. This overview can then be leveraged, e.g. in process monitoring and prediction. Our taxonomy adds descriptive knowledge to the field of BPM, specifically PPM, and strengthens the conceptual foundation of context-sensitive process monitoring and prediction.**

***Index Terms*—Taxonomy, Business Process, Context, Context-Sensitivity, Predictive Process Monitoring.**

I. Introduction

Business Process Management (BPM) has come a long way as an academic discipline [1] and by now established itself as good practice in various industries. The overarching goal of BPM is to manage business processes. Ideally, organizations wish to manage these processes proactively instead of fixing the non-ideal process setups or outcomes in a reactive manner. Process mining has emerged as a popular BPM technology, which allows process analysts to discover, check for conformance and ideally improve existing business processes [2]. The utilization of process mining represents reactive management of business processes since the information that is analyzed are event logs of historic business processes instances. A novel branch of process mining goes one step further and deals with real-time analysis of currently running process instances and also with the prediction of its future behavior. This branch is often referred to as Predictive Process Monitoring (PPM) and represents a proactive approach to BPM [3]. A number of PPM techniques have been developed that address a variety of goals and apply several different technologies [4]. In its simplest form, PPM techniques try to learn a technological representation of a business process from historic event logs of the log producing information systems (IS). This step often resembles classical process mining techniques. In the next step, the techniques analyze a currently running process step and the events that this instance of the process has already passed through (monitoring). Then, the goal is to assess or predict - given the knowledge about historic process runs - the next activity, remaining processing time, pre-determined risks, etc.

Most of the (early) PPM techniques base their analyzes and predictions solely on the control-flow characteristic of a business process, i.e. the process events. Since then, researchers continuously attempt to conceptualize and incorporate additional process-related information, also known as the process context, into their predictive models [5]. From our point of view, this makes a lot of sense. For example, if we take the example of a credit loan application process, one would assume that the loan amount, the customer's risk rating and even the employee who processes the application all have significant impacts on how the application is further processed. These context information can either directly stem from the event log of the IS or can even be gathered from external sources like sensors [6] or news [7].

The advance of Big Data, the Internet of Things (IoT) and with it the integration of various systems - generally referred to as digitalization - provides huge amounts of context information that can be leveraged for so-called context-sensitive PPM. Along with the first available PPM techniques, which support context, we have learned that a big, if not the biggest, challenge in context-sensitive PPM is to characterize context attributes, identify the correct ones and process them the right way (e.g. [8]). Simply using as many context attributes as possible will most often not improve predictive quality, similar to using the wrong ones.

As context-sensitivity is a novel concept in PPM, we do not yet have an accurate understanding of what context actually means, which types of values a context attribute can consist of or which implications different characteristics have on the technical application. This poses a problem if we want to research context-sensitivity in a structured way. Therefore, the goal of our research is to provide this structure through a taxonomy on business process context information for process monitoring and prediction. For the taxonomy development, we analyze the current body of knowledge regarding business process context, the available context-sensitive PPM methods,

2378-1971/20/$31.00 ©2020 IEEE
DOI 10.1109/CBI49978.2020.00012

and datasets, which are applied within the domain.

The taxonomy should serve researchers in the field of PPM to select the right context information, identify gaps in tool-support for context-sensitive PPM and create a common theoretical understanding of context information. Additionally, practitioners that consider the application of PPM in their business can apply the taxonomy to identify relevant context information to their business processes in preparation for their PPM endeavor.

The remainder of this paper is structured as follows. Section II gives a short overview about the research background of context-sensitive prediction of business processes, as well as taxonomy research. Section III describes the applied research method. The results of the taxonomy development process as well as the taxonomy itself are presented in sections IV to VI. Finally, section VII concludes the results and gives an outlook on future research steps.

II. Research Background

A. Context-Sensitive Prediction of Business Processes

The BPM research field concerns the discovery, analysis, re-design, and monitoring of business processes with the ultimate goal of improving them [9]. Historically, BPM research had a focus on modeling the to-be processes that depict the ideal process flow. However, nowadays the availability of big data and large volumes of event logs from information systems, which support business processes, has given rise to a focus on analytics of the as-is process. Given this development, process mining emerged as a research field. During process mining, specialized data mining algorithms are applied to event log data, which are data logs that contain the history of a process instance, in order to identify trends, patterns and details of the underlying business process. These logs are then used for discovery, conformance checking and enhancement of these business processes [2].

Concurrent with higher availability of computing resources and power, the activity in predictive analytics and machine learning has increased and PPM evolved based on traditional process mining [3]. As mentioned above, PPM deals with a proactive approach to BPM and tries to predict future process behavior to minimize process risks and improve process performance [10]. For this purpose, PPM techniques deliver insights into the future activity of a running process instance [4]. The research field of PPM is very active. For example, [11] recently presented an empirical comparison of several classification techniques for next event prediction and [12] introduced a process prediction reference process, which describes the basic steps to be followed in a PPM endeavor (see Fig. 1). Following this reference process, an exemplary process prediction objective could be the next activity of the loan application procedure of a financial institution (Step 1). This means that the value of the activity attribute of, e.g., the upcoming time step is to be predicted (Step 2 & 3). Once one of the available next activity prediction methods has been selected (Step 4), e.g. based on [4], the input data for the prediction method is gathered. This is where context information and context-sensitive prediction methods come into play. Historically, most existing methods only based their predictions on the control-flow information (i.e. the activity flow) of a business process to determine the next activity. In this case, it would include the way of first contact, the application document and information gathering steps, etc.

However, more and more developed PPM techniques are now also able to incorporate context information (e.g. [13], [14]), which is also referred to as context-sensitivity. The context of business processes is important because it can have a great influence on the process outcome and can add relevant information to the process. In a business process setting, the context can be defined as the minimum set of variables that contains all the important information that impacts their design, implementation and execution [15]. In our example of a loan application, the type of the loan application, i.e. the extension of an existing credit line or a new application, or later in the process the difference between requested loan amount by the applicant and the offered amount by the financial institution will have a strong influence on the future activity flow of the business process. These different input data are relevant in Step 4 and 5 of the reference process, before the selected prediction method is executed (Step 6).

Some existing works on context-sensitive PPM methods consider a priori context information (e.g. [8]), which means that the information is known before the execution and is expected to stay constant throughout the process instance. Others also include dynamic context attributes [13], [14], which may change throughout a process runtime. However, context information comes in various forms and sizes, as our above example already suggests, and more research towards context-sensitivity in PPM and its various dimensions is necessary.

B. Taxonomy Research

A taxonomy is a form of classification of objects and is often used analogously to the terms typology and framework [16]. To avoid confusion, essential terms for grouping objects are defined and delimited. The term classification is used to describe both the system or the process of organizing objects and the organization of objects according to a particular system. In the following, the term classification system is used for the abstract groupings or categories into which objects can be classified. The definition of a framework is closest to that of a classification system and can be used as a synonym. The term classification is used for the concrete result of classifying objects into groups or categories. The term typology is usually limited to a system of conceptually derived groupings. In addition, typologies are usually multidimensional and often more complex than classification systems. In part of the literature, taxonomy is limited to empirically derived groupings (through cluster analyses or similar). Most literature, however, uses taxonomies for grouping systems that are derived conceptually and/or empirically. As with classification, the taxonomy is sometimes used for the system or process and sometimes for the application of the system. In the remainder of this work, we follow the

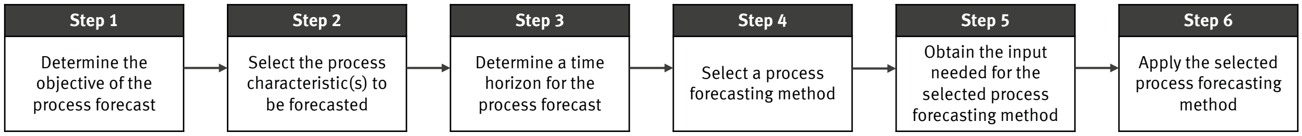

Fig. 1. Process Prediction Reference Process [12].

common practice of using the term taxonomy for both and the respective application context makes clear which is referred to [16]. Formally, [16] defined a taxonomy as follows:

$$T = \{D_i, i = 1, ...n | D_i = \{C_{ij}, j = 1, ..., k_i; k_i \geq 2\}\}$$

A taxonomy T consists of a set of n dimensions $D_i(i = 1, ..., n)$. Each dimension consists of $k_i(k_i \geq 2)$ mutually exclusive and all-embracing characteristics $C_{ij}(j = 1, ..., k_i)$. Thus, each viewing object has exactly one C_{ij} for each D_i.

III. Research Method

To reach our above-mentioned goal we apply the taxonomy development procedure by [16]. The systematic and rigor method has already been widely applied in the field of IS and has proven to be a useful tool to structure and explain novel concepts (e.g. [17]–[19]). The academic literature (conceptual-to-empirical), as well as empirical data (empirical-to-conceptual), can be investigated with this method through a number of iterations towards the creation of a taxonomy. In the beginning, the method requires one to define the purpose and goal of the taxonomy and a set of objective and subjective ending conditions that define at what point the iterative development process ends [16].

The aim of the taxonomy is to provide guidance and structure for describing, analyzing and implementing context information in context-sensitive PPM. All dimensions and characteristics have to be derived from this purpose to manifest the structural differences of context information in business processes.

We adopt three subjective and five objective ending conditions of [16]. The objective ending conditions include: All objects or a representative sample of objects have been examined, every dimension is unique and not repeated, every characteristic is unique within its dimension, each cell (combination of characteristics) is unique and is not repeated and at least one object is classified under every characteristic of every dimension. The subjective ending conditions include that the taxonomy should be concise, robust and extendable.

For the taxonomy development we executed four iterations in total. The first two iterations represent extensive literature analyses (conceptual-to-empirical) based on [20]. For both literature reviews, the databases Springerlink, Scopus, IWEE Xplore, ACM Digital Library, JSTOR and Web of Science were used and the results were limited to publications since 2000, as the domain of PPM has only evolved since then. The first literature review focused on published PPM techniques that use context information for their predictions. Following this goal, the above mentioned databases were queried for the key words *context, predictive, prediction, process mining, process monitoring* or *business*. After removing duplicates, 510 publications remained, from which 51 were selected based on the title and finally 24 remained after considering the abstracts. Five additional works were identified through a backward search. The results of the first literature review enabled us to understand the current status-quo in the domain and to see which kinds of context information have already been leveraged in which way. In the second iteration, existing context information classifications from the literature of business process management were considered. Here, the databases were queried for *contextualisation, contextualization, business* or *process*. After duplicates, we identified 614 publications from which we selected 45 based on the title and finally 20 remained after the consideration of the abstracts. These works extended our implementation-oriented understanding of the first iteration with a theoretical perspective.

The third and fourth iteration followed the empirical-to-conceptual approach and had the goal to identify a subset of objects (here: context information of business processes) with which the researcher is most familiar or which are most easily accessible [16]. We decided to use the datasets of the Business Process Intelligence (BPI) Challenges for these iterations. The BPI Challenge datasets form the basis of the evaluation of many process mining and prediction methods and are among the most comprehensive process logs that are freely available. We used the dataset from the year 2019 as a starting point (third iteration) and then iterated through suitable previous years, until all the ending criteria were reached. This was the case in the fourth iteration, where the dataset from year 2017 was investigated. Each data set went through the following steps:

1) Identify context information from the process log: Which attribute in the process log is context information?
2) Identify other relevant context: Which information outside of the process log influences the process?
3) Classification and evaluation: How can the context information be classified in the taxonomy?
4) Conception: Is there a need to add or remove dimensions/characteristics?

Table I gives an overview of the four taxonomy development iterations. Section IV summarizes the related work on context information in the BPM and PPM literature. Section V presents the final state of the taxonomy, its dimensions, and properties. In section VI we show how the taxonomy can be applied in the context of PPM to attributes of a real-world

TABLE I
TAXONOMY DEVELOPMENT ITERATIONS

#	Iteration
1	**Conceptual-to-empirical:** The first iteration uses the conceptual-to-empirical approach, since the domain is still an emerging field [16]. In this iteration, we build on a literature review on context-sensitive PPM methods (see section IV).
2	**Conceptual-to-empirical:** Since the results of the first iteration are technique-oriented, we additionally review existing classifications and structures of context information in the BPM and PMM literature (see section IV).
3	**Empirical-to-conceptual:** After the two literature-based iterations, the first empirical-to-conceptual iteration examines the developed taxonomy against actual process logs. Therefore, we analyze context information of the BPI Challenge 2019 dataset within this iteration.
4	**Empirical-to-conceptual:** In a second empirical-to-conceptual iteration, we examine the taxonomy against the process logs of the BPI Challenge 2017 dataset. Within this iteration we meet the ending conditions.

dataset, in this case the BPI Challenge 2017 and 2019 datasets from iteration three and four.

IV. CHARACTERISTICS OF BUSINESS PROCESS CONTEXT INFORMATION IN THE LITERATURE

This section summarizes the findings from the literature reviews of the first two taxonomy development iterations. The section is structured in dimensions of context information, as they will be relevant for the final taxonomy description.

Time. In existing PPM techniques, researchers differentiate between two types of time dimensions of context information. The first type of context information refers to those that are known at the beginning of a process. This is referred to as a priori information [21]. The second characteristic concerns the context information, which can be collected and processed during the process runtime [22]. The idea behind the use of a priori context information is that the performance characteristics of a specific activity or an entire process instance depend strongly on the application context. For example, previous tasks, limited resource availability, or even the weather can have a large influence on the process flow [23]. There are different methods to process a priori information. Some researchers [24]–[28] use clustering and regression analysis methods. [22], [26], [29], [30] not only use a priori context information, but also those that occur during the process runtime. The exclusive use of a priori context information implies that a certain combination of context information reflects the process behavior in such a way that each combination can be assigned to a scenario that is characterized by homogeneous properties. However, if context information occurs or changes during runtime, this data can be used to determine the scenario to which the currently running process instance belongs [26]. Additionally, if a suitable prediction model for context information is available, future context information can also be included in the process prediction. Currently, there are only a few approaches that take predicted context information into account. [7] have developed a method to analyze the moods in online media and on the basis of this forecast the mood in the population, which can have an impact on certain (business) processes. For example, in Brazil, the occurrence of an unknown disease, albeit with symptoms similar to a harmless cold, had triggered an alarm in the population. As a result, hospitals were overcrowded and the subsequent case processing of health insurance companies was delayed. Monitoring these kinds of developments and assessing the impact on the process provides an opportunity to respond to these trends as early as possible [7]. Another good example of future context information, which could also provide useful information for a predictive model, are weather forecasts.

Structure. Context information of business processes can be accessible in a structured format, e.g. an executing resource of an activity; in an unstructured format, e.g. text messages exchanged between the executing resources during the process; or in a semi-structured format, e.g. seasonal changes or a priori knowledge for a particular process instance about external events. The structure of the context information has an influence on how it can be processed, or whether preprocessing may be necessary [7]. Semi-structured and unstructured context information usually require enrichment or preprocessing before it can be assigned to participating process entities (process instance, event, activity, etc.) [22]. Most context-sensitive prediction methods are based on a structured context [7]. Both clustering and regression methods, as well as prediction methods based on hypothesis tests, require structured data. Structured context information is characterized by the fact that it has a fixed data schema [7]. Here, the scale level of the characteristics, whether scaled nominal, ordinal or metric and discrete or continuous, can have an influence on the applicability within a certain method. Semi-structured context information is data that itself does not have a fixed structure but contains hidden, implicit structural information. Firstly, they offer the possibility to use data sources, such as the web, which cannot be directly described by a data schema, and secondly, they offer a flexible format (e.g. XML or JSON) for the use of data. In addition, it can be helpful even when dealing with structured data, for example when merging data objects, to treat them as semi-structured context information [31]. If the implicit structure differs from case to case, the prediction model would have to be re-trained for each case, which affects the scalability of the method. [21] have presented a method for this problem, using techniques from the domain of Natural Language Processing, to generate knowledge about the structure and thus use semi-structured (unexpected but known) a priori context. Additionally, [28] present an approach to consider seasonal trends. In practice, structured data is often associated with unstructured (textual) data, such as e-mails or comments. For example, in a loan entry process, a free text description can take up the purpose of the loan or a loan officer can add comments to the loan application after a personal meeting [32]. If many generic operations are recorded at a low level, they can only be interpreted as meaningful business activities in the context in which they were applied [33]. The framework by [32] combines text mining techniques to extract features from texts with classification techniques for structured data. In addition, [7] argue that today due to the wide distribution of information, the mood in the news can provide relevant contextual information for a business process. They have therefore developed a method to enrich the process log with the mood of news. [26] have discussed an algorithm for learning decision trees that aim to capture the relationships between structured process variants and unstructured process characteristics.

Origin. Context information can stem from different sources. Most often, the existing context information in the event log is used. So far, only a few methods integrate external sources. Also, the use of context information from the event log also differs. [34], [35] only use process-instance context, such as customer status (premium customer, normal customer, ...) or cumulative information (number of running process instances). Similarly, [24], [25], [28], [36], [37] also consider process instance-related and cumulative information. External contexts can also be used in a variety of ways. [38] use information sources, such as enterprise social media, financial market data or weather reports. The aim is to establish a correlation between the timestamps of the information sources and the process instance in order to derive a wealth of information about the context by executing the process instance. [22] particularly use external information from the IoT in relation to the transport process of critical goods. [7] use online intelligence services as an external source. In some approaches, such as those of [35], [39], which so far only use internal context information from the process log, an extension to also include external context, such as weather or strike, is considered for future work.

[15] distinguish between four categories with regard to the origin of contextual information. Direct context information, without which the execution of a process instance would not be possible; internal context information within an organization; external context information about the business network and environmental context information outside the business network. The immediate context is embedded in the internal context, which is surrounded by the external context. The external context is in turn embedded in the environment. Between the elements on the different levels there can be interrelations with other elements on the same or the further inward layer. The immediate context of a business process includes those elements that go beyond the constructs that represent the pure control flow and include those elements that directly facilitate the execution of a process. Since these elements often play a central role, they are already well-considered in existing business process modeling techniques. These elements are typically essential for understanding and executing a business process. Examples of immediate contextual information include the data needed for the process; the organizational resource responsible for an activity; or the application software that supports a process step. The internal context contains information about the internal environment of an organization. The business process, including the immediate context, is embedded in an organization. Various elements have an indirect influence on the business process. This category includes, for example, the corporate strategy and the associated process objectives. In collaborative business processes, the internal context includes all other organizations involved. The environmental context is outside the business network in which the company is embedded. These environmental variables can be described by the categories society, nature, technology and economy. These include factors such as weather (e.g. increasing call volumes during the storm season), time (e.g. different business models on Sundays or before Christmas), or employee-related factors (e.g. shortages or strikes). While some of this environmental context information is dynamic and can have a major impact on the business process (e.g. weather conditions), much of this information can remain constant over time (e.g. availability of natural resources in a country).

Relevance. There are different ways to identify context information or to classify already identified context information and their relevance for the business process. Data-driven methods, such as data mining approaches, can be used to determine the quantitative influence of a context attribute on the business process [40]. In addition, technical experts can contribute their expertise to identify and classify contextual information. Both methods can also be combined [41]. Two concrete techniques to identify relevant context information are the ORGANON and the BPCREL method [40]. The ORGANON method aims to support the identification of context information in business processes through the analysis of process models. The method comprises two main phases: (i) identify the key activities of the process and (ii) analyze the impact of their attributes on a business process objective [40]. The BPCREL method supports the identification of external and environmental context information through a data-driven approach that also draws on technical experts.

Relevance represents the importance level of a contextual element in relation to the prediction goal. Relevance can be low, medium, or high [5]. Highly relevant context information is characterized by the fact that it influences the structure of the process model, for example, the control flow, the organizational resources involved or the data required. Context information of low relevance has only a minor influence on the process flow. The consideration of such context information only makes sense if an even more accurate prediction is desired. The greater the amount of useful context information that is considered, the more accurately a situation can be predicted [5]. However, the inclusion of an additional feature to a predictive model always caters to the risk of over-complicating the model.

Process relation. Another important aspect to consider is the relation of a context attribute to the business process. Process modeling methods usually try to model a system by describing different activities that can be performed. Here, the activity-related context includes elements, such as the executing or responsible resource, safety regulations, quality guidelines or environmental context variables. An event is something that happens in the course of a business process. The event-related context can also be modeled in relation to different context attributes to better respond to such events. The control flow determines the branching, splitting, merging or linking of process paths. The control flow-related context includes the information that influences the flow control as well as the scope (e.g. number of process instances) [42]. To design detailed business processes, some additional artifacts (data, exchanged messages, etc.) must be taken into account during the modeling of the entire process or individual process components [33]. For example, the artifact-related context contains a series of information to characterize data objects that are exchanged between business activities. This can also include a resource that is involved in the creation or manipulation of a data object [42]. In order to formalize the relation to the business process, it makes sense to connect the business process with the context information by means of semantics [43]. [5] formalized a multi-layered context-aware meta-model, which facilitates the maintenance and development of such a business model.

V. Taxonomy of Business Process Context Information

The taxonomy presented in Tab. II covers the most important dimensions of business process context information for PPM, which we identified throughout the four iterations (Tab. I) of the taxonomy development process. In the following, we briefly describe the dimensions and their characteristics.

Time - Point in time at which the context information is known.

- *A priori*. Context information is known at the start of the process instance execution.
- *Runtime*. Context information that emerges during the runtime of a process instance execution.
- *Future*. Context information, which lies in the future and is predicted before its availability.

Structure - Data model of the context information.

- *Structured*. The context information has a predefined, static data schema.
- *Semi-structured*. The context information has an implicit, possibly also dynamic data schema.
- *Unstructured*. The context information does not have a structure.

Origin - Source of the context information [15].

- *Immediate*. Context information that is directly relevant for the execution of a process instance. Without changes to the internal, external, and environmental context, the immediate context is sufficient to perform a business process.
- *Internal*. Context information that describes the surroundings of the business process within an organization.
- *External*. The external context includes elements whose design and behavior are outside the control of the organization. Nevertheless, these elements are still within the business network in which the organization operates.
- *Environment*. The environmental context lies outside of the business network in which the organization is embeded. Exemplary environmental aspects include society, nature, technology and the economy.

Relevance - Importance of the context information to the business process.

- *High*. Highly relevant context information is characterized by the fact that it influences the structure of the process model. Especially all context information, which determines the success of the process objectives, falls into this category.
- *Middle*. Context information that has an influence on the business process, which is neither classified as high nor as low.
- *Low*. Context information of low relevance has only a small influence on process success and performance. The consideration of such context information only makes sense if an even more accurate prediction is desired and achievable.

Process relation - The object of a process to which the context is related to [42].

- *Activity*. The context influences one or multiple activities of a business process.
- *Event*. The context influences one or more events.
- *Control flow*. The context influences the process flow or extent of a process instance.
- *Artefact*. The context influences an artefact (resources, data, etc.).

Runtime behavior - Whether the context information changes throughout a process instance execution or not.

- *Static*. Static context information does not change (or change only in a not relevant extent) during process execution.
- *Dynamic*. Dynamic context information can change their value one or multiple times during process execution.

VI. Application of the Business Process Context Information Taxonomy

In the third and fourth iteration of the taxonomy development process, we examined the taxonomy against real-world process logs. In the following, we illustrate this process and show how

TABLE II
TAXONOMY OF BUSINESS PROCESS CONTEXT INFORMATION

Dimension	Values			
Time	A priori	Runtime	Future	
Structure	Structured	Semi-structured	Unstructured	
Origin	Immediate	Internal	External	Environment
Relevance	High	Middle	Low	
Process relation	Activity	Event	Control flow	Artefact
Runtime behavior	Static	Dynamic		

the final version of the taxonomy can be used to classify a selection of context attributes from the BPI Challenge 2017 and 2019 datasets.

A. BPI Challenge 2019

The data set of the BPI Challenge 2019 [44] refers to the Purchase-to-Pay process from the procurement of a multinational corporation with more than 60 subsidiaries in the field of coatings and paints. Procurement processes are central to the value chain of companies. The execution of these processes includes business risks, such as the risk of long delivery times, diminishing production efficiency, rising costs, or the risk of potential fraud [45]. The process log contains 1,595,923 recorded events, which belong to 251,734 process instances. The events refer to 42 activities, executed by 627 users, including 607 human users and 20 batch users. A process instance refers to a single purchase order item.

TABLE III
SELECTED BPI 2019 CONTEXT ATTRIBUTES

Context	Potential Influence on the Business Process
Vendor	Different delivery times (e.g. due to different distances).
Document type	Certain document types require additional approval.
Item category	Has a major influence on the process flow (3-way with GR-based invoicing, 3-way-without, etc.).
Cumulative net worth	Larger amounts may require a higher inspection effort at goods receipt.

Table III shows an exemplary selection of context information from the BPI 2019 dataset, which potentially has an impact on the business process. Due to paper length restrictions, presenting all available context information of the dataset and its environment is out of scope for this paper. In the following, we apply the developed taxonomy to the selected context attributes and classify them according to the six dimensions.

Different vendors may have different delivery times. The delivery time influences the throughput time of the process. We assume that the vendor is known *a priori* for a specific order object. The concrete values represent nominal data with a fixed data schema and therefore represent *structured* data. Since the behavior of the supplier is outside of the direct influence of the organization, the vendor provides *external* context information. As the vendor does not influence the control flow of the process, we do not classify it as highly relevant context information. However, a company requires reliable suppliers who can deliver a quality product on time. Therefore, the vendor is considered to be a context information of *medium* relevance for this process. Additionally, the vendor is related to the *event* ”Goods Receipt” and the associated waiting times in the business process.

The document type (Standard PO, EC PO, Framework order), as well as the item category, influence the process flow since they determine which activities are necessary through compliance rules. Both are also known *a priori* and represent *structured* attributes. Since the course of the process instance is primarily defined by these two attributes and they are essential for understanding the process, these are part of the *immediate context* and have a *high* relevance. The process relation corresponds to the *control flow* of the business process.

The cumulative net amount is known at the beginning of a process instance (*a priori*) but can change during execution (*dynamic*). The values are euro amounts, which is *structured* (metric) data. The net amount is directly linked to the process instance, or more precisely to the order item, and thus belongs to the *immediate context*. We think that the cumulative net amount is only of *low* relevance, since it only affects individual activities, such as checking of the purchase order or goods receipt. For example, a high purchase order amount could require additional approval by a superior or lead to more precise controls at goods receipt. In this way, the cumulative net amount influences individual *activities*. Table IV summarizes the classification of the selected context attributes into the developed taxonomy.

B. BPI Challenge 2017

The dataset from the BPI Challenge 2017 [46] comes from a global financial institution. The process represented in the event log is the application procedure for personal loans or overdrafts. In the BPI Challenge 2017 there were three focus areas of interest for the financial institution: Firstly, the analysis of throughput times, in particular, the time where a customer is waiting to be processed by a user and the time spent waiting on input from an applicant. Secondly, to find out the relation between the number of requests for completion and the likelihood of an applicant not accepting an offer. Thirdly, how the number of offers a customer asks for influences the conversion. The process log contains 1,202,267 events in 31,509 traces representing the loan applications. For these applications, a total of 42,995 quotations are registered. There

TABLE IV
EXEMPLARY TAXONOMY APPLICATION TO SELECTED BPI CHALLENGE 2019 CONTEXT INFORMATION

Dimension	Vendor	Document type	Item category	Cumulative net worth
Time	A priori	A priori	A priori	A priori
Structure	Structured	Structured	Structured	Structured
Origin	External	Immediate	Immediate	Immediate
Relevance	Medium	High	High	Low
Process relation	Event	Control flow	Control flow	Activity
Runtime behavior	Static	Static	Static	Dynamic

are three overarching types of events: request status changes, changes in quotation status and workflow events. The data contains 149 Users, i.e. employees or information systems of the company. The following additional data is recorded for all credit applications: Requested loan amount, application type and the reason for the request. The following information is available for all offers: The offered amount, initial withdrawal amount, the number of payback terms agreed to, the monthly costs, the credit score of the customer, employee who created the offer, whether the offer was selected and whether the offer was accepted by the customer. In addition to this information, various events are recorded in the log, each with the respective employee who caused the event. The overall process looks as follows: A loan application is submitted via the website. Some automatic checks are then carried out. Further information is gathered by contacting the customer by telephone. If a customer is eligible, he will receive the first offer by mail. After the customer has responded to the offer, it is evaluated. If information for the evaluation is missing, the customer must be contacted again. This is followed by a final evaluation, whereupon the request is approved and activated [46].

TABLE V
SELECTED BPI 2017 CONTEXT ATTRIBUTES

Context	Potential Influence on the Business Process
Amount difference	A big difference between requested and offered amount might result in a decline of the offer.
Application type	An extension of an existing credit line might be accepted easier than a new application.
Central interest rate	A low central bank rate leads to low interest, often to higher economic activity and more loan requests.

Table V shows an exemplary selection of context information from or regarding the BPI 2019 dataset, which potentially has an impact on the business process. In the following, we apply the developed taxonomy to these context attributes and classify them according to the six dimensions.

The difference between the desired amount and the offered amount can have an influence on customer decisions. If the difference is high, the offer may be rejected. The difference is calculated during the *runtime* of a process instance since it is only calculated when the concrete offer is created. The amount represents *structured* context information since it can simply be calculated from the difference between the desired amount and the offered amount, given the existing data in the process log. Statistical tests need to determine the actual relevance of the value to the process. But at the very least, we assume a *medium* level of relevance. As the amount is based directly on the central process data it belongs to the *immediate* context. The customer decision, which is potentially impacted by the difference in amounts, influences the further process *control flow*. Since the offer is no longer changing, the attribute - once defined - is *static*.

The application type (e.g. new loan or loan increase) also affects the customer decision. The request type is known at the beginning of a process instance (*a priori*) and is also a *structured* (nominally scaled) context information. It refers directly to to the process instance and thus belongs to the *immediate context*. According to [47] the application type has a large influence on the process *control flow* as well as the process result and thus has a *high* relevance. The value of the request type is *static*.

The central interest rate (e.g. the main deposit rate of the European Central Bank) and economic trends (e.g. development of the gross domestic product) can have strong effects on credit conditions and consequently on the demand for credits. For example, if the key interest rate is low, loans usually become ”cheap” and demand will rise. Since interest rate data usually only changes in the medium to long term, it can be assumed that it behaves *static* in relation to a process instance and is therefore also known as *a priori*. We assume that the key interest rate is known and *structured* (metrically scaled) in the overall banking system. The evaluation of the relevance of the two indicators for the process requires the assessment of a domain expert. Table VI summarizes the classification of the context attributes into the taxonomy.

VII. CONCLUSION AND FURTHER RESEARCH STEPS

The emergence of the IoT, as well as developments around Big Data and massive data processing, foster new possibilities in the area of BPM, specifically in process mining and PPM. To support PPM research regarding the inclusion of context information, context-sensitive tool-support development, as well as PPM implementation in practice, this taxonomy on business process context information was developed in accordance with a systematic development approach [16]. To the best of our knowledge, this is the first taxonomy on business process context information.

In relation to the process prediction reference process (Fig. 1), which was presented in section II, the taxonomy supports the user especially in steps two, four and five. In step two, the prediction goal is determined. As lined out before, context information can have a strong influence on a running process

TABLE VI
EXEMPLARY TAXONOMY APPLICATION TO SELECTED BPI CHALLENGE 2017 CONTEXT INFORMATION

Dimension	Amountdifference	Application type	Central interest rate
Time	Runtime	A priori	A priori
Structure	Structured	Structured	Structured
Origin	Immediate	Immediate	Environment
Relevance	Medium	High	-
Process relation	Control flow	Control flow	Artefact
Runtime behavior	static	static	static

instance and therefore has explanatory properties towards a business process. At this point, the taxonomy provides the user with a reference framework to systematically capture related and influencing context information of the business process. There are two ways to identify relevant context information: First, the current process can be analyzed, especially the current process log, in order to recognize dependencies. Second, a domain expert that understands the process can name relevant context. Depending on the kind of context information that is selected in step two, the taxonomy supports the user in the classification and selection of available prediction methods in step four. For example, if a certain context attribute, such as a dynamic semi-structured one, has a major impact on the business process and should be included in the prediction, the capability of processing this type of context is essential during method selection. Towards this end, the available method overviews, such as [4], should be updated with new context-sensitive PPM methods and could be extended by, e.g., a context dimension that depicts which kinds of context information the method can process. Finally, in step 5 of the reference process, the relevant (contextual) input data needs to be gathered and possibly pre-processed for actual prediction.

Our research provides practical and theoretical implications in the domain of BPM. Practitioners, who plan to implement PPM in their business, can leverage the taxonomy as described alongside the six steps of the reference process. For research, we structured the phenomena of context information with a specific focus on the domain of PPM and its methods. However, there is room for improvements and alternative and extended versions of the taxonomy exist. To further evaluate and test the applicability of the taxonomy, we plan to evaluate and revise it together with domain and industry experts (e.g. through case studies, workshops, and interviews). The process logs of the BPI Challenge datasets represent the de-facto standard regarding datasets and are regularly applied in the PPM domain. For this reason, we selected the datasets for the empirical-to-conceptual iterations. Yet, these logs do not represent the entirety of empirical data. In future research, additional data sets should be investigated and possibly extend or improve our taxonomy. Nevertheless, the fourth iteration fulfilled all subjective and objective ending conditions of the structured taxonomy development process [16].

The research field of PPM is very active and new methods are constantly being developed. [4] developed an extensive overview of available PPM methods, some of which also consider context information. In future work, we want to combine the explanatory power of our taxonomy with an extended and updated overview of PPM methods as well as an extended PPM reference process. This way, the strengths of both contributions can be combined to further support data and method selection.

In conclusion, the taxonomy adds descriptive and explanatory knowledge to the field of BPM and PPM. It strengthens the conceptual foundations for future research on context-sensitivity and PPM through the assessment of the various dimensions of the business process context. We hope that our research will foster and encourage more work on the important topic of context-sensitivity and business process monitoring and prediction.

REFERENCES

[1] J. Recker and J. Mendling, "The state of the art of business process management research as published in the bpm conference," *Business & Information Systems Engineering*, vol. 58, 11 2015.

[2] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer Publishing Company, Incorporated, 2011.

[3] A. Marquez-Chamorro, M. Resinas, and A. Ruiz-Cortés, "Predictive monitoring of business processes: a survey," *IEEE Transactions on Services Computing*, vol. PP, pp. 1–1, 11 2017.

[4] C. Di Francescomarino, C. Ghidini, F. M. Maggi, and F. Milani, "Predictive process monitoring methods: Which one suits me best?" in *International Conference on Business Process Management*. Springer International Publishing, 2018, pp. 462–479.

[5] T. d. C. Mattos, F. M. Santoro, K. Revoredo, and V. T. Nunes, "A formal representation for context-aware business processes," *Computers in Industry*, vol. 65, no. 8, pp. 1193–1214, 2014.

[6] M. Borkowski, W. Fdhila, M. Nardelli, S. Rinderle-Ma, and S. Schulte, "Event-based failure prediction in distributed business processes," *Information Systems*, vol. 81, pp. 220 – 235, 2019.

[7] A. Yeshchenko, F. Durier, K. Revoredo, J. Mendling, and F. M. Santoro, "Context-aware predictive process monitoring: The impact of news sentiment," in *On the Move to Meaningful Internet Systems. OTM 2018 Conferences - Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, October 22-26, 2018, Proceedings, Part I*, 2018, pp. 586–603.

[8] S. Schönig, R. Jasinski, L. Ackermann, and S. Jablonski, "Deep learning process prediction with discrete and continuous data features." in *Proceedings of ENASE*, 2018, pp. 314–319.

[9] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers *et al.*, *Fundamentals of business process management*. Springer, 2013, vol. 1.

[10] A. E. Márquez-Chamorro, M. Resinas, and A. Ruiz-Cortés, "Predictive monitoring of business processes: A survey," *IEEE Transactions on Services Computing*, vol. 11, no. 6, pp. 962–977, 2018.

[11] B. A. Tama and M. Comuzzi, "An empirical comparison of classification techniques for next event prediction using business process event logs," *Expert Systems with Applications*, vol. 129, pp. 233 – 245, 2019.

[12] R. Poll, A. Polyvyanyy, M. Rosemann, M. Röglinger, and L. Rupprecht, "Process forecasting: Towards proactive business process management," in *International Conference on Business Process Management*. Springer, 2018, pp. 496–512.

[13] J. Evermann, J.-R. Rehse, and P. Fettke, "Predicting process behaviour using deep learning," *Decision Support Systems*, vol. 100, pp. 129–140, 2017.
[14] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with lstm neural networks," in *International Conference on Advanced Information Systems Engineering*. Springer, 2017, pp. 477–492.
[15] M. Rosemann, J. Recker, and C. Flender, "Contextualisation of business processes," *International Journal of Business Process Integration and Management (IJBPIM)*, vol. 3, no. 1, pp. 47–60, 2008.
[16] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," *European Journal of Information Systems*, vol. 22, no. 3, pp. 336–359, 2013.
[17] C. Kollwitz and B. Dinter, "What the hack?–towards a taxonomy of hackathons," in *International Conference on Business Process Management*. Springer, 2019, pp. 354–369.
[18] J. H. Beinke, D. Nguyen, and F. Teuteberg, "Towards a business model taxonomy of startups in the finance sector using blockchain," in *International Conference on Information Systems, San Francisco, CA, USA, December 13-16, 2018*, 2018.
[19] U. Paukstadt, T. Gollhardt, M. Blarr, F. Chasin, and J. Becker, "A taxonomy of consumer-oriented smart energy business models," in *ECIS*, 2019.
[20] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS quarterly*, pp. xiii–xxiii, 2002.
[21] C. Di Francescomarino, C. Ghidini, F. M. Maggi, G. Petrucci, and A. Yeshchenko, "An eye into the future: leveraging a-priori knowledge in predictive business process monitoring," in *International Conference on Business Process Management*. Springer, 2017, pp. 252–268.
[22] R. Mousheimish, Y. Taher, and K. Zeitouni, "Toward the support of challenging service level agreements (slas) in manual and context-dependent activities," in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2. IEEE, 2016, pp. 38–43.
[23] B. F. Hompes, J. C. Buijs, and W. M. van der Aalst, "A generic framework for context-aware process performance analysis," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2016, pp. 300–317.
[24] E. Cesario, F. Folino, M. Guarascio, and L. Pontieri, "A cloud-based prediction framework for analyzing business process performances," in *International Conference on Availability, Reliability, and Security*. Springer, 2016, pp. 63–80.
[25] A. Bevacqua, M. Carnuccio, F. Folino, M. Guarascio, and L. Pontieri, "A data-driven prediction framework for analyzing and monitoring business process performances," in *International Conference on Enterprise Information Systems*. Springer, 2013, pp. 100–117.
[26] F. Folino, G. Greco, A. Guzzo, and L. Pontieri, "Mining usage scenarios in business processes: Outlier-aware discovery and run-time prediction," *Data & Knowledge Engineering*, vol. 70, no. 12, pp. 1005–1029, 2011.
[27] A. Cuzzocrea, F. Folino, M. Guarascio, and L. Pontieri, "A predictive learning framework for monitoring aggregated performance indicators over business process events," in *Proceedings of the 22nd International Database Engineering & Applications Symposium*. ACM, 2018, pp. 165–174.
[28] F. Folino, M. Guarascio, and L. Pontieri, "Discovering context-aware models for predicting business process performances," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2012, pp. 287–304.
[29] J. M. E. van der Werf, H. Verbeek, and W. M. Van Der Aalst, "Context-aware compliance checking," in *International Conference on Business Process Management*. Springer, 2012, pp. 98–113.
[30] M. De Leoni, W. M. Van der Aalst, and M. Dees, "A general framework for correlating business process characteristics," in *International Conference on Business Process Management*. Springer, 2014, pp. 250–266.
[31] P. Buneman, "Semistructured data," in *Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*. ACM, 1997, pp. 117–121.
[32] I. Teinemaa, M. Dumas, F. M. Maggi, and C. Di Francescomarino, "Predictive business process monitoring with structured and unstructured data," in *International Conference on Business Process Management*. Springer, 2016, pp. 401–417.
[33] A. Cuzzocrea, F. Folino, M. Guarascio, and L. Pontieri, "A multi-view multi-dimensional ensemble learning approach to mining business process deviances," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 3809–3816.
[34] S. Ferilli and S. Angelastro, "Activity prediction in process mining using the woman framework," *Journal of Intelligent Information Systems*, pp. 1–20, 2019.
[35] A. Pika, W. M. Van Der Aalst, C. J. Fidge, A. H. Ter Hofstede, and M. T. Wynn, "Profiling event logs to configure risk indicators for process delays," in *International Conference on Advanced Information Systems Engineering*. Springer, 2013, pp. 465–481.
[36] F. Folino, M. Guarascio, and L. Pontieri, "Mining predictive process models out of low-level multidimensional logs," in *International conference on advanced information systems engineering*. Springer, 2014, pp. 533–547.
[37] ——, "Context-aware predictions on business processes: an ensemble-based solution," in *International Workshop on New Frontiers in Mining Complex Patterns*. Springer, 2012, pp. 215–229.
[38] K. Ponnalagu, A. Ghose, and H. K. Dam, "Leveraging regression algorithms for process performance predictions," in *International Conference on Service-Oriented Computing*. Springer, 2018, pp. 524–531.
[39] A. Senderovich, C. Di Francescomarino, C. Ghidini, K. Jorbina, and F. M. Maggi, "Intra and inter-case features in predictive process monitoring: A tale of two dimensions," in *International Conference on Business Process Management*. Springer, 2017, pp. 306–323.
[40] F. M. Santoro, F. Baião, K. Revoredo, and V. T. Nunes, "Modeling and using context in business process management: A research agenda," *Modélisation et utilisation du contexte*, 2017.
[41] P.-A. Masse, N. Laga, and J. Simonin, "A contextual data selection tool for an enhanced business process analysis," in *2016 IEEE 13th International Conference on e-Business Engineering (ICEBE)*. IEEE, 2016, pp. 1–8.
[42] K. Boukadi, A. Chaabane, and L. Vincent, "Context-aware business processes modelling: Concepts, issues and framework," *IFAC Proceedings Volumes*, vol. 42, no. 4, pp. 1376–1381, 2009.
[43] P.-A. Masse, N. Laga, M. O. Kherbourche, and J. Simonin, "An approach based on ontology for discovering data impacting the execution of a business process," in *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*. IEEE, 2016, pp. 216–221.
[44] B. F. van Dongen, "Dataset bpi challenge 2019. 4tu.centre for research data," 2019.
[45] K. Diba, S. Remy, and L. Pufahl, "Compliance and performance analysis of procurement processes using process mining," in *International Conference on Process Mining*, 2019.
[46] B. F. van Dongen, "Dataset bpi challenge 2017. 4tu.centre for research data," 2017.
[47] A. Neira, S. F. Gonzalez, and W. P. Fernandes, "Stairway to value: mining a loan application process," in *13th International Workshop on Business Process Intelligence*, 2017.