# Standardizing Process Data Exploitation by means of a Process Instance Metamodel

Antonio Cancela, Antonia M. Reina Quintero, Alejandro García-García, María Teresa Gómez-López

Universidad de Sevilla, Sevilla, Spain
{acancela,reinaqu,maytegomez}@us.es, alegargar1@alum.us.es

**Abstract.** The analysis of data produced by the enterprises during processes execution is key to know how these processes are working and how they can be improved. These data may be consumed to make different types of analysis, for example, data could be used as input for process discovery, decision-making and even process querying tools. However, each type of analysis needs data in different formats because they use different techniques and tackle the problem from a different point of view. Fortunately, if we look at the data exploitation problem from a higher level of abstraction perspective, we can realize that all the points of view share a common ground: the business process model and its instantiation are in the kernel of all of them. In this paper, we propose the use of a Business Process Instance Metamodel, which serves as a common interface to make independent the applications producing business process data from those applications that consume and exploit it. A tool has been implemented as a proof of concept to facilitate the matching between data from different data sources and the metamodel.

**Keywords:** Process Instance Metamodel, Data Model, Model Mapping, Domain Knowledge, Process Data

## 1   Introduction

Activities undertaken by companies can be choreographed by business processes to achieve their goals. Many companies use Business Process Management Systems (BPMSs) to support the automation of business processes, but others use software applications that are implemented ad hoc, having their own data sources with a domain-specific structure. As a consequence, business data cannot be exploited in a general way and, the analysis carried out using the stored data must also be developed ad hoc.

Some of the most common ways of exploiting data generated during process execution are: the creation of execution traces for process discovery algorithms to obtain the system behavior [1, 4]; querying process data to help decision-making in business process scenarios [9]; using ontology-based reasoning techniques [5, 3]. These different scenarios consume data in different formats, and some formatting tasks can be tedious and complex, for example, the creation of traces

from business process executions may be a tedious and complex task [2], since data can be stored in heterogeneous repositories.

On the one hand, the need for a data representation model that let applications that generate business data be independent from those that consume them was detected in [7] and a RDF data model was proposed as an interoperability mechanism. On the other hand, the necessity to define a Business Process Instance Metamodel was introduced in [6]. However, none of the previous contributions offers the possibility of a reusable data model valid for different data sources, different BPM techniques and even though different domains without customization.

In this paper, we propose the use of a Business Process Instance Metamodel as an intermediate layer to bring closer the domain specific data produced by business processes and how they can be exploited. The definition of an intermediate metamodel helps us to abstract domain specific knowledge from the analyzed data, being easier to apply different process analytics techniques. The approach is based on the definition of mappings between the data source and the expected elements in a Business Process defined as a Business Process Instance Metamodel. The benefits produced by the use of an intermediate metamodel are the reduction of the analysis time, and the exploitation of data in a more appropriated way [8]. In fact, the use of the intermediate metamodel is a benefit itself, since provides a standard way of accessing business process data and improving the interoperability among different organizations.

The paper is organized as follows: Firstly, Section 2 introduces the main ideas of our approach and describes briefly the Business Instance Metamodel. Secondly, Section 3 presents a tool implemented as a proof of concept to define the matching between an Oracle database and the Process Instance metamodel. And finally, the paper is concluded and some further work is pointed out.

## 2 Proposal

Business process data exploitation depends highly on the technology that supports business process execution as well as the way data is structured. Some companies have BPMSs to support the execution, but others have software applications that have been implemented ad hoc. As a consequence, there is no standard approach to exploit these data. Thus, for example, the data conversion needed to generate an event log for a process mining tool is totally different if data are stored in a relational database or the data storage of a SAP system.

To make data exploitation technology-agnostic, our approach is inspired by the guidelines provided by Model-Driven Architecture to structure specifications, in such a way that we propose a Business Process Instance Metamodel that helps us to separate the technological details and the structure of the information from the data itself. In other words, the business process instance metamodel can be seen as an intermediate artifact that is not dependent on the business domain nor focused on a subset of tools. Thus, it allows us to make applications that produce business process data independent from those applications that consume
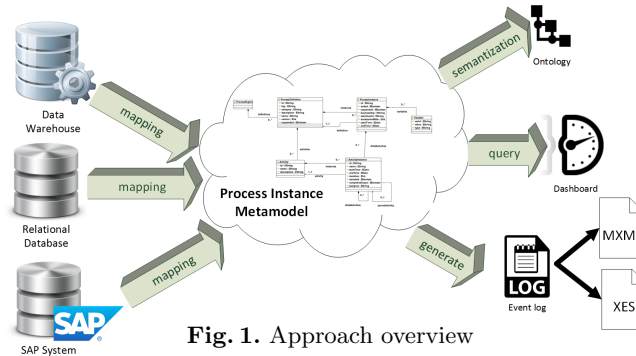
**Fig. 1.** Approach overview

them. Figure 1 depicts how the process instance metamodel acts as an interface for both, data producers and consumers.

The Business Process Instance Metamodel depicted in the center of Figure 1 is detailed in Figure 2. The metamodel has been specified with EMF [10]. Note that it is a very simple model which is mainly centered on the most basic entities related to business process instances together with their attributes.

The root of the metamodel is the **ProcessEngine** metaclass and represents the BPMS or software application that is in charge of processes execution. The process engine can be in charge of different processes. The **ProcessDefinition** metaclass represents the formal definition of the process, that is, what we call the Business Process Model. One business process can be executed many times and the **ProcessInstance** metaclass models these executions or instances. A business process is composed of different activities and the **Activity** metaclass models them. Finally, the **ActivityInstance** metaclass represents the execution of an activity and it is related to the **Activity** metaclass (note that an activity may be executed many times) and to the **ProcessInstance** metaclass (an activity may be executed in the context of different business processes). This metamodel allows us to exploit business data in different contexts independently of the storage technology and structure. We only need to define mappings from the concrete technology to the Process Instance Metamodel. Then, the information stored as a BP instance Metamodel may be used to generate event log traces (both in XES or MXML format), to be queried for decision-making or to be semantized and to reason about it.

## 3 Proof of Concept Implementation

We have implemented a tool that assists business experts in defining the mapping between the data repository and the Process Instance Metamodel, facilitating the later data analysis. The tool allows, once the data repository and metamodel are connected by the mapping definition, to analyze and use the data from a Business Process point of view and exploit those data by applying any of the techniques used in the different business analytics contexts. The tool has been developed as a web application. Figure 3 shows a screen-shot that captures
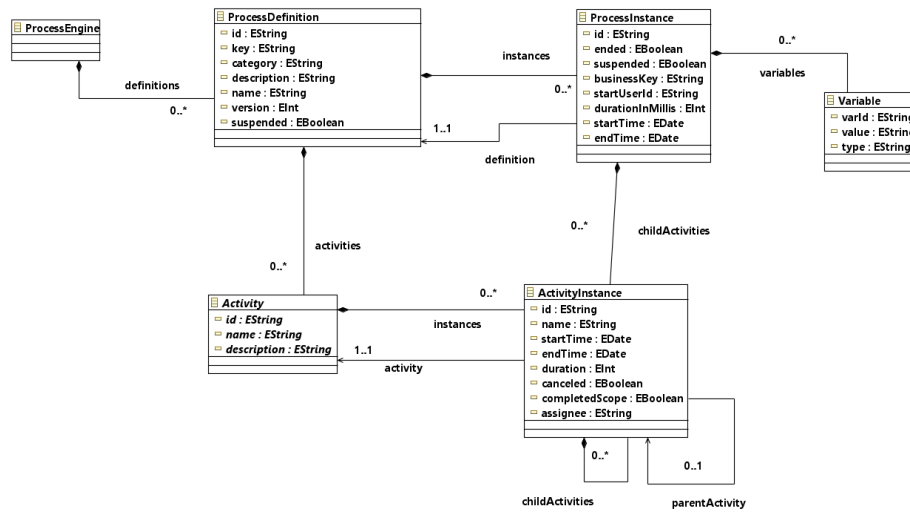
**Fig. 2.** Process Instance Metamodel

the mapping definition process. Further details about the tool can be found in
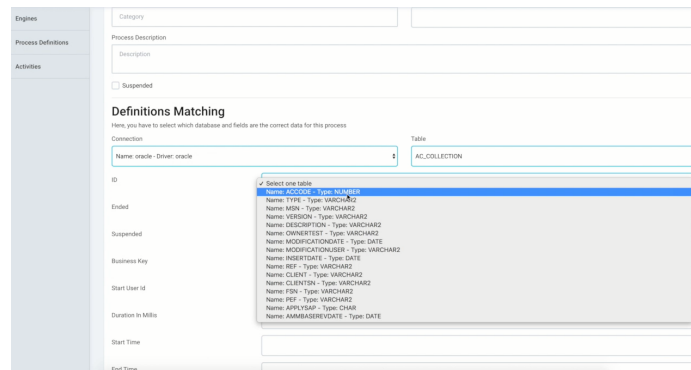`http://www.idea.us.es/portfolio-item/process-data-matching-tool/`.



**Fig. 3.** Developed tool as proof of concept

## 4   Conclusions and further work

This paper shows how the use of an intermediate metamodel can help to standardize the exploitation of business process data by defining a common infrastructure that may be used in different business process analytics contexts.

As further work, we consider interesting to enrich the way of defining the matching, making more flexible the tool and allowing the building of more complex processes and exploitation of more complex data sources.

## Acknowledgments

## References

1. van der Aalst, W.M.P.: Process discovery from event data: Relating models and logs through abstractions. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery 8(3) (2018), `https://doi.org/10.1002/widm.1244`
2. Buijs, J.: Mapping data sources to xes in a generic way. Department of Mathematics and Computer Science, Eindhoven University of Technology (2010)
3. Cairns, A.H., Ondo, J.A., Gueni, B., Fhima, M., Schwarcfeld, M., Joubert, C., Khelifa, N.: Using semantic lifting for improving educational process models discovery and analysis. In: SIMPDA. pp. 150–161 (2014), `http://ceur-ws.org/Vol-1293/paper11.pdf`
4. Calvanese, D., Montali, M., Syamsiyah, A., van der Aalst, W.M.P.: Ontology-driven extraction of event logs from relational databases. In: Business Process Management Workshops. pp. 140–153 (2015), `http://dx.doi.org/10.1007/978-3-319-42887-1_12`
5. Giordano, L., Dupré, D.T.: Enriched modeling and reasoning on business processes with ontologies and answer set programming. In: Business Process Management Forum - BPM Forum 2018, Sydney, NSW, Australia, September 9-14, 2018, Proceedings. pp. 71–88 (2018), `https://doi.org/10.1007/978-3-319-98651-7\_5`
6. Gómez-López, M.T., Reina Quintero, A.M., Parody, L., Pérez Álvarez, J.M., Reichert, M.: An architecture for querying business process, business process instances, and business data models. In: Teniente, E., Weidlich, M. (eds.) Business Process Management Workshops. pp. 757–769. Springer International Publishing, Cham (2018)
7. Leida, M., Majeed, B., Colombo, M., Chu, A.: A lightweight rdf data model for business process analysis. In: Cudre-Mauroux, P., Ceravolo, P., Gašević, D. (eds.) Data-Driven Process Discovery and Analysis. pp. 1–23. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
8. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: Guided process discovery - A pattern-based approach. Inf. Syst. 76, 1–18 (2018), `https://doi.org/10.1016/j.is.2018.01.009`
9. Polyvyanyy, A., Ouyang, C., Barros, A., van der Aalst, W.M.P.: Process querying: Enabling business intelligence through query-based process analytics. Decision Support Systems 100, 41–56 (2017), `https://doi.org/10.1016/j.dss.2017.04.011`
10. Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF: Eclipse Modeling Framework 2.0. Addison-Wesley Professional, 2nd edn. (2009)