Experiences from the Internet-of-Production: Using "Data-Models-in-the-Middle" To Fight Complexity and Facilitate Reuse

Wil M.P. van der Aalst[1,2][0000−0002−0955−6940]

[1] Process and Data Science, RWTH Aachen University, Aachen, Germany
wvdaalst@pads.rwth-aachen.de
http://vdaalst.com
[2] Celonis, Theresienstrasse 6, 80333 München

Abstract. Data-driven approaches play a key role in improving operational processes and production is no exception. The *Internet-of-Production* (IoP) is an ambitious initiative aiming at cross-domain collaboration in production while exploiting semantically adequate and context-aware data at different levels of granularity. The *Internet-of-Things* (IoT), in the context of production also referred to as *Industry 4.0* or the Industrial Internet of Things, provides a wide range of data assets. However, these are often handled in an ad-hoc manner with little support for reuse. Data pipelines convert machine- or system-specific data into a format suitable for data-science techniques such as machine learning. Based on an analysis of the data used in IoP, we developed so-called "Data-Models-in-the-Middle" (DMMs). Two such models are described in this paper: *Measurement and Event Data* (MAED) and *Object-Centric Event data* (OCED). OCED enables Object-Centric Process Mining (OCPM), allowing organizations to view their operational processes from any perspective using a single source of truth. However, OCED is not suitable for low-level machine data that contain a mixture of continuous measurements (e.g., time series data describing position, temperature, force, speed, etc.) and discrete events. Therefore, we also propose MAED as a data format. The combination of both "Data-Models-in-the-Middle" (MAED and OCED) provides a good coverage of many production-related use cases.

Keywords: Internet-of-Things · Internet-of-Production · Process Mining · Business Process Management · Reference Models.

1 The Internet-of-Production

The *Internet-of-Production* (IoP) is a so-called *Cluster of Excellence* at RWTH Aachen University funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy. Over 200 researchers from 35 organizational units in production technology, production management, materials science, data science, computer science, social science, and management science have joined

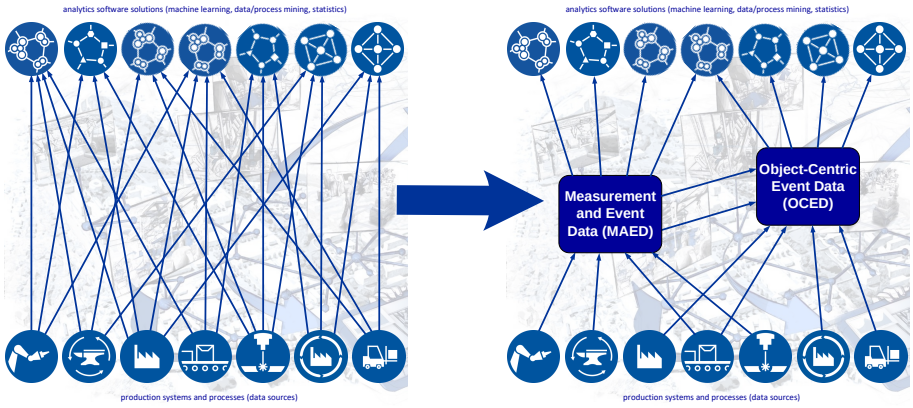

Fig. 1. Production machines and systems produce specific data sets without explicit semantics and data pipelines are often ad-hoc with little reuse (left-hand side). We aim to tackle the problem using "Data-Models-in-the-Middle" (right-hand side) such as Measurement and Event Data (MAED) and Object-Centric Event Data (OCED).

forces to create an infrastructure enabling a new level of cross-domain collaboration by providing semantically adequate and context-aware data related to production [3].

The two main concepts used in IoP are (1) the creation of a *World Wide Lab* (WWL) in which production engineering techniques, data, and software can be used across domains and (2) the use of so-called *Digital Shadows* (DS) capturing condensed knowledge by tightly connecting data and models [3]. To realize the ambitious IoP vision, there needs to be a unification of data and reuse of software. However, a systematic detailed analysis of over 80 data sets created and used in IoP, revealed that data are collected in ad-hoc one-of-a-kind formats and different organizational units tend to create "fragile data pipelines" to leverage data-science and machine learning tools. Therefore, we decided to focus on the creation of "Data-Models-in-the-Middle" as explained in the remainder.

2 Data-Models-in-the-Middle (DMMs)

Figure 1 illustrates both the problem and the envisioned solution. For each machine or system considered in the context of IoP, data are collected and analyzed using a range of techniques and software tools. However, these data sets tend to be one-of-a-kind and end-to-end pipelines tend to be created from scratch (left-hand side of Figure 1). Note that the many connections on the left often correspond to complex manual data-transformation workflows. We advocate the use *Data-Models-in-the-Middle* (DMMs) as shown on the right-hand side of Figure 1. These serve as a *decoupling point* between the application-specific data and focused analytics capabilities. The idea to unify data is not new. For example, the survey presented in [4] shows the usage of ontologies in such environments.

However, the adoption of ontologies is limited, because it takes additional effort and the benefits are not immediately visible for a team analyzing a specific machine or system. In our view, a DMM should enable *immediate analytics capabilities*. Once the data is stored in the format of the DMM, it should be possible to visualize and analyze the data without further transformations and configurations. The idea is inspired by the recent uptake of *process mining* [1]. Given event data in the proper format, one can automatically generate process models, detect process-related problems (e.g., bottlenecks and anomalies), and predict performance (e.g., remaining processing time). By agreeing on basic concepts such as event, activity, case, timestamp, resource, etc. one gets access to a range of analytics capabilities. Note that event data have much more structure than tabular data fed into general-purpose tools. This allows for instant analysis results for a wide spectrum of operational processes in IoP and beyond.

We envision a set of 3-5 DMMs in the context of IoP. Here we briefly describe the two shown in Figure 1: *Object-Centric Event data* (OCED) and *Measurement and Event Data* (MAED).

3 Example: Object-Centric Event data (OCED)

Figure 2 shows a meta-model for *Object-Centric Event data* (OCED). Events are typed and may involve any number of objects. Also objects are typed and may be involved in any number of events. Events have a timestamp and any number of attributes. Objects do not have a timestamp, but can also have time-stamped attributes (e.g., price changes). Objects may be related (e.g., a part-of relation). Both the Event-to-Object (E2O) relations and Object-to-Object (O2O) may be qualified.

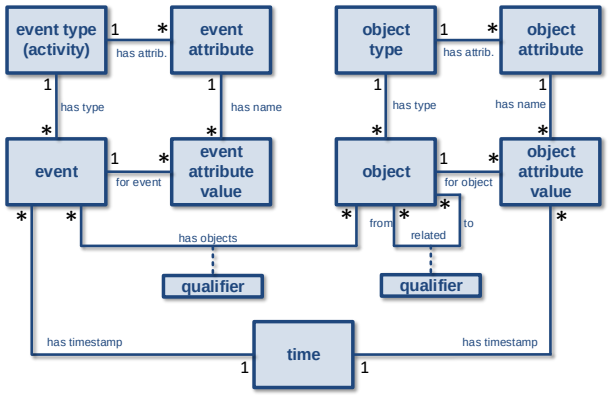

Fig. 2. Meta-model describing Object-Centric Event data (OCED).

Compared to traditional event data (stored for example in XES format [1]), OCED overcomes several limitations. For example, one event can involve any

number of objects of different types (instead of a single case). However, it is still possible to automatically discover process models, expose performance and compliance problems, and predict dynamic behavior [2]. OCED enables *Object-Centric Process Mining* (OCPM) supported by tools such as OC-PM (www.ocpm.info), OCpi (ocpi.ai), and Celonis Process Sphere [2]. It is expected that the whole field will transition to OCPM.

4 Example: Measurement and Event Data (MAED)

OCED is limited to discrete events. However, in production, there are often continuous *measurements*, e.g., position, force, and temperature are monitored using a sampling rate. It does not make sense to see such measurements as events. Their occurrence carries no information; only the values measured do. Analyzing the data of many machines, we noted that they often contain a mix of discrete events and continuous measurements. This triggered the development of the *Measurement and Event Data* (MAED) meta-model shown in Figure 3.

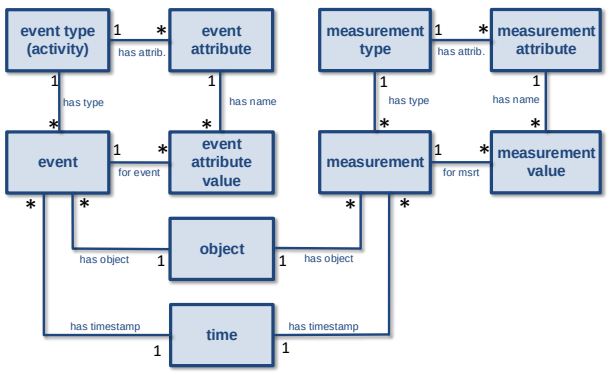

Fig. 3. Meta-model describing Measurement and Event Data (MAED).

Events and measurements are typed and have attributes. Both events and measurements are timed and refer to precisely an object (i.e., more restrictive than OCED). Although events and measurements look similar, they are handled completely different during analysis. Measurements are continuous and do not correspond to specific actions. For example, position, force, and temperature may be measured every 10 milliseconds. Only the values matter. Events are discrete in nature and do not need to happen at specific times (e.g., the machine overheating or restarting). Therefore, we need both.

5 Conclusion

In this paper, we shared experiences from the Internet-of-Production (IoP) and suggested using "Data-Models-in-the-Middle" (DMMs) to fight complexity and facilitate reuse. We proposed two DMMs: Object-Centric Event data (OCED) and Measurement and Event Data (MAED). The former is mature and well-supported and the latter is still under development. Although the idea is not very original (see, for example, the many ontologies developed in the past [4]), we leverage experiences from process mining. Most ontologies do not provide dedicated types of analysis that can be used without further data transformations and configurations. This is in stark contrast with event data (stored in XES or OCEL format) which can be used to create process models immediately showing the actual process, including compliance and performance problems. Future work will also include supporting the transformation of MEAD into OCED. This aligns well with the challenges identified in [5], where the translation of low-level IoT data into events is seen as one of the key problems connecting IoT to Business Process Management (BPM). Moreover, in IoP we also want to standardize event types, object types, and measurement types. Compared to general ontologies, this our approach is more restrictive, but allows for immediate analytics capabilities without further data transformations.

Acknowledgements Supported by the Alexander von Humboldt (AvH) Stiftung and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC Internet of Production, 390621612.

References

1. Aalst, W. van der: Process Mining: Data Science in Action. Springer-Verlag, Berlin (2016)
2. Aalst, W. van der: Object-Centric Process Mining: Unraveling the Fabric of Real Processes. Mathematics **11**(12), 2691 (2023)
3. Brauner, P., Dalibor, M., Jarke, M., Kunze, I., Koren, I., Lakemeyer, G., Liebenberg, M., Michael, J., Pennekamp, J., Quix, C., Rumpe, B., Aalst, W. van der, Wehrle, K., Wortmann, A., Ziefle, M.: A Computer Science Perspective on Digital Transformation in Production. ACM Transactions on Internet of Things **3**(2), 15:1–15:32 (2022)
4. Ekaputra, F., Sabou, M., Serral, E., Kiesling, E., Biffl, S.: Ontology-based data integration in multi-disciplinary engineering environments: A review. Open Journal of Information Systems **4**(1), 1–26 (2017)
5. Janiesch, C., Koschmider, A., Mecella, M., Weber, B., Burattin, A., Di Ciccio, C., Gal, A., Kannengiesser, U., Mannhardt, F., Mendling, J., Oberweis, A., Reichert, M., Rinderle-Ma, S., Song, W., Su, J., Torres, V., Weidlich, M., Weske, M., Zhang, L.: The Internet-of-Things Meets Business Process Management: Mutual Benefits and Challenges. CoRR **abs/1709.03628** (2017)