

Process-Powered Business Intelligence: Contextualized Data for Real-time Data Analysis and Decision Support

Dominik Voigt¹, Juergen Mangler¹, Julian Wipfler², Sebastian Ullmann² and Stefanie Rinderle-Ma¹

¹ Chair of Information Systems and Business Process Management,
TUM School of Computation, Information and Technology,
Technical University of Munich
85748 Garching, Germany

² Technical Methods, Liebherr-Aerospace Lindenberg GmbH,
88161 Lindenberg im Allgäu, Germany

Abstract

Aerospace Manufacturing greatly benefits from data-driven methods aligned with Industry 4.0 to improve manufacturing quality and efficiency, given the high cost of manufacturing and stringent quality requirements. However, data analysis for manufacturing is challenging. Data is distributed across different silos with heterogeneous interfaces and requires expert knowledge to merge consistently. Furthermore, existing data pipelines provide raw, real-time manufacturing data collected from machines. Thus, the collected data needs to be contextualized ex post to be meaningful, which makes the development of data-driven methods challenging. To enable Business Intelligence, we developed and deployed a Process-Powered Data Pipeline that collects and contextualizes manufacturing data from multiple heterogeneous sources in real time. Using a graphical process language, the approach remains easily modifiable by domain experts who are not process design experts and allows integration with online analysis models. Furthermore, by utilizing a transpilation-based execution approach, we achieve real-time performance. This paper demonstrates the viability and advantages of developing Process-Powered Data Pipelines for complex and high-performance scenarios.

1. Introduction

Liebherr-Aerospace is a leading manufacturer of a broad range of high-precision aerospace components, operating in an industry where the cost of raw materials, machinery, and production time demands the highest standards of manufacturing efficiency. Each machine is equipped with specialized software that not only enables high-precision manufacturing but also provides access to a wide range of hardware parameters and sensor information, enabling detailed insight into the current state of operation. Collecting this information at high frequencies opens the possibility of applying data-intensive methods aligned with Industry 4.0 principles to detect quality issues early, optimize processes, and ultimately improve both yield and resource utilization.

However, applying these methods, especially at runtime, relies on high data quality and detailed knowledge of the scenario and context, i.e., how individual data points can be interpreted in relation to one another. Making the contextualized collection of data essential. For example, a sensor data stream only provides insight when the exact machine and its tooling, the exact product variation and its features, the raw material, and a multitude of other factors are known. The prototypical data pipeline already deployed at Liebherr-Aerospace does not provide these capabilities, as it was developed to explore and enable the initial wide-scale collection of manufacturing data.

To address these challenges, we use the Design Science Methodology proposed by Hevner et al. [1] hyperop two artifacts. The first artifact already deployed at Liebherr-Aerospace is a process-controlled

Proceedings of the Industry Forum (BPM 2026), Toronto, Canada, September 27th to October 2nd, 2026.

*Corresponding author.

✉ dominik.voigt@tum.de (D. Voigt); juergen.mangler@tum.de (J. Mangler); julian.wipfler@liebherr.com (J. Wipfler); sebastian.ullmann@liebherr.com (S. Ullmann); stefanie.rinderle-ma@tum.de (S. Rinderle-Ma)

ORCID: 0009-0009-6589-8951 (D. Voigt); 0000-0002-6332-5801 (J. Mangler); 0000-0001-5656-6108 (S. Rinderle-Ma)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

system that collects and contextualizes data. The second artifact, still in development, will leverage the collected data for predictive measurement and outlier detection.

The remainder of this paper is structured as follows: Section 2 further discusses the challenges this project addresses. Section 3 discusses the process we followed, our initial findings, and the requirements we identified for our artifacts. Section 4 discusses our first artifact in detail and describes our plan for the second artifact. Section 5 summarizes the project's core findings and their significance beyond this application case. Section 6 clearly outlines the limitations of the designed artifact and the applicability for the findings. Section 7 summarizes the paper.

2. Problem Statement

Aerospace manufacturing is a complex industry with many challenges. Aerospace parts are produced in small volumes and have high requirements with regard to material properties and geometric precision. Thus, the required raw materials and manufacturing machines are expensive. This, in turn, makes the machining process expensive and the machines' throughput a bottleneck in the production chain. Therefore, to maximize operating profit, failures during the machining process need to be minimized and detected early to reduce material and manufacturing costs and increase throughput. To address these challenges, a data-driven system should be designed to detect deviations and other outliers during the machining process, thereby reducing failure rates and wasted machining time on failed parts. However, the existing data pipeline collects the data streams emitted by the machines into a data lake. While this is a robust solution, it necessitates the design of custom ETL pipelines for any downstream analysis task. We can summarize the data collection-related challenges into three main ones:

- **(C1) Overhead:** Machines deliver sensor data streams into a data lake, connecting this data with other structured information from heterogeneous data sources (e.g., feature-, measurement, and CRM-databases), which introduces significant load on the IT infrastructure, and also widens the time between actual events and when contextual data is available. Also, when multiple methods are to be applied to the data, information should not be aggregated multiple times.
- **(C2) Low Volume:** Many data-intensive methods require a large number of events, but aerospace parts typically have low manufacturing volumes. This can be generalized across all manufacturing domains with low volumes or a high number of product variations.
- **(C3) Conflicting Operational Goals:** While machine operation and product design are handled by mechanical engineering practitioners, data collection technology is operated by the IT department. Individual stakeholder groups have conflicting objectives, e.g., machine operation stakeholders might want to limit the number of data points and the data collection frequency to speed up the machining process and the machine's performance (i.e., machines become less responsive when many data points are collected at a high frequency, and production process duration increases when too many contact probe measurements are taken during production).

To address these challenges, we designed a process-controlled data collection system that aggregates and contextualizes the data stream based on the machining process and the feature currently being manufactured.

3. Approach

The project described in this paper aims to collect contextualized data from raw data streams (i.e., directly from heterogeneous sources such as machines, external sensors, ...) generated during the manufacturing process and to leverage this data to improve the efficiency and precision of the manufacturing process. To realize this objective, we applied the Design Science Research principles proposed by Hevner et al. [1]. Fig. 1 depicts the steps we followed when applying the methodological framework.

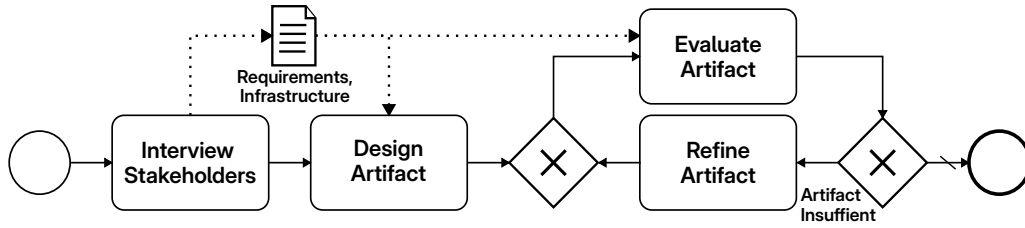


Fig. 1: Adapted from Hevner et al. [1]: Design Science Research

Initially, we identified three relevant stakeholder groups within Liebherr-Aerospace and conducted workshops with experts from these groups to elicit information on the technical requirements and the existing infrastructure. We identified stakeholders within the organizational units “Shopfloor Operations”, “IT Operations”, and “Technical Methods”. Workshop participants hold different positions within the organization, including CNC Programmers, Data Scientists, Data Engineers, and Technical Managers. Each participant provided knowledge and requirements based on their area of expertise. The results of these workshops, which cannot be disclosed as transcripts, as they contain sensitive information about products and infrastructure, are discussed in the remainder of this section.

The main objective of this project was to evaluate the predictive quality and outlier detection algorithms on CNC Mills already owned by the company. Each CNC mill includes a set of out-of-the-box sensors that provide fine-grained measurements of various aspects, such as the current tool position, temperature, individual motor energy consumption, and acceleration. Furthermore, some machines are equipped with tactile probes that allow measuring the part’s geometry during machine operation. This enables the collection of large volumes of geometric data, which is essential for predictive measurement methods. Existing Quality Assurance methods utilize Coordinate Measurement machines with extremely high precision to detect geometric deviations from the geometric feature specifications. However, these measurement processes have significant limitations. Measurements can only be done on a limited number of available coordinate measurement machines, due to their size, cost, and personnel requirements. Furthermore, data access, aggregation, transformation, and association can only be realized ex post and require significant time and expertise.

Currently, a unified data pipeline for the collection of Manufacturing Data has been created and exists in its prototype stages. The Pipeline employs the Siemens Brownfield Connector (BFC) to stream the data generated by the machines into an existing data lake for processing. Furthermore, standalone solutions for managing producible parts, orders, and associated QA data exist.

Based on the knowledge gained during the workshops, we identified 5 mission-critical objectives for this project.

- O1: Contextualized Data Collection:** Collection and contextualization of multimodal data emitted by machines and other sensors during the manufacturing process.
- O2: Per-Feature Data Collection:** Contextualization of data at the level of individual features, e.g., bore holes. This allows for smaller and faster analysis of data frames, making it easier to discover relations between the collected data streams and the manufacturing outcomes, e.g., the relations between vibration data and geometric deviations.
- O3: Availability:** Contextualized data packages are available in a standard format and can be readily used for analysis tasks with minimal additional effort. Contextualized Data packages also capture the context of a specific point in time, staying relevant and analyzable even when the infrastructure evolves; no complex data migrations are necessary to keep aggregation mechanisms working.
- O4 Usability and Evolution:** Create an adaptable and **flexible system** to enhance shared understanding among stakeholders and to provide the flexibility to evolve data collection, without requiring custom programming code.

O5 Predictive Measurement and Outlier Detection: Utilize different machine learning methods for online prediction to improve the efficiency and precision of the manufacturing process.

O1-O4 could not be solved using the company's existing technology, as data streams are collected and stored individually. While **O2** could be addressed with a static programmatic solution, it does not allow non-IT personnel to easily adapt the behavior of the data collection process, as any modification would require changing the program code and redeploying the solution. Based on **O1-O4**, we identified that process technology can be utilized hyperop an efficient solution that achieves both objectives. We designed a process-controlled system artifact that addresses these challenges. We discuss our process-controlled system artifact in detail in Section 4. Subsequently, we will use the artifact to collect contextualized data during the manufacturing operation and leverage the collected contextualized data to evaluate different predictive measurements and outlier detection approaches, using the same process depicted in Fig. 1.

To evaluate the fitness of both artifacts, we use different approaches. We evaluate the performance of the process system in a field study using a selection of different parts and data-generation loads. Here, we mainly observed whether the process-controlled data processing would bottleneck the data pipeline. Additionally, we introduced stakeholders in the technical operations department to the process notation used and conducted workshops with them on its understandability and modifiability. For evaluating predictive measurement and outlier detection approaches, we plan to use standard machine learning methods and metrics.

4. Key Findings

In this section, we discuss the findings and results of following the design science process, depicted in Fig. 1, for the process-controlled data collection system. For the predictive measurement and outlier detection artifact, which currently still in development, we discuss how we address the challenges we identified, the approaches we plan to evaluate, and the evaluation metrics we will use.

During our initial workshops with stakeholders in the Technical Methods department, we identified the objective of leveraging manufacturing data to improve the performance of the manufacturing process. Subsequently, we conducted a workshop with stakeholders from the IT Operations department and identified the existing IT Infrastructure Components, including the existing Data Pipeline. This data pipeline delivers the data streams into an existing data lake. While this approach streamlines the data pipeline and makes it more robust, it requires designing a custom ETL pipeline for every data analysis project that relies on data stored in the data lake. However, constructing an ETL pipeline requires in-depth knowledge of the available data streams and their semantics, the production process, including the involved machines, and other heterogeneous data sources and their structures (**C1**). Also, ETL pipelines are prone to change: when the process or any data source changes, the ETL step must change as well. Developing these ETL steps often take >80% of the project duration. We thus formulated the Objectives **O1-O4** and identified that a process-centered system can achieve them. Processes provide a flexible basis for specifying data analysis pipelines (**O1**) and provide contextualized data packages for data-analysis projects without the need for further ETL steps.

We designed a workflow-based, process-controlled system that follows the process, interacts with existing services and data sources (including existing databases and the BFC for machine data), and collects and contextualizes manufacturing data. Fig. 2 depicts the architecture of the designed system artifact.

Data produced by the machine is collected by the Siemens BFC and published via MQTT on a configurable set of topics as message streams. Since the Cloud Process Execution Engine (CPEE), the process engine we utilize, only supports HTTP-based communication, we designed an MQTT-to-HTTP protocol gateway that allows workflow activities to subscribe to published data streams, thus connecting the existing push-based data pipeline to the process-controlled data collection system.

The artifact contains two processes executed by the process engine to achieve the desired behavior (see Fig. 3). The top-level process ① runs continuously. This process spawns data collection process

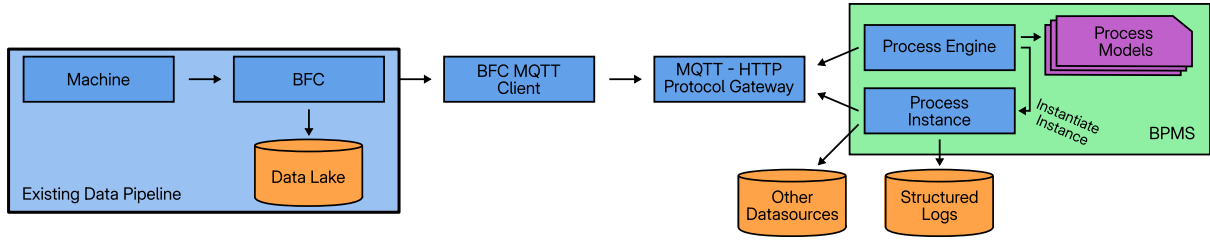


Fig. 2: Component Architecture diagram of the System Artifact

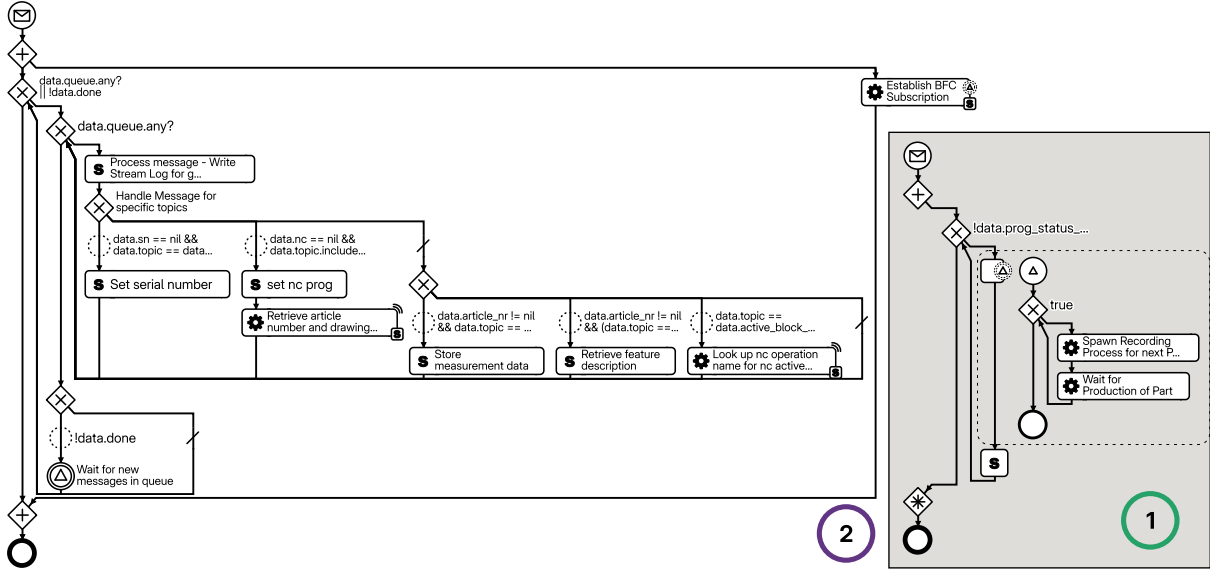


Fig. 3: ①: Top-Level Process that spawns process instances for each manufacturing process ②: Data Collection Process Model for a single Manufacturing Process Instance

instances whenever a machine starts manufacturing a part. To ensure no data is lost, instances are created before the manufacturing process begins. The core of the process is the instance spawning loop (dashed box). For each machine registered, the process spawns a separate instance of this initialization loop. Thus, for each manufacturing process instance, a data collection process instance is created.

The data collection process ② is shown in Fig. 3. The depicted process model was iteratively designed in cooperation with the Technical Methods Department. This process is responsible for collecting and contextualizing the data produced during every part manufacturing process (the execution of an NC Program on the CNC Mills). Thus, each machining process is handled by a process instance of the data collection process, which produces a single event log containing contextualized machining data.

The *Establish BFC Subscription* activity in the right parallel branch invokes the external MQTT-to-HTTP protocol gateway and asynchronously pushes received messages into a queue. The left parallel branch processes the received MQTT messages and writes the data to the log. This is done via the *Process* event activity. Specific events carry metadata related to the machining process, such as the NC program being run or the article and serial number of the part being manufactured. This metadata is used within the data collection process in two major ways:

1. Aggregate additional data related to the part and the current machining process from other data sources, e.g., the tolerances for the specific features or the currently manufactured (geometric) feature (O1).
2. Contextualize each data point by recoding the feature that was manufactured while the data point was emitted (O2).

Thus, the produced data package is the event log for a single instance of the data collection process.

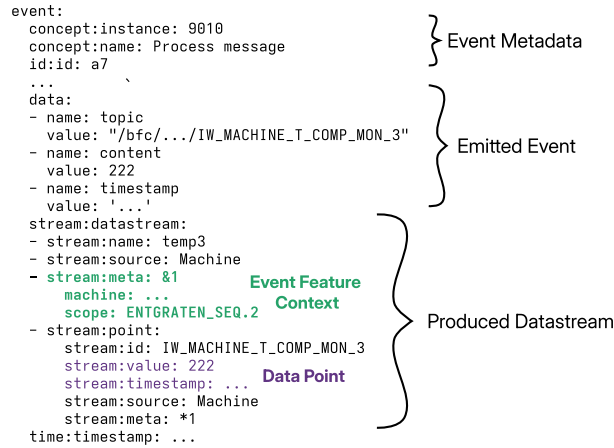


Fig. 4: Event excerpt from an XES Event Log collected by the process artifact. Depicts the collection of a temperature data point during the ENTGRATEN_SEQ. 2 machining process step

We utilize the XES log format with the data stream extension [2] to ensure compatibility with existing tools (O3). Fig. 4 shows an example entry collected using the artifact.

To evaluate the fitness of our designed system, we tested two properties:

- Performance: Does the introduction of a process system create a bottleneck in the data pipeline?
- Understandability: Do stakeholders understand the process notation and semantics sufficiently to adapt process behavior?

We evaluated the performance by deploying the developed artifact and observing its behavior under load with real-world data. In the deployed environment, the CNC mill was emitting data from more than 10 sensors, with individual sensors producing events at up to 500 Hz. In this experiment, we identified a bottleneck during spikes in the machine’s event emission, resulting in data loss. We identified that the execution of the hot path was not sufficiently fast in the transpiled Ruby Process Instance. To address this bottleneck, we introduced two improvements. First, a message queue within the Gateway to buffer spikes in event emission frequency. Second, a Rust-based drop-in replacement proposed by Voigt et al. [3].

During the iterative design workshops with stakeholders from the Technical Methods department, we introduced the process notation used by the CPEE, which closely resembles BPMN 2.0, and discussed different iterations of the process model. During the workshop, stakeholders demonstrated their understanding of the process notation and semantics by identifying semantical gaps in the process model and missing data sources. To summarize our findings collected during the design of the first artifact:

- Process-based data collection can replace ETL steps in a flexible, adaptable manner and produce standardized XES logs.
- Process Technology can be used to collect and contextualize data in an efficient and adaptable manner.
- The collected data can be stored with XES event logs by utilizing the Data-stream extension.
- Protocol Gateways are a flexible way of connecting Event Streams to process instances. Message Queues in the Gateway can buffer arrival spikes of the event stream.
- In comparison to ETL, data collection processes are artifacts that Stakeholders understand and can modify themselves.

- Workflow-based Processes are understandable to relevant stakeholders and allow independent evolution of processes, given changes to the data collection Process are necessary.
- Transpiled Processes achieve the high-performance requirements of high-frequency data collection.
- Process orchestration supports a Microservice-based Architecture that uses modular interfaces to the existing IT infrastructure, using REST.

To address our initial objective (O1), we are currently realizing an artifact for predictive measurement and outlier detection and expose an API to be invoked by the process system during runtime. To this end, we plan to evaluate different approaches for both applications. To collect sufficient data despite low production volumes across different parts, we decided to collect data not on a per-part or per-batch basis, but on a per-feature basis. Thus, we collect multiple data points for each part being manufactured. While this requires training individual models for each class of features (e.g., channels, bore holes), it provides us with significantly more data. Furthermore, we argue that the model will perform better as the data (e.g., vibration and power consumption) and the end result (e.g., the hole diameter) have a higher correlation. Additionally, per-feature predictions are easily interpretable and verifiable; i.e., if the model predicts a deviation at the feature level, the deviation can be checked. In contrast, if either approach is done at the part or machining process level, it is hard to determine whether the prediction is correct.

For the predictive quality measurement method, we plan to test different regression models based on features extracted from the time series data and the planned properties of the geometric feature. We want to test classical approaches such as Random Forests and Bayesian Neural Networks. For outlier detection, we want to evaluate Gaussian Mixture Models on the extracted features and Matrix Profiles on the collected time series. We plan to evaluate these models using classical evaluation metrics, such as Root Mean Squared Error for regression and F1-score for outlier detection.

5. Significance and Relevance

The project demonstrates that process technology can be utilized in complex, performance-critical applications, such as high-volume data collection. Furthermore, it demonstrates that process-controlled systems can be adapted by domain experts, and that microservice architectures encouraged by process-based orchestration simplify the integration of existing IT infrastructure.

Furthermore, it shows how data can be contextualized during collection to address the challenges of designing custom ETL pipelines for every data analysis project and the low production volume of aerospace manufacturing. The need for contextualized data collection also exists in other domains where event and sensor data are collected and potentially influence each other, for example, in the healthcare domain. From an academic point of view, process-based data collection (event data streams with sensor data) supports the simplified use of process mining tools, streamlined adaptation of process structure, and the application of in-process prediction and visualization techniques, e.g., [4, 5].

6. Scope and Limitations

While this case study demonstrates the advantages of process technology, it did not compare this process-oriented approach against a pure event-based approach. Furthermore, the designed system artifact requires application-specific modifications. While the microservice architecture used allows a homogeneous integration approach for the existing IT infrastructure, it still requires programming effort to implement these Interfaces. Additionally, the system's scalability has not yet been fully evaluated. While the experiments demonstrate the workflow system's ability to handle events from up to two machines simultaneously, individual services might become bottlenecks as the system is adopted

across more and more machines. While the individual process instances are implemented as operating system processes, which makes a bottleneck unlikely, bottlenecks in other components, such as the logging service or the MQTT-to-HTTP protocol gateway, might still occur. Lastly, the advantage of contextualized data collection was argued, but could not yet be evaluated due to the second artifact's in-development status.

7. Conclusions

Process technology provides an easily adaptable integration platform for existing IT infrastructure. The flexibility offered by process technology enables adaptation to changing requirements, i.e., changes to the data collection process in our case. The transpiled process instances demonstrate their ability to handle significant loads, processing and contextualizing thousands of events a second while maintaining an event log that also collects the contextualized data. Thus, workflow technology offers a viable option for data processing while being adaptable for practitioners who are not process experts.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools in writing this paper, creating its figures, or for any research presented in this paper.

References

- [1] Hevner, March, Park, Ram, Design science in information systems research, *MIS Quarterly* 28 (2004) 75.
- [2] J. Grüger, L. Malburg, J. Mangler, Y. Bertrand, S. Rinderle-Ma, R. Bergmann, E. S. Asensio, *Sensorstream: An xes extension for enriching event logs with iot-sensor data*, 2022.
- [3] D. Voigt, L. Podszun, J. Mangler, S. Rinderle-Ma, *Hiprox: Highly efficient process execution on iot and edge devices*, in: *BPM Process Technology Forum*, 2025, pp. 184–199.
- [4] F. Stertz, S. Rinderle-Ma, J. Mangler, *Analyzing process concept drifts based on sensor event streams during runtime*, in: *BPM*, 2020, pp. 202–219.
- [5] M. Ehrendorfer, J. Hebstreit, J. Mangler, S. Rinderle-Ma, *Interactive drift visualization in sensor data streams for explainable process outcome prediction*, in: *Business Process Management Forum*, 2024, pp. 162–178.