Data Science meets BPMN: A Taxonomy of Data Objects Modeling

Inês de Almeida Barata
University of Coimbra, CISUC/LASI, DEI
Coimbra, Portugal *inesbarata@student.dei.uc.pt*

Vítor Hugo Machado Ribeiro
University of Coimbra, CISUC/LASI, DEI
Coimbra, Portugal *vhribeiro@dei.uc.pt*

João Barata
University of Coimbra, CISUC/LASI, DEI
Coimbra, Portugal *barata@dei.uc.pt*

Paulo Rupino da Cunha
University of Coimbra, CISUC/LASI, DEI
Coimbra, Portugal *rupino@dei.uc.pt*

Abstract

Data science is transforming business processes. Therefore, process modeling approaches must evolve to capture the entire lifecycle of data transformations. Business process modeling and notation (BPMN) offers a possible solution. However, data objects in BPMN are usually relegated to a secondary role in process flows, missing the complex interactions between data sources and pipelines. This paper presents a systematic literature review and a taxonomy of five types of modeling approaches for data objects in BPMN. The work is conducted in the context of a green and digital transformation project in ports and logistics. Data scientists and process owners may find our proposals interesting for adopting BPMN in their data-driven projects, detailing in a transparent way how (1) data inputs are obtained, (2) processed, and (3) used at a process level of analysis. Theoretically, our work contributes to BPMN literature, comparing five types of modeling approaches for data objects.

Keywords: BPMN, Data Objects, Data Science, Systematic Literature Review.

1. Introduction

Business process modeling and notation (BPMN) is a popular standard for the graphical representation of business processes [11]. It was initially published in 2004 and continuously improved to capture the most relevant aspects of a process, including events, tasks, gateways, messages, and process participants. The BPMN specification also includes extension mechanisms [37], enabling, for example, the creation of tailored elements for specific cases, as exemplified by [53].

Data can also be represented in BPMN diagrams, clarifying key inputs, outputs, and storage in workflows. However, recent studies state that "*techniques such as ER diagrams, UML, DFDs and BPMN provide valuable frameworks for capturing business processes and relationships [but] often fall short in supporting the complex, unstructured, and multidimensional data formats prevalent in modern AI applications and data science workflows*" [35].

The transformative role of data science in business process reengineering is unquestionable [2]. Yet, despite the growing interest in BPMN data objects, as exemplified by BPMN extensions for machine learning (ML) workflows [47], a domain-specific visual model inspired on BPMN and UML for representing data analytics operations and solutions [22], a catalog of sub-process templates for modeling distinct ML scenarios supported on an extension [46], and proposals to identify different versions of data objects [23], BPMN cannot yet represent the entire lifecycle of data associated with business processes (e.g., data lifecycle details [32]). Therefore, we formulated two research objectives:

RO1 – Understand the current landscape of data objects research in BPMN.

RO2 – Identify possible solutions to model data science pipelines associated with BPMN data objects.

We conducted a systematic literature review (SLR) [36, 52] on BPMN's data objects, aiming to understand its applicability, use, and existing pains for modeling processes and related data assets. Moreover, the research team identified a taxonomy of modeling approaches for data objects, exploring alternatives that utilize extensions or complementary solutions, as well as standard BPMN 2.0 elements. We demonstrate their use in several modeling exercises in the digital and green transformation of ports.

Our paper is structured as follows: Section 2 describes the background, including process models, BPMN, and data objects. Next, Section 3 details the SLR, including the corresponding methodology and findings. The taxonomy to model data objects is described in Section 4. Section 5 presents the discussion. Lastly, Section 6 reports the main conclusions, limitations, and future work opportunities.

2. Background

Business process models provide graphical and textual descriptions that enable analyzing and improving processes, and are a valuable communication tool [43] to increase the stakeholder's (e.g., process engineers, workers) knowledge and awareness of business process workflows [3, 41]. BPMN [37] is a well-accepted and widely adopted standard within the industry and academia [12], being supported by multiple business process engines (BPEs) and modeling tools (e.g., Camunda, Bizagi, IBM Blueworks Live) [43]. The notation is known for its versatility and understandability [41], making it suitable for users with distinct backgrounds (e.g., domain users and engineers).

BPMN structures its elements in five categories [37]: flow objects model the processes' behavior, including events, activities, and gateways (e.g., decision points or parallel tasks); sequence flows link the flow objects (e.g., activities, events) to represent process progression and order (e.g., the consecutive steps in the process according to a defined order), message flows (e.g., information exchange between stakeholders), and associations (e.g., model data outputs); pools and lanes represent organizations and departments, respectively, clarifying the responsibility for task execution; data focuses on the resources and objects that are used, manipulated, and transferred throughout the process flow (e.g., data objects representing an application decision document); artifacts add additional context, information, and details to the process model, including groups (e.g., signal tasks that face potential security issues) and text annotations (e.g., add text in the model to specify the version of the software that is being used) [37].

However, BPMN data objects are limited in terms of modeling. BPMN is conceived as an activity-centric notation, missing the details to model object-centric setups (e.g., data lifecycle and dependencies in process runtime) [32]. These are serious concerns as organizations collect and analyze large volumes of data while executing their business processes (e.g., sensor data in logistics scenarios), providing an insightful source for conducting process monitoring, improvement, and optimization [13].

Data objects in BPMN describe the distinct categories of data used in a business process [32, 37]. BPMN process models can feature "*persistent*" and "*non-persistent*" process-related data objects [32]. Data objects contribute to representing the interactions between specific business process tasks or events and available data resources [23]. Nevertheless, existing work highlights multiple limitations in BPMN-based process models to represent data-related operations. For example, it does not address data status and classes [23], and "*the type of a write access on data objects*" [32]. Additionally, the notation does not represent the multiple instances of the same data object that can be manipulated in parallel in distinct business processes, nor the relationship between its multiple instances [15, 32], and does not consider mechanisms to address data interoperability [17]. Therefore, there is a need to address the limitations of using data objects in BPMN-based modeling.

3. Systematic Literature Review

3.1. Methodology

Our SLR focuses on a concept-centric analysis [52] to understand how researchers and practitioners use, manipulate, extend, and apply data objects in BPMN-based process modeling. The review was performed in April 2025, following: (1) "*Purpose of the literature review*", (2) "*Protocol and training*", (3) "*Searching for the literature*", (4) "*Practical screen*", (5) "*Quality appraisal*", (6) "*Data extraction*", (7) "*Synthesis of studies*", and (8) "*Writing the review*", as proposed by [36].

The research team selected Scopus and Web of Science (WoS) as the primary sources since they are two predominant bibliographic sources [16]. Our search string aimed to understand the literature landscape on data objects in BPMN and to identify existing gaps in the available approaches to model data-related operations in business processes. For this purpose, the research team applied the search string ("BPMN" AND ("data object" OR "data objects")). We obtained 146 matches – 17 in WoS and 129 in Scopus. The first exclusion criteria (EC) aimed at removing duplicates, with a total of 131 papers remaining. Next, the second EC focused on screening the abstracts and keywords to assess the contribution's relevance (e.g., priority for data objects), resulting in 38 papers for full text analysis.

3.2. Systematic Literature Review Results

Our concept-centric analysis addresses contributions that (1) extend BPMN to enhance data-related operations, and (2) use data objects for multiple possibilities.

Extending BPMN's Capabilities for Modeling Data

The literature highlights limitations in BPMN's data objects and some solutions. For example, in [5], the authors extend BPMN's data objects with foreign keys and states to model data dependencies. The works of [17, 18] propose a BPMN extension that represents data interoperability issues. In [23], the authors extend the BPMN's data objects to model "*dynamic references to different objects of the same class*" employing variables – an identifier bound with a specific data object for referring and accessing the same instance in upcoming tasks. Other approaches propose annotations to include lifecycle information, an identifier, and potential dependencies with other data assets [32]. There is also a framework to model data-aware business processes, extending BPMN's data objects with annotations on preconditions and effects [40]. The extension in [19, 20] links data objects with event monitoring points – creating object state transition events - to enhance data logging in process-related occurrences. The literature confirms that data object extensions can address some limitations in data science modeling. However, such an option may not be sufficient to model all operations and data manipulation in complex processes.

BPMN extensions may derive domain-specific languages, focusing on data objects and the full scope of the notation's structure and concepts. An example is included in [42] with elements tailored for personal and non-personal data, data governance-related "shadow" operations, and data-flow transmission. The work of [7, 8] suggest new types of tasks (e.g., data preparation) and data objects (e.g., result object) for quantum modeling, while [28] propose a set of BPMN extensions for object-centric representation, addressing data lifecycle representation, access control, and location information in inter-organizational processes. Finally, the study of [26] proposes a BPMN extension that creates a new type of data object (e.g., has access to runtime application information) for cloud applications. BPMN extensions contribute to the usability of data objects, but they are typically deployed for domain-specific contexts. Moreover, modeling data science activities using extensions may still lead to complex and excessively saturated process models, considering that the integration of additional icons and text may rise difficulties for users to understand their meaning.

The literature demonstrates that complementing BPMN models with other notations is another possibility. UML models can embed additional information layers about data characteristics and dependencies. This is the approach selected by [29], using UML profiles to complement process models with additional information layers regarding data use. The works of [50, 51] propose a BPMN-inspired notation complemented with UML for discrete events modeling that extends data objects to incorporate system state changes and preconditions. UML can incorporate details on data structures and systems; however, it requires users to know the notation, to manage additional documentation, and to deal with UML's increased complexity.

Another possibility is to explore Petri-nets, complementing the process models with additional documents, or to embed Petri-nets semantics in the BPMN. The study of [24] expands BPMN with extended Petri-nets semantics to include "*variable identifiers for data objects*" supported on a query language to verify and check compliance with the notation's standards. Colored Petri-nets can also be used, for example, for modeling cross-case and case-specific data object use [15]. The research of [25] proposes a process modeling semantics supported on Petri-nets to represent the split and merge of BPMN's data objects. Petri-nets are viable for modeling states, manipulations, and transitions over data assets. However, similarly to UML, it can lead to increasingly complex process models and difficulties in maintaining them.

Three or more modeling techniques can be combined at the same time. For example, some approaches combine UML, Petri-nets, and BPMN choreographies to detail data exchange [31]. In a similar line, [44] propose a modeling solution that combines BPMN, UML, Petri-nets, and fragment-based case management, incorporating concurrent object flows, relations, synchronization, and reference. However, combining modeling techniques may make models' management and communication more difficult. Moreover, the users must learn the dynamics of the combined modeling techniques to understand the representation. This option may lead to separated and longer process modeling activities, which may give rise to issues in understanding the data science pipelines, alongside the main process execution.

BPMN's Data Objects Applications and Use

Data objects' role in process models varies, ranging from purely technical purposes to supporting management concerns in business processes. Software engineering, for example, has explored them to collect functional requirements (e.g., data collection, processing, storage) [45]. The work of [1] includes an approach that analyzes data objects and their dependencies to identify microservices. Other studies focus on defining decision models, using BPMN-based process analysis with a focus on data objects as a primary source to identify decision-making sources and inputs [6].

Data-related risks and issues become visible in BPMN models. The work of [9] presents an approach to check potential data leaks in collaborative processes. In a similar line, [38] proposes and evaluates a workflow model to control access to data objects in the process. Other approaches focus on inter-organization processes, identifying and allocating data objects' responsibility in inter-organizational BPMN-based process models [21] or the coordination of inter-organizational data management in choreography models [27]. The analysis can also be automated, as in [2] which deploys ML algorithms that identify and categorize documents, participants, and systems.

Some studies offer solutions to ensure consistent use of data objects. The work of [39] describes a framework to discover data state and verify consistency, compliance, and completeness, while [34] proposes an approach that verifies data flow correction in BPMN models relying on data objects analysis.

The synergies of data and activity flows are significant. On the one hand, data can be used to improve business processes. Data objects provide critical indicators for process synchronization as revealed by [14]. [48] performs data impact analysis to understand the influence of data assets in business processes (e.g., decision-making, thresholds) modeled using BPMN (e.g., data objects and related dependencies), while [49] presents an approach that analyses BPMN's data objects to determine potential business process orchestration issues. Data objects must be accessible at runtime to avoid potential deadlocks [30]. On the other hand, it is possible to extract data requirements from BPMN activity-centric artifacts [10]. Additional examples include [33] that capture data objects and their connections to obtain the corresponding data models, and [4] that derives and models the data lifecycle.

4. A Modeling Taxonomy of Data Objects for Data Science

The research was conducted in the context of a 90M€ project to optimize ports' digitalization and sustainability (NEXUS). BPMN was selected as one of the notations that project partners can use. The consortium extensively uses data science to support decision making, requiring practical modeling approaches for data transformation (e.g., specifications, documentation, and audit). The research team collected and analyzed documentation (e.g., reports) to obtain an overview of the process architecture involved in the port operation. Subsequently, multiple modeling exercises were done to propose a taxonomy for using data objects in business process models, focusing on the port management operational scenarios (e.g., smart berth planning and smart gate operations). For example, participating port management entities use data science to improve slot allocation management for incoming ships and automatically detect container damages.

In the context of port digital transformation, we identified five main types of modeling approaches for data objects, separated into two groups. The first group, prevalent in literature and emerging from the SLR, includes solutions that extend or complement the BPMN notation:

- **Rich representation type**: BPMN elements (e.g., data objects) tailored for data science activities with specific symbols [5, 7, 8, 17–20, 26, 28].
- **Complementary type**: adopt other languages/solutions to improve data modeling [15, 24, 25, 29, 31, 44, 50, 51] or include data objects presenting the different versions of a data object, their dependencies, and relationships (e.g., Dataset v1, Dataset v2) as the project evolves [23, 32, 40].

However, BPMN simplicity is also a key advantage and one of the reasons for its popularity among practitioners. Therefore, the research team analyzed BPMN's metamodel and structures to create a second group of modeling alternatives that rely exclusively on standard BPMN elements to support data science-related information.

- **Subprocess type**: subprocess describing the data science activities associated with the data objects in the process.
- **Side pool type**: a specific pool concentrating all the data science tasks.
- **In-process integration type**: lifecycle of data objects explained in parallel with the process.

We will now illustrate the second group of alternatives for modeling data objects, based on modeling exercises made by the research team in the context of smart berth and logistics terminals. As an example, we consider a smart gate process designed to manage a container entrance at a port, via road transport, and automatic analysis of possible damages using computer vision and machine learning techniques. This process involves both data preprocessing and the application of a predictive model, which identifies possible damages. However, since this data-driven evaluation constitutes only a part of the broader workflow, there are multiple options available for its representation within the BPMN diagram, using exclusively the standard elements in the notation.

Subprocess Type

Fig. 1 illustrates the embedded subprocess approach.

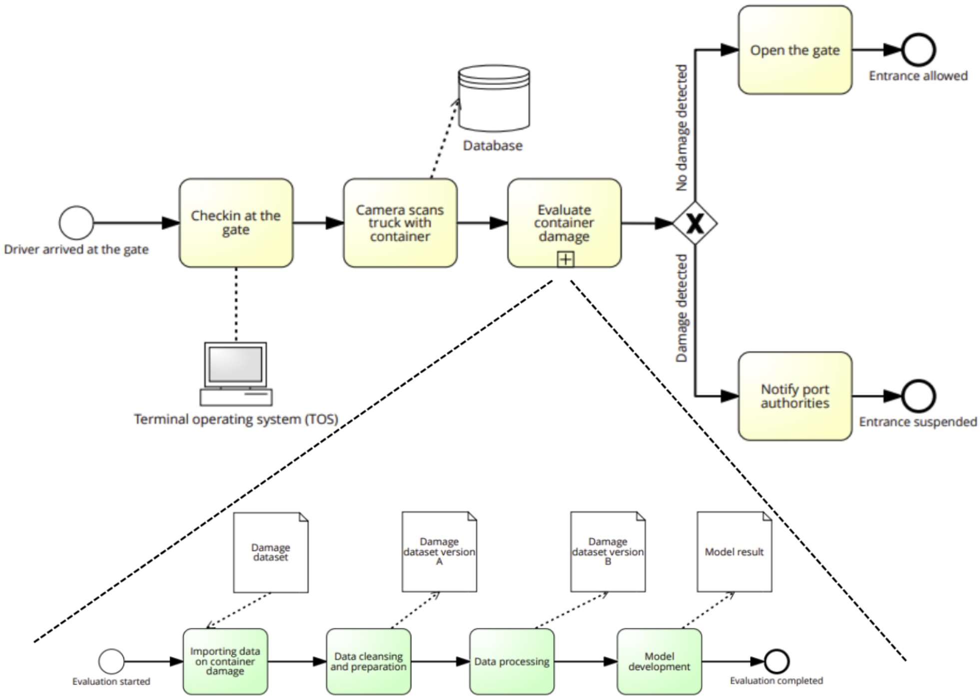

Fig. 1. Subprocess describing the data science activities associated with the data objects.

In the first option, we can represent a subprocess (below) within a main process by embedding it inside the diagram, accessible through a "+" marker. This approach is particularly effective when the subprocess is complex enough to merit its own detailed view, yet closely related to the overall workflow, ensuring clarity without overwhelming the main process diagram.

Side Pool Type

Fig. 2 shows the use of a separate pool for the data process.

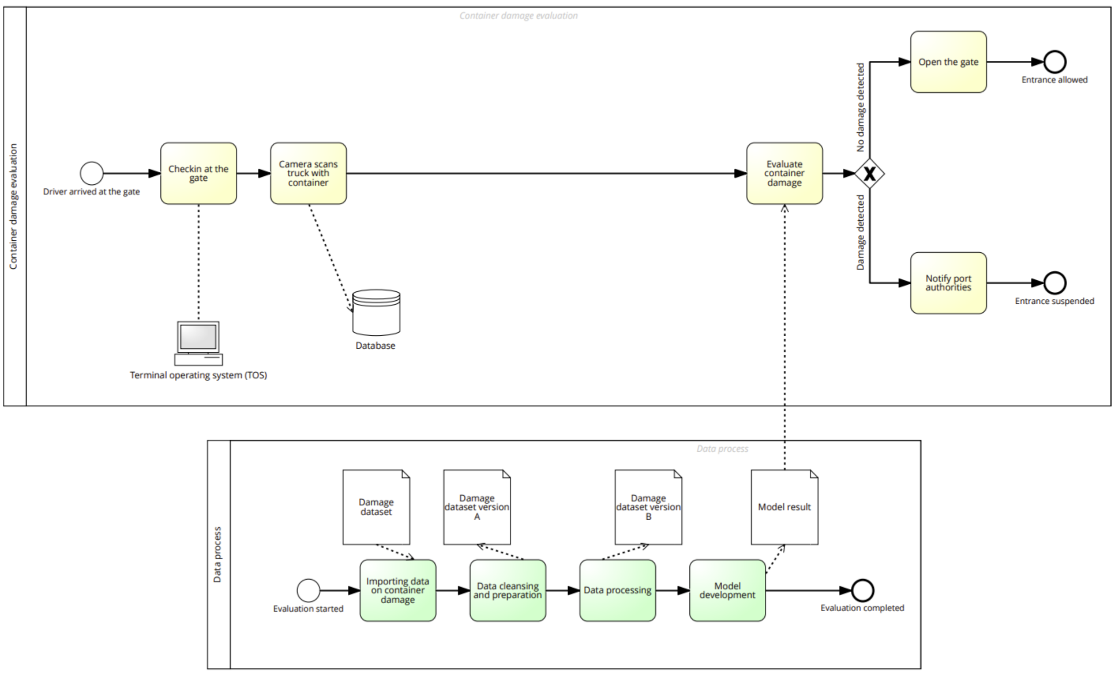

Fig. 2. Side pool with the data science tasks.

Alternatively, the data process can be represented using a separate pool within the BPMN diagram (Fig. 2). This method is especially appropriate when the data process is handled by an independent participant or system, emphasizing autonomy and clarifying the interaction between distinct organizational entities or subprocesses.

In-process integration type

Fig. 3 presents the in-process integration of data object lifecycles.

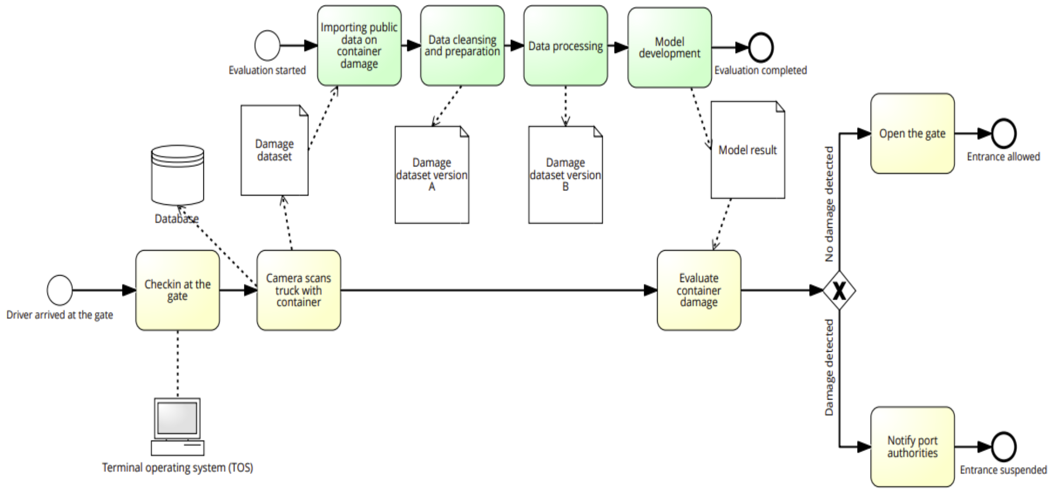

Fig. 3. In-process integration with the lifecycle of data objects evolving in parallel with the process.

A third option involves side-by-side integration, where the tasks of the subprocess or side pool (previously presented types) are placed in parallel with the primary process. This approach treats data objects as first-class entities explicitly modeled within the process flow itself. This method is advantageous when the process progression is tightly coupled with changes in the data, allowing for a more holistic and synchronized view of both control and information flows.

The following section discusses the literature's findings and provides a comparative analysis of the six approaches for data object modeling that address data science needs.

5. Discussion

Different approaches are used to deal with data-driven business process modeling. Some authors suggest additional tools/languages (e.g., UML) to solve the limitations of BPMN, and there are also BPMN extensions (e.g., capturing the state of data objects). These contributions reveal the need to go beyond the narrow perspective of representing data as supplementary information in process models. When business processes use advanced data science techniques like machine learning or other artificial intelligence enhancements (e.g., computer vision, process automation), showing the data inputs and outputs is relevant, but insufficient. Explainable AI is increasingly important, requiring an understanding of the state of the lifecycle of data objects and the operations involved (e.g., data cleaning, pre-processing, data provenance, versioning, and others).

After understanding the current landscape of data objects research in BPMN (RO1), five types of modeling approaches emerged, separated into two main groups that (1) extend or complement, or (2) adapt the BPMN standard to the needs of data science. Table 1 compares the different alternatives when extending or complementing BPMN concepts.

Table 1. Taxonomy comparison for extending or complementing BPMN elements.

Type	Advantages	Disadvantages
Additional BPMN Elements	✓ *Use a domain-specific language* ✓ *Highlight particular details that are relevant to the context*	✓ *More difficult to memorize* ✓ *Requires a legend* ✓ *More complex*

Type	Advantages	Disadvantages
Rich representation type	• Experts' familiarity with the language • Improved communication of details	• Reduced comparison with other domains • Requires training / cognitive overhead
Complementary type	• Flexibility to use the advantages of different languages/notations • Keeps the standard original specification • Solves the challenge of versioning of artifacts • Focus only on the data objects (limited extension)	• Does not allow a single representation • Maintenance can be more complex with different artifacts for the same purpose • Models are more complex • Requires an explanation to the reader

Table 2 depicts the multiple alternatives for adapting standard BPMN concepts, that we illustrated for port digital transformation.

Table 2. Taxonomy comparison for adapting standard BPMN elements.

Type	Advantages	Disadvantages
Adapting Standard BPMN Elements	✓ *Simplicity of the model* ✓ *Coherence* ✓ *Reduced training*	✓ *Limited elements* ✓ *Reusing the elements for different purposes may lead to confusion*
Subprocess type	• Creates layers of process interpretation, keeping each one simple • Allows a quick vision of the process, only presenting details if needed	• Navigation is more complex • Does not provide a complete picture unless the user makes a drill-down of the model
Side pool type	• Data science tasks become separate (interesting for this particular stakeholder) • Highlights the data science pipeline	• The side pool with different pipelines can be more complex to read
In-process integration type	• Integration of the process flow and data flow near the use of data	• Complexity can increase

Our comparison can be used by practitioners in selecting the best approach, according to the characteristics of the process and the data science requirements. Extending BPMN with additional elements in a rich representation type allows practitioners to derive domain-specific languages that incorporate both (1) data use and manipulation concepts and (2) context-specific activities and events – as an example, modeling port movements (e.g., extend tasks) and corresponding data categories. This option can be suitable for scenarios in which practitioners want to address data science operations and domain-specific contexts. However, additional information layers may lead to difficulties in applying the notation in other domains, requiring practitioners to learn new domain-specific elements.

Complementing BPMN with a state model, users can address BPMN's limitations in data objects use and applicability by detailing datasets' characteristics and version – as an example, detail data versioning in a data pipeline, considering raw data (Dataset v1) and data cleaning (Dataset v2) while manipulating the same data object. The state information is suitable for scenarios in which the focus is solely on generic data manipulation and usage operations. Still, enriching data objects with state information may lead to increased complexity in model readability. BPMN can also be complemented with other notations or modeling techniques (e.g., Petri nets) to add new information layers while complying with the standards, thereby extending the power of a single approach. For example, combining BPMN and Archimate can expand details on IT systems used in process runtime. The complementary type can be a valuable strategy when modeling complex business processes with parallel data science operations (e.g., representing data lineage and provenance). Nevertheless, the complementary type requires more documentation and may lead to difficulties in maintaining the process models.

Exploiting standard BPMN elements is another path to modeling data science scenarios. Notably, this approach is scarcely mentioned in the literature and appears to be more straightforward to implement in many situations. A subprocess type (as demonstrated in Fig. 1) incorporates additional data science activities using layers. This type can be useful to maintain compliance with BPMN standards when practitioners' have limited knowledge of other modeling techniques. However, such strategy may lead to difficulties in navigating through the processes and sub-process's structure and may underestimate process analysis and interpretation. The side pool type incorporates a separate structure that specifically addresses data science operations (as shown in Fig. 2). This can be useful for representing data science operations relevant to process runtime tasks that can be represented in a single process model. However, this strategy may lead to an increased number of pools due to the use of different datasets, increasing the model's complexity and saturation. Lastly, the in-process integration type (Fig. 3) embeds data science operations directly in the process flow. This type is valuable for modeling simple business processes that heavily depend on data usage and manipulation for decision-making and execution. Yet, the in-process integration may lead to more complex and saturated process models since they embed both operational and data-related layers' representation.

The project in which this research takes place involves over 30 organizations, including port authorities, logistics operators (by road, sea, or rail), IT providers, importers/exporters, technology transfer institutes, and universities, and extensively utilizes data science. BPMN is an excellent option to model data, but model coherence and completeness are crucial to ensure proper stakeholder communication. The proposed taxonomy guides data scientists in selecting the best alternative for each development scenario.

6. Conclusions

This paper presented a SLR on data representation in BPMN literature. Several approaches for extending, complementing, or adapting BPMN were identified, and additional ones were proposed. Our review also found multiple use cases for data governance associated with business processes, including quality verifications, approaches for modeling, or extracting value from data objects. However, we also found a gap in the representation of data object processing activities relevant to a data scientist audience, who need to understand the details of data pipelines in modern processes that utilize AI. Therefore, our second contribution includes modeling exercises that adapt BPMN to the purpose of data scientists' representation, providing a taxonomy, a comparison, and recommendations for selecting the best approach according to the context for modeling data objects. For the upcoming stages, the research team plans to conduct a formal evaluation using a questionnaire based on the System Usability Scale and open-ended questions targeted at business process modeling experts, aiming to compare the performance, utility, and applicability of several modeling strategies.

Several limitations need to be stated. First, the review focused on two of the most relevant databases (Scopus and WoS), but other sources can be used. Second, our analysis of data representation is restricted to the scope of BPMN. Third, while the literature supports our taxonomy, we also propose new types of data object representations: a BPMN adaptation strategy. A subsequent stage of our work includes a system usability scale questionnaire to understand the view of the practitioners. Fourth, although we adopted the taxonomy in a real project, we have not conducted a formal evaluation of the multiple adaptation strategies and their corresponding applicability. Another possibility is to compare the proposed solution with other modelling languages, such as UML (e.g., UML Activity Diagram to explain the process, UML State Diagram, UML Class Diagrams for data, UML Object Diagram for objects and class instances), more popular among IT professionals but perhaps less familiar to business experts. Such an experiment would consider users with distinct backgrounds, aiming to derive potential alternatives that could use or combine different modelling notations. Fifth, other languages may provide additional details about data (e.g., structure, states) that are not yet incorporated in our proposal.

BPMN is one of the most popular solutions for modeling business processes. It is accessible to technology experts and non-experts, which increases the potential for discussing process reengineering supported by data science techniques. The identified taxonomy may open new avenues of research in data object representation that do not require the overhead of BPMN extensions (which are more challenging for non-experts to interpret) and are sufficiently flexible to be adapted in practice. Future work will include developing a set of guidelines for selecting and using the most adequate adaptation strategy according to the context and purpose of modeling.

Acknowledgments

This project was funded by Agenda "NEXUS—Pacto de Inovação—Transição Verde e Digital para Transportes, Logística e Mobilidade", financed by the Portuguese Recovery and Resilience Plan (PRR), with no. C645112083-00000059 (investment project n° 53). This work is also partially financed through national funds by FCT - Fundação para a Ciência e Tecnologia, I.P., in the framework of the Project UIDB/00326/2025 and UIDP/00326/2025. The second author is funded by FCT - Fundação para a Ciência e Tecnologia, I.P., under the Ph.D. grant 2023.00740.BD and DOI identifier https://doi.org/10.54499/2023.00740.BD.

References

1. Amiri, M.J.: Object-aware identification of microservices. In: 2018 IEEE International Conference on Services Computing (SCC). pp. 253–256. (2018) *
2. Avila, D.T., de Moura, V.C., Thom, L.H.: Using machine learning to classify process model elements for process infrastructure analysis. In: Proceedings of the XIX Brazilian Symposium on Information Systems. pp. 45–52. (2023) *
3. Bandara, W., Gable, G.G., Rosemann, M.: Factors and measures of business process modelling: model building through a multiple case study. European Journal of Information Systems. 14 (4), 347–360 (2005)
4. Bano, D., Zerbato, F., Weber, B., Weske, M.: Enhancing discovered process models with data object lifecycles. In: 2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC). pp. 124–133. (2021) *
5. Bao, R., Cai, H.: Management of complex data objects in ship designing process. In: 2014 Int. Conf. Data Sci. Adv. Analytics. pp. 534–540. (2014) *
6. Bazhenova, E., Zerbato, F., Oliboni, B., Weske, M.: From BPMN process models to DMN decision models. Inf Syst. 83 69–88 (2019) *
7. Beisel, M., Barzen, J., Bechtold, M., Leymann, F., Truger, F., Weder, B.: Metamodel and Formalization to Model, Transform, Deploy, and Execute Quantum Workflows. In: Int. Conf. Cloud Comput. Serv. Sci. pp. 113–136. (2022) *
8. Beisel, M., Barzen, J., Bechtold, M., Leymann, F., Truger, F., Weder, B., others: QuantME4VQA: Modeling and Executing Variational Quantum Algorithms Using Workflows. In: CLOSER. pp. 306–315. (2023) *
9. Belluccini, S., De Nicola, R., Dumas, M., Pullonen, P., Re, B., Tiezzi, F.: Verification of privacy-enhanced collaborations. In: Proceedings of the 8th International Conference on Formal Methods in Software Engineering. pp. 141–152. (2020) *
10. Cabanillas, C., Resinas, M., Ruiz-Cortés, A., Awad, A.: Automatic generation of a data-centered view of business processes. In: CAiSE 2011. pp. 352–366. (2011) *
11. Chinosi, M., Trombetta, A.: BPMN: An introduction to the standard. Comput Stand Interfaces. 34 (1), 124–134 (2012)
12. Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F., Vandin, A.: A formal approach for the analysis of BPMN collaboration models. Journal of Systems and Software. 180 (2021)
13. Fosso Wamba, S., Mishra, D.: Big data integration with business processes: a literature review. Business Process Management Journal. 23 (3), 477–492 (2017)
14. Gómez-López, M.T., Pérez-Álvarez, J.M., Varela-Vaca, A.J., Gasca, R.M.: Guiding

the creation of choreographed processes with multiple instances based on data models. In: BPM 2016 International Workshops. pp. 239–251. (2017) *
15. Haarmann, S., Weske, M.: Cross-case data objects in business processes: Semantics and analysis. LNBIP. 392 (392), 3–17 (2020) *
16. Harzing, A.W., Alakangas, S.: Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics. 106 (2), 787–804 (2016)
17. Heguy, X., Zacharewicz, G., Ducq, Y., Tazi, S.: Interoperability Markers for BPMN 2.0 Making Interoperability Issues Explicit. In: Dubey, S.C., Anh, H.P.H., and Namazi, H. (eds.) EAME 2017. pp. 330–333. , Paris, France (2017) *
18. Heguy, X., Zacharewicz, G., Ducq, Y., Tazi, S., Vallespir, B.: A performance measurement extension for BPMN: One step further quantifying interoperability in process model. In: Enterprise Interoperability VIII: Smart Services and Business Impact of Enterprise Interoperability. pp. 333–345. (2019) *
19. Herzberg, N., Meyer, A., Khovalko, O., Weske, M.: Improving business process intelligence with object state transition events. In: ER 2013. pp. 146–160. (2013) *
20. Herzberg, N., Meyer, A., Weske, M.: Improving business process intelligence by observing object state transitions. Data Knowl Eng. 98 144–164 (2015) *
21. Kassahun, A., Tekinerdogan, B.: Adopting Workflow Patterns for Modelling the Allocation of Data in Multi-Organizational Collaborations. In: DATA. pp. 110–118. (2016) *
22. Khalajzadeh, H., Simmons, A.J., Abdelrazek, M., Grundy, J., Hosking, J., He, Q.: An end-to-end model-based approach to support big data analytics development. J Comput Lang. 58 100964 (2020)
23. König, M., Lichtenstein, T., Seidel, A., Weske, M.: Data objects with variables in BPMN. In: BPM. (2024) *
24. König, M., Lichtenstein, T., Seidel, A., Weske, M.: Introducing Variables to Data Objects in BPMN. In: International Conference on Enterprise Design, Operations, and Computing. pp. 137–154. (2024) *
25. König, M., Weske, M.: Multi-instance data behavior in BPMN. In: ER. (2023) *
26. Kopp, O., Binz, T., Breitenbücher, U., Leymann, F.: BPMN4TOSCA: A domain-specific language to model management plans for composite applications. In: International Workshop on Business Process Modeling Notation. pp. 38–52. (2012) *
27. Lichtenstein, T., Weske, M.: Data Consistency as a Criterion for Process Choreography Design. In: ER Forum/PhD Symposium. (2022) *
28. Lohmann, N., Nyolt, M.: Artifact-centric modeling using BPMN. In: Service-Oriented Computing-ICSOC 2011 Workshops, International Workshops WESOA, NFPSLAM-SOC, and Satellite Events. Revised Selected Papers 9. pp. 54–65. (2012) *
29. Marcinkowski, B., Gawin, B.: Beyond BPMN Data Objects–Method Tailoring and Assessment. In: Information Systems: Development, Applications, Education: 8th SIGSAND/PLAIS EuroSymposium 2015, Gdansk, Poland, September 25, 2015, Proceedings 8. pp. 89–99. (2015) *
30. Meyer, A., Polyvyanyy, A., Weske, M.: Weak conformance of process models with respect to data objects. In: Proceedings of the 4th Central-European Workshop on Services and their Composition. pp. 74–80. (2012) *
31. Meyer, A., Pufahl, L., Batoulis, K., Fahland, D., Weske, M.: Automating data exchange in process choreographies. Inf Syst. 53 296–329 (2015) *
32. Meyer, A., Pufahl, L., Fahland, D., Weske, M.: Modeling and enacting complex data dependencies in business processes. In: BPM 2013. pp. 171–186. (2013) *
33. Meyer, A., Weske, M.: Extracting data objects and their states from process models. In: 17th IEEE Int. Enterprise Distrib. Object Comput. Conf. pp. 27–36. (2013) *
34. Mülle, J., Tex, C., Böhm, K.: A practical data-flow verification scheme for business processes. Inf Syst. 81 136–151 (2019) *
35. Napolitano, E.V., Masciari, E., Ordonez, C.: Integrating Flow and Structure in Diagrams for Data Science. In: 2024 IEEE Int. Conf. Big Data. pp. 5769–5774. (2024)
36. Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of

Information Systems Research. SSRN Electronic Journal. (2012)
37. OMG: Business Process Modelling Notation Version 2.0. OMG Specification, Object Management Group. 19 52–60 (2011)
38. Peng, L.: A BPMN BASED SECURE WORKFLOW MODEL. In: International Conference on Enterprise Information Systems. pp. 268–272. (2009) *
39. Pérez-Álvarez, J.M., Gómez-López, M.T., Eshuis, R., Montali, M., Gasca, R.M.: Verifying the manipulation of data objects according to business process and data models. Knowl Inf Syst. 62 (7), 2653–2683 (2020) *
40. Proietti, M., Smith, F.: Reasoning on data-aware business processes with constraint logic. In: 4th International Symposium on Data-Driven Process Discovery and Analysis. pp. 60–75. Citeseer (2014) *
41. Recker, J., Rosemann, M., Indulska, M., Green, P.: Business process modeling-a comparative analysis. J Assoc Inf Syst. 10 (4), 1 (2009)
42. Ribeiro, V., Barata, J., da Cunha, P.R.: Modeling inter-organizational business process governance in the age of collaborative networks. Electronic Markets. 34 (1), 51 (2024) *
43. Rosenthal, K., Ternes, B., Strecker, S.: Business Process Simulation on Procedural Graphical Process Models: Structuring Overview and Paths for Future Research. Business and Information Systems Engineering. 63 (5), 569–602 (2021)
44. Seidel, A., König, M., Weske, M.: Towards Object-centric BPMN Process Models. (2024) *
45. Sholiq, S., Yaqin, M.A.: BPMN diagram dataset: Comprehensive collection of functional requirements. Data Brief. 57 110882 (2024) *
46. Take, M., Becker, C., Alpers, S., Oberweis, A.: Modeling the Integration of Machine Learning into Business Processes with BPMN. In: International Congress on Information and Communication Technology. pp. 943–957. (2023)
47. Tetzlaff, L.M.: BPMN4sML: A BPMN Extension for Serverless Machine Learning. Technology Independent and Interoperable Modeling of Machine Learning Workflows and their Serverless Deployment Orchestration. CoRR. abs/2208.02030 (2022)
48. Tsoury, A., Soffer, P., Reinhartz-Berger, I.: Towards impact analysis of data in business processes. In: BPMDS 2016, 21st International Conference, EMMSAD 2016, Held at CAiSE 2016. pp. 125–140. (2016) *
49. Tuyishime, A., Basciani, F., Di Salle, A., Izquierdo, J.L.C., Iovino, L.: Streamlining Workflow Automation with a Model-Based Assistant. In: 50th Euromicro Conf. on Softw. Eng. and Adv. Appl. (SEAA). pp. 180–187. (2024) *
50. Wagner, G.: Business process modeling and simulation with DPMN: Resource-constrained activities. In: 2020 Winter Simulation Conference. pp. 45–59. (2020) *
51. Wagner, G.: Process design modeling with extended event graphs. In: Proceedings of the 2019 Summer Simulation Conference. pp. 1–12. (2019) *
52. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly. 26 (2), xiii–xxiii (2002)
53. Zarour, K., Benmerzoug, D., Guermouche, N., Drira, K.: A systematic literature review on BPMN extensions. Business Process Management Journal. 26 (6), 1473–1503 (2019)

(*) papers included in the SLR sample