A Comprehensive Data Schema for Digital Twin Construction

Schlenger J.[1] , Yeung T.[2] , Vilgertshofer S.[1] , Martinez J.[2] , Sacks R.[2] , Borrmann A.[1]
[1]Technical University of Munich, Germany, [2]Israel Institute of Technology, Israel
jonas.schlenger@tum.de

Abstract. Lean construction, originating from lean management, aims to proactively improve the overall efficiency of construction processes. This requires continuous assessment of performance to identify key problems, which allows continuous improvement of construction processes. Novel digital twin approaches form an excellent technological foundation for performance assessment through regular updates with product and process information directly from the construction site. While some publications favour a schema-less approach to digital twins, we argue that well-defined data structures are required to represent complex information reliably and transparently. Existing process and product models are inadequate with respect to the requirements of a digital twin of the construction phase. As a result, we introduce a new process-oriented model that provides an improved basis for advanced process evaluation in the digital twin environment. This data schema, which is the main outcome of this paper, is presented in UML format together with a first approach to transfer it to an ontology usable in the Semantic Web context.

1. Introduction

In the last few decades, digital twins have become a rapidly growing field of interest in industry and academia in several domains. Just recently, they found their way to the construction sector. Here, a digital twin is understood as the digital replica of a real-world physical construction asset, updated at regular intervals to reflect any changes to the asset (Bolton et al. 2018). Many existing approaches focus on either the design phase, where BIM tools are applied to create digital prototypes, or the operational phase, where sensor data available on IoT platforms is evaluated to monitor parameters relevant to building operations (Jones et al., 2020). The construction phase, however, was missing a digital twin that gives a comprehensive overview of the entire project. Sacks et al. (2020) were among the first researchers to approach this topic holistically. They envision a digital twin of the whole construction project to gain situational awareness during the complete construction phase and thus, to support a full-cycle lean model of planning and control.

Although monitoring construction products is essential in building situational awareness, capturing the construction processes is at least as critical, and perhaps more so. Due to the dynamic nature and dependence of production activities on multiple input flows (e.g., components and materials, information, equipment, and availability of space), their execution in practice is often far from optimal. Accurate situational awareness is crucial for good production planning and control that can manage the input flows reliably (Koskela, 2000).

While some publications favour a schema-less approach to digital twins (El-Diraby, 2021; Miloslavskaya and Tolstoy, 2016), we argue that well-defined data structures are required to represent complex information in a reliable and transparent manner. This is indispensable for construction performance evaluation and assessment of the execution of construction processes. Currently, there is no standard data model established in the context of digital twins in construction (Akanmu, Anumba and Ogunseiju, 2021). Existing product and process model data schemata (buildingSMART International, 2021; Rasmussen et al., 2020; Bonduel, 2021; Zheng, Törmä and Seppänen, 2021) lack completeness with respect to the requirements of a digital twin of the construction phase.

The overall aim of this study is to introduce a model schema that allows easy comparison of the intent and status of a digital twin of a construction site, which sets the basis for calculating performance-related indicators like cycle times, work in progress, throughput, and others (Sacks, 2016). In contrast to existing models, the differences between project intent and status should be carefully addressed. The process model schema does not purport to give a complete view of the construction project but instead proposes a set of core classes that can be used for a wide range of digital twin use cases. In this way, it shall be used as a foundation to be further extended by partial models for domain-specific use cases. With the help of real-world monitoring data organized in this model, one can detect or even anticipate patterns of activities and thus work proactively to initiate timely countermeasures. The scope of the developed schema is limited to building projects. Linear infrastructure projects differ significantly in their process workflows and related data schema requirements. For this reason, they are outside the scope of this paper.

The theoretical background of the paper is set with an introduction to the use of Semantic Web Technologies in the civil engineering domain and the state of the art of process and product models in Section 2. Subsequently, the methodology applied to develop a process-oriented data schema is described in Section 3. This schema is presented in close detail in Section 4, comprising both the planned project intent and the project status of the actual execution, including the first steps on how to transfer the process-oriented model to Semantic Web Technologies. The paper closes off with a discussion on the model's limitations in Section 5 and a conclusion containing a summary and future works in Section 6.

2. Theoretical Background

2.1 Semantic Web in the Realm of Civil Engineering

The Digital Twin Construction, which refers to a mode of construction management supported by digital twins, defined by Sacks et al. (2020) adopts a holistic approach to the digital twin concept. This entails capturing data from multiple sources that vary highly in their nature and the way they are captured. Useful information about project status is obtained by fusing and interpreting several data streams simultaneously. In such a multi-data environment, interconnecting these various types of data in a meaningful way is challenging. Interoperability, in general, is a well-known issue in the civil engineering domain. The great number of stakeholders in a singular construction project presents a significant challenge in communication and data exchange and makes interoperability an essential factor in further optimization and improvement of the whole construction workflow. As a possible solution for data connection and interoperability, Semantic Web Technologies (SWT) comprising Linked Data are an active area of research in the construction sector (Pauwels et al., 2017).

Tim Berners-Lee (2009) introduced the Semantic Web concept with Linked Data as a subpart in the early 2000's. His idea lies in overcoming the disadvantages of the decentralized structure of storing data on the web by linking and sharing it. This is done using structured, directed graphs. Since its invention, the Semantic Web has established a set of standards and functionalities. Included are, e.g., OWL, SHACL, and SPARQL that allow knowledge inference, reasoning, rule checking, and data querying. Furthermore, RDF provides the basis to store, interlink, and exchange data in graph form. All of these standards are not only useful in the web context but can also be applied to use cases that require another form of interlinking data belonging to different sources (W3C, 2019).

In their literature review, Pauwels et al. (2017) discussed how the Semantic Web concept is applied in the AEC industry. Here, linking data across domains is identified as one of the three main advantages that justify SWT usage. The Semantic Web standards could provide a good foundation for interoperability in the digital twin environment that requires interconnected data. With the use of ontologies, data becomes machine-readable and machine-interpretable, which is a great advantage when working with big data.

2.2 Existing Process and Product Data Models

There are various already existing product and process data models tailored for the civil engineering domain. The most suited ones are introduced briefly and compared against the requirements for application in the process-oriented digital twin context.

The Industry Foundation Classes (IFC) developed by buildingSMART International (2021) are a well-known and internationally used standard for data exchange of construction-related information. It supports multiple serializations, with the most common one being the EXPRESS format for the definition of the IFC schema and STEP files for instantiation. Among others, a mapping to IfcOWL was developed, replicating the EXPRESS schema in an OWL ontology (Pauwels and Terkaj, 2016). One of the characteristics of IFC is the rich variety of geometry representations ranging from boundary representations to procedural descriptions including CSG and sweeping operations. The representation chosen for a concrete exchange depends on the purpose of the data handover: explicit representations are used for pure checking and analysis tasks, while procedural representations allow modifications on the receiving side.

The large number of possibilities to represent geometries make the IFC format a powerful, but extensive schema and increases its complexity significantly. Its complex structure and lack of modularity are some of the most criticized aspects (van Berlo et al., 2021). Regarding semantic information, IFC contains classes for comparatively fine-grained categorizations. A minimal set of classes is provided for process information, e.g., the *IfcTask*, which can be understood as any kind of construction process. Such a task can be connected to a respective building element, a set of other tasks in a specific order, or *IfcResources* representing the required resources of the construction task. According to the authors' experience, the process part of the IFC format is, however, barely used in practice. Furthermore, the IFC format exclusively considers information about the planning phase, so to say as-designed and as-planned data. There is no straightforward way to include as-built and as-performed data, which is essential for the digital twin context. Although simple processes can be represented in an ordered sequence, the construction digital twin environment demands representation of dynamic schedule changes, where task prerequisites play an essential role in rescheduling (Dori, 2015). In the current version of IFC, such task dependencies cannot be modeled appropriately. As the building element subpart of IfcOWL with some minor extensions, the Building Element Ontology (BEO) allows more flexible use but is still limited to product information (Pauwels, 2018).

Getuli (2019) developed an extension for IFC to support the scheduling construction processes better. His work is based on IfcOWL, which he extended with four new sub-ontologies for construction time, construction workspace, buildings, and construction scheduling. Even though they significantly improve the modeling of construction processes, the fundamental IFC issues - its complexity and focus on the as-planned side - persist. Additionally, the IfcOWL extension is not available online, making the reuse of the defined classes challenging.

In contrast to IFC, the Building Topology Ontology (BOT) by Rasmussen (2020) is a minimal ontology, exclusively focusing on the building elements and their subcomponents and the ways in which they relate to one another. With fewer than 25 classes and properties, it is meant to be combined with domain-specific data models. Its most important classes are the zones to define spatial areas, like buildings, spaces, and stories, the building elements contained in them, and the interfaces between them. It is not specifically designed for either as-designed or as-built construction, but can be used in both contexts. For Digital Twin Construction, it is clearly missing the process-related portion and further requires additional distinctions between as-designed and as-built information. Nevertheless, it provides high-level classes that can be the basis for a wide range of use cases.

Both BOT and IFC are stand-alone models, whereas Bonduel (2021) introduces an extensive ontology network. His framework builds a common foundation for combining data from different stakeholders belonging to build heritage use cases. He combines various existing ontologies with newly created ontologies to cover many aspects of built heritage-related information. Contrary to the IFC model, the focus lies on representing the as-built context and its decay over time. For building elements and zones, the BOT ontology is coupled with ontologies that allow to add geometric information to the building elements. Additionally, national and international taxonomies for furniture, building elements, and MEP elements are used for further specification. The Construction Tasks Ontology (CTO) is used for modeling processes, defining five types of tasks (maintenance, repair, inspection, removal, and installation). These classes are designed for maintenance tasks during the operational phase but are not well suited for the construction phase. The absence of definitions of construction resources and task prerequisites further affirms this statement. To differentiate between project intent and status, the context class is provided to group entities accordingly. However, it is highly questionable if the same set of classes can adequately describe both intent and status.

Finally, the Digital Construction Ontologies (DiCon) (Zheng 2021) are an ontology looking at construction and renovation tasks, with a focus on construction planning during the complete building lifecycle from construction to demolition. It considers the different types of flows that are relevant for construction processes. These are the labor performed by agents, construction equipment, workspace, construction components like building elements, information entities, external conditions, and prerequisite tasks. Overall, this enables a detailed description of process input and output. It also uses the context concept to represent as-planned and as-performed indirectly. On the other hand, the processes themselves are represented only in a generic way with so-called object activities, which can be any activity related to any type of entity. For more detailed process types, DiCon refers to OmniClass, Uniformat, and Talo2000. Nevertheless, this does not allow structuring of construction processes hierarchically. Like the built heritage BIM framework, the context class allows grouping object instances to either the as-planned or the as-performed side of the construction data.

To summarize, there are various already existing data models specific to the civil engineering domain. While the majority of them focus on construction products, some also represent construction management and processes. Although there are possibilities to assign data to a specific context (e.g., as-planned and as-performed context), they fail to address differences that do exist between data from the design and the construction phase. Currently, there is no data model that represents construction processes considering digital twin use cases and copes with varying requirements for project intent and status.

3. Schema Development

A schema development approach similar to Zheng, Törmä and Seppänen (2021) has been applied as a research methodology. It includes four steps for developing a data model: specification, knowledge acquisition, implementation, and validation. Although this approach is tailored to developing ontologies, it also suits data modelling purposes. First, the overall goals and requirements of the process model were specified. There are already product models in various complexities specific to the civil engineering domain like the Building Topology Ontology (BOT) and IFC, which is not the case for process models. Currently, no existing schema covers the core classes required for construction performance evaluation, including the construction intent as well as the construction status. Therefore, the goal was to fill this gap with a lightweight schema that is meant to be extended with domain-specific classes.

Furthermore, the authors decided to organize the overall data in three layers, according to Ackoff (1989). These three layers: data, information, and knowledge, can be seen as horizontal layers of a pyramid where the end-user value and conciseness of the data increase from bottom to top. For a digital twin in construction, the data layer will include point clouds, sensor data, pictures from the construction site, and other raw data, as it is generated by various devices. Information is then gained mainly through the statistical and arithmetical analysis of data, e.g., by extracting the as-built geometry of a building element from a large point cloud. Through further evaluation and interpretation, one acquires knowledge like key performance indicators (KPIs), delays in the construction schedule, and prominent issues that help construction managers to get a condensed overview of the project. The model presented below concerns the middle layer, the information layer. For the raw data layer, simple data structures already suffice. For this purpose, existing ontologies like SOSA and SSN can be reused to define monitoring devices together with the data they capture and additional metadata (Janowicz et al., 2019). The knowledge layer, however, requires global reasoning on the lower-level information, which first requires an adequate representation of the information layer.

With the overall goal set, a literature review was conducted to identify existing process and product models in the realm of civil engineering, as described in Section 2.3. Furthermore, the schema development was heavily based on the results from the online questionnaires and expert interviews executed by Torres et al. (2021). They identified an optimal construction workflow enabled by digital twin construction by interviewing a wide range of personnel involved in construction projects. The questions targeted the main inefficiencies of construction processes, possible ways to counteract these, and the data required from construction sites to monitor the status of construction processes in real-time. Their results were further used by Mediavilla et al. (2021) to derive main semantic concepts (with their relationships) in a top-down approach. We developed the digital twin construction data schema introduced in Section 4 in a bottom-up approach based on these two reports. During the development, close attention was put on identifying overlaps and differences with existing process and product models to align the new model with the current state-of-the-art.

4. Digital Twin Process Model as a Basis for Advanced Performance Evaluation

The data model for Digital Twin Construction is divided into two parts. On the one hand side, there are the classes related to the as-planned information (Figure 1). These describe the intent, which can be understood as the future state of the construction project formulated in plans and schedules. On the other hand, there are the classes relevant for the as-performed status of the project (Figure 2). Here, only the information about the present status of the executed processes are considered. Where as-planned and as-performed refer to processes, equivalent terms exist for product information. In this case, as-designed represents the product intent and as-built the product status. Since products result from the corresponding processes, they form a subpart of the respective as-planned or as-performed model side (Sacks, 2020).

Some classes are used on both sides of the model since they do not differ in their attributes, but only in the attribute values of the class instances. Others are represented through separate distinctly named classes because status and intent require differing attributes. Furthermore, additional classes only exist on one side of the model, e.g. defects of building elements are only part of the project status since they are not planned upfront. For this reason, the two sides of the model should be understood as two information containers that partially share classes.

Although the model is presented in two parts, there are clear connections between them that are omitted in Figures 1 and 2 for clarity. The classes *site*, *building*, *storey* and *space* form a bridge between as-planned and as-performed because they are not expected to be influenced by the process execution on-site but hold true for project status and intent. Furthermore, *resource* and *zone* form common parent classes for the context-specific subclasses. Moreover, every class from the as-planned side directly connects to its equivalent class on the as-performed side. Any deviation between the two would be stored in the knowledge layer because global reasoning could be required to assess the difference between an as-planned product and its corresponding as-performed process.

4.1 Core Data Model Classes

All in all, the model classes can be grouped into four main categories. First, there are the classes that represent the construction processes. Second, the resources that are used by the processes. Third, the products that are the end result of the processes, and fourth, the working zones where the processes are executed. This again shows the strong orientation for processes of the developed model. The following section explains the four groups of classes in more detail and discusses their differences between the as-planned and the as-performed model sides.

Processes: The processes are the central part of the model and can be found in the middle part of Figures 1 and 2. They are organized on three levels, starting on the left side with the most general level up until the most low-level information on the right-hand side. On the most general level, there is the *work package*. It holds information about the used construction method and can be seen as an aggregation of more detailed processes. On the level below, there is the *activity*. Every *work package* consists of multiple *activities,* where the *activity* describes one construction step to build one or a group of objects, like placing formwork or pouring concrete. Each *activity* is further dissected into *tasks*, where each *task* represents an *activity* related to a singular *building element* or *building element part*. *Preconditions* can be connected to *activities* and *tasks* that describe the requirements of a process to support the proactive make-ready steps of lean construction. There are various types of *preconditions*, e.g., a *zone* that needs to be available or another process that needs to be finished beforehand. All types of preconditions can be found in Figure 1 directly to the right of the process classes. Where the as-planned processes hold details about long-term averaged performance factors dependent on the construction company, the construction method, and the project's particularities (Hofstadler, 2007), the as-performed processes (*construction*, *operation*, and *action*) need to support short-term performance evaluation. Fine-grained insight into performance variation and process disruption allows the development of timely countermeasures to improve the overall construction performance.

Products: In terms of products, the model contains *building elements* and *building element parts* (see lower right corner of Figures 1 and 2). This possibility of decomposition should be highly oriented towards the corresponding processes, e.g. a wall with multiple layers that are constructed in separated steps should have its layers represented as *building element parts*. On the contrary, a window that might consist of various parts but is entirely installed by a single *task* should be represented by a single *building element* without multiple parts. However, the level of granularity of the just mentioned decomposition is one of the main modeling challenges and needs to be carefully chosen according to the present use case. The same applies to the granularity of the construction processes. The *building elements* are logically organized according to the overall spatial structure of the building with the classes in the lower left corner of both figures. Here the project is broken down into the *site*, one or multiple *buildings*, their *storeys*, and their *spaces*. Depending on its type, a *building element* is either associated to a *space*, e.g., a specific room or to a complete *storey*. None of these classes hold geometric information but only indicate the general building breakdown. There are no significant differences between the as-designed and as-built sides regarding required attributes. Therefore, the same classes are used for both sides of the model. To tell the class instances apart nevertheless, the Boolean attribute *IsAsDesigned* is introduced.

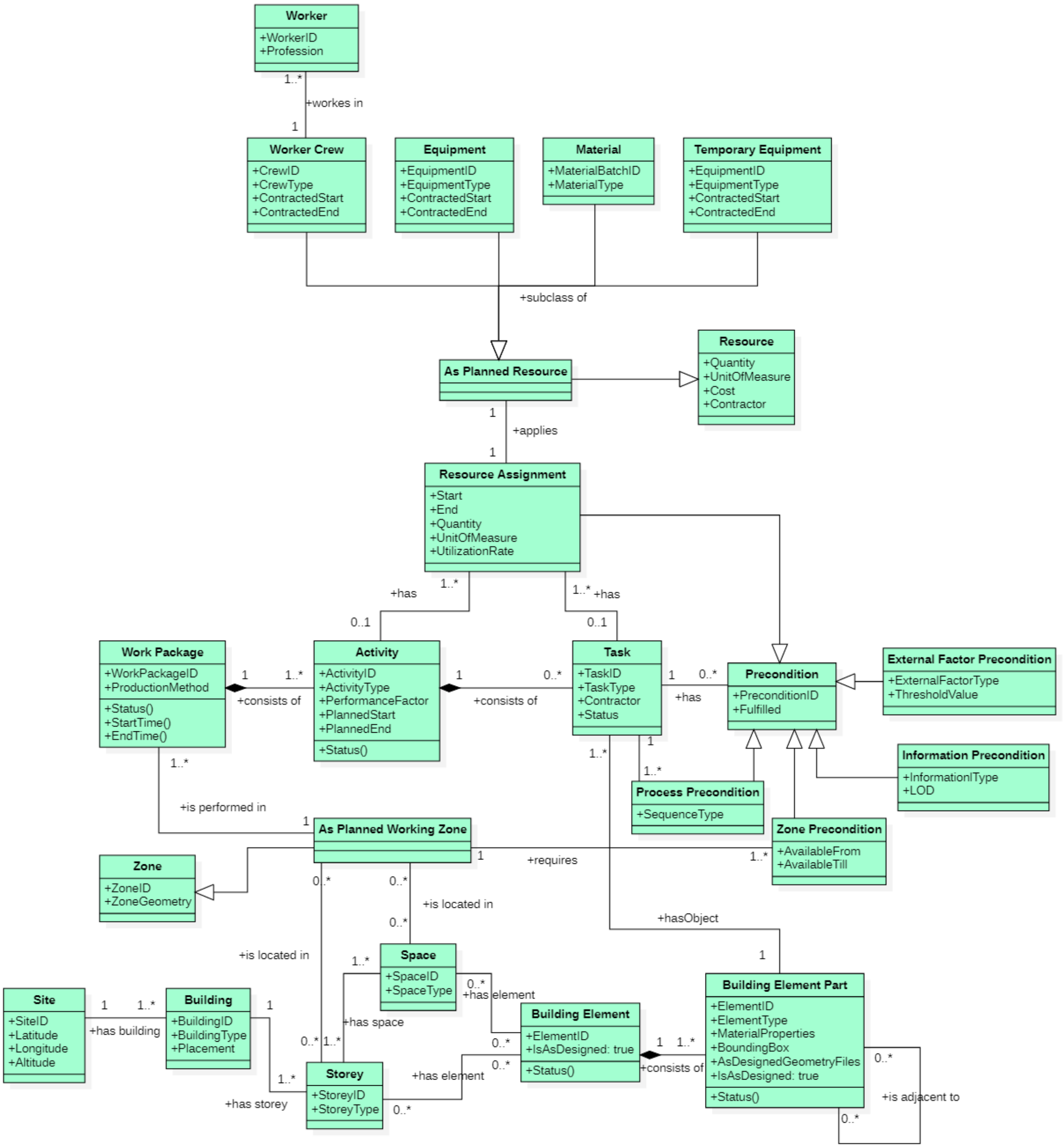

Figure 1: UML model of the project intent information (as-designed and as-planned).

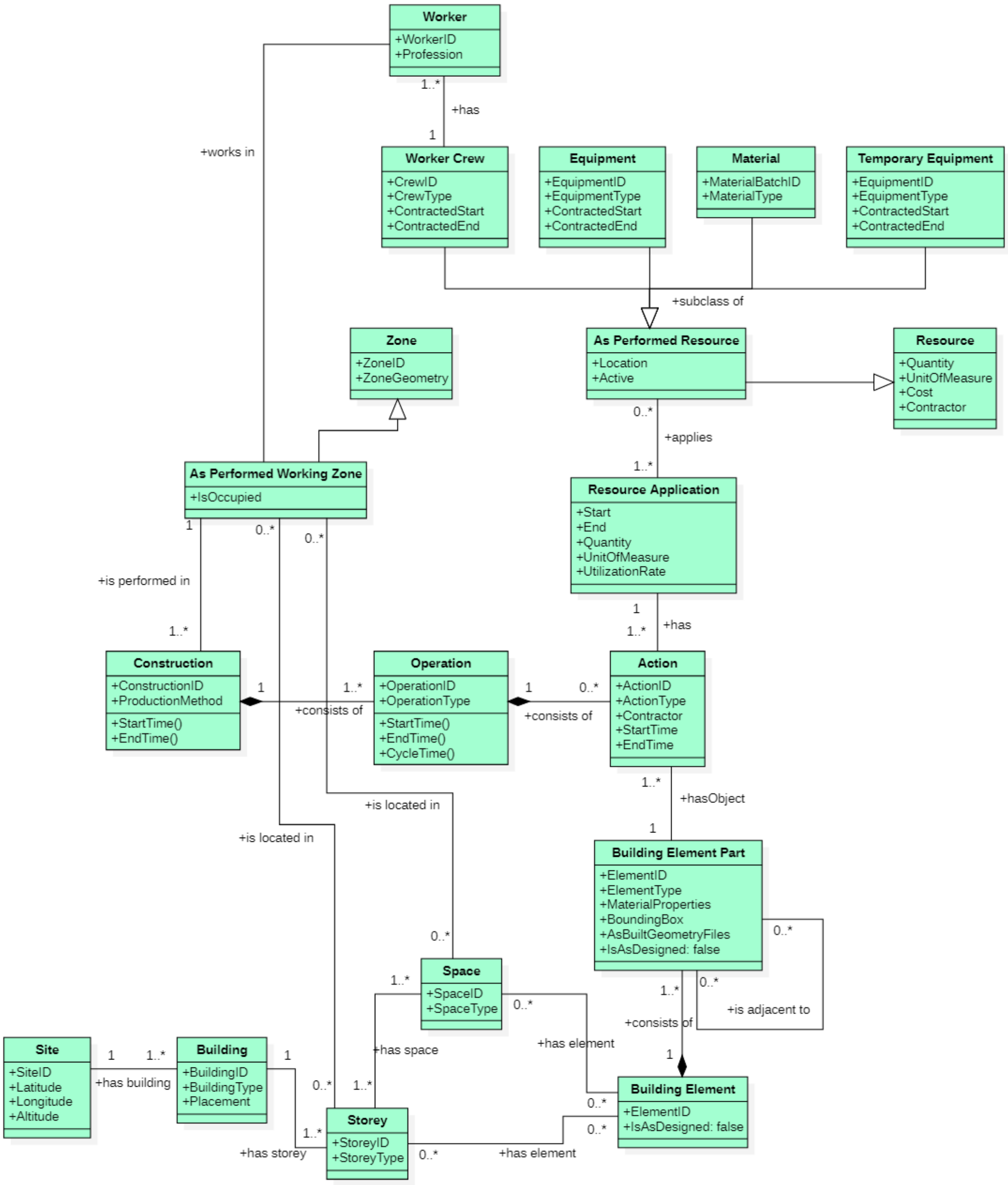

Figure 2: UML model of the project status information (as-built and as-performed).

Resources: In the presented model, the *resources* are all persons or physical things that are required during a process to build a specific *building element*. They are grouped into four classes. These are the labor force in the form of *worker crews*, *equipment* like heavy machinery and small tools, *materials*, and *temporary equipment* like formwork and guardrails. All resource classes, together with their parent classes, are located in the top section of Figures 1 and 2. It is essential to note the difference between the *resource assignment* and the *resources*. The *resource assignment* represents the *resources* required by a specific process, whereas the *resource* classes are used to model the amount of resources existing on the construction site (planned or performed). In this way, a *resource* instance can be connected to multiple *resource assignments*. The *resource assignment* is replaced with the *resource application* on the as-performed side of the model. The *resources* also differ between as-planned and as-performed because, during execution, exact location and current activity/ inactivity are available, which are not planned in this detail ahead of time.

Zones: Finally, the *zones* enable modeling of the location breakdown structure of the construction project. Every *zone* represents a spatial area where construction is executed. For this reason, the *zone* is directly connected to the processes. Unlike the *space* and *storey*, the *zones* do have geometric information. In some cases, a *zone* might be equivalent to a *storey*, or a *space*, but a direct relationship is not always given. Overall, the *zones* are an essential indicator of the construction flow because flow can be judged on the occupation rate of worker crew and flow of materials but also on the occupancy of working locations (Sacks, 2016).

4.2 BIM2TWIN Core Ontology

As discussed in Section 2.1, the Linked Data concept is suited well for the digital twin context, which requires uniting data from various sources. In addition, when sharing data across digital twins, the interoperability provided by the Semantic Web Technologies gains even greater importance. For this reason, the first steps for converting the presented data model into an ontology that can be used in the Linked Data context are provided. Reusing existing ontologies is a central idea in the Semantic Web. Therefore, one should evaluate thoroughly which classes can be reused from existing ontologies and which should be created newly.

The BIM2TWIN Core Ontology was developed completely based on the UML diagrams presented in Section 4.1. All classes, relationships, and class attributes were translated into the corresponding ontology classes, object properties, and data properties. Only domains and ranges of object and data properties were defined to allow other researchers and practitioners to reuse the B2T Core Ontology without restricting them too much in the application. The cardinalities were not integrated into the ontology to broaden the application range.

Within the B2T Core Ontology, several existing ontologies are reused. Most importantly, the lightweight BOT ontology, which closely resembles the structure of building elements organized in spaces, storeys, and buildings, is completely integrated into B2T Core. However, the process classes did not coincide with the definitions in existing ontologies and were modelled with entirely new classes. Finally, the resources required a mixed approach, with some classes reused from DiCon and others added newly. In terms of object and data properties, several other ontologies could be reused: Basis Geo WGS84 for spatial referencing, QUDT for representing units of literals, and OWL Time for all time-related attributes. While the first preliminary version of the B2T Core ontology is finished, further refinement regarding alignment with existing ontologies is required before the ontology can be published. Publishing the ontology in a well-documented form is the next goal, which will be part of future work.

5. Discussion

Even though a data model that serves as a basis for advanced performance evaluation was presented, this forms only the first step in reaching the goal of facilitating process assessment. The actual data structure to organize performance indicators is part of the knowledge layer of Ackoff's data pyramid, which the present paper has not touched. However, a solid foundation has been set in presenting a set of core classes that represent the essential parts of the information layer in the required level of detail. A data model alone, without any data, is useless. Since the data model breaks down processes in a fine-grained process network, it will be a significant challenge to collect sufficiently detailed process-related data from construction projects. Doing this in a mainly automated way without interrupting the ongoing construction process is still a great challenge that is not yet resolved in the current state-of-the-art.

Furthermore, finding the right granularity of processes and products is a task that the model leaves to the data modeler. Regarding the proposed ontology, an initial concept was introduced. To be used properly, it still needs further refinement and thorough documentation. According to adhere to Semantic Web principles only once this is done, can the ontology be published online.

6. Conclusion

Digital twin concepts together with lean construction principles are promising approaches to raise a project's efficiency and lift the design and operation phases but also the execution phase to the next level. Tackling the execution phase of building projects in a holistic approach will require clear data structures to be able to extract meaningful information and knowledge. Existing data models lack relevant aspects of the process analysis with a digital twin of construction. A thorough process description with input and output flows and process prerequisites is required for dynamic rescheduling. Also, the distinction between as-planned and as-performed processes is an essential part of it.

The present paper introduced a novel process model tailored to building projects that serves as a foundation for advanced construction performance evaluation, fulfilling all of the mentioned requirements. It defines a core set of classes that can be accompanied by domain-specific extensions. Moreover, a concept was developed to translate the process-oriented model into an ontology that can be used in the semantic web context. However, the data model has not been applied yet in a real-world digital twin scenario. Therefore, it will be the highest priority to carry out a test study that thoroughly evaluates the model's performance and compares it to other alternatives. Additionally, the model is limited exclusively to the execution phase. During the operational phase, a similar performance evaluation can be conducted. Extending the model by classes specific to the operational phase and allowing a smooth transition from the execution to the operation phase could also be interesting for future work. Further extension and adaption of the model should also be dedicated to linear infrastructure projects like rails, roads, and tunnels whose process workflows vastly differ from conventional high-rise buildings and are currently not covered. Finally, the first concepts to apply the data model to the semantic web context were executed. Still, they will require further work to result in a publicly available ontology that complies with linked data standards.

7. Acknowledgement

The research described in this paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 958398, "BIM2TWIN: Optimal Construction Management & Production Control". The authors thankfully acknowledge the support of the European Commission in funding this project.

References

Ackoff, R. (1989). From data to wisdom - Presidential address to ISGSR, June 1988. *Journal of Applied Systems Analysis*, Vol. 16, pp. 3–9.

Akanmu, A.A., Anumba, C.J. and Ogunseiju, O.O. (2021). Towards next generation cyber-physical systems and digital twins for construction. *J. Inf. Technol. Constr.*, Vol. 26, pp. 505–525.

Berners-Lee, T. (2009). Linked Data - Design Issues. Available at: https://www.w3.org/DesignIssues /LinkedData.html (Accessed Jan. 05, 2022).

Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A. and Makri, C. (2018). Gemini Principles. Apollo – University of Cambridge Repository, Cambridge, UK, https://doi.org/10.17863/CAM.32260.

Bonduel, M. (2021). A Framework for a Linked Data-based Heritage BIM. PhD thesis, Katholieke Universiteit Leuven – Faculty of Engineering Technology, Leuven, Belgium.

buildingSMART International (2021). IFC Schema Specifications Database Website, https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/, accessed November 2021.

Dori, G. (2015). Simulation-based methods for float time determination and schedule optimization for construction projects. PhD thesis, Technical University of Munich – Chair of Computational Modeling and Simulation, Munich, Germany.

El-Diraby, T. (2021). Can IFC (mentality) be the basis of digital twins? No. *Keynote talk at the 38th Int. Conference of CIB W78 2021.*

Getuli, V. and Capone, P. (2019). Ontology-based modeling for construction site planning: Towards an ifcOWL semantic enrichment. *36th Int. Conference of CIB W78*, pp. 701–713.

Hofstadler, C. (2007), Bauablaufplanung und Logistik im Baubetrieb. Springer-Verlag Berlin Heidelberg, 2007.

Janowicz, K., Haller, A., Cox, S., Le Phuoc, D. and Lefrançois, M. (2019) SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *J. Web Semant.*, Vol. 56, pp. 1-10.

Jones, D., Snider, C., Nassehi, A., Yon, J. and Hicks, B. (2021). Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.*, Vol. 29, pp. 36-52.

Koskela, L. (2000). An exploration towards a production theory and its application to construction. PhD thesis, Helsinki University of Technology, Espoo, Finland.

Mediavilla, A., San Mateos, R. and Torres, J. (2021). Definition of the digital workflows for the construction process (Deliverable 1.2). EU Horizon 2020 BIM2TWIN.

Miloslavskaya, N. and Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. *7th Annual Int. Conference on Biologically Inspired Cognitive Architectures*, vol. 88, pp. 300–305.

Pauwels, P. and Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, vol. 63, pp. 100–133.

Pauwels, P. (2018). Building Element Ontology. Available at: https://pi.pauwel.be/voc/buildingelement/index-en.html (Accessed Feb. 02, 2022).

Pauwels, P., Zhang, S. and Lee, Y.C. (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, Vol. 73, pp. 145–165, doi: 10.1016/j.autcon.2016.10.003.

Rasmussen, M.H., Lefrançois, M., Schneider, G.F. and Pauwels, P. (2020) BOT: The building topology ontology of the W3C linked building data group. *Semantic Web*, Vol. 12(1), pp. 143-161.

Sacks, R. (2016). What constitutes good production flow in construction? *Construction Management and Economics*, Vol. 34(9), pp. 641-656.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S. and Girolami, M. (2020). Construction with digital twin information systems. *Data-Centric Engineering*, Vol. 1(e14), pp. 1-26.

Torres, J., San Mateos, R., Lasarte, N., Gonzalo Pinto, H., Noiray, F., Alhava, O., Tual, M. and Velasquez, S. (2021). D1.1 - As-is analysis and end-user requirements. EU Horizon 2020 BIM2TWIN.

Van Berlo, L., Krijnen, T., Tauscher, H., Liebich, T., van Kranenburg, A. and Paasiala, P. (2021). Future of the Industry Foundation Classes: towards IFC 5. *38th Int. Conference of CIB W78*, pp. 123-137.

W3C (2019). Semantic Web Wiki. Available at: https://www.w3.org/2001/sw/wiki/Main_Page (Accessed Jan. 05, 2022).

Zheng, Y., Törmä, S. and Seppänen, O. (2021). A shared ontology suite for digital construction workflow. *Automation in Construction*, Vol. 132, pp. 1-23.