Developing a standardised approach of Asset Lifecycle Information Modelling for Semantic Digital Twins in the built environment

Rizal Sebastian, rizal.sebastian@tno.nl
TNO – The Netherlands Organisation for Applied Scientific Research, Delft, The Netherlands

Michel Böhms, michel.bohms@tno.nl
TNO – The Netherlands Organisation for Applied Scientific Research, Delft, The Netherlands

Bart Luiten, bart.luiten@tno.nl
TNO – The Netherlands Organisation for Applied Scientific Research, Delft, The Netherlands

Abstract

A Digital Twin comprises three key aspects, i.e. data management, integration of expert models, and simulation-based decision-making. This paper focuses on the research challenges regarding data management where the lack of interoperability is acknowledged as a key constraint for the wider uptake of Digital Twins. In ongoing research, a standardized approach of Asset Lifecycle Information Modelling (ALIM) is proposed to tackle the issues of data variety and uncertainties, big data repositories, and semantic data mining. ALIM facilitates the creation of Semantic Digital Twins (SDT) by capitalizing on Semantic Web and Linked Data technologies as well as open standardisation for modelling data across the lifecycle of building and civil infrastructure assets. This paper describes the ongoing development of ALIM and the outcomes of an exploratory case study where a set of ALIM ontologies for SDT have been developed.

Keywords: Digital Twin, Asset Lifecycle Information Modelling, Semantic Web, Linked Data

1 Introduction

1.1 Origin and rise of Digital Twins

The development and application of Digital Twins have been growing rapidly in research and practice. The term “Digital Twin” became widely known after the presentation of the concept in 2002 within the context of Product Lifecycle Management (Grieves et al. 2019). The origin of Digital Twins can be traced back to NASA’s Apollo space program where the twin concept was used for two identical space vehicles, one on earth and other one in the space. The space vehicle on earth can mirror, simulate and predict the conditions of the other one in space (Liu et al. 2021; Jones et al. 2020). Later, NASA gave a more detailed definition of Digital Twins concerning simulation-based systems engineering as part of the roadmap for modelling, simulation, information technology and processing (Shafto et al. 2012).

Digital Twins rise and penetrate various domains along with the fourth industrial revolution, better known as the Industry 4.0. The term “Industry 4.0” was coined at the industrial trade fair at Hannover Messe in 2011 and became globally acknowledged at the 2016’s World Economic Forum’s meeting in Davos, Switzerland (Pfeiffer 2017). Industry 4.0 followed a series of disruptive innovations that started with the first industrial revolution in late 18th century when steam power combined with mechanical production led to the industrialisation; the second revolution in late 19th century when electricity and assembly lines resulted in mass production;

and the third revolution in 1970s when electronics and IT combined with the globalisation greatly accelerated industrialisation (Davies 2015; European Parliament 2016). One of the emerging technologies in Industry 4.0 that becomes more available and affordable; hence, helps the growth of Digital Twins, is Cyber-Physical Systems (CPS). CPS allow for remotely sensing, real-time monitoring and controlling of devices; and therefore, provide a direct integration and synchronisation between the physical and virtual worlds. A Digital Twin leverages the synchronicity of the cyber-physical bi-directional dataflows and it can represent the physical reality at a level of accuracy suited to its purpose (Sawhney et al. 2020; Kritzinger et al. 2018; Boje et al. 2020; CDBB 2018; Lim et al. 2020).

1.2 Research challenges of Digital Twins in the built environment

The concept and applications of Digital Twins in the built environment have gained interest from researchers and practitioners. However, the further development and wider applications of Digital Twins face several key challenges, among them is the absence of a widely agreed definition of Digital Twins (Brilakis et al. 2019). Although a common definition is yet to be established, there is a shared understanding of a Digital Twin as a virtual model that is connected to the physical object – allowing for monitoring, analysing, predicting and controlling the condition, behaviour and performance of the object.

Digital Twins can be comprehended through its main usages and characteristics. Liu et al. (2021) and Jones et al. (2020) performed comprehensive reviews of 240 and 92 academic publications respectively. Based on this, three primary usages of Digital Twins have emerged: 1) for monitoring and analysing the conditions of the physical object; 2) for improving decision-making through engineering, numerical analyses and artificial intelligence; and 3) for managing the whole lifecycle of the physical object, including planning of maintenance activities. Within this context, a Digital Twin is characterized by the two ways of its relationship with simulations, i.e. a) a Digital Twin as the model that represents a system upon which varieties of simulations can be based; and b) a Digital Twin as the simulation of the system itself. As such, a Digital Twin is more than just a virtual model that contains static and dynamic data. It also integrates algorithms that describe the physical counterpart and can support decision-making on actions in the real world based on processed data and simulations (Kritzinger et al. 2018).

Based on this understanding, the technology architecture of a Digital Twin comprises: 1) data-related technologies; 2) high-fidelity modelling, and 3) model-based simulations (Liu et al. 2021). These three key aspects correspond to the three main layers of information value-chain in Digital Twins as envisioned by the Gemini Principles, i.e.: 1) data management; 2) sense making; and 3) decision-making (CDBB 2018). In the light of this, the research challenges of Digital Twins can be identified regarding:

1. Data management, which addresses quality, interoperability, sharing and processing of the data on which Digital Twins are based.
2. Integration of expert models, which addresses fidelity of the algorithms in the knowledge domain where Digital Twins are used.
3. Simulation-based decision-making, which addresses visualisation of the data and the objects along with decision support for the experts and end-users, as well as automated decision-making and feedback to the physical objects through actuators.

This paper focuses on the research challenges of Digital Twins in the built environment regarding data management. There is an urgency for tackling data management challenges since a number of digital twins at different scales and for a variety of purposes have already begun to appear, and many more will follow; however, at present only a few digital twins share data across the different objects' lifecycle phases, organisations, sectors, and geography locations. The lack of interoperability is acknowledged as a key constraint for a wider uptake of Digital Twins (CDBB 2018).

The data management challenges are inherent to the characteristics of Digital Twin data with large variety and uncertainties as well as the needs for big data repositories and semantic data mining. On data variety, numerous types of data from different sources are generated across the lifecycle of the object and its digital twin. The data is often structured in different formats. On uncertainties, Digital Twin data covers multiple aspects of uncertainty regarding the accuracy of virtualisation and continuous updates between the physical and virtual objects. The close one-to-one mapping between the physical and virtual object will be incomplete without addressing the uncertainties involved both in the physical observation and the digital modelling. Therefore, uncertainties quantification is not only important for delivering reliable results, but also important for the evolution of the Digital Twins over time. Finally, on the needs for big data repositories and semantic data mining, the data collected from various sources needs to be stored in data repositories to be accessed and processed to generate valuable information for analysis, simulation and decision-making. The large variety and amount of data over the objects' lifecycle result in even bigger and more complex data repositories that make data mining difficult (Singh et al. 2020).

1.3 Objective and structure of this paper

This paper addresses an applied research challenge in developing and implementing innovations in Digital Twins at the crossing of Building Information Modelling (BIM) and semantic technologies. It aims to introduce the Asset Lifecycle Information Modelling (ALIM) concept relying on W3C Semantic Web and Linked Data technologies and open standardisation for modelling data across the lifecycle of building and civil infrastructure assets.

This paper focuses on the vision of Semantic Digital Twins (SDT) in the built environment. It describes ongoing research based on the ALIM approach to tackle the data management challenges to create an SDT. It shows an example how linked data ontologies are developed based on Dutch (NEN) and European (CEN) standards.

The next section of this paper sets out a conceptual review of SDT and ALIM. Subsequently, this paper describes an exploratory case study on a Digital Twin of a highway bridge and presents the ongoing development of a set of ALIM ontologies for an SDT. Finally, conclusions are drawn based on the analysis, and the direction for follow-up research is discussed.

2 Conceptual review of SDT and ALIM

2.1 Progress towards Semantic Digital Twins (SDT)

The innovations leading to Digital Twins can be classified according to the data integration level. From this perspective, a progress can be observed from a Digital Model to a Digital Shadow, till a Digital Twin (Kritzinger et al. 2018; Liu et al. 2021):

- A Digital Model is a digital representation of a physical object without automated data exchanges between the physical object and its digital representation. The digital representation might include the description of the physical object and the simulation models; however, the data exchange, integration and update do not occur automatically. Therefore, a change in the state of the physical object has no direct effect on the digital counterpart, and vice versa.
- A Digital Shadow contains an automated one-way dataflow between the physical object and its digital counterpart. A change in the state of the physical object leads to a change in the digital object, but not vice versa.
- A Digital Twin integrates bi-directional dataflow. In this way, a change in the state of the physical object leads to a change in state of the digital object, and vice versa. The Digital Twin might also act as a controlling agent to the physical object or support improved decision-making, which creates the opportunity for imposing positive feedback onto the physical twin (CDBB 2018).

A clear understanding of the distinction between Digital Model, Digital Shadow and Digital Twin is important since their different natures and levels of data integration require a different set of technologies. The often-referred technologies for Digital Twins include, but not limited to, real-time communication protocols, Internet of Things, continuous simulations, cloud computing, and other technologies for Cyber-Physical Systems (CPS) that are addressed in the Industry 4.0 (Kritzinger et al. 2018).

Based on this understanding, this paper aims to clarify the progress of Digital Twin beyond Building Information Modelling (BIM). Until now, there is no commonly agreed answer to the question of 'What is the difference between BIM and Digital Twin?'; yet, most researchers agree that BIM fulfils the definition of Digital Model (Brilakis et al. 2019). Typically, BIM consists of a three-dimensional model of a building or civil infrastructure asset, containing the information of the properties of the objects or elements in the model. A BIM model as semantically rich digital representation of a physical object provides a good basis for setting up a Digital Twin. Relevant BIM-based information to be integrated and enriched in a Digital Twin can include, for instance, geometric models and changes in the building layout; monitoring data over the condition and degradation of structural components; and occupancy, usage, operational and performance information of the building or civil infrastructure. Extending BIM to Digital Twin requires the means and the solutions for capturing and processing real-world data and feeding it back into the model to create a circular information loop.

An important aspect here is the management of information throughout the asset's lifecycle. Thanks to the bi-directional dataflows, Digital Twins can overcome the limitations of BIM to capture dynamic information, update the digital model automatically, and perform simulations for decision-making throughout the lifecycle of a building or civil infrastructure asset. In some industrial sectors, like airplane engine manufacturing, the Digital Twin of the engine is updated immediately because the engineers need to know the exact condition of an engine in real-time. However, in the field of civil engineering, the crucial changes on the building or civil infrastructure assets usually occur in a much longer period. The update rate and frequency of Digital Twins should, therefore, be adjusted to suit its application purpose (Brilakis et al. 2019).

This paper builds further on the research by Boje et al. (2020) that envisioned Semantic Digital Twins (SDT) after analysing 196 academic publications in construction information technology. Through semantic modelling, SDT crosses beyond BIM and enables a Digital Twin to exploit the greater potentials of Internet of Things (IoT) and Artificial Intelligence (AI). SDT breaks through the constraints of static and closed data with recursive interoperability issues and opens the way towards a Linked Data paradigm. By relying on Semantic Web and Linked Data solutions, SDT can comprise the whole asset lifecycle information and make it machine-interpretable for AI.

2.2 Enhancement by Asset Lifecycle Information Modelling (ALIM)

Aligned with the paradigm shift from BIM to Semantic Web based approach, the Asset Lifecycle Information Modelling (ALIM) concept is introduced. The description and the term of "ALIM" were first used by TNO to emphasize the whole-lifecycle coverage of the information modelling for assets in the built environment, which stands in contrast to the usual division between information modelling during the planning until project delivery phase (also known as the Project Information Management or PIM) and information modelling of the asset during the operational and maintenance phase (also known as Asset Information Management or AIM) (Luiten and Böhms 2019; Luiten, Koehorst, Böhms 2020). ALIM underlines the integrated view in the ISO 19650, the international standard for managing information over the whole life cycle of a built asset using building information modelling. ALIM capitalizes on the power of Semantic Web and Linked Data technologies for the built environment in its three main thrusts:

1. Applying Semantic Web and Linked Data to the data, information and knowledge for whole lifecycle asset management of buildings and civil infrastructure.
2. Contributing to open standardisation at national, European and international level.
3. Incorporating the FAIR data principles to improve the findability, accessibility, interoperability, and reuse of digital assets (Wilkinson et al. 2016).

The term "Semantic Web" was coined by Tim Berners-Lee with the idea about the capability of analysing the data on the web –the content, links, and transactions between people and computers (Berners-Lee et al. 1999). The term "Linked Data" was coined in 2006, also by Tim Berners-Lee, in response to the fact that a surprising amount of data on the web was not linked. Berners-Lee (2006) proposed several rules to develop Linked Data. These rules have evolved into "the Five Stars to qualify Linked Data", i.e.

- 1 Star: data available on the web with an open license to be Open Data.
- 2 Stars: data available as machine-readable and structured.
- 3 Stars: data available as machine-readable, structured, in a non-proprietary format.
- 4 Stars: all the above plus the use of open standards from W3C (e.g. RDF and SPARQL) to identify things.
- 5 Stars: all the above plus linking one's data to the others' data to provide context.

Pauwels et al. (2017) made an extensive review of Semantic Web technologies in the Architecture, Engineering and Construction (AEC) industry. The use and increase of these technologies in AEC are driven by a desire to: 1) overcome the interoperability issues among software tools used in different disciplines, or at least to improve information exchange processes; and 2) to connect to various domains of application that have opportunities to identify the untapped valuable resources closely linked to the information already obtained in the AEC industry. There is a significant progress towards the development of a robust semantic structure and a well-organized semantic connectivity map beyond adopting software applications that simply display the geometric perspectives and three-dimensional views of a building or presenting the lengthy textual descriptions and spreadsheets of unstructured data. Industrial players and software developers have become more and more interested in organizing and sharing the semantics of building and civil infrastructure during their entire lifecycle.

ALIM goes further to define, tailor and manage the semantics of asset information, and to link this information to the geometry of a building or civil infrastructure. With Semantic Web and Linked Data technologies, various data can be combined by representing the information in structured graphs. This approach allows for efficiently linking and sharing of information of entirely different natures, for example BIM data, Geographic Information System (GIS) data, Facility Management (FM) data, material repositories, regulation data, cadastre data, and urban data. As a result, the development of software applications that rely on multiple data sources is within reach (Pauwels et al. 2017). Moreover, with the increased applications of sensing technologies in the built environment, Semantic Web and Linked Data technologies are bringing an added value to enrich the existing information models with sensing and monitoring data.

ALIM significantly contributes to the innovations in data modelling for Semantic Digital Twins (SDT). Data modelling has been used in many industries for more than three decades; however, the conventional approach to data modelling often lacks semantics during the development process. For this reason, the typical issues of Digital Twins, which are related to heterogeneous databases and interoperability, are difficult to be dealt with by the conventional data models. In various industries, including in the manufacturing, there is ongoing research on Semantic Web and Linked Data technologies for SDT which addresses: 1) the information flow and the need of efficient data management for digital twins; 2) a novel method to propose a minimum data structure to model the digital twin data; 3) the use of ontologies to define semantics, restrictions and data structures to apply digital twins in a certain domain; and 4) the user accessibility and database queries using digital twin domain ontology. Semantic Web introduces new ways of managing data and metadata, and maintaining a higher order of logical and conceptual schemas. Properties and values of assets can be defined by shared schema or ontology. An ontology is the explicit formal specification of concepts and their relations in a domain. When used in data management, ontologies can guide a Digital Twin to validate the domain data models by allowing interactions between various data which is held in different formats. As ontologies are semantically richer than databases, a Digital Twin ontology model will maintain the semantics of data and the concept definition throughout the object's lifecycle (Singh et al. 2020).

An SDT ontology model needs to contain the conceptual knowledge from a certain domain of application. In the built environment, until now the role of Semantic Web has only been considered as complementary or supporting to BIM (Pauwels et al. 2017). This paradigm will change with the development of an SDT where an ontology approach is considered more suited for the future compared to older standard file formats. The Semantic Web based approach has, therefore, become part of the UK's government strategy for defining and developing BIM Level 3 and beyond (Boje et al. 2020). As such, it is logical to adopt the Semantic Web approach to progress beyond BIM towards SDT.

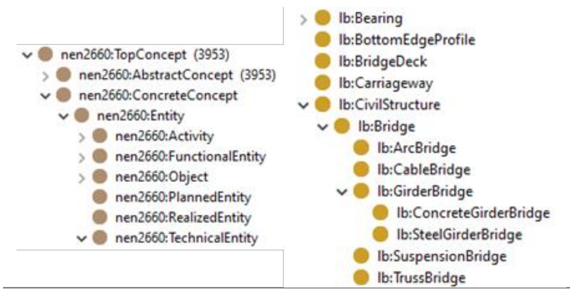
ALIM promotes and contributes to open standardisation for semantic modelling and linking data in the built environment (NEN 2660-2, 2021; CEN/TC 442, 2021). Within this context, interoperability is addressed by: 1) Model Standards to specify data structure for entities, geometry and related properties as well as classification for exchanging data models; 2) Data Dictionary Standards to specify data structure for defining data-semantic concepts (entity, property, classification, etc.) and relations between them; and 3) Process Standards to specify how to describe the required information supporting a given process (CEN-CENELEC 2021). Furthermore, in order to facilitate the widescale adoption of SDT, ALIM ensures the information in an SDT is findable, accessible, interoperable and reusable (FAIR) by relevant users and organisations in compliance with the EU General Data Protection Regulation (GDPR).

3 Exploratory case study

Using an exploratory case study of a highway bridge in the Netherlands, the ALIM approach was applied and a bridge ontology was developed according to the upcoming Dutch and European standards. The broader objective of the case study was to develop a Digital Twin prototype of the bridge for Structural Health Monitoring (SHM), predictive maintenance simulations, and impact analysis of the changing traffic loads and vehicle configurations (such as truck platooning where two or more trucks move in convoy using connectivity technology and automated driving support systems).

Through the case study, this paper shows how ALIM approach was used to define an ontology to structure Digital Twin data of the bridge and to make the data eligible for further simulations and analysis. The case methodology contained the following steps: 1) Choosing the appropriate linked data language and format (also called 'serialisation'); 2) identifying and model the relevant concepts; 3) relating those concepts in a specialisation hierarchy (also known as assigning the taxonomy); 4) identifying and modelling the relevant attributes; 5) identifying and modelling the relevant (inter)relations; 6) identifying and modelling the relevant constraints especially the decomposition constraints (giving the meronomy); and 7) populating the ontology for the real bridge. In this case study, it was decided to use Ontology Web Language (OWL) as linked data modelling language, and Turtle as the human-friendly format. The preliminary results from the case study are summarized in Table 1.

Table 1. Preliminary results of ALIM in the highway bridge case study

Result representation / example	Result description
	A small fragment of the defined girder bridge ontology is shown here; as an example, the SteelGirderBridge is a subclass of respectively a GirderBridge, a Bridge and a CivilStructure. The CivilStructure itself is a specialisation of the NEN 2660 top level model concept 'TechnicalEntity'.

`lb:designLifespan`
`rdf:type owl:ObjectProperty ;`
`rdfs:range nen2660:QuantityValue ;`
`nen2660:hasQuantityKind quantitykind:Time ;`
`nen2660:hasUnit unit:YR .`
	All relevant attributes and relations are defined; an OWL code example for a designLifespan (the planned lifetime of the bridge) is shown here.

https://w3id.org/liggerbrug/def/designLifespan or, in abbreviated from: 'lb:designLifespan'. The prefix 'lb' stands for: <https://w3id. org/liggerbrug/def#> making the total id/name for the property designLifespan.	The Turtle code is a convenient shorthand for a set of triples of the form 'subject-predicate-object'. The predicate object part is shared for the same object. Each component of the triple is an Internet/WWW URI consisting of a 'name space URI' represented by a prefix and an actual id/name.

The implementation of open standards in the Digital Twin ontology modelling can be summarised as follows. The draft Dutch NEN2660 standard (NEN 2021) aligned with the European TC442 Semantic Modelling and Linking (SML) standard (CEN/TC 442 2021) were then used for defining both the ontology and its data instances. The modelling patterns were applied for:

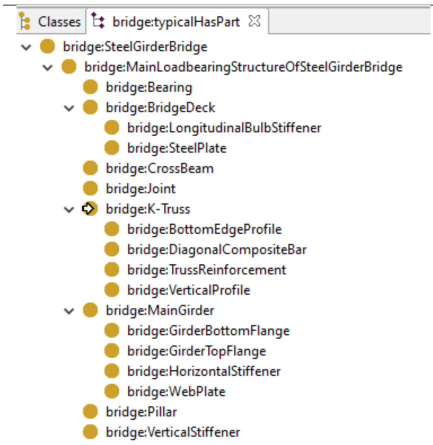
- modelling the quantities with their quantity kinds and units according to Quantities, Units, Dimensions and dataTypes (QUDT - public repository, https://github.com/qudt/qudt-public-repo);
- modelling the enumeration types with allowed items;
- defining the approach for identifiers (including an URI-strategy) and annotations like labels and definitions.

The languages from W3C were applied according to the approach envisaged by NEN and CEN. In the case study, this approach was applied for measurement and other datasets, and it resulted in standardised data for the bridge, comprising a standard format (Turtle) and standard semantics (bridge ontology) specified in standard languages (OWL) according to NEN2660. This standardised data is accessible through the standard W3C Linked Data Query Language (SPARQL) and can be used for various Digital Twin functionalities. The attribute modelling was done according to NEN 2660 as a relation (or in terms of OWL: an object property). Its quantity-kind was 'Time' and its measuring unit was a year ('YR'). Quantity kinds and units were fully reused from the most recent QUDT ontology. The preliminary results are summarized in Table 2.

Table 2. Preliminary results of ALIM in the highway bridge case study

Result representation / example	Result description

`prob:StochasticValue`
`a owl:Class ;`
`rdfs:subClassOf [`
`a owl:Restriction ;`
`owl:cardinality "0"^^xsd:nonNegativeInteger ;`
`owl:onProperty prob:posteriorMean ;`
`] ;`
`rdfs:subClassOf [`
`a owl:Restriction ;`
`owl:cardinality "0"^^xsd:nonNegativeInteger ;`
`owl:onProperty prob:posteriorStdev ;`
`] ;`
`rdfs:subClassOf [`
`a owl:Restriction ;`
`owl:cardinality "1"^^xsd:nonNegativeInteger ;`
`owl:onProperty prob:priorDistributionType ;`
`] ;`
`rdfs:subClassOf [`
`a owl:Restriction ;`
`owl:cardinality "1"^^xsd:nonNegativeInteger ;`
`owl:onProperty prob:priorMean ;`
`] ;`
`rdfs:subClassOf [`
`a owl:Restriction ;`
`owl:cardinality "1"^^xsd:nonNegativeInteger ;`
`owl:onProperty prob:priorStdev ;`
`] ;`
`skos:prefLabel "StochasticValue"@en .`
	The range of this attribute is a QuantityValue (defining a complex value that can hold an actual value plus extra metadata). Because of the application of Bayesian methods, a neutral stochastic value is also defined and used to show how some constraints look like in in this instance. A stochastic value is a complex value involving prior and posterior mean-values with standard deviations.
	The various constraints are modelled in OWL by making a concept a subclass of the set of concepts having the constraint. So, for instance, a StochasticValue is a subclass of the class of things having exactly one prior mean value (in bold lines shown here).

	When applying such a constraint to decomposition relations, a typical decomposition structure also referred to as a meronomy can be modelled and visualized. The taxonomy and the meronomy are together often seen as the 'backbone' of the ontology. With a simple SPARQL-query, explicit typical decomposition relations can be inferred from the instance decomposition constraints resulting the hierarchy shown here.

```
ib:IJsselbridge_1-main_detailed
  rdf:type lb:SteelGirderBridge ;
  lb:amountOfSpans 5 ;
  lb:constructionMethod lb:WithGirders ;
  lb:designLifespan [
      rdf:type nen2660:QuantityValue ;
      rdf:value 80.0 ;
      nen2660:hasUnit unit:YR ;
    ] ;
  lb:length [
      rdf:value "295.0"^^xsd:float ;
      nen2660:hasUnit unit:M ;
    ] ;
  lb:materialType lb:Steel ;
  lb:residualLifespan [
      rdf:value "20.0"^^xsd:float ;
      nen2660:hasUnit unit:YR ;
    ] ;
  nen2660:hasPart ib:MainLoadbearingStructureOfIJsselBridge .
```

When all concept, attributes, relations and constraints are modelled, one can instantiate the ontology for instances of bridges, called individuals in 'linked data speak'. Here is some global data as example that is fully compliant (i.e. logically consistent) to the ontology with all its constraints.

The concept or ontology design decision that was taken in order to implement and maintain the live data connection between the digital and physical twins can be summarised as follows:

- A dataset was selected to describe the design breakdown, attributes and relations of the elements in the semantic digital twin case study of the steel bridge. Anomaly data values were then generated to contain the deviations between the real conditions and the design data. These values were based on the comparison between the prior assumptions and the posterior values derived from the repeated finite element analyses after the actual data was taken into account. Both prior and posterior data as well as the mean and the standard deviations were stored in the digital twin datasets. For the Bayesian 'learning' ('updating of the knowledge') the actual measurements under specific traffic load conditions were used.
- An example of the ontology design in the case study for the analysis for 'estimating the thickness of the bottom flanch of the steel bridge that has been reduced by corrosion' is shown below:

```
aib:ThicknessReduction_1
  a anomaly:ThicknessReductionOfBottomFlange ;
  anomaly:forPhysicalObject ib:GirderBottomFlange_west ;
  anomaly:globalEffect false ;
  anomaly:hasCause aib:Corrosion_1 ;
  anomaly:parameterType anomaly:Deterministic ;
  anomaly:thickenssReductionBottomFlange [
     prob:posteriorMean [
        rdf:value "8.2E5"^^xsd:float ;
      ] ;
     prob:posteriorStdev [
        rdf:value "1.1E5"^^xsd:float ;
      ] ;
     prob:priorDistributionType prob:NormalDistribution ;
     prob:priorMean [
        rdf:value "0.03"^^xsd:float ;
      ] ;
     prob:priorStdev [
        rdf:value "0.0006"^^xsd:float ;
      ] ;
   ] ;
.
```

- In a intended next phase of the case study, the sensing data from different measurement points at the steel bridge will be stored as linked data according to the standard ontologies like W3C's SSN/SOSA.

At the time of writing of this paper, a further study on the ALIM approach is ongoing, especially regarding the developed ontologies and datasets for the integration of the relevant software functionalities. The neutral data of the bridge in this case study typically requires data transformations to and/or from the processing software applications through translations (transforming the format) and conversions (supporting the classification towards alternative semantics offered by the existing software applications). In the intended follow-up the case study, the linked data technologies will be combined with a standard cloud-native web service development, Docker containers, GraphQL interfacing and orchestration. They should be implemented in a Common Data Environment (CDE) with open-source and commercial-of-the-shelf tools, including for the semantic RDF Tripe Store. As such, the ALIM-based structured data can be used for a digital twin and for multiple inter-connected digital twins. Along with this, the concept of data lakes for data ingestion, transformation, federation and discovery is potential to resolve the big data issues at a low cost.

A CDE is recommended to support the semantic-based concept with proper data governance and technology integration reforms. The reference architecture for the CDE would facilitate seamless workflows between the data from the physical objects, and the software services and applications to perform analyses, simulations and decision-makings. The CDE development can reflect on the progressive view of Boje et al. (2020) that includes a 3-tier generation: 1) starting from 'monitoring platforms' as the first step in the real-time integration of sensing data and digital models; 2) going on to 'intelligent semantic platforms' which are enhanced monitoring platforms with Semantic Web technologies; and 3) moving forward towards 'agent-driven socio-technical platforms' that accommodate fully semantic digital twins, leveraging acquired knowledge with the use of AI-enabled agents.

4 Conclusions and recommendations

4.1 Conclusions

This paper clarifies that a Digital Twin is not merely a digital copy containing the data of the physical object, but it also incorporates the expert knowledge required for analysing the condition, behaviour and performance of the object. The extent of the data, the detail of the information, and the frequency of the bi-directional updates through the inter-connectivity between the physical and virtual worlds are determined by the purpose and application area of Digital Twins.

At present, the first and foremost challenges of Digital Twins in the built environment are concerned with data management. These challenges are associated with the large variety and high uncertainty of the data, the difficulties of big data repositories, and the complexities for semantic data mining. These challenges must be resolved before the other challenges regarding the encapsulation of expert knowledge and the activation of simulation-based decision-making can be dealt with.

This paper endorses the vision of Semantic Digital Twins (SDT) and proposes a standardized approach to data management in SDT known as Asset Lifecycle Information Modelling (ALIM), which is based on Semantic Web and Linked Data technologies and open standardisation. ALIM brings a unique added value to tackle the data structuring and modelling complexities in Digital Twins, especially for the purpose of managing building and civil infrastructure assets throughout their whole lifecycle. After describing the ALIM concept and its progress beyond BIM, this paper presents the preliminary outcomes of a case study where ALIM ontology is developed to create a Semantic Digital Twin (SDT) of a bridge.

4.2 Recommendations

As Singh et al. (2020) pointed out, semantic data mining for the converging behaviours of the physical and virtual objects in a digital twin is still an open area of research. SDT with its semantic-based data modelling and knowledge graphs, is considered as the direction for further development. Graph databases store entities and their relationships as nodes, edges and attributes. As such, retrieving the edge between two entities do not involve time-consuming 'join operations' that are typical for relational databases where data is structured in a table format.

Simultaneously, pursuing open standardisation is essential both for the further development as well as for the wider uptake of SDT in the built environment. Standardisation is among the preconditions in the Industry 4.0 (Davies 2015; European Parliament 2016). Standardisation will stimulate openness and collaboration between the stakeholders in the built environment. Along with open standardisation, a multidisciplinary effort to build up and strengthen the scientific body of knowledge of SDT is needed to avoid the risks of fragmentation and vendor lock-in if the public knowledge of Digital Twins is only driven by market opinions. The scientific underpinning of SDT will also contribute to the positioning of the future SDT within the anticipated Industry 5.0 where the SDT will suit a people-centric approach (European Parliament 2021).

References

Boje, C., Guerriero, A., Kubicki, S., Rezgui, Y. (2020). Towards a Semantic Construction Digital Twin: Directions for future research. *Automation in Construction.* 114 (x).

Berners-Lee, T., Fischetti, M. (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor.* Harper, New York.

Berners-Lee, T. (2006). W3C Design Issues – Linked Data (date 2006-07-27), https://www.w3.org/DesignIssues/LinkedData.html, last accessed 2021/03/21.

Brilakis, I., Pan, Y., Borrmann, A., Mayer, H.G., Rhein, F., Vos, C., Pettinato, E., Wagner, S. (2019). Built Environment Digital Twinning. *Report of the International Workshop on Built Environment Digital Twinning.* TUM Institute for Advanced Study and Siemens AG.

buildingSMART International (2020). Enabling ecosystem of digital twins. Positioning paper, buildingSMART International.

CDBB (2018). *The Gemini Principles.* Centre for Digital Built Britain, Cambridge.

CEN-CENELEC (2021). Webinar: Building Information modelling (BIM) supports digitalization of standards for the Construction sector, 2021-02-15.

CEN/TC 442 (2021). prEN 17632:2021 Building Information Modelling (BIM) – Semantic Modelling and Linking (SML). Draft for Public Enquiry. CEN, Brussels.

Davies, R. (2015). Industry 4.0: Digitalisation for productivity and growth. Briefing document. European Parliamentary Research Service, Brussels.

European Commission Directorate-General for Research and Innovation (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. Policy brief. European Commission, Brussels.

European Parliament Directorate General for Internal Policies (2016). *Industry 4.0. Report.* European Parliament, Brussels.

Grieves, M. (2019). Virtually Intelligent Product Systems: Digital and Physical Twins. In: Flumerfelt, S., Schwartz, K.G., Mavris, D., Briceno, S. (eds.). *Complex Systems Engineering: Theory and Practice.* American Institute of Aeronautics and Astronautics, Reston, VA.

Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology.* 29 (x). pp. 36–52.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. *IFAC PapersOnLine.* 51(11). pp. 1016–1022.

Lim, K. Y. H., Zheng, P., Chen, C.H., Huang, L. (2020). A digital twin-enhanced system for engineering product family design and optimization. *Journal of Manufacturing Systems.* 57(x). pp. 82–93.

Liu, M., Fang, S., Dong, H., Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems.* 58 (x). 346–361.

Luiten, G.T, Böhms, H.M. (2019). The CEDR-INTERLINK approach towards Asset Information Management. In: Workshop in Life Cycle Management, 23-29 October 2019, Rotterdam.

Luiten, G.T, Koehorst, B., Böhms, H.M. (2020). Beyond BIM: multi-standard digitalization. In: InfraBIM Open Conference, 3-5 February 2020, Tampere.

NEN 2660-2 (2021). *Rules for information modelling of the built environment - Part 2: Practical configuration, extension and implementation of part 1 (in preparation).* NEN, Delft.

Pauwels, P., Zhang, S., Lee, Y.C. (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction.* 73(x). pp. 145–165.

Pfeiffer, S. (2017). The Vision of "Industrie 4.0" in the making – a case of future told, tamed, and traded. *Nanoethics.* 11(x). pp. 107–121.

Sawhney, A., Riley, M., Irizarry, J. (eds.) (2020). *Construction 4.0: An Innovation Platform for the Built Environment (1st ed.).* Routledge, London.

Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., Wang, L. (2012). *Modeling, Simulation, Information Technology & Processing Roadmap.* NASA, Washington DC.

Singh, S., Shehab, E., Higgins, N., Fowler, K., Reynolds, D., Erkoyuncu, J.A., Gadd, P. (2020). Data management for developing digital twin ontology model. *Proceedings of the Institution of Mechanical Engineers.* SAGE, New York.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data.* 3(1).