Developing Digital Twin Data Structure and Integrated Cloud Digital Twin Architecture for Roads

Weiwei Chen, PhD.;[1,2], Ioannis Brilakis[1]
1 Laing O'Rourke Centre of Construction Information and Technology Laboratory, University of Cambridge, UK. Email: wc349@cam.ac.uk, ib340@cam.ac.uk
2 Bartlett School of Sustainable Construction, University College London, London, UK.
Email: weiwei.chen@ucl.ac.uk

ABSTRACT

The systems that support construction and maintenance planning for delivering the road digital transformation lack the up-to-date road data and digital processes to leverage them. Digital Twins recently disrupted the infrastructure sector and can now serve as a fundamental building block to help address this problem. Digital twin differs significantly from a static design-stage BIM (Building Information Modeling) model. It can span across the road's lifecycle for reflecting the physical twin's properties, updating data continuously, and improving connectivity. However, no implementation of such digital twin solutions exists on the horizon for infrastructure. The important item that has not been defined yet is workable digital twin data structures and cloud architecture. This study aims to develop a proof-of-concept digital twin data structure and an integrated cloud architecture that are able to support road products and processes. The outcomes of this research are the proposed data structures and developed road digital twin integrated cloud architecture. As reliable references, these data structures benefit people by applying digital twin technologies to facilitate road construction and O&M processes and improve the decision-making processes.

KEYWORDS: Digital Twins, Data Structures, Integrated Cloud Architecture, Roads

1. INTRODUCTION

Recently, Digital Twins (DT) is now able to serve as fundamental building blocks to help address this problem. The DT is not a new concept but has recently started to break through the infrastructure sector. According to the Centre for Digital Built Britain (CDBB), a Digital Twin is "a realistic digital representation of assets, processes or systems in the built or natural environment" (Department for Transport, 2019). DT technology enables to build of a digital representation of road assets/highways for operation and maintenance management. Technology providers have rushed to capitalize on this and retool their building information modelling (BIM) systems into DT platforms after some improvements (e.g. Bentley Systems iTwin) or made their cloud platforms DT compatible (e.g. Microsoft Azure DTs). A DT differs significantly from a DT-friendly cloud service or a static design-stage BIM model; it is meant to span across the asset's lifecycle, be dynamic (always updated), reflect the physical twin's properties (replication), and be connected to it (connectivity) (Sacks, et al., 2020). CDBB and others have recently defined the built environment DT concept and its principles at a high level (CDBB, 2020).

Road authorities are aiming to tackle this challenge in part by preparing to start upgrading A road into a highway by 2040 (Design Manual for Road and Bridge, 2020). A highway is a high-speed dual carriageway that has at least two lanes in each direction with grade-separated junctions and technology to support operational regimes. Authorities have systems to support expansion and maintenance planning to deliver the conversion to highway, but their challenge lies in the lack of up-to-date road data along with the digital processes and automation needed to leverage them.

Many researchers have done studies on data collection and geometry generating for digital twins. Various devices are used to collect data from projects in the physical world to their digital entities. In the field of construction and civil engineering, laser scanners, LiDAR devices, RGB cameras, near-infrared cameras, sensors, and GPS devices are employed to collect geometric and semantic information from a target project in the physical world to create the digital twins (Lu and Brilakis, 2019), detect defects (Omer et al., 2019; Ariyachandra and Brilakis, 2020) , assist the asset management and monitoring (Xie et al., 2020) , maintain the existing infrastructure (Angieliu et al., 2020), including roads/highway, buildings, bridges, and others. However, most DT applications focus on data collection, data segmentation, data classification, and DT model generating, rather than data structures, data transmission, data fusion and DT cloud architecture. Currently, we cannot properly leverage digital tools to service our infrastructure needs without a resilient data backbone. The key considerations are (i) understanding what exactly is a DT in this context and how should it be designed, and (ii) understanding how should a digital twin be constructed, maintained and operated.

In addition, there is no working implementation of such a DT solution on the horizon for infrastructure (Davila et al., 2018). DT data structures and integrated cloud architecture have not yet been defined. Therefore, at the fundamental level, we need to derive a proof-of-concept DT data structure and cloud architecture able to support road product data and road process data.

The paper illustrated the developed digital twin architecture and the data structure for data transfer between different layers. Section 2 presents the methodology for the developed digital twin architecture and the data management services, which is followed by the presentation of its data structure in Section 3. Section 4 presents the data model including the ontology and the UML data model. Finally, the developed approach is summarised in the conclusions.

2. METHODOLOGY

2.1 THE DEVELOPED CLOUD-INTEGRATED DIGITAL TWIN ARCHITECTURE

In order to develop the digital twin cloud architecture, the paper analysed the technical requirements defined to derive the Digital Twin (DT) representation to be developed in the road project. The short name of the road project is OMICRON. Furthermore, the work presented in this paper considered the different sources of inspection data (including legacy systems and inspection systems) and the interfaces with the Decision Support Tool (DST). In addition, the deliverable considered other inputs from the different information flow in many Use Cases in the OMICRON project. In summary, this document defines a conceptual architecture for the research project as a whole.

In order to derive the final specifications of the DT, a formal information analysis process was conducted to analyse the requirements defined. Two methods were used to analyse the information. In a top-down approach, the task used the outputs to consider innovative DT workflows and define the partial data models of the DT that mirror real road assets in all relevant aspects of the road asset geometry and enable its comprehensive analysis. In a bottom-up approach, the output generated was considered.

For the processing of low-level data into high-level information, the necessary processing chains were defined. The outcomes include the specification of the Digital Twin representation (UML data models), the specification of workflows and processing chains (BPMN), the specification of different layers of abstraction (data, information, knowledge) as well as a description of the diverse required sub-models on these layers and the processing paths between the layers.

Based on the above This section presents OMICRON's concept for road Digital Twin architecture. Figure 1 shows the proposed Cloud Integrated Digital Twin Architecture for the Digital Twin. This includes information from all the technology use cases in OMICRON (including the Decision Support Tool) and hence it represents architecture at the project level.

The architecture is organized according to seven different layers:

a. The physical layer.
b. The integration layer.
c. The database management layer.
d. The digital twin layer.
e. The service layer.
f. The user interface layer.
g. The user layer.

The **Physical Layer** aims to collect data during the whole lifecycle of the road, from design and construction to operation and maintenance. The data is collected in different ways:

a. Inspection. Including both evolution of the conventional inspection methods and advanced inspection methods, such as laser scanning inspections using LiDAR scanners, photogrammetry methods using RGB cameras, near-infrared (NIR) cameras, Unmanned Aerial Vehicles (UAV), GPS devices, etc.
b. Monitoring. Through the use of sensors, CCTV and GPS to monitor the real-time condition of roads.
c. Maintenance tasks information from the control centre and other work packages (WP4).
d. Traffic information connecting to V2X communications.

The **Integration Layer** oversees the management of static and dynamic data, data coming from the asset management databases and geometric data and semantic data from BIM models.

In the Data Management Layer, different types of data are managed to support the Digital Twin instances and classes:

a. Asset management databases are used to store construction records and maintenance records in relational databases.
b. Documents such as design documents, construction records, periodical survey reports, maintenance records, and legacy data are stored in the asset management database in different formats such as pdf, excel, word, dwg, etc.
c. Static data is stored in relational databases.
d. Dynamic data measured is to be integrated into the DT by an MQTT Broker while the data stored in the time-series database will be managed by an Azure Function. Data that is not measured over time will be stored in relational databases (RDBMS).
e. Ontology instances, represented in RDF, XML or OWL formats, have to be stored in a semantic repository database.

f. Meanwhile, 3D models created based on point cloud data and RGB images, which are collected using LiDAR scanners and RGB cameras respectively, are to be stored in the Internet of Things (IoT) database or Cloud Platform (Microsoft Azure). This data is visualized following Machine Learning (ML) and computer vision algorithms which are applied to segment the point cloud to identify the elements and objects of roads for 3D reconstruction.

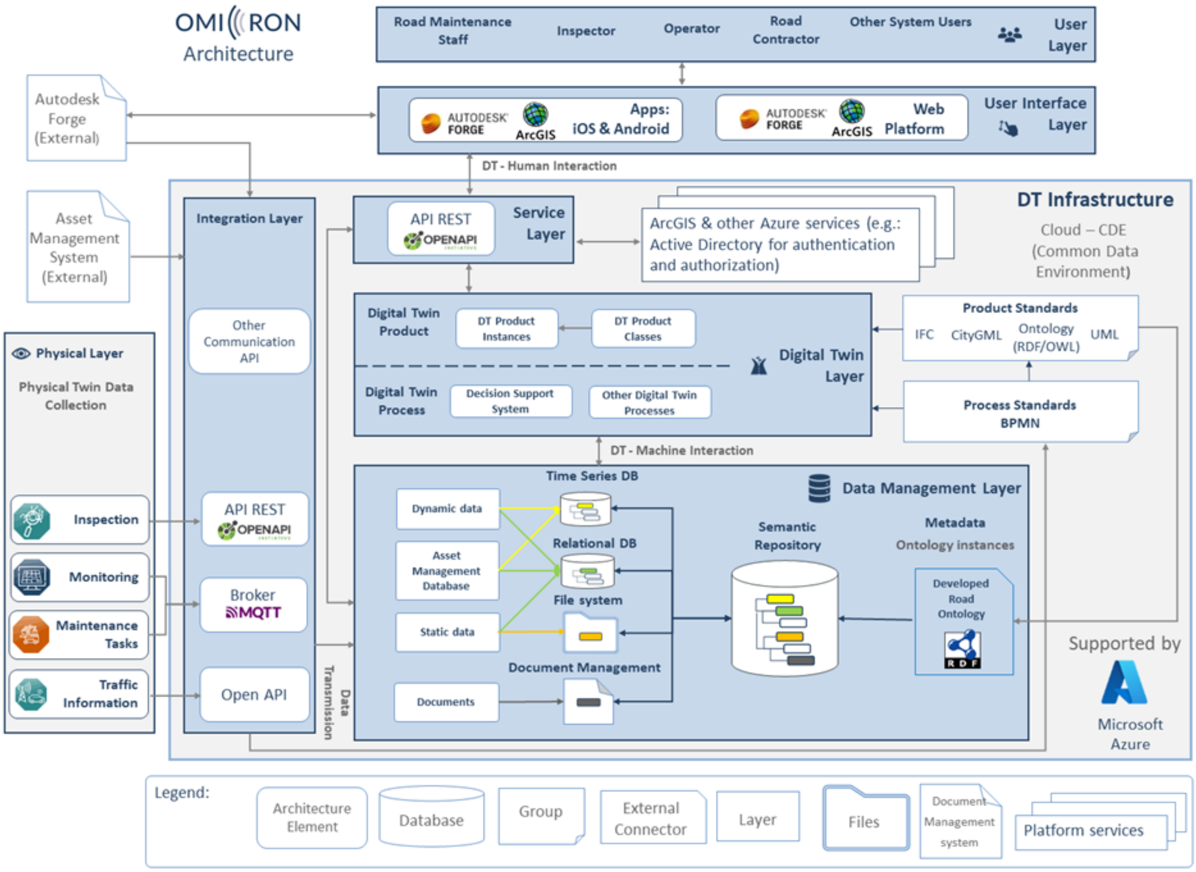

Figure 1 The Cloud Integrated Digital Twin Architecture

In the **Digital Twin Layer**, the fundamental issue is the standards of DT products and DT processes. Various data standards for data representation were considered, including IFC, CityGML, Ontology and UML to represent the specification of data. Based on the product standards, the platform is able to represent the DT product instances and DT product classes. In the process standards section, the specification of workflows and processing chains, using BPMN (Business Process Modeling Notation) to visualise the Digital Twin processes, will be detailed.

The **Decision Support System (DSS)** analyses data from the DT to find patterns using Machine Learning algorithms, providing the decision-making functions for road maintenance. In particular, the intelligent decision support tool addresses infrastructure condition analysis and asset management planning optimization for road infrastructure maintenance. The workflow for the DSS includes:

a. Processing data from the DT.
b. Assessing infrastructure degradation (and failure) risk.
c. Analysing and predicting infrastructure conditions.
d. Suggesting maintenance plans.

The DSS takes into account different sources of uncertainty, such as asset status prediction and other real operational environment factors. The DSS processes the data from the DT with AI techniques to predict future conditions and set predictive maintenance planning. Two modules conform to the DSS: the infrastructure condition analysis and the asset management plan optimisation module.

The infrastructure condition analysis module represents a methodology to define and predict road asset states. The methodology is based on AI techniques and segmentation methods focusing on three main stages:

a. Analysis and prediction of infrastructure condition.
b. Road sectioning according to condition state.
c. Evaluation of asset degradation risk, based on severity levels and criticality.

This infrastructure condition analysis module assigns to road sections a related degradation risk in order to be used in the optimisation of the asset management plan.

The asset management plan optimisation module has the objective of developing a maintenance planning functionality to improve the availability and reliability of the infrastructure while maintaining the levels of service and the safety of the road infrastructure. It also enables a higher level of automation in decision-making through advanced optimisation techniques. The maintenance plans adapt to short, medium and long-term planning procedures, considering emergency, routine and extraordinary intervention levels and different road assets. This module computes:

a. Prioritisation of interventions, considering the previous results from the infrastructure condition analysis.
b. The road sections in terms of safety and impact on traffic.

Based on this information, the DSS uses advanced optimisation techniques, maintenance plan simulations and comparison via what-if scenarios to optimise asset conditions given monetary and resource constraints.

Finally, the **User Interface Layer** provides the graphical user interface 3D and 2D model visualization and information interoperability. The platform can ensure user-friendly and convenient remote control.

2.2 THE DIGITAL TWIN LAYER DEVELOPMENT

In the cloud DT architecture, the core is the development of a digital twin layer to support the connection of different layers. The creation of the Digital Twin Layer within the road project follows the methodology shown in Figure 2 and explained below.

Step 1. The reference data models for roads are compared to select the proper reference model for the project. The IFC 5 road reference model, CityGML reference model, land and infrastructure conceptual model standard and iCity transport planning standard were compared.

Step 2. After comparison, it was found that the IFC 5 and CityGML data models were the appropriate ones to be used as the baseline.

Step 3. Based on the data collected from the Physical Layer, including traffic information, maintenance tasks, monitoring data, and inspection data, the BPMN models were developed to illustrate the specification of workflows and processing chains of relevant activities in the Digital Twin.

Step 4. According to the BPMN process, the entity, properties, relationship, and data properties needed to be clarified, so as to create the road ontology in Protégé.
Step 5. Turn the created ontology into RDF and OWL data format.
Step 6. Populate the RDF/OWL data to enable a link to other databases, and then publish the RDF into a graph database.
Step 7. The ontology enables the system to query the information from the graph DB, because it is linked to other databases, including time series databases, relational databases, PostgreSQL and Influx databases.

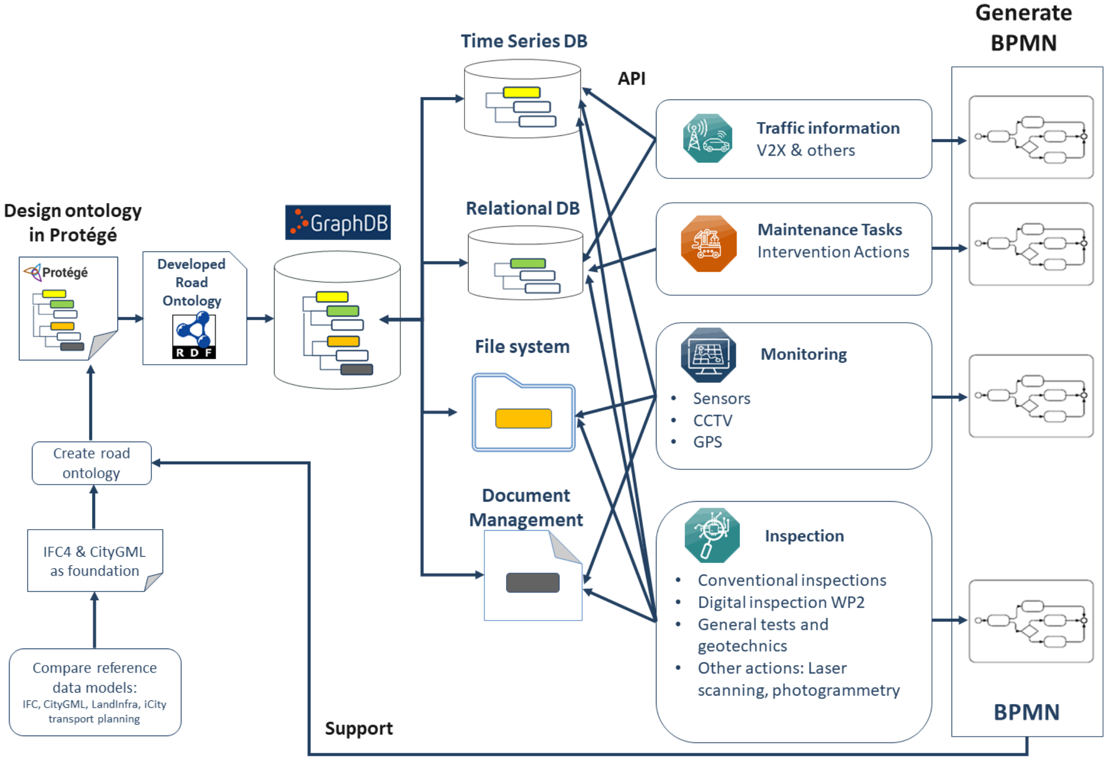

Fig 2 The Methodology for the Creation of the Digital Twin Data Process

The methodology is comprehensive and can be used for other similar construction projects, especially complex megaprojects for the digital twin application analysis and data process analysis.

3 DATA MANAGEMENT AND DATA PROCESS

3.1 Data Management Layer

In the digital twin architecture, the data management layer is very important to add value to decision-making. The Data Management Layer of the Digital Twin defines the type of data and the repositories to store the data, based on the different sources of inspection, monitoring, maintenance and traffic data to be integrated into the Digital Twin and identified in the analysis of the Use Cases in the research project.

The types of data identified are the following: Metadata, Static data, Dynamic data, Asset Management data, and Documents.

The Digital Twin has to process heterogeneous data that come from different sources and therefore have different natures and formats. Depending on these formats there are optimised technologies for their management. Data management aims to use the appropriate technologies to facilitate the management of each type of data while obtaining optimum performance. The types of repositories defined to store the project data are the following ones: 1) Semantic Repository for metadata, 2) Time Series DB for dynamic data, 3) Relational DB for static, dynamic and asset management data, 4) File System for static data wrapped in files, and 5) Document management for documents.

The following sections describe in more detail what each data type is and why the repository types have been selected. Figure 3 shows on the left the type of data identified coming from the Physical Layer, in the middle the repositories to store the data, and on the right, it is represented the Metadata and the Ontology in charge of modelling the project Digital Twin.

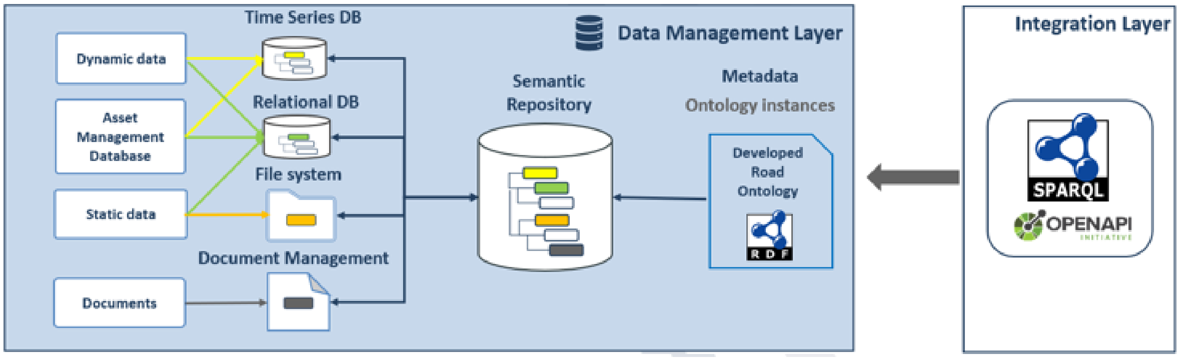

Figure 3 Data Management Layer Architecture and Repository relationship with API Layer

The Integration Layer will oversee the two main methods of data access that are performed against the repositories, data insertion and data access. On the one hand, the data insertion will be performed by the respective protocol, and libraries provided by each of the repositories. The data access will be a two steps process; the first will be a request to the semantic repository by a SPARQL query and the second one will be a request to the specific repository.

3.2 Semantic Repository in Use

The integration between the data provided by Physical Layer and the Data Management Layer is implemented by the Integration Layer, which will oversee the data insertion and the data request.

The procedure for accessing the stored data is defined in the following sequence (Figure 4):

a. The OpenAPI method in charge of getting the data to be provided to the DT executes the corresponding SPARQL defining for answering the Competency Question in charge of providing the information for the specific method.
b. The semantic repository returns the data associated with the requested SPARQL. One of the fields returned is the identifier of the repository and the entity where the raw data is stored. The OpenAPI will use the UML model corresponding to the data requested to compose the data coming from the different repositories.
c. The OpenAPI method requests data to the specific repository the data. For example, if the raw data is stored in a Relational database, then a SQL command is executed against the relational DB server. In the case of data stored in the time-series DB, then the method executes the request to the database by the use of the corresponding API provided by the database.

d. Finally, the OPENAPI method with all the data acquired, on the one hand from the semantic repository and on the other hand from the corresponding repository composes the full response message to be delivered to the application that executes the request.

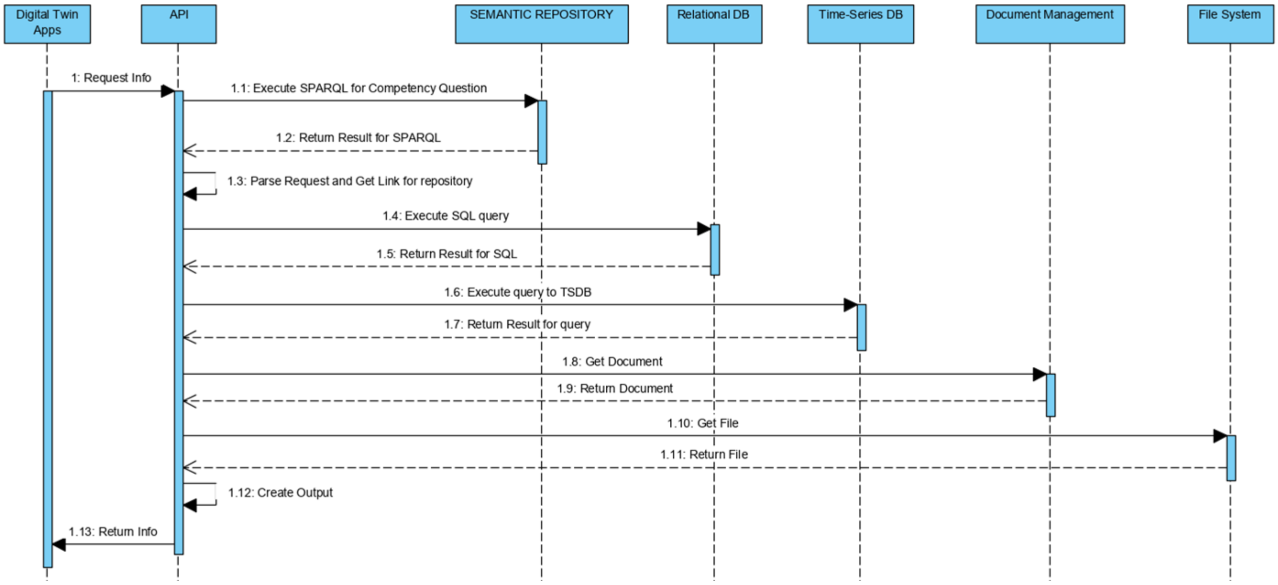

Figure 4 Access Data Sequence Diagram

4. DATA MODEL CREATION

As stated earlier, it is important to note that the ontology is in progress. The project is working on entities, objects, properties and relationships, and hence this paper only shows a partial representation of the ontology. After the creation of the ontology, it is to be converted to RDF/OWL/UML data format and then set in a graph database. The system will query information from the DT graph database to get the required information. The graph database will be linked to other databases to get the information needed. In this respect, Figure 5 shows a preliminary ontology graph description for road assets in a graph database.

The ontology data model needs to be represented in the UML data model for information querying and management. IFC 4.3 for Road Elements defines UML data models based on the specification of ISO 6707-1 (ISO 6707-1, 2017). IFC 4.3 defines the basic elements of the road such as the road itself (IfcRoad), the pavement (IfcPavement), signals (IfcSignal), some elements for security, road guard elements, such as the safety barrier, and line markings. IfcRoad is defined as a route built on land to allow travel from one location to another, including highways, streets, cycle, and footpaths, but excluding railways. As a type of Facility, a road provides the basic element in the project structure hierarchy for the components of a road project (i.e. any undertaking such as design, construction or maintenance). In this case, the Road element (IfcRoad) is a subclass of the facility element defined (IfcFacility). This package defines the Annotations, Spatial elements, Physical elements and Systems that could exist in a road infrastructure. Figure 6 illustrates the UML model is converted from ontology data model and based on the IFC standards.

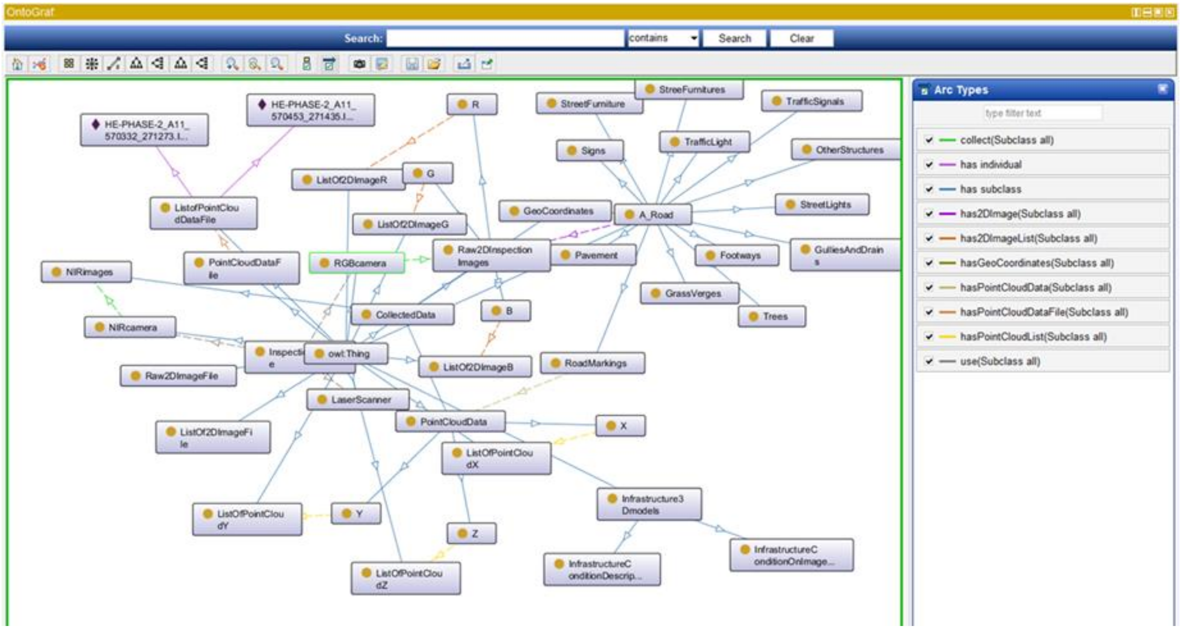

Figure 5 Preliminary road asset ontology representation for road assets

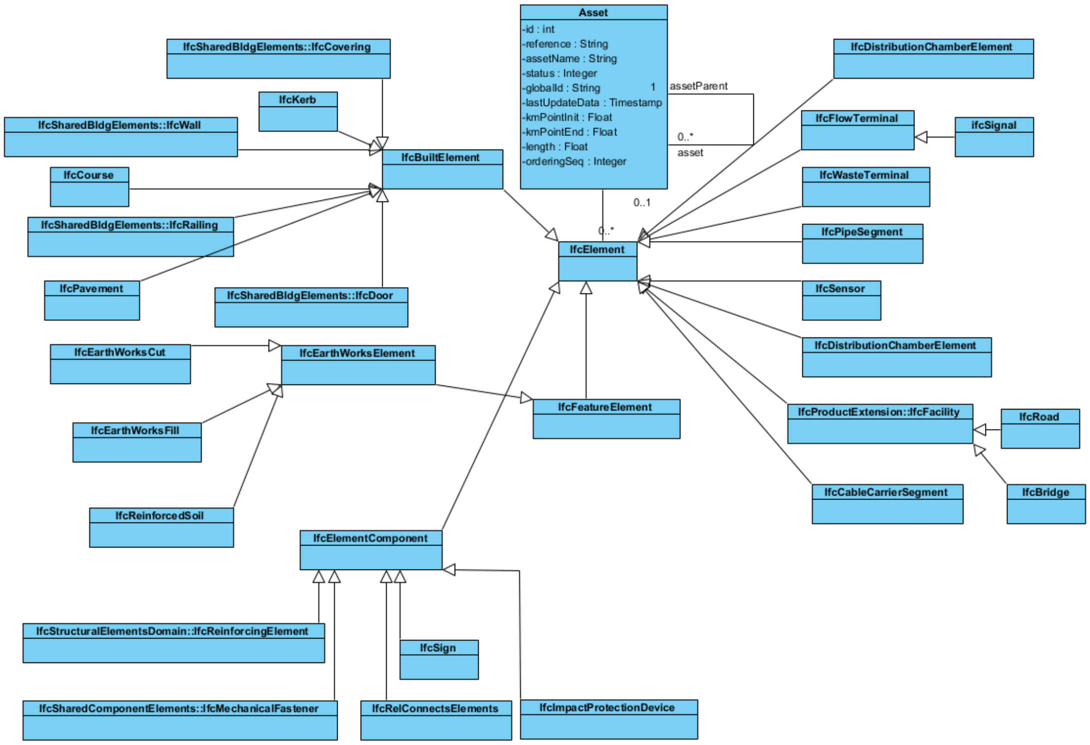

Figure 6 The UML data model based on the ontology and IFC

CONCLUSIONS

This paper developed the cloud-based integrated digital twin architecture and data models for data management in digital twin applications. The Digital Twin architecture has been presented showing connections to all the technologies of the project, serving as a project-level architecture.

Different levels have been defined and explained in the architecture, including a physical layer, a data management layer, an integration layer, a Digital Twin layer, a service layer, a user interface layer and a user layer. In particular, the paper described the digital twin layer and data management layer and proposed the data model and data structure for digital twin implementation. In the future, the implementation of the DT architecture into a real functional system would be studied for validation and improvement.

ACKNOWLEDGE

Thanks to the research project H2020 OMICRON and grant (No. 955269) for supporting this study. Thanks to my research collaborators and colleagues for their support and collaboration throughout this project.

REFERENCES:

Centre for Digital Built Britain (2020). The approach to delivering a National Digital Twin for the UK.

Davila Delgado, J.M., Butler, L.J., Brilakis, I., Elshafie, M.Z.E.B. and Middleton, C.R. (2018). Structural Performance Monitoring Using a Dynamic Data-Driven BIM Environment. Journal of Computing in Civil Engineering, 32(3), May 2018, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000749

Design Manual for Roads and Bridges (2020). GD300 Requirements for new and upgraded all-purpose truck roads (expressways). Highways England.

Department for Transport (2019). Transport Statistics: Great Britain 2019.

G. Angjeliu, D. Coronelli, G. Cardani, Development of the simulation model for digital twin applications in historical masonry buildings: the integration between numerical and experimental reality, Comput. Struct. 238 (2020), https://doi.org/10.1016/j.compstruc.2020.106282

ISO 6707-1:2017, Buildings and civil engineering works https://www.iso.org/standard/72244.html

M. Omer, L. Margetts, M. Hadi Mosleh, S. Hewitt, M. Parwaiz, Use of gaming technology to bring bridge inspection to the office, Struct. Infrastruct. Eng. 15 (10, 2019) 1292–1307, https://doi.org/10.1080/15732479.2019.1615962.

M.R.M.F. Ariyachandra, I. Brilakis, Detection of railway masts in airborne LiDAR data, J. Constr. Eng. Manag. 146 (9) (2020), https://doi.org/10.1061/(ASCE) CO.1943-7862.0001894.

R. Lu, I. Brilakis, Digital twinning of existing reinforced concrete bridges from labelled point clusters, Autom. Constr. 105 (2019), https://doi.org/10.1016/j. autcon.2019.102837.

Sacks, R., Girolami, M. and Brilakis, I. (2020). Building Information Modelling, Artificial Intelligence and Construction Tech.

X. Xie, Q. Lu, D. Rodenas-Herraiz, A.K. Parlikad, J.M. Schooling, Visualised inspection system for monitoring environmental anomalies during daily operation and maintenance, Eng. Constr. Archit. Manag. 27 (8) (2020) 1835–1852, https:// doi.org/10.1108/ECAM-11-2019-0640