Federated data modelling for built environment Digital Twins

Nicola Moretti[1], Ph.D

Xiang Xie[2], Ph.D

Jorge Merino Garcia[3], Ph.D

Janet Chang[4], MBA, M.St

and Ajith Kumar Parlikad[5], Ph.D

ABSTRACT

The Digital Twin (DT) approach is an enabler of data-driven decision making in Architecture, Engineering Constructions and Operations. Various open data models that can potentially support the DT developments, at different scales and application domains, can be found in the literature. However, many implementations are based on organisation-specific information management processes and proprietary data models, hindering interoperability. This article presents the process and information management approaches developed to generate a federated open data model supporting

[1]Reseach Associate at the Institute for Manufacturing, Dept. of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS; and part of the Centre for Digital Built Britain Centre for Digital Built Britain E-mail: nm737@cam.ac.uk

[2]Reseach Associate at the Institute for Manufacturing, Dept. of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS; and part of the Centre for Digital Built Britain Centre for Digital Built Britain E-mail: xx809@cam.ac.uk

[3]Reseach Associate at the Institute for Manufacturing, Dept. of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS; and at the Centre for Digital Built Britain Centre for Digital Built Britain E-mail: jm2210@cam.ac.uk

[4]Ph.D student at the Institute for Manufacturing, Dept. of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS. E-mail: jc2019@cam.ac.uk

[5]Professor of Asset Management and Head of the Asset Management research group at the Institute for Manufacturing, Dept. of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS; and part of the Centre for Digital Built Britain Centre for Digital Built Britain E-mail: aknp2@cam.ac.uk

DT applications. The Business Process Modelling Notation, Transaction and Interaction modelling techniques are applied to formalise the federated DT data modelling framework, organised in three main phases: requirements definition, federation, validation and improvement. The proposed framework is developed adopting the cross-disciplinary and multi-scale principles. A validation on the development of the federated building-level DT data model for the West Cambridge Campus DT research facility is carried out. The federated data model is used to enable DT-based Asset Management applications at the building and built environment levels.

INTRODUCTION

The Digital Twin (DT) concept has been broadly defined in Architecture, Engineering Constructions and Operations (AECO) as a digital replica of an asset, enabling the control and operation of the physical twin via real-time data flows (Bolton et al. 2018a). Though a standard conceptualisation has not been defined yet (Sacks et al. 2020; Boje et al. 2020; AIAA Digital Engineering Integration Committee 2020), many agree that DTs are dynamically characterised by the presence of sensors, actuators and the live measurement of a physical asset's performance. These features characterise the DT implementations in several disciplines (Al-Sehrawy and Kumar 2020) and are used to improve the decision making process, permitted by the higher volume of updated asset information. In fact, the information management is a primary function in all the industries and is one of the fundamental areas within the broad Asset Management (AM) discipline (The Institute of Asset Management 2015). In this context, the definition of an effective data modelling and exchange strategy is crucial to allow a seamless data transfer across domains and digital tools used during the life cycle of the assets. This provides the foundation for the development of DT-based Asset Management (AM) applications and tools.

Innovative digital technologies

The built environment is undergoing a progressive digital innovation affecting products and processes (Nambisan et al. 2020). The large use of digital technologies (e.g., within the Software as a Service - SaaS model) impacts the way the products are designed and produced, and the services provided (Laine et al. 2017). However, the AECO is characterised by high complexity and fragmentation of the products supply and by high demand variability (Baldini et al. 2019). Thus, it is known that innovation often present high customisation and slow technological adoption (Davies and Harty 2013). Elaborating the classification proposed by Baldini et al. (2019), some disruptive technologies can be identified in AECO:

- Internet of Things, including sensing, actuation technologies and wearable;
- Communication technologies including mobile internet and, in general, wireless systems;
- Data acquisition technologies e.g., Light Detection and Ranging - LiDAR and remote sensing;
- Robotics;
- Augmented, mixed and virtual reality;
- Artificial intelligence;
- Additive Manufacturing and Construction;
- Distributed Ledger Technologies;
- Building Information Modelling.

In addition, Papadonikolaki et al. (2022) identify some main digital technologies in BIM, Augmented Reality (AR), Virtual Reality (VR), Internet of things (IoT), cloud computing and Big Data. They discuss how the digital technologies contribute in reshaping and catalysing the digital transformation in the built environment, arguing that connected technologies such as BIM, big data and smart cities contribute to the evolution of the concept of digital in the built environment. Within the scope of our research, BIM, IoT and Artificial Intelligence (AI) are among the most relevant digital innovations.

BIM is considered one of the core drivers for digitalisation of the AECO. The body of knowledge on its use and the benefits in improving process, information management, design and collaboration has grown remarkably and its application are widespread along the assets life cycle (Sanchez et al. 2016). However, some blockers in BIM-based digitalisation can be identified in the cost, effort and changes needed in organisational processes; shortage of skilled professionals and the definition of clear demand from clients (RICS 2022).

The Internet of Things (IoT) technologies offer new possibilities of real-time monitoring and control of buildings systems and assets (Tang et al. 2019). Some benefits concern the possibility of providing real-time warnings to react to unexpected disruptions and the use of live data to inform preventive maintenance strategies. Generally, the stream and the uses of this information, allow to operate in a smart way the building and its parts, to increase, for instance, the internal comfort, indoor air quality, safety, and occupancy monitor (Poli et al. 2020). Custom sensors systems can be potentially extended with the Building Management Systems (BMS), installed in many commercial buildings, to increase the effectiveness in the control and operation. However, some drawbacks in implementing IoT technologies such as possible poor sensor quality and measurement capabilities; placement and calibration issues; communication issues in range and bandwidth; sensors maintenance and data quality are highlighted in (Merino et al. 2022).

Interoperability and data modelling problems

Appropriate data modelling and management enables system's interoperability and reuse of information across different applications. This can be leveraged to develop data-driven decision making tools, supporting a more reliable assets' operations. Interconnected technologies provide the data sources enabling a better knowledge of the dynamics governing the functions and operations of the systems in the built environment, allowing assets' performance control, prediction and data-centric inference on the possible causes related to specific assets' behaviour. For example, the great data volume offered by sensing technologies coupled with Machine Learning (ML) techniques as reinforcement learning, can be used for cross-building transfer learning. The expert knowledge on the building operation can be embedded in automated and data-driven approaches. Data from other buildings can be used to predict energy consumption, faults patterns in similar assets configurations and space utilisation. However, data comes with costs. The data management operations are expensive and the disruptions caused by the COVID-19 pandemic caused enormous loss of significant information, potentially undermining the applicability of historical data as the training basis (Xie et al. 2021a).

A crucial issue in the building and infrastructural asset information management is the need for a clear interoperability approach (Jetlund et al. 2020). This can be achieved if data is exchanged and used across software platforms and domains, minimising losses and errors. In fact, the interoperability refers to the ability of an organisation to exchange information (ISO 2015b). This concerns the processes realised both within the organisation and outside of it, and it is a mean to achieve better business outcomes. Data interoperability can be enhanced using open data models that standardise format and exchange. A data model is the representation of data within an organisation in an understandable way describing the structure and meaning of information (West 2011). It comprises the documentation on the abstract representation of a physical object, of the relevant vocabulary, taxonomy and relationships, that provides a structured hierarchy and methods to communicate data requirements (DAMA International 2017). Usually, a data model addresses a specific scale and domain of application (Moretti et al. 2022). Several examples can be found in the literature, most of them supporting particular ontologies (Partridge et al. 2020). Given the concept and technologies on which they are based, DT applications require access to data which can be structured according to a variety of ontologies and data models.

Adequate data accessibility allows better and new knowledge on the performances of the built assets. However, seamless data access and exchange capabilities need to be granted to unlock the potential of data-driven methods capable of triggering these new outcomes. An example is the need to access and consume real-time sensor readings; geometry data of the built assets; system hierarchies; properties of the assets for fine grain energy monitoring: a service that enables better understanding and control of the energy usage in buildings (Xie et al. 2021b). However, the required data is generally handled according to organisation-specific or proprietary formats, models, and requirements that hinder the data interoperability and reuse in multiple applications. A standard approach to address this problem in developing DT-based AM applications is still to be agreed. More specifically, the following research gaps are identified:

(RG1) there is lack of clarity in the information and process requirements enabling foundational AM built environment DT-based applications;

(RG2) there is the need to define the methods for accessing and interrelate the fragmented data that enable to align with the organisational business needs and to deliver DT-based services;

(RG3) it is unclear how the existing systems can be effectively interconnected to exchange data, while providing an updated and reliable knowledge base for data-driven decision making, based on advanced digital tools;

Aim and impact of the research

This article presents an information and process management framework able to support the DT data modelling phase, through a federation approach. The federation enables the reuse of data within its own domain, preserving the capabilities of each selected data model, as opposed to physical data integration. The proposed framework defines the rules and methods for accessing data in different domains enabling their interrelation and unlocking the value of information for DT-based applications. The aim is to formalise the way AM applications are created through the federated approach, empowering the interconnection and seamless management of assets' properties, real-time data changing over time, AM and contextual data.

This information can be better exchanged and used:

- in different software platforms, maintaining the characterising metadata and relations (Tang et al. 2020);
- across different domains, e.g., technical and functional characterisation of the building and its components, building management, Facility Management (FM), Energy Management (EM), and AM (Pauwels et al. 2017b);
- at different scales, making possible the asset portfolio information management tasks (Moretti et al. 2021a).

The management and integration of different data sources formulate a data-centric mode for supporting informed decision-making, in accordance with the DT paradigm (Lu et al. 2020). This can impact the future development of DT-based applications in research and practice, providing guidelines and reducing the ambiguity in the data and ontological modelling tasks.

RESEARCH BACKGROUND

To scope the research, it is worthily to provide some details on the data modelling process, a crucial phase in the definition of an enterprise architecture. It involves a series of organised steps, with increasing level of detail and complexity, from the conceptual to the physical modelling of the data structures enabling the organisation's business processes. The highest level of abstraction corresponds with the conceptual modelling, providing the broad view of the abstract concepts, used in an organisation to achieve its business objectives. It aims at connecting different subject area models, identify their relationships and vocabularies. The conceptual model can be further detailed in the logical models, able to represent the attributes and specifications of each entity with a higher level of granularity. However, the logical model is relatively abstract and there are no direct relations with the technology used. The logical model, can specify also less-relevant domain-specific entities. The more detailed level is the physical modelling, usually application or project-specific (DAMA International 2017). The physical data model allows to relate the data structures to the assets' classification systems and registers used in the operations. Moreover, it is usually related to a specific software and technology implementation and provides a detailed representation of the physical entities.

Data modelling in management of the built environment

Several approaches can be found in AECO (Plume et al. 2017; Gilbert et al. 2020; BuildingSMART International 2020). At the international level, the Open Geospatial Consortium (OGC) and buildingSMART International (bSI) work actively for standardisation in the domains of built environment theories and practice. Thus, recently the two institutions collaborated for studying the most prominent data models in the AECO, identified in the Industry Foundation Classes (IFC), CityGML, and LandInfra (Gilbert et al. 2020). The study highlights the main features and capabilities of the analysed data models and proposes some actions for overcoming issues associated with interoperability, application development, and wider adoption in industry and research. These actions concern: (i) conducting of illustrative cases specifying software applications, input requirements and data representations; (ii) developing of a best practice on 3D georeferencing with an appropriate level of precision and accuracy; (iii) making publicly available dictionaries and shared resources for identifying synonyms; (iv) creating a system of standard unique identifiers of the real-world entities' to be used during the life cycle; (v) agreeing on a shared method for standardisation of conceptual representation at the thematic level.

Despite the three data models analysed are only some of those which can be found in literature, the matters mentioned above are extremely relevant. In fact, in DTs implementations, the real-time data (e.g., the sensor readings) need to be enriched with contextual and reference data, which is produced during the life cycle of the assets. This can generate new insights on the assets (e.g., buildings and their parts), across different domains (e.g., Energy Management, Facilities Management, Space Management etc.). However, this data is often not directly usable, due to the lack of a clear process in its generation, classification and maintenance (Moretti et al. 2020). In particular, actions iii, iv and v, could boost the standardisation and generalisation of DT-based applications. The Centre for Digital Built Britain (CDBB), within the context of the development of the Information Management Framework (IMF), supporting the national DT program development (Hetherington and West 2020a), has promoted the investigation of the available ontologies and data dictionaries in different sectors. An extensive survey of the existing industry data models has been accomplished (Leal et al. 2020) and provides a baseline for data model design enabling DT-based applications. However, an ontology able to support the whole complexity of the data modelling of the DT-based applications has not been defined yet. In the context of management of built environment, at different scales, the data models presented in Table 1 are among the most documented in the literature.

Their implementation still depends mainly on the scale of the developed applications (i.e., building components, buildings, BE, etc.) and the business context. But they can be potentially used to transfer relevant information according to a multi-scale and cross-disciplinary approach. To achieve this objective, towards the development of DT-based applications, the semantic web approaches can provide further support for data interoperability. Through the linked data methods, information belonging to different domains, can be represented and connected using the Resource

Description Framework (RDF) model. The RDF organise data in directed labelled graphs, enabling an effective semantic enrichment and interconnection (Pauwels et al. 2017a). However, an integrated approach, able to leverage the capabilities of these models and enabling the use of data beyond the boundaries of the domain-specific data siloes in DTs has not been defined yet.

Digital Twin implementations

Several applications at different scales can be found in literature. The DT is the digital replica of an asset, a system or a process that, through the use of data, enables better to achieve better knowledge and control the performances of the physical assets (Bolton et al. 2018b). DTs are tools enabling better decision making, control and improvement of performance in all the life cycle stages of the assets. With a focus on the use phase, the DT is applied to support the information management processes at different levels and addressing several vertical areas; for instance systems, buildings, infrastructures and the wider campus and city environment.

At the building level, it consumes the data of smart and cognitive buildings. For example, the Sydney Opera House project aims at developing a cloud-based web interface, that combines the functions of a Building Management System (BMS), various building information databases and Building Information Modelling (BIM) (Cooperative Research Centre for Construction Innovation 2007; Schevers et al. 2007). The CARTIF-3 INSITER project aims at leveraging self-inspection and virtual reality methods for construction, refurbishment and maintenance of energy-efficient buildings (Hernández et al. 2018). The Energy Laboratory as University eXpo (eLUX) initiative enhances the automation of building operations, through sensors and smart technologies, in a wider smart district context (Tagliabue et al. 2021). Although IFC and derivative platform-neutral BIM formats can somehow address the information exchange and management complications at the building scale, the rigid data structure, to a very large extent, limits the flexibility in accommodating uncertainties of building components and their relationships, and also the applicability of the representation for solving generalised problems.

At the campus or district level, the DT implementations leverage the potential of the 3D city models. Eicker et al. (2020) present the software architecture and the data modelling approach to develop a service-oriented urban platform, improving the energy and waste/wastewater issues. Aleksandrov et al. (2019) propose a Precinct Information Model (PIM) system architecture for integrating BIM, 3D Geographic Information System (GIS) and sensors data, allowing also to visualise and manipulate information, through a front-end layer. The Virtual London Platform (ViLo) is an interactive and collaborative digital platform used to visualise and analyse both real-time and offline space-time urban data (Dawkins et al. 2018). Within the context of the Amsterdam Smart City, the Amsterdam smart port initiative represents one example of smart technologies implementation, at the district level (Amsterdam Smart City 2016). A further example aimed to improve energy management and operations is the City-zen smart grid solution: an EU-funded project aiming at identifying possible solutions for better management of the urban energy infrastructures. The project involves two cities: Amsterdam and Grenoble and supported the development of applications ranging from the use of digital technologies for smart infrastructures, to the enhancement of the citizens' participatory processes (European Commission 2009). With the intention of becoming an open and platform-independent geographic information model, additional application-specific information in various granularities needs to be represented by adding features and attributes that are not readily available in data models like CityGML.

At the city level, the DT concept overlaps to the smart city paradigm. For instance, the Vivacité platform was conceived to monitor the energy and water consumption from the city system operators. This data is then provided, with an informative and educational purpose, to property managers, governmental actors and residents (Koch-Mathian et al. 2019). The Herrenberg city level digital twin aims to propose a domain-specific application for smart cities and simulations, based on a DT, linking and combining heterogeneous urban data coming from simulations, modelling, analysis and crowd sourced data (Dembski et al. 2020). Smart Cambridge is a collaborative initiative focusing on the transformation of transportation in the Greater Cambridge area by exploring embedded digital solutions and emerging technology combined with better quantity, quality and use of data (Smart Cambridge 2020). The digital twin of the city of Newcastle aims to enable experts to test the infrastructure's potential in hypothetical situations including rising sea levels, drought, freak weather events and energy shortages (Newcastle University 2019). Smart Nation is a Singapore Government initiative aims at harnessing digital technologies to improve living, build a closer community, boost the job market and encourage businesses to innovate and grow (Singapore Smart Nation 2021). Bristol is Open (BiO) is a strategic initiative with the objectives of strengthening the city's digital foundations so that it becomes well-connected and better placed to deliver the technological innovation needed to keep the city healthy and safe, in a sustainable way (Bristol City Council 2021). Table 2 summarises the described DT implementations, focusing on the proposed data modelling strategy. However, only in few of them this is clearly documented, particularly, when open data models are used, e.g., IFC and CityGML schemas. In other cases, either it is not described or it is not clearly identified. Despite the DT-based services can be delivered without a clear and open data modelling strategy, this hinders their generalisations and so, the repeatability in other scenarios. On one hand, in the perspective of the development of a city-scale DT, which requires to leverage a great data veracity, the adoption of a scalable approach, fostering interoperability and data re-use in several applications is needed. On the other hand, many open data models already exist in literature and can be interconnected to represent the built environment assets. Still, an approach able to leverage this wide potential has not been defined yet.

METHODS AND TOOLS

Buildings, and more widely, the built environment assets are infrastructures comprehending various decentralised, loosely organised components (e.g., systems, spaces, etc.). The UK National Infrastructure Commission realised a report (National Infrastructure Commission 2017), indicating that "*the UK needs a digital framework for data on infrastructure to harness the benefits from sharing better quality information about its infrastructure; how it is used, maintained and planned*". A well-designed data model can represent both the geometrical and contextual information of components and facilitate better understanding of data in the BE. The entities and their characteristics (i.e., geometry, location and semantic properties) are defined through Information Requirements (IR) with a standardised naming convention, which reduce ambiguity. On one hand, this standard representation together with timely update of information enables the traceability throughout the whole life cycle for each component. On the other hand, entities can be queried, grouped, and related to each other based on their interdependencies. These capabilities are critical for integrating dynamic data to specific objects or groups.

Figure 1 represents the proposed three-tier method three phases enabling the creation and testing of the proposed federated data model and the way it is applied to develop DT-based applications. A description of the early version of the overall approach can be found in Moretti et al. (2021b), with reference to the West Cambridge Digital Twin research facility. In this article the methodology is generalised, contributing to the development of further DT applications. The requirements definition phase allows to identifying the relevant data to be transferred and used for activating the DT functions. The DT federated data model definition phase allows to identify the existing ontologies and related data models able to represent the data with the appropriate level of detail, precision and the semantic, geometrical, and topological interfaces.

The federation phase provides also the conceptual and logical data models, building the foundations for the development of DT-based AM applications. The DT data model testing and validation phase allows to define the application-specific IR definition, either leveraging data hierarchies and structure of the underpinning data models, or informing their semantic enrichment. Thus, through the DT-based applications development the federated data model can be tested in practice, providing also guidance for improvements, according to a Plan/Do/Check/Act cycle (ISO 2015a). To develop the proposed framework, a mixed research method is adopted. This combines the bibliographic research, evidence-based, semi-structured interviews and case study research that bring from the formalisation of the three-tier federated data modelling approach, to the implementation of the proposed framework in the West Cambridge Digital Twin research facility.

To generalise the development of the three phases, the Information Delivery Manual (IDM) approach is used (ISO 2017). The IDM enables a BIM-based workflow to model and manage processes, according to different level of abstractions. The process mapping of the key phases implemented to achieve the business objectives is modelled according to the Business Process Modelling Notation (BPMN) (ISO 2013). These schemes represent the set of activities encompassed in the process, the roles of the participants, and the sequence of operations to be implemented. The interaction mapping allows the second level of abstraction focusing on the information exchange among parties. This techniques is implemented for the crucial data drops. When the roles of the initiator and executor of a specific action are verified, then a transaction happens. The third modelling technique, namely the transaction mapping, represents the data exchange occurring among the parties involved in a transaction. Through the IDM, it is possible to generalise different aspects of the procedures, events and interactions realised to achieve a given objective. This approach is originally aimed at supporting the Model View Definitions (MVD) in the BIM domain, enabling the identification of the data and process requirements. However, the IDM comprehends a set of strong information management methods, that can be used to support built environment DT-based implementations as well. Also, the BIM data is a fundamental part of the federated DT model. For this reason, the IDM offers good capabilities to formalise the main phases of the proposed approach.

THE FEDERATED DATA MODELLING APPROACH

An incremental identification and interconnection of the data models capable to represent the requirements of the DT implementations is developed. This allows to leverage the vast number of ontological and data modelling techniques in the AECO (Partridge et al. 2020). The federation approach allows to avoid that a particular model, vocabulary or method is imposed over others (ISO 2011). The interconnection is enabled through the identification and conceptualisation of the interface entities, allowing to navigate across different (federated) ontologies. In this way, the representation, classification, and query capabilities of each data model are exploited in their specific domain. Thus, the semantic interoperability is enabled through the identification of these few core entities, enabling the connection of multiple data models. The ontology and data model mapping is reduced to the minimum to control the data redundancy and enabling a flexible and modular inclusion of additional data models.

Requirements definition

At first, the requirements needed for implementing data-driven AM applications are identified. This concerns:

- the definition of the business needs expressed by the senior management within an organisation, namely the "client" in Figure 2) to achieve better organisational outcomes;
- the identification of the information requirements comprehending the static and dynamic, geometric and non-geometric information at the building and built environment levels;
- the characterisation of the data modelling strategy supporting the supply and update of timely and quality data, related inter-dependencies, and hierarchies.

The requirements definition phase is implemented following the process in Figure 2 and involves three roles: the client expressing the organisational objectives, the asset manager for transferring the business objectives into asset operation practices, and the DT development team responsible for the delivery of the DT solutions (Figure 3). The parties are interacting, in this phase, in two transaction steps. The first one (T1), corresponds to the communication, by the client, of the organisational and business needs to the asset manager, which identifies and develops the asset classification and the Asset Information Requirements (AIR). The second (T2) corresponds with the transfer of the AIR to the DT developer, which designs the data model requirements. The main outcomes are the definition of the asset classification, AIR and the data model requirements. This information is used in the following stage for mapping existing data models and for the identification of the entities composing the core DT data model.

The definition of the AIR is a crucial to translate the needs identified in the organisational objectives into a set of operational information categories. The AIR definition should be accomplished by adopting a clear assets classification to characterise hierarchically the abstract and physical entities according to different criteria (e.g. functional, product, construction phase, etc.) (Afsari and Eastman 2016). An effective approach to align the client's requirements with the functional outputs and the related AIR is described in Heaton et al. (2019). The AIR can be defined through the analysis of existing standards and scientific literature, following an evidence-based method, through interviews and workshops with experts (Dumas et al. 2018). A list of fundamental AIR categories is collected in Table 3. The first two categories, *General* and *Technical* comprehend the information that is produced during the design and construction phases or when a substantial modification occurs on the assets. They include, for example, drawings, calculations (e.g. energy and structural) and authorisations. They are particularly relevant for the quantification and representation of the assets and it is the basic set of data that can be used to acquire the overall knowledge (e.g. dimensions, technical performances, locations, etc). The *Asset Information, Operations&Maintenance - O&M, People* categories comprehend the AIRs used to store and manage data related to the operation of the physical assets. These requirements allow to manage information in the AM domain and can be used to operate effectively the assets. However, the AIR definition is known to be case-specific. Therefore, Table 3 only list the minimum set of sample information.

The last outcome of this step concerns the definition of the DT data model requirements. These can be defined according to existing approaches (Crockett et al. 1991; West 2011) and including the guidelines and principles of the Information Management Framework (IMF) (Hetherington and West 2020b) and the Gemini Principles (Bolton et al. 2018a).

Federation

Federation refers to the provision of data "*without additional persistence or duplication of source data*". (DAMA International 2017). A data model is often the result of the ontological modelling providing the representation of the entities composing a physical or abstract system, of their characteristics, their relationships and hierarchies. The ontology, is a formal abstract system able to represent a domain of knowledge. However, in the proposed approach a top ontology is not defined in advance. It is prioritised, instead, the use of a set of existing data models able to represent domain-specific information. The identified data models are then federated to cover the complete set of requirements identified in the previous phase, reducing the customisation for achieving a simple and fundamental conceptual model. The main purpose of this phase (represented in Figure 4) is to design the data model underpinning the DT-based applications in AM. For generalisation purposes, a catalogue of available data models can be developed, so the related entities able to represent the DT requirements can be identified. The information requirements can be either fully or partially matched, thus a custom definition is needed (however, custom information requirements should be minimal if a comprehensive built environment ontology library is developed). This process leads to the development of the entities of the foundation data model, which is quality-checked to ensure the correct representation of the whole set of functional, information and data modelling requirements. The mapping among the final set of data models identified leads to the definition of the interface entities, permitting the interrelation and access of the federated data models (i.e., enabling Master Data Management). Some issues to be controlled in this process are identified (Salheb et al. 2020) and summarised in Table 4.

The final outcome of this phase is the conceptualisation of the federated foundational data model, and it is entirely conducted by the DT development team. Some general semantic definitions of the built environment assets can be identified. The spatial and systems' hierarchies and classification are two of these. It is crucial that, at this conceptualisation level, these features are correctly represented, either through geometric or semantic data modelling techniques. In AM, the explicit geometric representation might not be needed. However, it is important that the location and topology are consistent at the different allowed scales.

Validation and improvement

The third and last phase concerns the use of the federated data model for application development (Figure 5 and Figure 3, T3). Two roles are involved: the asset manager and the DT developer. The definition of the service needs informs the development of the detailed applications specific information requirements. These are checked against the federated data model capabilities and, if the requirements are fully matched, a verification step with the asset manager is accomplished. This can be the case of canonical AM application as, for instance, the maintenance scheduling, the work order management and the life cycle costing. The federated data modelling approach is developed to address the complexity multi-scale and interdisciplinary applications. Therefore phase 2 (Federation) is a recursive process, triggered whenever the definition of new information requirements is needed (e.g., for the improvement of the foundational model and for the definition of application-specific entities). Once the application IR are defined and approved by the asset manager, the data pipelines can be design to provide data to AM applications. Data pipelines will interface with the federated data models to acquire information, meeting application-specific requirements like data frequency, formats, and quality. This enables the development of APIs for the delivery of the digital services empowering DT-based AM applications. On one hand, in real-time AM applications, high-paced data like sensor readings, needs to be integrated with semi-static data like geometries. On the other hand, to represent contextual data, as the geometries the objects or the spatial hierarchies of a building (i.e., site, building, floors, rooms), a low velocity and a less frequent update can be foreseen. The federated data models enable integrated data flows from independent information sources with specific requirements. In this phase, the rules enabling the data sharing among the involved parties, and the related responsibilities, trough the users' profiling and authorisation should also be defined. Useful tools to enable these capabilities are offered by the Common Data Environment (CDE) techniques, within the BIM approach (Patacas et al. 2020). Lastly, through the applications development, the data modelling approach can be periodically tested and validated, highlighting potential issues in terms of:

- dynamic and static data models validation;
- validation of the data collection, storage and update methods;
- geometric, topological, semantic validation;
- accessing, querying and visualising information;
- managing the missing information;
- Applications Programming Interfaces (APIs) development.

CASE STUDY

The proposed approach is applied to the West Cambridge DT, the main objective of which is to understand how to develop built environment DTs (aligned with the CDBB's framework) impacting on better operation, performance and organisational productivity. An insight on the AM

IR definition and on the development of the federated data model enabling canonical DT-based applications is presented in this section. The managed assets are of different types and vary in typology and scale, spreading from the buildings' systems to the whole infrastructural and built environment. Multiple information layers at various scales have been considered. Therefore, the IRs have been defined considering not only the building and its components, but also the outdoor spaces and infrastructures. The requirements are defined through the systematic literature review and semi-structured interview-based approach and form the basis for developing the DT-based AM and FM applications. The process in Figure 2 was followed at this stage. In acquiring the assets information (Input specification in Figure 6) both static and dynamic data needed to effectively operate and maintain the assets of the Alan Reece building, part of the West Cambridge DT facility, have been considered.

A variety of dynamic data (i.e., data changing over time) is available from different resources, as the Energy Management System (EMS), Computerised Maintenance Management System (CMMS), Space Management System (SMS), and financial records. In managing energy efficiency, a Building Management System (BMS) is installed to operate the building at an optimal level and to remove wasted energy uses by remotely controlling Heating Ventilation Air Conditioning (HVAC), lighting, and power systems. A CMMS is utilised to control inventory of spare parts, while managing maintenance work orders of the university assets. The descriptive asset information is organised by buildings; and the asset registry includes a component level. Such registry contains the name of components, the manufacturer, model number, serial number, New Rules of Measurement (NRM3), SFG20, etc. Also, the CMMS keeps records of all asset management activities, such as monitoring cost-to-date information related to the assets, planned and preventive maintenance work, scheduling regulatory compliance inspections and service work. A Property Management cloud-based system, is adopted to manage space planning and provide spatial analytics based on the room bookings to manage the university space. This system also maintains CAD floor plans and building service reports. Table 5, lists the fundamental IRs used for developing the DT-based Asset and Energy Management applications in the West Cambridge

Digital Twin research facility. Only basic and non aggregated data is considered, to streamline the IR definition process.

The DT data model is conceptualised according to a multi-scale and interdisciplinary approach. In fact, data needs to be exchanged and reused in multiple applications within the DT ecosystem, with the right timing and level of detail. The proposed federation approach can support the data curation over time and enhance the data availability for further DT-based applications, though not be bounded to any proprietary software and data model. The federated DT data model, represented in Figure 7 has been developed following the proposed federated data modelling framework.

At the building level, the Industry Foundation Classes (IFC) schema has been used to implement the BIM-based IR requirements, allowing to structure and classify data, through an openBIM approach (Moretti et al. 2020). The IFC schema allows the information exchange in several domains. However, the full complexity of the schema is not needed, nor the geometric information is always required for AM implementations. Therefore, a Model View Definition (MVD) representing the relevant classes and their relations has been defined (Figure 7 IFC section). The IFC schema allows to transfer relevant data across assets' life cycle phases (e.g., from the design and construction to the use phase). However, as-built BIM data are often incomplete or not updated, hampering interoperability and effective information exchange. Figure 8a represents the case study building and how the IR related to IoT sensors used in the West Cambridge DT research facility can be handled in an openBIM environment, enriching the original BIM data.

In fact, some of the IR (i.e., the sensor type, the sensor identifier, and its description) can be created in the BIM editing software. Therefore, a new sensor object is developed and enriched with the relevant IFC parameters (version 4 ADD2 TC1), that can be mapped and exported as attributes of the IfcSensor class. Data is then be post-processed and queried to access to relevant technical properties of the installed sensors as their location, topological, and semantic relationship. However, other relevant information is not directly available in IFC data. This is the case of the IRs describing the relationships among sensors and the measured asset and spaces in the building. For this purpose, the class IfcAssset can be used to group the building components based on their association with a homogeneous set of elements forming an asset, namely an entity that can be managed and maintained consistently. For this re-classification of the building components, it is necessary to further process the IFC data (Figure 8b). This can be done according to the workflow described in Moretti et al. (2020). After Once this task is accomplished, it is possible to access the AM IR (as for isntance, the cost-related parameters, the corresponding maintenance schedule, etc.) through the IfcAsset attributes. However, the IFC data model is effective for building-related static data representation, while it does not provide sufficient support for agile real-time data query. For instance, the semantics of the building systems, to which the real-time data refers, can be represented through more flexible schemes allowing agile modelling of non-geometric building features (e.g.: the interrelations of non-visible building elements). This is a particularly helpful, for existing buildings, where often data is incomplete and the non-visible parts of the systems cannot be modelled geometrically. This is the reason why the IFC data model is federated with the BrickSchema (Balaji et al. 2016), instead of extending it. According to the approach described in Xie et al. (2021b), the IFC and the Brick data models are leveraged to access both the spatial and system hierarchies, characterising the spaces of the Alan Reece building (Figure 7, Brick section adn Figure 8b).

The real-time data is collected through a custom IoT network, formed by Radio Frequency (RF) and LoRaWan sensors measuring indoor parameters like temperature, humidity, CO2, and light detection. The real-time information is also retrieved from the BMS. Data is handled through the Adaptive City Platform (2021) and is structured according to the Crate data model. A crate is a 2D object, enclosed in a perimeter, representing a whole building or part of it (e.g., a floor, a space etc.). Crates can be nested using a parent-child relationship to represent the spatial hierarchy of the building (Figure 7 Crate section). Thus, the hierarchical spatial structure of the building (i.e., building, floor, zone, space) is adopted in the DT data model to federate the three domain-specific data models (IFC, BrickSchema and Crate). Through this approach, the sensor readings are enriched with the contextual information and can be used for different purposes, through the development of the APIs, enabling the DT services (Figure 8c). The federated approach enables the use and integration of different-paced data (original source information requirements impose that pace), and can effectively support the integration for analytic tools in AM applications

DISCUSSION AND CONCLUSIONS

The proposed federated data modelling approach aims to support data interoperability enabling better information management. To address the RG1, the formalisation of the processes for its implementation is realised through the IDM approach and is structured in three phases: requirements definition, federation, validation and improvement. Though the proposed approach, it is possible to develop the foundational data model, enabling built environment DT-based AM applications which, at the building level, is hinged on the spatial hierarchies.

The information and process requirements for data exchange in the previous DT implementations listed in Table 2) is seldom addressed and the fundamental definitions, data dictionaries and relations, enabling DT-based applications are not explicit. This is, for instance the case of the eLUX DT (Tagliabue et al. 2021), ViLo (Aleksandrov et al. 2019) and the Amsterdam smart port (Amsterdam Smart City 2016). In other cases, as the Herrenberg smart city (Dembski et al. 2020), Singapore Smart Nation (Singapore Smart Nation 2021) and BiO (Bristol City Council 2021) there is no explicit description of the adopted data modelling strategy. While in the Smart Cambridge (Smart Cambridge 2020) there is high reliance on a custom data modelling approach.

Despite these issues do not prevent to stress the benefits of the application of the DT concept, they undermine the possibility of generalise and leverage the existing data modelling approach in different scenarios (other buildings, districts, cities). In the perspective of managing an ecosystem of DTs, which can be modelled as a complex network (Herrera et al. 2020), this minimise the potential benefits of data availability and semantic interoperability. For example, a poorly documented data modelling forces to re-build the data classification and hierarchies from scratch, and leads to time consuming ontology mapping processes. This results in inefficient application development and poor data reuse.

The process and information modelling proposed in this article can be used for unlocking the potential of multiple interconnected open data models that enable the generalisation and repeatability of DT implementations. Adopting such approach it is possible to address the data fragmentation and heterogeneity that characterise the AEC sector (RG2). In fact, the three-tiered process modelling enables the alignment of the asset information requirements with the service needs, during the life cycle of the assets. This allows the data re-use and integration at the application level, empowering timely, reliable and quality DT-based AM service developments.

Buildings can be considered as systems of systems. And each system is aimed at delivering a set of services. For instance, the building spatial functionalities are delivered through the space system, the comfort through the HVAC system, illumination and power through the electrical systems, and so on. The DT-based data modelling framework proposed in this article empowers the interconnection between them through a seamless information flow. This is the fundamental enabler for wider scale interconnected DTs, at the city and national level (Wee et al. 2022).

The case study presents how the proposed federated data model can be leveraged for a semantically enriched and better dynamic APIs development (RG3). The APIs are used to support the decision making process in the building operations. For instance they provide rapid data access for assets anomalies detection, fault detection and diagnosis and sensor readings visualisation in a dynamic building model. An insight on the use of the IFC, BrickSchema and Crate is used for validation and for demonstrating how the IR at the building level can be developed and used in DT-based AM applications. The implementation of the proposed framework on the Alan Recee building, shoes how to handle the multi-scale, interdisciplinary and temporal dimensions of the DT application from a data perspective. Static and real-time data (e.g., the building spatial hierarchy, condition inspection data and sensors readings) are different by nature. In the case study we demonstrated that the federated approach can be used to drive the integration of diverse variety and velocity data. Comparing to other interoperability models like IFC, and all-in-one ontologies, management of high-velocity data is done directly rather than through reasoners, which are inherently slower (Kritikos and Plexousakis 2008)

The proposed federation approach addresses the issues of data heterogeneity, interoperability and ontological modelling for DT-based applications in AECO. However, the mapping across different ontologies is not an trivial task and becomes more and more complex, with the increasing number of federated data models. For instance, the interconnection of built and infrastructural assets representation in different domains (e.g., civil engineering, asset management, planning etc.), their geometric and topological definition, attributes definition at different scales are issues that need to be further explored in research; as well as the development of an automated mapping and translation methods. This can bring to the definition of an extended interface ontology that can support the data exchange in the city level DTs. Moreover, further research is needed to identify and develop the integration methods able to connect data and processes feeding data-driven AM applications.

To conclude, being the DT concept and implementations relatively at its infancy in AECO, it is still unclear how to generalise the application of the DT-related technologies for the development of data-centric services. The proposed data modelling approach shades light on how the data needs to be processed and organised to support better interoperability and data exchange in built environment DTs. However, the needed architecture and integration methods are only partially addressed. The formalisation of additional DT-based AM application development will allow to further demonstrate the value of the open and federated data modelling approach.

DATA AVAILABILITY STATEMENT

Some data models used during the study are available in a repository or online in accordance with funder data retention policies. The Industry Foundation Classes data model can be found at https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2_TC1/HTML/. The BrickSchema can be found at https://brickschema.org/. The description of the Crate data model can be found in (Brazauskas et al. 2021). The IFC processing approach is described in (Moretti et al. 2020).

ACKNOWLEDGEMENTS

This research forms part of the Centre for Digital Built Britain's (CDBB) work at the University of Cambridge within the Construction Innovation Hub (grant number: NMZM/429). The Construction Innovation Hub is funded by UK Research and Innovation through the Industrial Strategy Fund.

REFERENCES

Afsari, K. and Eastman, C. M. (2016). “A Comparison of Construction Classification Systems Used for Classifying Building Product Models.” *52nd ASC Annual International Conference Proceedings*, (2001), 1–8.

AIAA Digital Engineering Integration Committee (2020). “Digital twin: Definition & value an aiaa and aia position paper, <https://www.aiaa.org/docs/default-source/uploadedfiles/issues-and-advocacy/policy-papers/digital-twin-institute-position-paper-(december-2020).pdf>.

Al-Sehrawy, R. and Kumar, B. (2020). “Digital Twins in Architecture, Engineering, Construction and Operations. A Brief Review and Analysis.” *International Conference on Computing in Civil and Building Engineering*, Springer, Cham, 924–939.

Aleksandrov, M., Diakité, A., Yan, J., Li, W., and Zlatanova, S. (2019). “SYSTEMS ARCHITECTURE for MANAGEMENT of BIM, 3D GIS and SENSORS DATA.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(4/W9), 3–10.

Amsterdam Smart City (2016). “Amsterdam Smart City ~ Vehicle2Grid, <https://amsterdamsmartcity.com/organisations/amsterdam-smart-city http://amsterdamsmartcity.com/projects/detail/id/72/slug/vehicle2grid>.

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærgaard, M. B., Srivastava, M., and Whitehouse, K. (2016). “Brick: Towards a unified metadata schema for buildings.” *Proceedings of the 3rd ACM Conference on Systems for Energy-Efficient Built Environments, BuildSys 2016*, Association for Computing Machinery, Inc, 41–50, <http://dx.doi.org/10.1145/2993422.2993577> (nov).

Baldini, G., Barboni, M., Bono, F., Delipetrev, B., Brown, N. D., Macias, E. F., Gkoumas, K., Joossens, E., Kalpaka, A., Nepelski, D., de Lima, M. V. N., Pagano, A., Prettico, G., Sanchez, I., Sobolewski, M., Triaille, J.-P., Tsakalidis, A., and

Brancati, M. C. U. (2019). *Digital Transformation in Transport, Construction, Energy, Government and Public Administration*. Publications Office of the European Union, <https://publications.jrc.ec.europa.eu/repository/handle/JRC116179>.

Boje, C., Guerriero, A., Kubicki, S., and Rezgui, Y. (2020). “Towards a semantic Construction Digital Twin: Directions for future research.” *Automation in Construction*, 114(November 2019), 103179.

Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., et al. (2018a). “Gemini principles.

Bolton, A., Enzer, M., and Schooling, J. (2018b). “The Gemini Principles.” *Report no.*, <https://www.cdbb.cam.ac.uk/system/files/documents/TheGeminiPrinciples.pdf>.

Brazauskas, J., Verma, R., Safronov, V., Danish, M., Merino, J., Xie, X., Lewis, I., and Mortier, R. (2021). “Data Management for Building Information Modelling in a Real-Time Adaptive City Platform.

Bristol City Council (2021). “Bristol is Open, <https://www.bristol.gov.uk/policies-plans-strategies/bristol-is-open>.

BuildingSMART International (2020). “Enabling an Ecosystem of Digital Twins. A buildingSMART International Positioning Paper. How to unlock economic, social, environmental and business value for the built asset industry.

Cooperative Research Centre for Construction Innovation (2007). *Adopting BIM for facilities management: Solutions for managing the Sydney Opera House*. Icon.Net Pty Ltd, Brisbane.

Crockett, H., Guynes, J., and Slinkman, C. (1991). “Framework for development of conceptual data modelling techniques.” *Information and Software Technology*, 33(2), 134–142.

DAMA International (2017). *DAMA-DMBOK Data Management Body of Knowledge*. Technics Publications, Basking Ridge, New Jersey, second edi edition.

Davies, R. and Harty, C. (2013). “Measurement and exploration of individual beliefs about the consequences of building information modelling use.” *Construction Management and Economics*, 31, 1110–1127.

Dawkins, O., Dennett, A., and Hudson-Smith, A. (2018). “Living with a Digital Twin: Operational management and engagement using IoT and Mixed Realities at UCL’s Here East Campus on the Queen Elizabeth Olympic Park.” *Giscience and Remote Sensing*.

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., and Yamu, C. (2020). “Urban digital twins for smart cities and citizens: The case study of herrenberg, germany.” *Sustainability (Switzerland)*, 12(6), 1–17.

Dumas, M., La Rosa, M., Mendling, J., and Reijers, H. A. (2018). *Fundamentals of business process management: Second Edition*. Springer Berlin Heidelberg, second edi edition (mar).

Eicker, U., Weiler, V., Schumacher, J., and Braun, R. (2020). “On the design of an urban data and modeling platform and its application to urban district analyses.” *Energy and Buildings*, 217, 109954.

European Commission (2009). “Climate Action. EU Action. Strategies. 2020 climate & energy package, <https://ec.europa.eu/clima/policies/strategies/2020_en>.

Gilbert, T., Rönsdorf, C., Plume, J., Simmons, S., Nisbet, N., Gruler, H.-C., Kolbe, T. H., van Berlo, L., Mercer, A., et al. (2020). “Built environment data standards and their integration: an analysis of ifc, citygml and landinfra.” *Report no.*, Lehrstuhl für Geoinformatik.

Heaton, J., Parlikad, A. K., and Schooling, J. (2019). “A Building Information Modelling approach to the alignment of organisational objectives to Asset Information Requirements.” *Automation in Construction*, 104(March), 14–26.

Hernández, J. L., Lerones, P. M., Bonsma, P., van Delft, A., Deighton, R., and Braun, J. D. (2018). “An IFC interoperability framework for self-inspection process in buildings.” *Buildings*, 8(2), 32.

Herrera, M., Pérez-Hernández, M., Parlikad, A. K., and Izquierdo, J. (2020). “Multi-agent systems and complex networks: Review and applications in systems engineering.” *Processes 2020, Vol. 8, Page 312*, 8, 312.

Hetherington, J. and West, M. (2020a). “The pathway towards an information management framework-a ‘commons’ for digital built britain.

Hetherington, J. and West, M. (2020b). "The pathway towards an Information Management Framework - A 'Commons' for Digital Built Britain." *Report no.*, <https://www.repository.cam.ac.uk/handle/1810/305579>.

ISO (2011). "EN ISO 11354-1:2011 - Advanced automation technologies and their applications — Requirements for establishing manufacturing enterprise process interoperability — Part 1: Framework for enterprise interoperability.

ISO (2013). "ISO/IEC 19510:2013 Information technology — Object Management Group Business Process Model and Notation.

ISO (2015a). "EN ISO 9001 : 2015 Quality management systems - Requirements (ISO 9001:2015) Systèmes.

ISO (2015b). "ISO 11354-2:2015 - Advanced automation technologies and their applications – Requirements for establishing manufacturing enterprise process interoperability – Part 2: Maturity model for assessing enterprise interoperability. ISO/TC 184/SC 5, <https://www.iso.org/standard/57019.html>.

ISO (2017). "BS EN ISO 29481-1:2017 Building information models – Information delivery manual. Part 1: Methodology and format.

Jetlund, K., Onstein, E., and Huang, L. (2020). "Ifc schemas in iso/tc 211 compliant uml for improved interoperability between bim and gis." *ISPRS International Journal of Geo-Information*, 9(4), 278.

Koch-Mathian, S., De Geoffroy, V., Gesmond, F., Morel, B., Robillard, J., and Freiser, D. (2019). "Report on Vivacité extension to metropolitan scale." *Report no.*

Kritikos, K. and Plexousakis, D. (2008). "Evaluation of qos-based web service matchmaking algorithms." *2008 IEEE Congress on Services - Part I*, 567–574.

Laine, E., Alhava, O., Peltokorpi, A., and Seppänen, O. (2017). "Platform ecosystems: unlocking the subcontractors' business model opportunities." *Proceedings for the 25th Annual Conference of the International Group for Lean Construction. Heraklion, Greece*, 177–184.

Leal, D., Cook, A., Partridge, C., Sullivan, J., and West, M. (2020). "A survey of industry data models and reference data libraries-to identify requirements for, and to provide input to, a foundation data model.

Lu, Q., Parlikad, A. K., Woodall, P., Don Ranasinghe, G., Xie, X., Liang, Z., Konstantinou, E., Heaton, J., and Schooling, J. (2020). "Developing a digital twin at building and city levels: Case study of west cambridge campus." *Journal of Management in Engineering*, 36(3), 05020004.

Merino, J., Sasidharan, M., Herrera, M., Zhou, H., del Castillo, A. C., Parlikad, A. K., Brooks, R., and Poulter, K. (2022). "Lessons learned from an iot deployment for condition monitoring at the port of felixstowe." *IFAC-PapersOnLine*, 55, 217–222.

Moretti, N., Ellul, C., Cecconi, F. R., Papapesios, N., and Dejaco, M. C. (2021a). "Geobim for built environment condition assessment supporting asset management decision making." *Automation in Construction*, 130, 103859.

Moretti, N., Xie, X., Garcia, J. M., Chang, J., and Parlikad, A. K. (2022). "Built environment data modelling: a review of current approaches and standards supporting asset management." *IFAC-PapersOnLine*, 55(19), 229–234.

Moretti, N., Xie, X., Merino, J., Brazauskas, J., and Parlikad, A. K. (2020). "An openbim approach to iot integration with incomplete as-built data." *Applied Sciences (Switzerland)*, 10(22), 1–17.

Moretti, N., Xie, X., Merino Garcia, J., Chang, J., Parlikad, A. K., Garcia, J. M., Chang, J., and Parlikad, A. K. (2021b). "Developing a federated data model for built environment digital twins." *International Conference on Computing In CIvil Engineering (I3CE 2021), 12-14 September 2021, Orlando, Florida USA*, American Society of Civil Engineers.

Nambisan, S., Lyytinen, K., and Yoo, Y. (2020). "Digital innovation: towards a transdisciplinary perspective." *Handbook of Digital Innovation* (7).

National Infrastructure Commission (2017). "Data for the public good." *Report no.*, <https://nic.org.uk/app/uploads/Data-for-the-Public-Good-NIC-Report.pdf>.

Newcastle University (2019). "Newcastle's 'digital twin' to help city plan for disasters, <https://www.ncl.ac.uk/press/articles/archive/2019/01/digitaltwin/>.

Papadonikolaki, E., Krystallis, I., and Morgan, B. (2022). "Digital technologies in built environment projects: Review and future directions." *Project Management Journal*.

Partridge, C., Mitchell, A., Cook, A., Sullivan, J., and West, M. (2020). "A survey of top-level ontologies-to inform the ontological choices for a foundation data model.

Patacas, J., Dawood, N., and Kassem, M. (2020). "BIM for facilities management: A framework and a common data environment using open standards." *Automation in Construction*, 120(June), 103366.

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

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017b). "Semantic web technologies in aec industry: A literature overview." *Automation in construction*, 73, 145–165.

Plume, J., Marchant, D., Mitchell, J., and Newhouse, O. (2017). "Proposal for an open data model schema for precinct-scale information management." *Procedia engineering*, 180, 822–831.

Poli, T., Mainini, A. G., Speroni, A., Blanco Cadena, J. D., and Moretti, N. (2020). "The effect of real-time sensing of a window on energy efficiency, comfort, health and user behavior." *Research for Development*, 291–296.

RICS (2022). *Digitalisation in construction report 2022*, <https://www.rics.org/globalassets/rics-website/media/knowledge/research/research-reports/rics0112-digitalisation-in-construction-report-2022-web.pdf>.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., and Girolami, M. (2020). "Construction with digital twin information systems." *Data-Centric Engineering*, 1(6).

Salheb, N., Arroyo Ohori, K., and Stoter, J. (2020). "AUTOMATIC CONVERSION OF CITYGML TO IFC." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV(September), 127–134.

Sanchez, A. X., Hampson, K. D., and Vaux, S. (2016). "Delivering value with bim: A whole-of-life approach." *Delivering Value with BIM: A Whole-of-Life Approach*, 1–344.

Schevers, H., Mitchell, J., Akhurst, P., Marchant, D., Bull, S., McDonald, K., Drogemuller, R., and Linning, C. (2007). "Towards digital facility modelling for Sydney Opera House using IFC and semantic web technology." *Electronic Journal of Information Technology in Construction*, 12, 347–362.

Singapore Smart Nation (2021). "Transforming Singapore trough technology, <https://www.smartnation.gov.sg/>.

Smart Cambridge (2020). "Smart Cambridge 2019 - 2020." *Report no.*, <https://www.connectingcambridgeshire.co.uk/wp-content/uploads/2019/09/Smart-Cambridge-brochure-2019-2020.pdf>.

Tagliabue, L. C., Cecconi, F. R., Maltese, S., Rinaldi, S., Ciribini, A. L. C., and Flammini, A. (2021). "Leveraging digital twin for sustainability assessment of an educational building." *Sustainability (Switzerland)*, 13(2), 1–16.

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., and Gao, X. (2019). "A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends." *Automation in Construction*, 101(February), 127–139.

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., and Gao, X. (2020). "Bim assisted building automation system information exchange using bacnet and ifc." *Automation in Construction*, 110, 103049.

The Institute of Asset Management (2015). "Asset Management – an anatomy (v3)." *Report No. December*, <www.theIAM.org/AMA>.

Wee, X. B., Herrera, M., Hadjidemetriou, G. M., and Parlikad, A. K. (2022). "Simulation and criticality assessment of urban rail and interdependent infrastructure networks." *https://doi.org/10.1177/03611981221103594*, 036119812211035.

West, M. (2011). *Developing high quality data models*. Elsevier.

Xie, X., Lu, Q., Herrera, M., Yu, Q., Parlikad, A. K., and Schooling, J. M. (2021a). "Does historical data still count? exploring the applicability of smart building applications in the post-pandemic period." *Sustainable Cities and Society*, 69.

Xie, X., Moretti, N., Garcia, J. M., Chang, J., and Parlikad, A. K. (2021b). "Ontology-Based Spatial and System Hierarchies Federation for Fine-Grained Building Energy Analysis." *Joint*

Conference CIB W78 - LDAC 2021, Luxembourg 11-15 October.

List of Tables

TABLE 1. Data models used in management of the built environment.

Institution	Data Model	Domain	Encoding	Standard
BuildingSMART	IFC	Buildings	EXPRESS, XML, OWL	ISO 16739-1
Onuma Inc.	BIMXML	Buildings	XML, Java	
Linked Building Data Community Group	Building Topology Ontology (BOT)	Buildings	RDF	
Brick Consorium	BrickSchema	Building systems	TTL	
BuildingSMART / OGC	LandInfra	Infrastructures	XML/InfraGML	
Land XML consortium	LandXML	Civil eng. design (survey)	XML	
CEN/TC 278	TN-ITS spec.	Road network and regulation	XML, XSD, WADL	TN-ITS CEN TS17267:2018
OGC	CItyGML	City/Territory	XML, GML, Model-Driven Architecture (MDA)	ISO 19103, ISO 19109 ISO 19136, ISO 19107, ISO 19111
EU	INSPIRE	Land cover	XML/GML	Directive 2007/2/EC
ISO/TC 211	LADM	Land administration		ISO/TC 211

* Table extended from (Moretti et al. 2022).

TABLE 2. Examples of DT implementation and the data modelling strategies adopted.

Level	Implementation	Goal	Data modelling strategy	General limitation
Building	Sydney Opera House (Schevers et al. 2007)	Developing a BIM-supported digital twin of the Opera House in Sydney, able to support life cycle operations	BIM proprietary files and IFC conversion for FM functions	Rigid data structure limits the flexibility in accommodating uncertainties and the applicability of the presentation in solving generalised problem
	CARTIF-3 INSITER (Hernández et al. 2018)	Using self-inspection techniques and virtual reality for construction, refurbishment and maintenance of energy-efficient buildings	IFC	
	eLUX Digital Twin (Tagliabue et al. 2021)	Increasing the automation level of building operation, through the use of sensors and smart technologies	3D model of the case study building, not specified	
District	New York Open Urban Platform (Eicker et al. 2020)	Developing an urban platform, for improving energy and waste/wastewater issues	Open Urban Platform (OUM) data model based on CityGML	Features and attributes need to be defined and added to the open and platform-independent data model, to represent application-specific information in various granularities.
	UNSW campus (Aleksandrov et al. 2019)	Developing a 'smart campus' with intelligent strcturing and management of geospatial data	Unified Modeling Language (UML) based architecture (called PIM)	
	ViLo (Aleksandrov et al. 2019)	Visualising and analysing both real-time and offline space-time urban data	Not specified	
	Amsterdam Smart Port (Amsterdam Smart City 2016)	Increasing the sustainability of the port area among other metropolitan issues in social, economic and ecological issues	Real-time 3D graph, not specified	
	City-zen smart grid (European Commission 2009)	Improving the energy management and operations at the urban infrastructure level	Not specified	
City	Vivacité energy platform (Koch-Mathian et al. 2019)	Monitoring energy and water consumption from the city system operators	Not specified	Except for smart Cambridge, most of the DT implementations do not explicitly describe the data modelling strategy.
	Herrenberg smart city (Dembski et al. 2020)	Enabling domain-specific applications for smart cities and simulations	3D model embeded in COVISE software	
	Smart Cambridge (Smart Cambridge 2020)	Driving the transformation of transportation in the Greater Cambridge area	Customised "crate" model in JSON	
	Newcastle digital twin (Newcastle University 2019)	Testing the infrastructure's potential in hypothetical situations like rising sea levels, etc.	IFC, CityGML	
	Singapore Smart Nation initiative (Singapore Smart Nation 2021)	Harnessing digital technologies to improve living, boost the job market and encourage businesses to innovate and grow	Not specified	
	Bristol is Open (BiO) (Bristol City Council 2021)	Strengthening the city's digital foundations to keep the city moving, healthy and safe in a sustainable way	Not specified	

TABLE 3. Possible AIR categories to be used in the definition of the Digital Twin federated data model.

Category	Requirements
General	Use, Easement, Authorisations, etc.
Technical	Energy, fire safety, structure, systems installation and conformity, etc.
Asset Information	Functional/object classifications, Id, manufacturer, model, warranty, service life, etc.
Operations&Maintenance	Planned interventions, frequencies, maintenance history, work orders, status, costs, etc.
People	Owner, responsible organisation, vendor etc.

TABLE 4. Data model mapping issues in most common data models

Issues	Description
Encoding	Different techniques in the data encoding and data structures. E.g., entity-relationship model, tree representation, graph, etc..
Semantics	Formal generalisation of the modelled entities according to different scales and granularity of the data.
Geometry and location	Different representation of the geometries (e.g, Constructive Solid Geometries - CSG, BRep and Sweeping); separation between semantics and geometries; local vs global coordinates systems, implicit or explicit topology representation.

TABLE 5. Digital Twin federated data model Information Requirements. Data Set (DS), Certificate (CE), Drawing (DG), Model – three-dimensional (M3), Database (DB), Data sheet (DT)

Id	Category	Sub-category	Group	Form of Info
IR.1	**Built Environment**			
IR.1.1.		**General information**		
IR.1.1.1			Use	CE
IR.1.1.2			Easements	CE
IR.1.1.3			Authorisations	CE
IR.1.1.4			Listing	CE
IR.1.2.		**Technical information**		
IR.1.2.1			Energy	DS
IR.1.2.2			Structures	DS
IR.1.2.3			Fire safety	DS
IR.1.2.4			Installation and conformity of systems	CE
IR.1.2.5			As built drawings	DG
IR.1.2.6			3D models	M3
IR.1.2.7			Functional classification	DB
IR.1.2.8			Object classification	DB
IR.1.2.9			Type of sensor	DT
IR.1.2.10			Generation	DT
IR.1.2.11			Sensor readings	DS
IR.1.2.12			Related assets	DB
IR.3	**Asset Management**			
IR.3.1.		**Asset Information**		
IR.3.1.1			Asset IDs	DS
IR.3.1.2			Description	DT
IR.3.1.3			Manufacturer Model	DT
IR.3.1.4			Manufacturer serial number	DT
IR.3.1.5			Acquisition Date	DS
IR.3.1.6			Asset warranty	DS
IR.3.1.7			Reference Service Life (RSL)	DS
IR.3.1.8			Actual Service Life (ASL)	DS
IR.3.1.9			Estimated service Life (ESL)	DS
IR.3.2.		**O&M Data**		
IR.3.2.1			Planned maintenance interventions	DB
IR.3.2.2			Maintenance frequencies	DB
IR.3.2.3			Maintenance history	DB
IR.3.2.4			Open work orders	DB
IR.3.2.5			Maintenance status	DS
IR.3.2.6			Maintenance cost	DB
IR.3.2.7			Installation cost	DB
IR.3.2.8			Replacement cost	DB
IR.3.2.9			Spare parts	DB
IR.3.2.10			Inspection type	DS
IR.3.2.11			Inspection frequencies	DB
IR.3.2.12			Inspection report	DB
IR.3.3.		**People**		
IR.3.3.1			Owner	DS
IR.3.3.2			Tenant	DS
IR.3.3.3			Responsible mgt group	DS
IR.3.3.4			Vendor/Supplier	DS

List of Figures

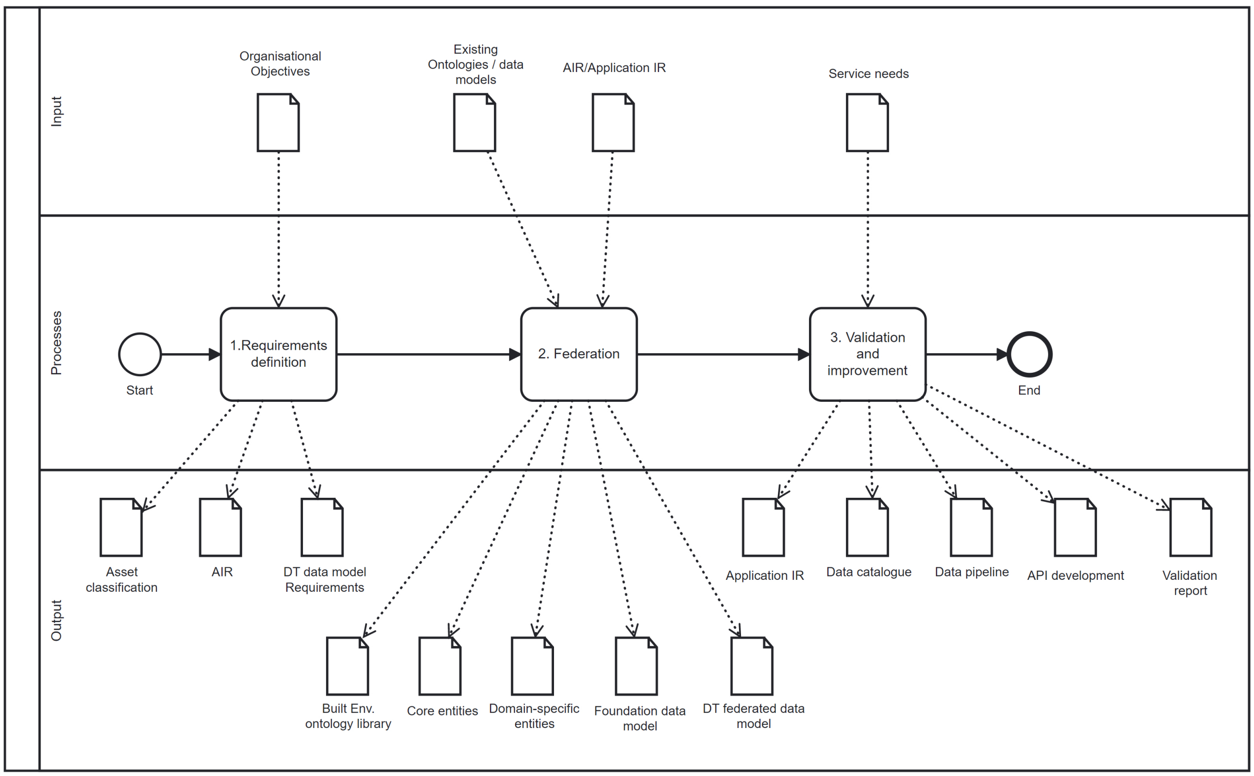

Fig. 1. Main phases implemented for developing and testing the federation approach.

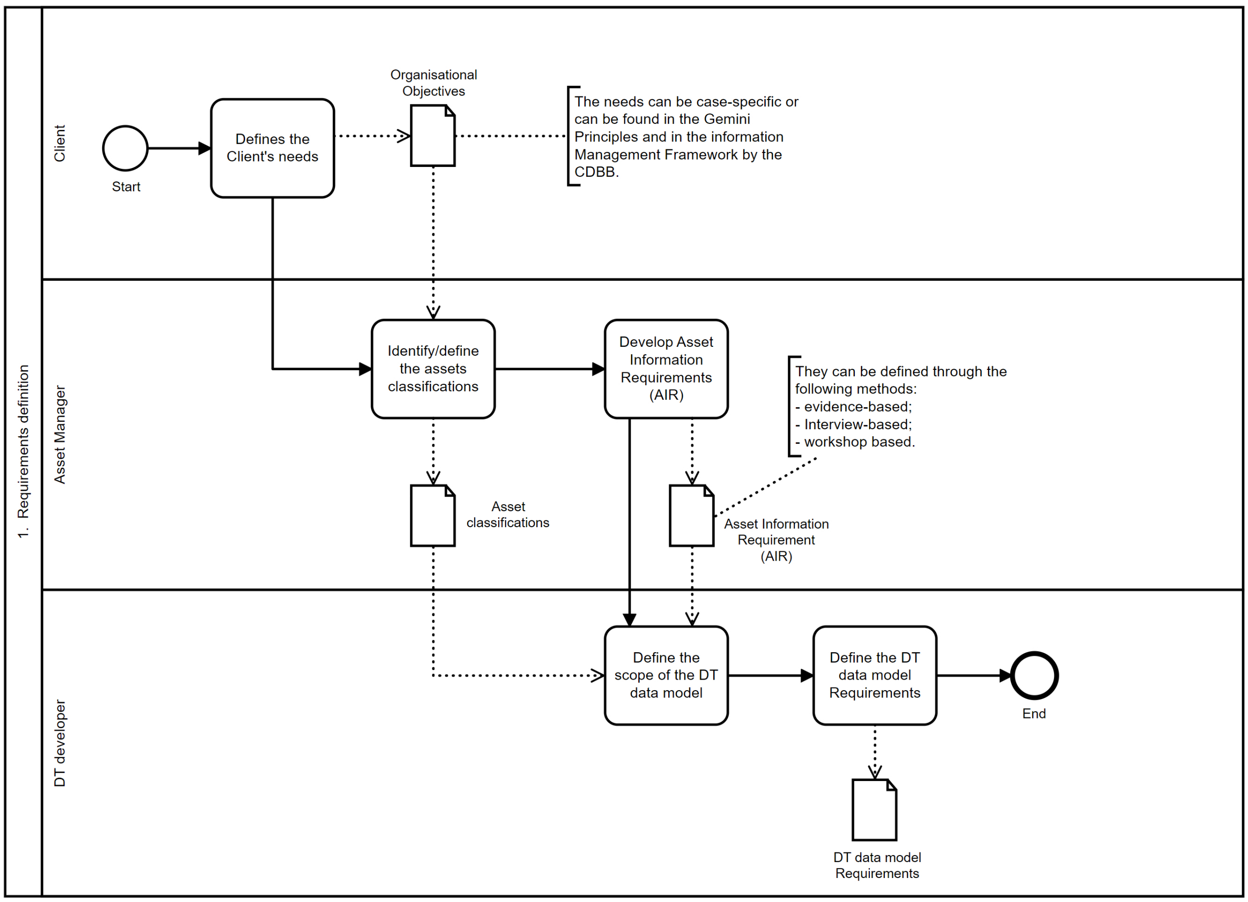

Fig. 2. Definition of the requirements process map.

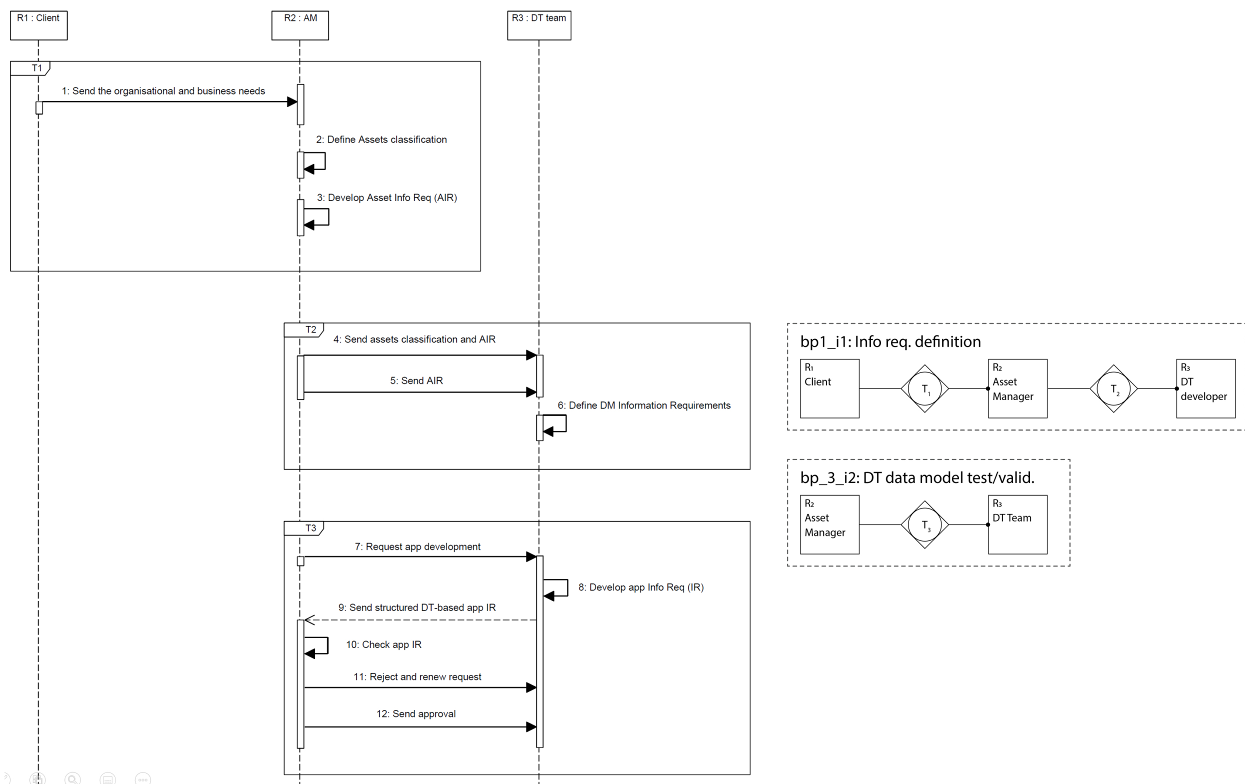

Fig. 3. Modelling of he transactions and interactions realised in the development of the built environment federated DT data model.

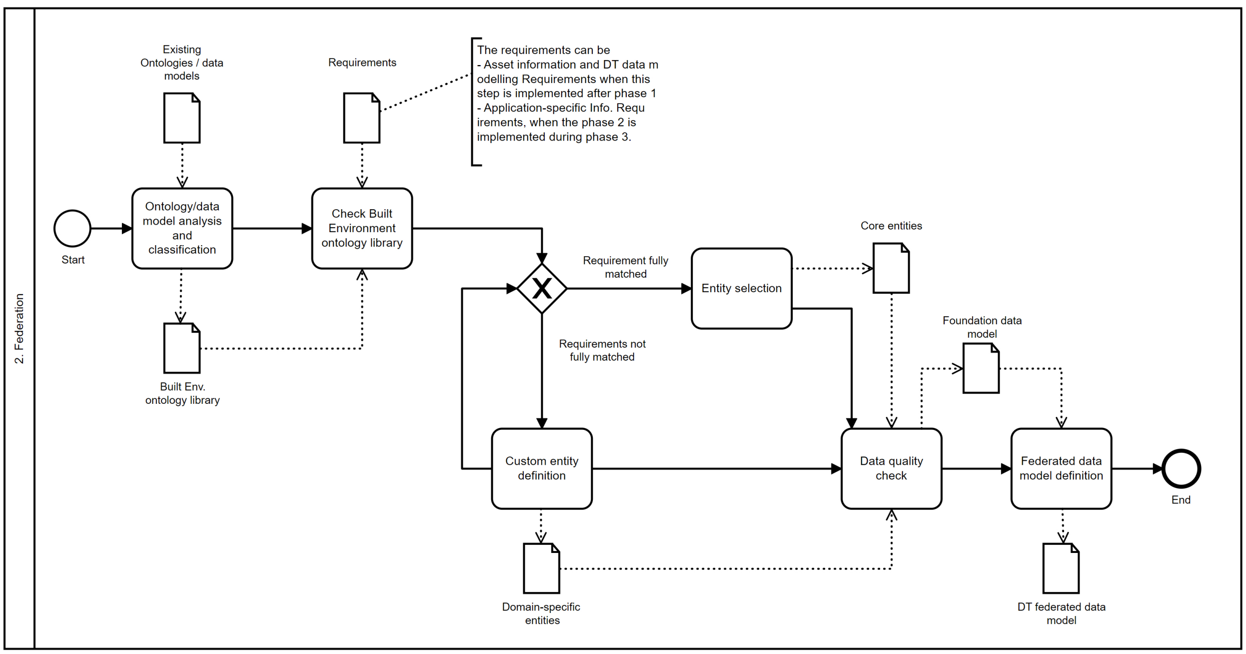

Fig. 4. Data model federation process map.

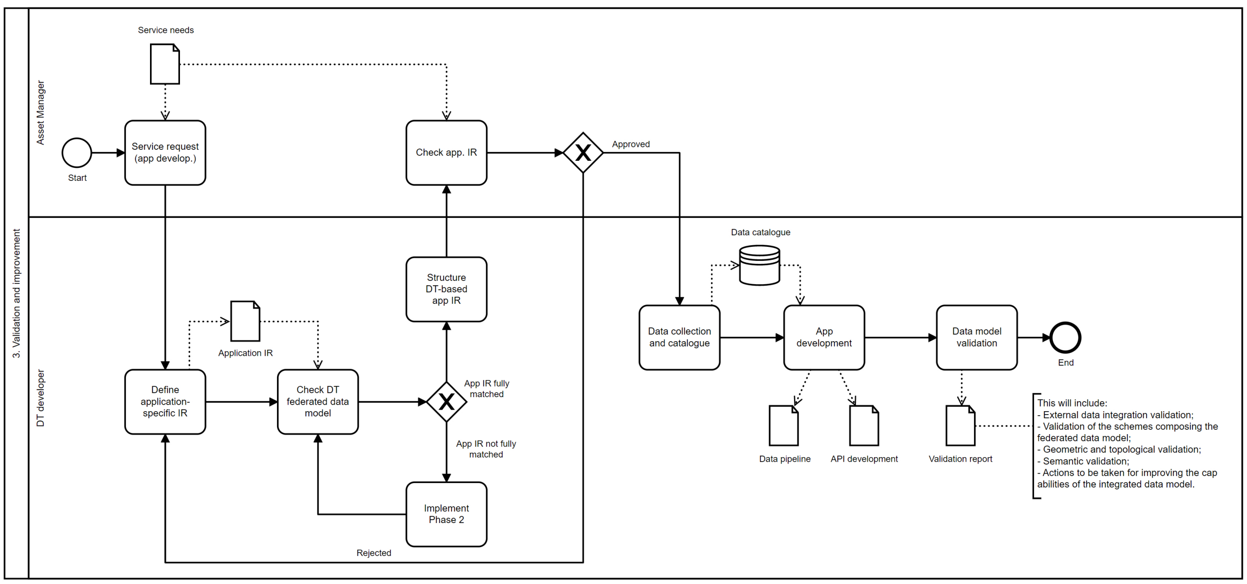

Fig. 5. Validation and improvement process map.

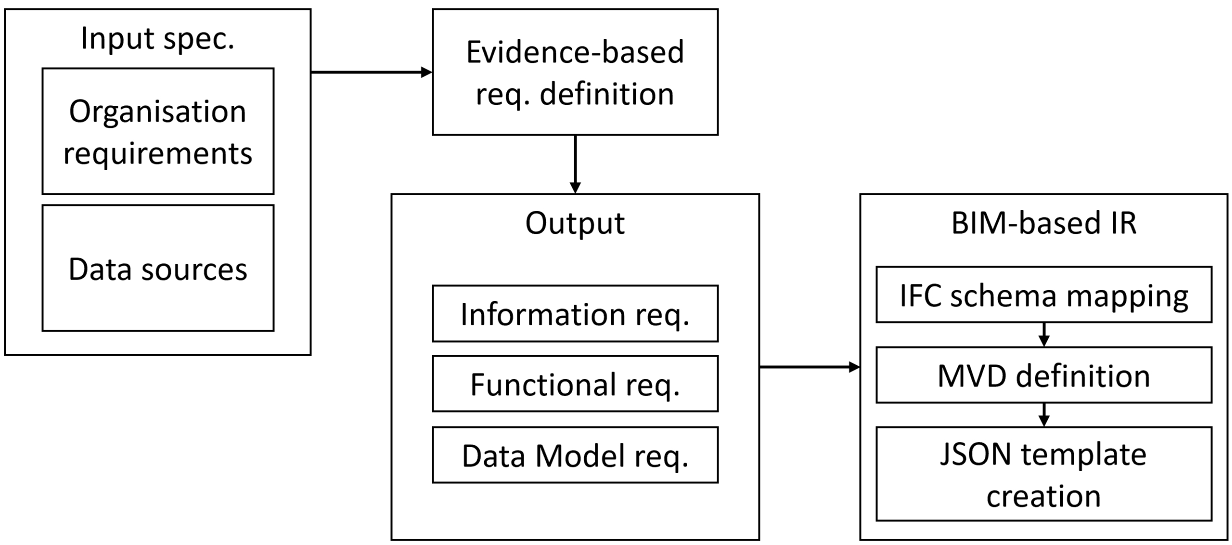

Fig. 6. BIM-based Information Requirements definition.

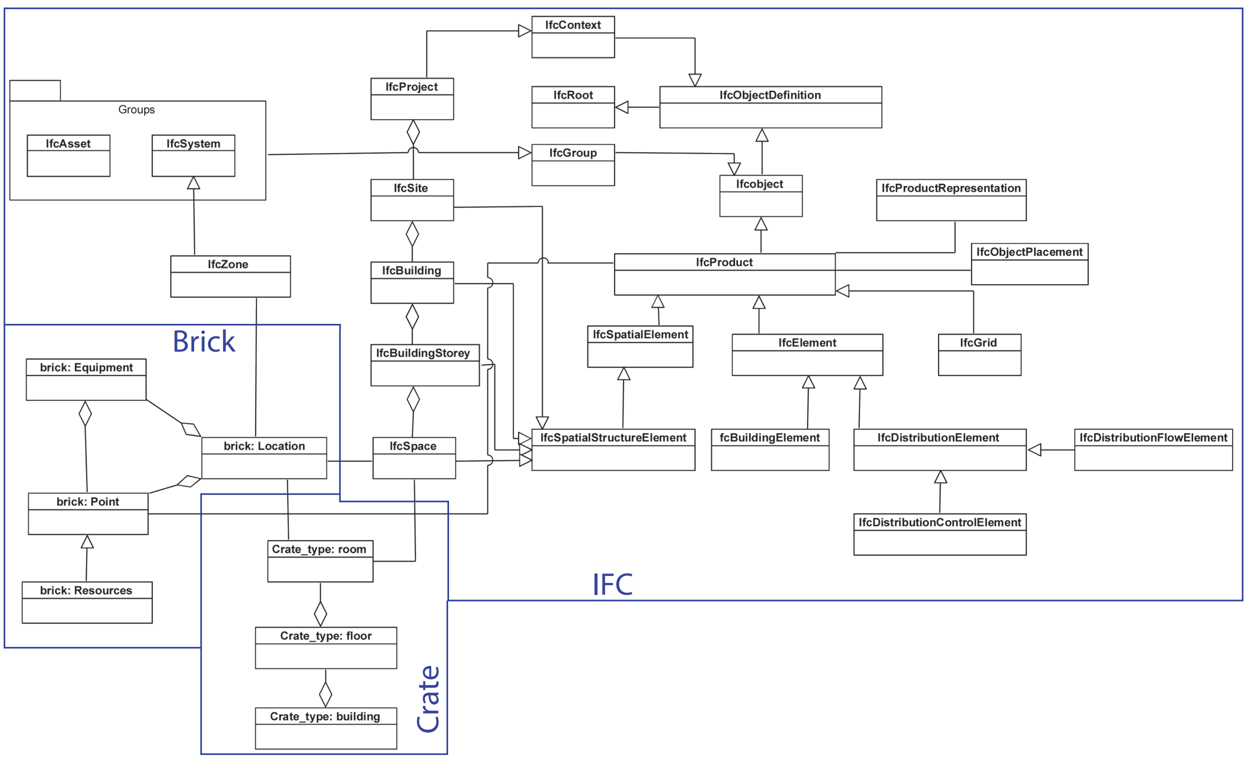

Fig. 7. West Cambridge Digtal Twin federated conceptual model.

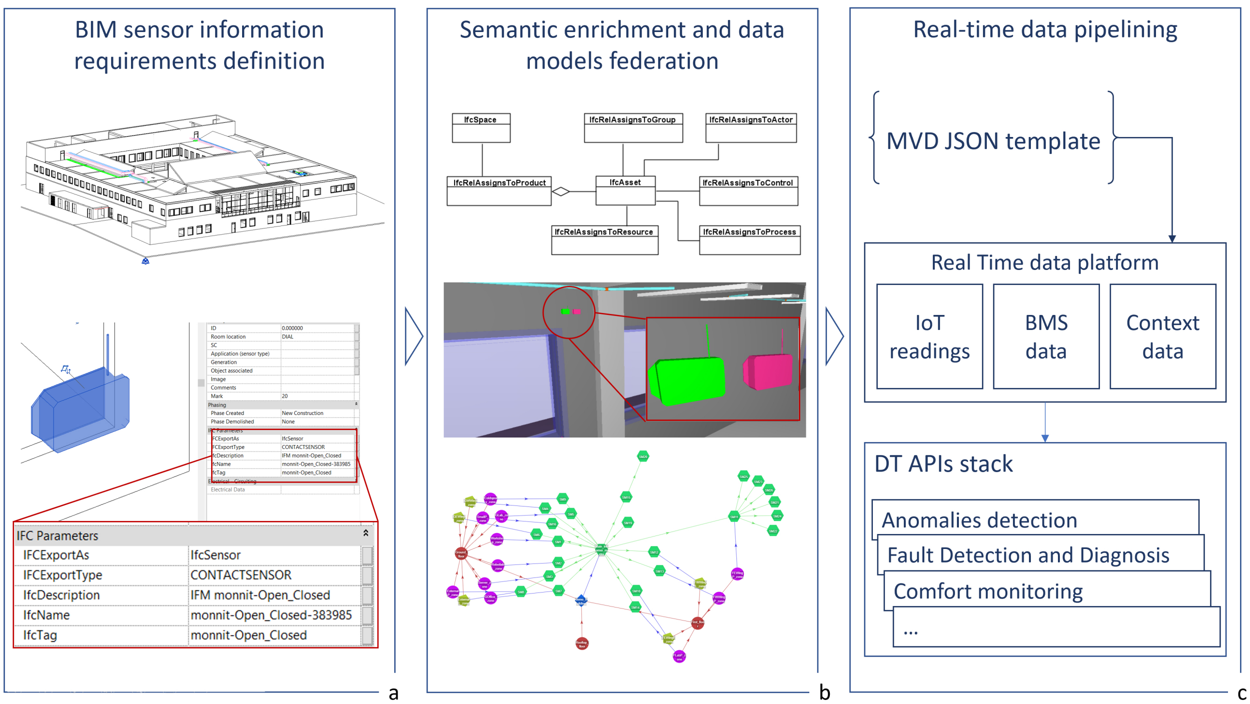

Fig. 8. openBIM-based sensors Information Requirements definition (a), federation (b) and pipelining (c) in DT-based APIs.