Developing a Digital Twin at Building and City Levels: A Case Study of West Cambridge Campus

Qiuchen Lu [1*], A.M. ASCE; Ajith Kumar Parlikad [2]; Philip Woodall [3]; Gishan Don Ranasinghe [4]; Xiang Xie [5];
Zhenglin Liang[6]; Eirini Konstantinou [7]; James Heaton [8]; Jennifer Schooling[9]

Abstract

A Digital Twin (DT) refers to a digital replica of physical assets, processes and systems. DTs integrate artificial intelligence, machine learning and data analytics to create living digital simulation models that are able to learn and update from multiple sources, and to represent and predict the current and future conditions of physical counterparts. However, the current activities related to DTs are still at an early stage with respect to buildings and other infrastructure assets from an architectural and engineering/construction point of view. Less attention has been paid to the operation & maintenance (O&M) phase, which is the longest time span in the asset life cycle. A systematic and clear architecture verified with practical use cases for constructing a DT would be the foremost step for effective operation and maintenance of buildings and cities. To this end, this paper presents a system architecture for DTs, which is specifically designed at both the building and city levels. Based on current research about multi-tier architectures, this proposed DT architecture enables integration of heterogeneous data sources, supports effective data querying and analysing, supports decision-making processes in O&M management, and further bridges the gap between human relationships with buildings/cities. Based on this architecture, a DT demonstrator of the West Cambridge site of the University of Cambridge was developed. This paper aims at going through the whole

---

[1] Research Associate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: ql284@cam.ac.uk (Corresponding Author)
[2] Senior Lecturer, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: aknp2@cam.ac.uk
[3] Senior Research Associate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: pw325@cam.ac.uk
[4] Ph.D. Candidate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: gd416@eng.cam.ac.uk
[5] Research Associate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: xx809@cam.ac.uk
[6] Research Associate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: zl284@eng.cam.ac.uk
[7] Research Associate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: ek415@cam.ac.uk
[8] Ph.D. Candidate, Institute for Manufacturing, 17 Charles Babbage Road, University of Cambridge, Cambridge, UK, CB3 0FS, Email: jrh212@cam.ac.uk
[9] Director of the Centre for Smart Infrastructure and Construction, University of Cambridge, Cambridge, UK, CB2 1PZ, Email: jms33@eng.cam.ac.uk

process of developing DTs in building and city levels from the technical perspective and sharing lessons learnt and challenges involved in developing DTs in real practice. Through developing this DT demonstrator, the results provide a clear roadmap and present particular DT research efforts for asset management practitioners, policymakers and researchers to promote the implementation and development of DT at the building and city levels.

Keywords: Digital Twin (DT), Asset Management, Operation & Maintenance (O&M), Building and City Levels

Introduction

Computerisation and digitisation are emerging to have a widespread impact on the way the lifecycle of physical/engineering assets being managed (Pärn et al., 2017). For instance, artificial intelligence (AI) is predicted to add 10% to the UK economy by 2030, and improved data sharing can result in lower consumer bills, reduce the impact on the natural environment and realize smart asset management (NIC, 2017). Advances in building information modelling (BIM) is likely to aid the reduction of the time taken for updating databases in operations and maintenance (O&M) phases by 98% (Ding et al. 2009). The necessary technologies and approaches, such as data integration and processing (Woodall 2017), information and communication technologies (ICTs) (Ahuja et al. 2009), BIM etc., are more or less already available. However, data needs to be stored and shared safely and securely, and technologies also need to be well-designed and ensure security and efficiency (NIC, 2017). Therefore, the concept of digital twins (DTs) has evolved as a comprehensive approach to manage, plan, predict and demonstrate building/infrastructure or city assets.

DTs align well with other related emerging paradigms such as Cyber-Physical Systems and Industrie 4.0, and it is predicted that half of the large industrial companies will use DTs by 2021, resulting in those organizations gaining a 10% improvement in effectiveness (Gartner 2017). In the architecture, engineering, construction and facility management (AEC/FM) sectors, DTs are examined in the context of smarter cities/buildings. For instance, Mohammadi and Taylor (2017) provided predictive insights into a city's smarter performance and growth based on virtualization and DT of the city. Ma et al. (2018) also explored the role of Big Data in urban physical, social and cyber spaces to construct smart cities. Moreover, Oliver et al. (2018) provided a practical investigation of developing DTs using the example of the new University College London campus. However, unified guidance and wider applications at different levels were still limited in their research. A number of studies also exist where only some of the DT's concepts have been implemented. For instance, Motawa and Almarshad (2013) proposed a Case-Based Reasoning (CBR)-integrated BIM system for building maintenance to improve the efficiency of decision making and communication among different stakeholders. The restoration team of the Sydney Opera House designed a unified central data repository integrating different resources to support effective O&M management (CRC Construction 2007). Clearly defined and well-organised principles and a system architecture to supervise the implementation will help identify the shortcomings of current approaches and provide roadmaps for future development. These are missing in current developments and the literature, and thus form the core focus of this paper.

Furthermore, NIC 2017 states 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 framework can benefit for better understanding the performance data and fitness for uses. In order to maximize the value of data, present DT development processes and further evaluate the value and challenges of DTs, this study firstly presents a system architecture for DTs at both building and city levels. This architecture is brought to life through the development of a DT demonstrator of the West Cambridge site, in the University of Cambridge.

Literature Review

Proposition of a DT development in building and city levels is raised due to research attempts and industry trends. This literature review firstly discusses the existing definitions related to DTs. Lessons can be learned through the review of current literature discussing limitations related to research efforts based on partial concepts of DTs in the AEC/FM sector. This section, therefore, aims at providing a well-grounded foundation for further system architecture and demonstrator development.

Definitions of DTs

In simple terms, a DT is a dynamic digital representation of an asset/system and mimics its real-world behaviour (GE Digital, 2017; Bolton et al., 2018). The concept of DTs originated from the aerospace industry when NASA published a roadmap on modelling and simulation, where they provided the first definition for DTs (Shafto et al., 2012). Although gaining popularity in the academic literature and industrial practice, there is no commonly accepted definition for it. A brief examination of the literature (Table 1 provides a few definitions from the perspective of different industry sectors) shows that although the precise definitions vary, the overall thrust should be similar.

Table 1: Definitions of Digital Twin

Reference	**Definition**	**Industry**
Shafto et al (2012); Glaessgen and Stargel (2012); Knapp et al. (2017)	An integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin. The digital twin is ultra-realistic and may consider one or more important and interdependent vehicle systems.	Aerospace
Grieves, M. and Vickers (2017)	A set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level. At its optimum, any information that could be obtained from inspecting a physically manufactured product can be obtained from its Digital Twin.	Complex systems
Bolton et al., (2018)	A realistic digital representation of assets, processes or systems in the built or natural environment.	Infrastructure
GE Digital (2017)	A dynamic digital representation of an industrial asset, that enables companies to better understand and predict the performance of their machines and find new revenue streams, and change the way their business operates.	Manufacturing systems, Industrial equipment
HVM Catapult (2018)	A model of the physical object or system, which connects between digital and physical assets, transmits data in at least one direction and monitors the physical system in real-time. In addition, they also should support analytics, control and simulation functions.	Manufacturing systems

This paper specifically focusses on the AEC/FM sector, which as we will show in this section later, currently lags behind the manufacturing and aerospace sectors in the maturity of development in digital twins. The National Infrastructure Commission – in their report ‘Data for the Public Good’ – set forth a number of recommendations for the government towards digital infrastructure (NIC, 2017). One of those key recommendations was to develop a so-called ‘National Digital Twin’. The ‘Gemini Principles’ published by the UK Digital Framework Task Group and the Centre for Digital Built Britain outlines a fundamental set of ‘properties’ a digital twin – and hence the national digital twin – should adhere to (Bolton et al., 2018).

Fundamental to this concept is that a national digital twin is not a single monolithic model of a whole nation's infrastructure, but consists of digital twins that are constructed in different scales (e.g., individual asset scale, network/system scale, city scale), built for various purposes, and using different approaches, that are connected together and all built on data. In the AEC/FM sector, a DT of a city, for instance, would be built on a hierarchical architecture and include a network of sub-DTs (e.g., building DTs). For the purposes of this study, a DT refers to "a dynamic digital replica of physical assets, processes and systems through involving internet of things (IoT) devices and information feedback from citizens" (Bolton et al., 2018; Sackey et al. 2014; Inyim et al. 2014). Dynamic city DTs integrate their sub-DTs and intelligent functions (e.g., AI, machine learning, data analytics etc.) to create digital models (e.g., simulation) that are able to learn and update from multiple sources, and to represent and predict the current and future condition of their physical counterparts correspondingly and timely.

The state-of-the-art of DT development in AEC/FM sector

The effectiveness of asset management in the O&M phase would heavily rely on continuous data on asset conditions and performances and properly documented professional knowledge (Pärn et al. 2017; France-Mensah and O'Brien 2018; Lu et al. 2015; Lu et al. 2018a). There have been a number of contributions by the academic community that enables the exploitation of BIM and digital technologies/tools in the through-life management of building and infrastructure assets. A summary of the key literature in this area along with their key contributions is provided in Table 2. Most of these studies focus on some of the concepts of DTs for developing high-performance BIM-enabled asset management systems (Farghaly et al. 2018; Giel and Issa 2015; Son et al. 2017; Song et al. 2017) or project management development (Cao et al. 2016; Taylor and Bernstein 2009; Ma et al. 2018). It can also be seen that these studies concentrate on specific applications such as enhancing collaboration, improved visualisation, optimising work orders. The review also reveals that current developments focus on and/or utilise limited data resources, and does not integrate all existing data sources to support their digital development. They lacked a comprehensive overview and a system architecture (i.e., DTs), which establishes the foundation (e.g., asset and data integration), organises the internal structure and further guides for continuous development.

Table 2: Brief summary of BIM-enabled asset management development

Author/year	Key technologies	Key algorithms/tools	Key contribution
Lee et al. (2013)	Sensor, BIM, GIS, Ubiquitous Sensor	Integration of facilities-related information and	Presents an intelligent urban facilities

	Network, Urban Object Identification	integration of management functions	management for real-time emergency response
Kang and Hong (2015)	GIS, BIM, IFC, CityGML	BIM/GIS-based information Extract, Transform, and Load (BG-ETL) architecture	Proposes a software architecture for the effective integration of BIM into a GIS-based FM system
Róka-Madarász et al. (2016)	CAFM, CAD, Database	Top-down object hierarchy; Geometric Description Language	Elaborates a methodology for gathering building O&M costs data
Shalabi et al. (2016)	BIM, IFC, BEMS, BAS	A schema that enables the integration of data; a process linking alarm reports of equipment failures with IFC BIM	Proposes an automated process that responds to alarms by retrieving alarms reported by FM systems for corrective maintenance
Peng et al. (2017)	Data warehouse, BIM	Clustering algorithm; Cluster-based frequent pattern mining algorithm	Proposes a BIM-based Data Mining approach for extracting meaningful patterns and detecting improper records
Suprabhas et al. (2017)	BIM, Sensor, COBIE	Data integration and visualisation	Develops an application that integrates sensor data and reports the data via the virtual model of the building.
Hu et al. (2018)	BIM, GIS, BAS, Web-service, QR code/RFID	Logic chain generation algorithm; Equipment identification and grouping algorithm	Develops a cross-platform Mechanical, Electrical and Plumbing (MEP) management system
Chen et al. (2018)	BIM, IFC, Facility management systems	A* algorithm used for optimal maintenance path planning; Dijkstra algorithm used for maintenance scheduling	Proposes a BIM-based framework for automatic scheduling of facility maintenance work orders

Note 1: GIS: geographic information system; RFID: radio frequency identification devices; BEMS: building energy management systems.

The state-of-the-art of multi-tier architectures development

Various researchers have proposed multi-tier architectures to support heterogeneous environments (e.g., multi-function and a large amount of data). This can be classified as (1) Cyber-Physical Systems (CPS), (2) IoT platform architectures, and (3) smart cities and big data architectures. Table 3 provides a summary of multi-tier architectures from related literature.

For the architecture of CPS, a new CPS science is still needed to integrate the theories of computing and communication systems, sensing and control of physical systems, and the interaction between humans and CPS (Rajkumar et al., 2010). For the architecture of IoT, big data techniques and cloud computing are suggested to improve its performances. For smart cities architectures, current research is limited to some specific applications (e.g., only considering city-level implementations) and more interaction with human users should be proposed in their architectures.

Table 3: Summary of multi-tier architectures

Architecture Classification	Key Layers	Challenges	Reference
Architecture of Cyber-Physical Systems (CPS)			
CPS for Electric Power Grid	Connection, conversion, cyber, cognition, configuration	Communications, computing and physical dynamics should be modelled at different levels of **scale**; Improved unified and core abstractions of computing are needed;	Lee et al. (2015); Lee (2008); Kleisslv and Agarwal (2010); Rajkumar et al. (2010)
CPS for Smart Building			
IoT Platform Architecture			
IoT-based Services	IoT architecture includes: three-layer; middle-ware based; SOA based; five-layer.	**Big data** analytics in support of the IoT are needed; **Cloud computing** for the IoT are needed; Fog Computing can act as a bridge between smart devices and large-scale cloud computing and storage services; The need for better **horizontal integration** between application layer protocols.	Al-Fuqaha et al. (2015); Krylovskiy et al. (2015)
IoT supported Smart Cities	Middleware		
Smart Cities and Big Data Analytics Architecture			
Urban Planning and Building Smart Cities based on the IoT using Big Data analytics	Bottom tier; intermediate tier 1&2; top tier.	Limited implemented areas; No interaction with citizens.	Rathore et al. (2016); Silva et al. (2017)

Note 1: SOA means Service Oriented Architecture.

To construct an effective digital architecture to exploit the benefits of sharing better quality services in the building and city level, the following challenges still need to be addressed:

a) The architecture should be developed using a unified, hierarchical and extensible approach, which can be implemented in different scales from assets (e.g., pump), buildings to cities.

b) Besides data collection and acquisition, assets need to be 'connected' and relevant information regarding their lifecycle (e.g., maintenance history) should be collected as well.

c) Interaction and communication channels with humans are needed to provide 'in-time' services.

d) Data or status visualisation is required for different groups of users to help them monitor 'as-is' condition and activities.

DTs can support many different applications such as from security and health management to energy management. Each application will have its own data requirements which need to be catered for. This is problematic when data comes from different systems, because the source system may have a different intended use of these data that does not fully match the requirements of all those applications. Dealing with these differences and repurposing data from the source systems poses a challenge (Woodall, 2017) especially for developing a specific architecture for DT development in the AEC/FM sector.

A DT System Architecture for Building and City Levels

A city is a comprehensive system connecting the physical, social and business aspects (Silva et al. 2018). Widespread deployment of ICT infrastructure in cities allows the extraction of intelligence from various datasets and allows it to connect different asset groups (Silva et al. 2018). A city can be thus considered as an asset that integrates different sub-assets such as buildings, utilities, transportation infrastructure, and people. Hence, a DT at the city level is a dynamic digital replica of a city that integrates each sub-DT (e.g., building DT, bridge DT) (see Figure 1). Figure 1 demonstrates the parent-child relationship of DTs at different levels. DTs in the upper level (e.g., city DT) interact with the sub-DTs (e.g., building DT) in a bidirectional way, by querying for the required information, responding to different stakeholder requirements and providing them with specific services without compromising data confidentiality at each individual DT. This study presents a hierarchical architecture at the building and city levels. This architecture (as shown in Figure 2) is comprised of five layers: data acquisition layer, transmission layer, digital modelling layer, data/model integration layer and service layer.

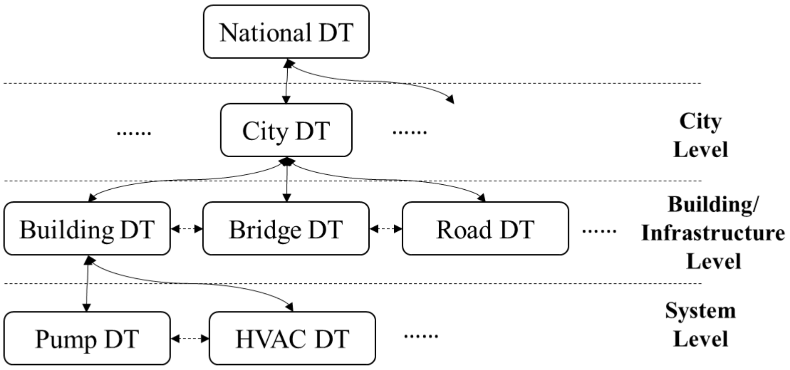


Figure 1. DTs connection and hierarchy among different levels

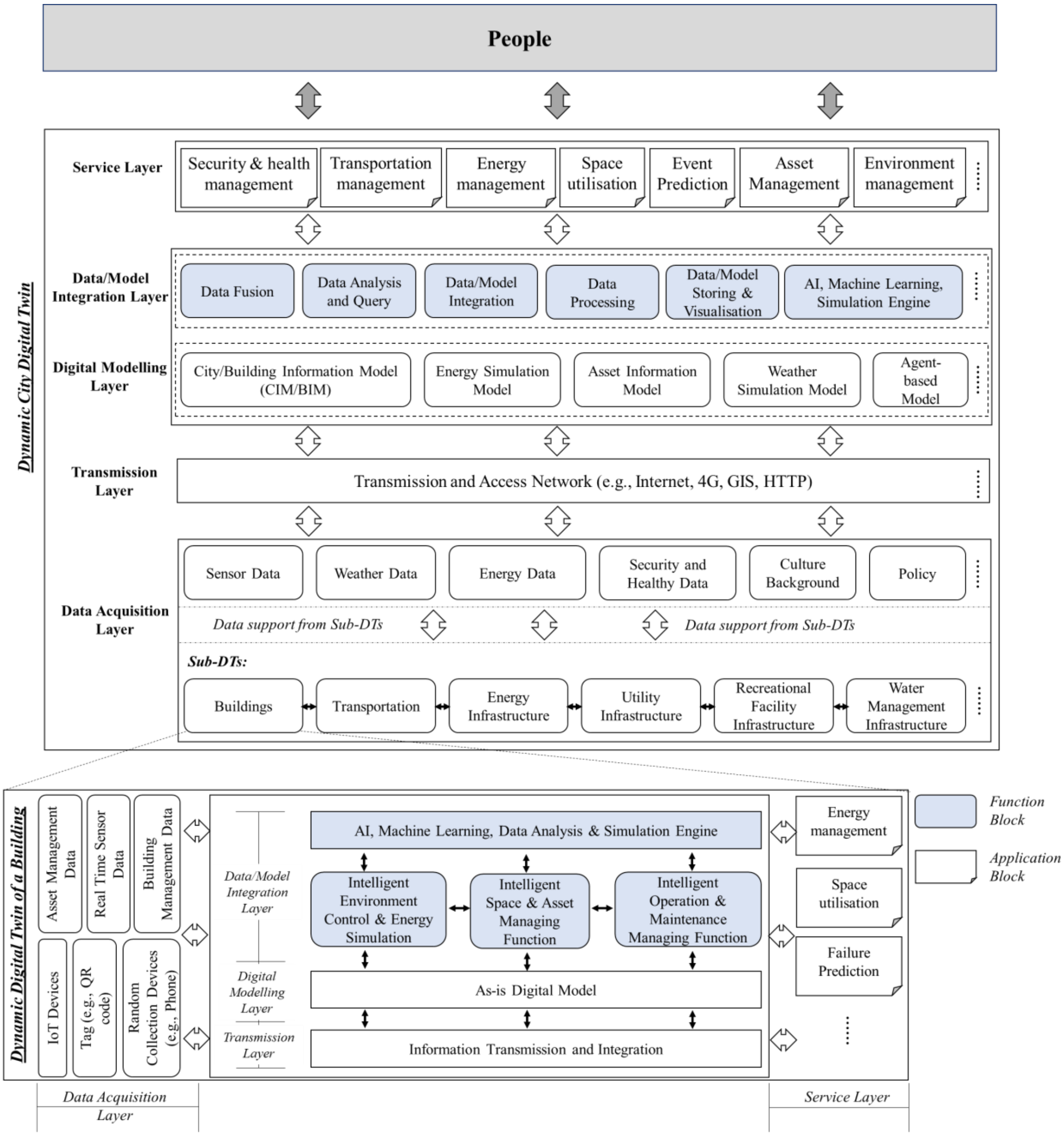


Figure 2. The system architecture of DT development in a city and building level

- Data acquisition layer is the foundation of each DT. Due to the heterogeneity and large volume of data in city levels, design of data acquisition mechanism and approach is a foremost and challenging task, especially when considering the type, format, source and content of data. Moreover, the sub-assets (e.g., buildings, transportations) will have their sub-DTs in terms of their functions in daily services and these sub-DTs will further provide necessary data/information/model when receiving a query from the city DT. Examples of data collection techniques include: contactless data collection (e.g., RFID, image-based techniques), distributed sensor systems, wireless communication, and mobile access (e.g., WiFi environment). Based on different levels of sub-DTs (e.g., buildings), each twin is designed based on the DT architecture, including real-time data collection, effective data management and integration (Fattah et al. 2017; Hu et al. 2018). For instance, the DT architecture is presented in Fig.2 (bottom part) and designed for buildings. Through sharing the same architecture with city DTs, building DTs also include data acquisition layer (e.g., using IoT devices and wireless sensor network or QR codes), digital modelling layer, transmission layer, digital modelling layer, data/model integration layer (e.g., simulation engine and data analysis functions) and service layer (e.g., space utilization and workplace design).
- The transmission layer aims at transferring the acquired data to the higher layers for modelling and analysis. Various communication technologies could be used in this layer, such as short-range coverage access network technologies (e.g., WiFi, Zigbee, near field communication (NFC), M2M, and Zwave) and wider coverage (i.e., 3G, 4G long-term evolution (LTE), 5G, and low-power wide-area networks (LP-WAN)) (Ge et al. 2016; Huang et al. 2012; Ohmura et al. 2013). With the increasing development of technologies, Wi-Fi is still the well-known wireless local area network (WLAN) technology and widely used. Although the most popular technology, the unlicensed spectrum band is a concern when developing city DTs using Wi-Fi (Lehr and McKnight 2003) due to security issues. Considering energy efficiency of networks and speed of transmission, light fidelity (Li-Fi) and LP-WAN are promising alternatives for wide-range coverage for developing DTs in building and city levels (Saini 2016; Silva et al. 2018).
- Digital modelling layer contains a set of digital models of the physical asset (e.g., BIM, City Information Modelling (CIM)) and supplements information (e.g., weather information, cultural backgrounds) that support the upper layers. The CIM shares similar

concepts with BIM, while describes information models in city levels. It extends the use of models, information, and techniques in urban levels (e.g., Geographic Information Systems (GIS)) in city applications (e.g., urban planning) as decision support tools (Gil et al. 2010; Gil et al. 2011). Different models/model types can be used for different purposes in DTs. Examples for these are: real-time status/control, managing assets (e.g., asset management model), planning infrastructure/cities (e.g., CIM), modelling scenarios and decision support (Bolton et al., 2018; Kim et al., 2018). When a DT at building and city levels is designed, a pre-defined schema and well-organised modelling processes are required to conform firstly and aligned to the target specific applications from a single infrastructure to entire cities and buildings.

- Data/model integration layer is the kernel in this architecture. This layer aims at integrating all the data resources based on the designed data structure. This layer also contains the functions required for data and model manipulating, storing, analysing, integrating, processing, and AI-supported knowledge learning which supports decision making (Glaessgen and Stargel 2012). In this architecture, real-time data analysis and processing functions would update as-is conditions of the city assets (including transportation conditions, energy consumption) and building assets (including work orders, up-to-date maintenance information, status) (Lu et al. 2018b). Where complex and massive amounts of data are collected and large-scale data storage & managing systems are needed in a city level, effective and hierarchical model/data storing, integration and query design are the most significant functions for guaranteeing the city DTs' performance. Here, cloud storage and computing, and data/model visualisation can be used to achieve dynamic and effective data management in a city and building level (Lin et al. 2013; Silva et al. 2017). Data is the core of the DT architecture. Based on all the available data resources, different intelligent functions (e.g., AI, machine learning module and simulation) could be realized for advanced decision-support, such as transportation prediction, energy usage optimisation or asset anomaly detection. These functions are essentially driven by different knowledge engines (KEs). By assimilating data continuously, live KEs for physical assets, processes and systems can be established, describing their dynamic conditions. The establishment of KEs is much dependent on domain knowledge. Hence, in a DT with multi-functions, a specific KE under a certain scenario would be developed and added under a target domain knowledge. It is crucial to recognize that embedded KEs play a key role in delivering better-informed services by utilizing the strong data integration capability of the proposed DT. This study provides a case of a pump to demonstrate an example of KE. Different KEs would depend on different domain knowledge, which will be proposed and designed by different DT developers based on their different purposes. In addition, intelligent functions can keep updating their embedded algorithms and supporting continuous applications in future development.

- The service layer is the top and implementation layer of the DT architecture that interprets the knowledge from KE and enables the interaction between people/society and the data/model integration layer. The service layer provides services for the society, evaluates performances of constructed DTs and can influence human satisfaction including sustainable community development, environmental management, and smart transportation. And the feedbacks from people should feed into KEs as external knowledge for improving overall satisfactions.
- Interaction with people: in the designed architecture, the service layer is targeted towards FM professionals and end-users, providing them with decision-making supports and interaction. To avoid compromising the operation performances especially in the early implementation stage, the optimized decisions should be checked and confirmed manually before implemented in practices. The designed smart building/city allows for flexible decision-making process and supporting the interactions with FM professionals/users.

Based on existing multi-tier architectures (e.g., IoT and Big Data architectures in table 3), this proposed DT system architecture is specifically designed for AEC/FM sectors. Four benefits can be summarised as implementing DTs in buildings and cities using this specific DT architecture:

i. Based on the research of existing multi-tier architectures and Gemini Principles, this architecture is designed in a five-layer format (see Figure 2) for various hierarchical levels from systems (e.g., pump), bridges, buildings to cities (see Figure 1), which keep unified and share federation (linking) among different levels.
ii. Integrating heterogeneous assets and data sources via linking with the digital models, this architecture supports for integrating 3D geometric and geo-referenced entities with other data resources in a distributed manner. For example, the IFC is used to integrate building digital models with daily management system, geo-coded sensor data etc.
iii. With the basis of cloud computing and IoT-based services, it enables compatibility with many protocols and environments with abilities to manage real-time sensors and distribute data in numerous formats.
iv. Interaction and communication channels with human users are added to bridge the gap

between human relationships with buildings/cities.

West Cambridge Digital Twin Demonstrator

Overview

The pilot evaluation study of the proposed DT was conducted at west Cambridge site of the University of Cambridge. The West Cambridge site includes more than 20 university buildings, sports centres, residence areas, main roads, parking places and restaurants. This can be therefore be considered as a small example of a city and a promising testbed. For the building level, this study used the Institute for Manufacturing (IfM) building, which is a 3-storey building at the West Cambridge site. This building includes teaching, study, office, research and laboratory spaces and stands over a 40000-square-foot comprehensive area. Five critical stakeholders were engaged in the development of the pilot (see Figure 3).

i. University Estate Management Team – responsible for the O&M requirements for the entire university.
ii. University Facility Management Team – responsible for the day-to-day O&M activities for a specific building or location within campus. In this DT project, we brought together two facility management teams, namely the team that manages the West Cambridge site and the facility management and technical support team of the IfM building.
iii. Modelling and data collection company – supports the data collection and model development of the DT, including UAV point cloud scanning, and localised laser scanning and photogrammetry.
iv. Consulting company – provides project management and collaborative expert supports. The core requirements of the consulting company are to provide the organisational progress of cost management, time schedule management, and resource management.
v. Academic team – provided overall leadership to the project and is responsible for the design and implementation of the architecture. Further, the academic team also ensured that the DT development architecture and methodology were correctly implemented and was repeatable and extensible.

Based on the developed system architecture, the DT in west Cambridge site integrated various data resources and included several applications. The objective of this pilot is to demonstrate how a dynamic digital twin of existing buildings and infrastructure can be developed and to explore the opportunities and challenges.

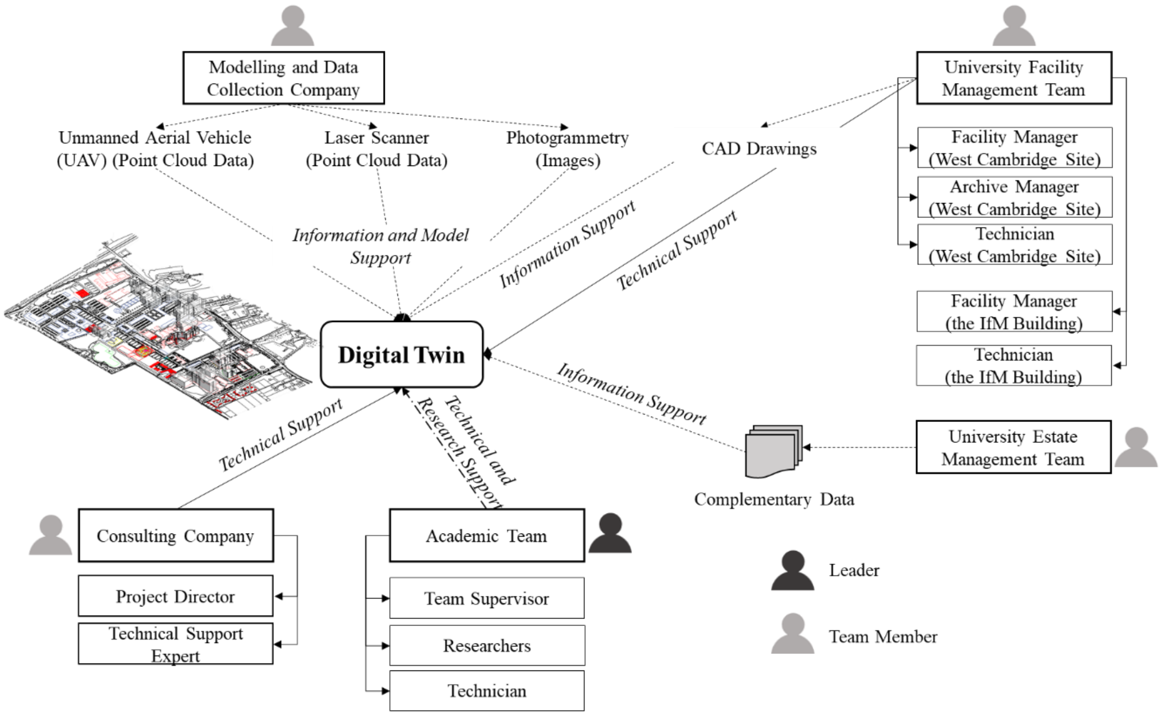


Figure 3. Stakeholders in the West Cambridge Digital Twin pilot

Data acquisition layer

In data acquisition layers, data from environments and physical assets is a fundamental requirement of the proposed system architecture of the DT. This presents several challenges for developing a data acquisition and transmission system: it needs to support data uploads from the sensors that are deployed at distributed locations since the assets are dispersed, and it also needs to be scalable to support a large number of assets and data resources in building and city levels. The West Cambridge DT is integrated with the data acquired from the Building Management System (BMS), Asset Management System (AMS) used in Cambridge, space management system (SMS), which are MySQL-based, as well as real-time sensors. The BMS is installed in each building and it controls the mechanical and electrical systems (e.g. power systems, heat ventilation and air conditioning (HVAC) systems and security systems). The AMS is a work-order management system that keeps records of all asset management activities and service carried out on the university assets. In this study, Planet is used for managing assets, such as asset register, preventative maintenance plan and storeroom stock (Planet 2019). The SMS manages room bookings and therefore provides space utilisation information. MiCAD space management system is used in this study. It is a cloud-based publishing system that holds

CAD floor plans, building condition records and room bookings for each building in the West Cambridge site (MiCAD 2019).

Transmission layer

In this work, challenges of ‘real-time’ data collection are overcome by developing IoT-enabled wireless sensor network (WSN) and QR code-based asset management network in the data transmission layers. Figure 4 provides an illustration of the data acquisition and transmission system developed for the pilot. WSN refers to a collection of distributed and dedicated wireless sensors for monitoring and recording conditions of environments and equipment (Lewis, 2004). The sensors in WSNs are called nodes and they measure the environmental conditions such as indoor temperature and relative air humidity, and HVAC equipment conditions such as component vibration, surface temperature and speed of the rotating parts. In addition to the sensor nodes, the WSN consists of gateway nodes that act as the bridge between the local sensors and the remote applications such as cloud-hosted databases and online web pages that visualise data. In recent years, WSNs gained attention due to the emergence of IoT and proliferation in Micro-Electro-Mechanical Systems (MEMS) technologies (Yick et al., 2008). These technologies allowed WSNs to be smarter by utilising computing capabilities yet cheaper and smaller (Yick et al., 2008). In this section, a discussion on the IoT-enabled WSN developed for the proposed system architecture of the DT is provided (Figure 4). Firstly, the IoT devices used as the nodes in the WSN are introduced and secondly a discussion on the overall WSN is provided.

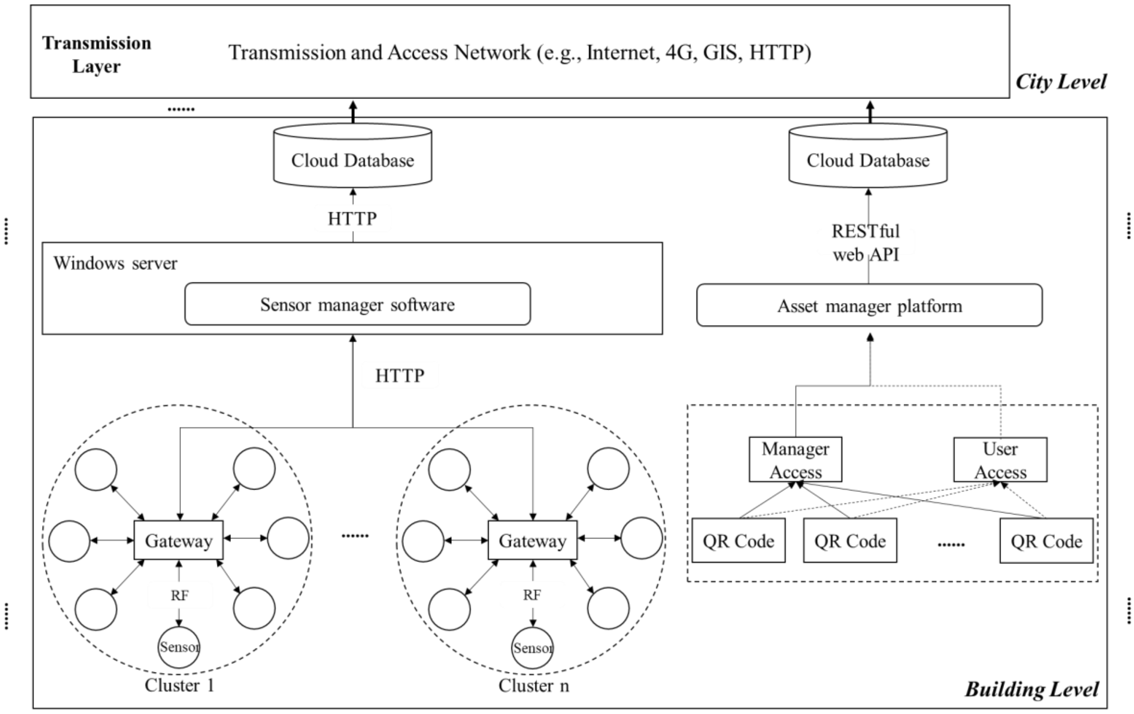


Figure 4. Schematic of the WSN for data acquisition and transmission from the assets

The IoT sensors used in this pilot are the Monnit wireless sensors (Monnit 2018a) with a one-minute heartbeat. The sensors and gateways communicate over the 868 MHz radio frequency (RF). The RF antenna in the sensors acts as the transmitter and the receiver, and it sends the measured data to the gateways. These sensors are capable of 250 – 300 feet non-line-of-sight (partially obstructed path for radio transmission) RF range (Monnit 2018b). The wireless communication capability of these sensors over RF is suitable for the distributed nature of the DT system architecture as RF is a low-cost communication medium (Lanzisera et al., 2011), and it supports the required range to connect the distributed set of sensors with the gateways. In this pilot, a wide range of sensors such as temperature, humidity and motion detection were used for capturing data from various locations and equipment in the IfM building. Monnit Ethernet gateways (Monnit 2018c) are used as the gateway nodes in the WSN. These devices are AC powered and consist of RF antennas that allow the communication with sensors. Moreover, gateway devices consist of ethernet ports which allow them to communicate with the remote applications over the internet and also provide scalability for a large number of assets in building and city levels.

The nodes in the WSN are grouped into different clusters depending on the distance between sensors and gateways. This allows robust connection between sensor nodes and gateway nodes as sensors can connect with the closest gateways which increase the RF signal strength between the two devices. During the initialisation phase of the WSN, the gateways are pointed to a virtual server (i.e., connected with the IP addressed of a virtual server) created by a Sensor Manager software. Sensor Manager is a custom-developed .NET software which is hosted on a Windows server and integrated with the Monnit Mine API, which is an interface that allows custom-developed applications to retrieve data from the Monnit gateways. Once the gateways are pointed to the server hosted by the Sensor Manager, the sensors are registered with the gateways by sending a command to the gateways over the internet using the HyperText Transfer Protocol (HTTP). This command contains the unique device identifiers (UDIDs) of the sensors a gateway needs to be connected with. After the initialisation phase, the sensor nodes are capable of monitoring environmental and equipment conditions, and uploading data over RF to the gateway nodes. Upon receiving the data, gateway nodes upload data into the Sensor Manager over the internet.

In addition, more than 200 assets within the IfM building and the site were tagged with QR codes to provide an individual profile that provides good quality information. QR codes were attached to the surfaces of different assets (e.g., refrigerators, street lights). A user-friendly mobile-phone app developed by Redbite Solutions (itemit, 2019) enables maintenance personnel to update information about maintenance and inspection based on their responsibilities and roles. Similar to the WSN, information collected through scanning QR codes can be sent to the Asset Manager platform via a RESTful web API.

Finally, the sensor manager and asset manager send the data and collected information to the DynamoDB NoSQL database supported by the Amazon Web Services (AWS). In two networks developed for the DTs, the whole process of sensing condition data to storing data in the cloud database occurs every minute to facilitate timeliness of the DT and QR code-based information collection creates communication channels between people and DTs.

Besides these two networks, challenges of various data resources integration were solved through well-designed transmission process. For instance, a BMS controller that collects data from the mechanical and electrical systems is integrated with hard-wired sensors. A Trend SIP interface (Synapsys, 2018) was deployed to allow the data captured by the BMS in 15-minute intervals to be uploaded as CSV files to an SMTP server every 1 hour (see Figure 5).

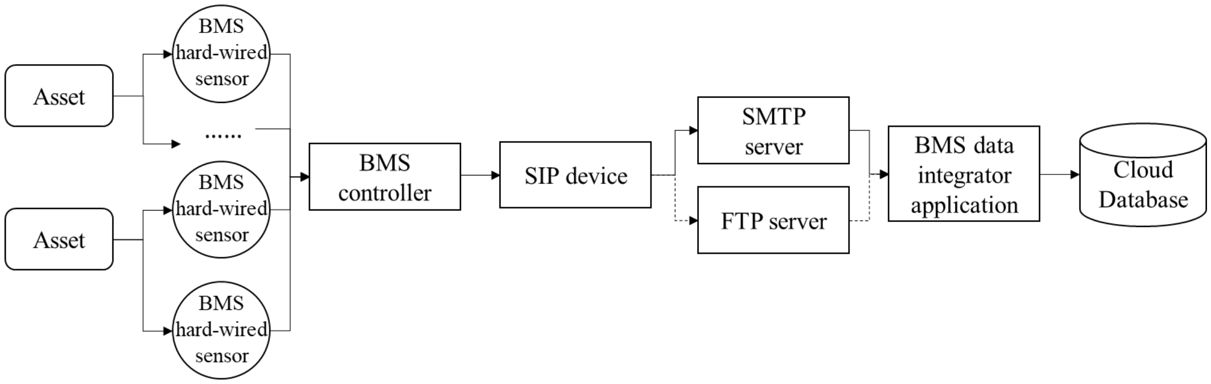


Figure 5. Diagram of the hard-wired sensors data transmission process

A BMS Data Integrator software was developed for reading the data stored in the CSV files in the SMTP server and uploading them into the AWS DynamoDB database.

Digital modelling layer

Information requirements are various at different scales. In this layer, a three-sublayer digital model was built based on different information levels (see Figure 6). This includes a geometry model of the West Cambridge site in a city level, the BIM model of IfM building with a medium level of detail (including architecture, structure and mechanical, electrical and pumping (MEP) components), and a BIM model of specific areas in IfM with highly detailed information (e.g., facilities and pipes in the plant room) in a building level. This layer aimed to establish a visualised model-based platform to support upper layers. The site-level photogrammetry data was captured using fixed-wings drones and vehicle-based scanning devices. The highly detailed 3D geometry scans of the interiors of the building were captured using laser scanners and digital cameras. The process and plan of generating digital model in a city level are presented in Figure 7. In addition, complementary data was further collected in this layer.

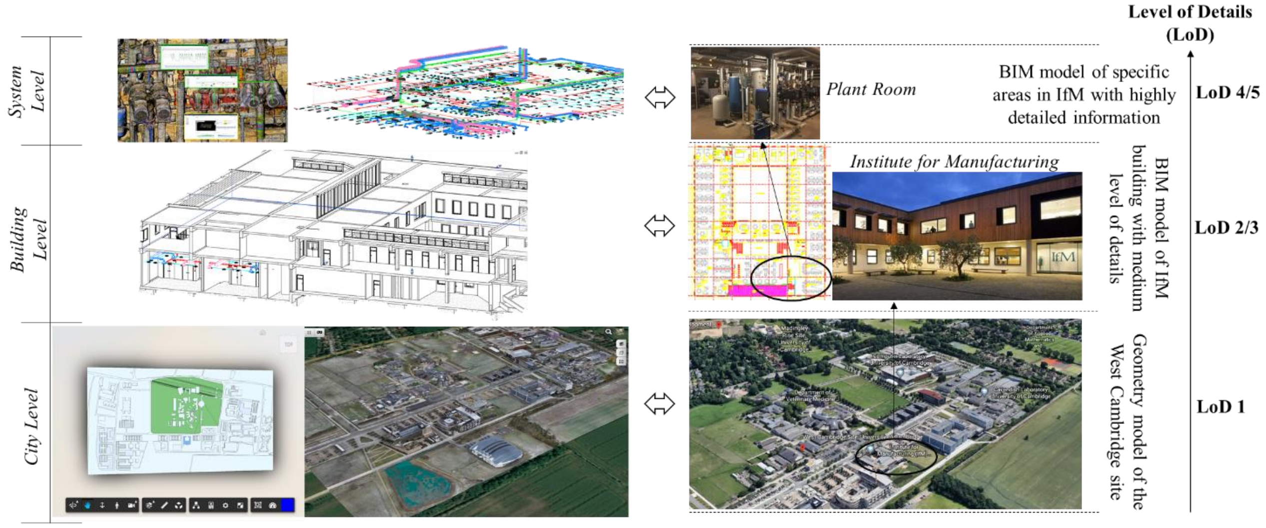


Figure 6. The digital modelling layer development of the city DT at the West Cambridge Site

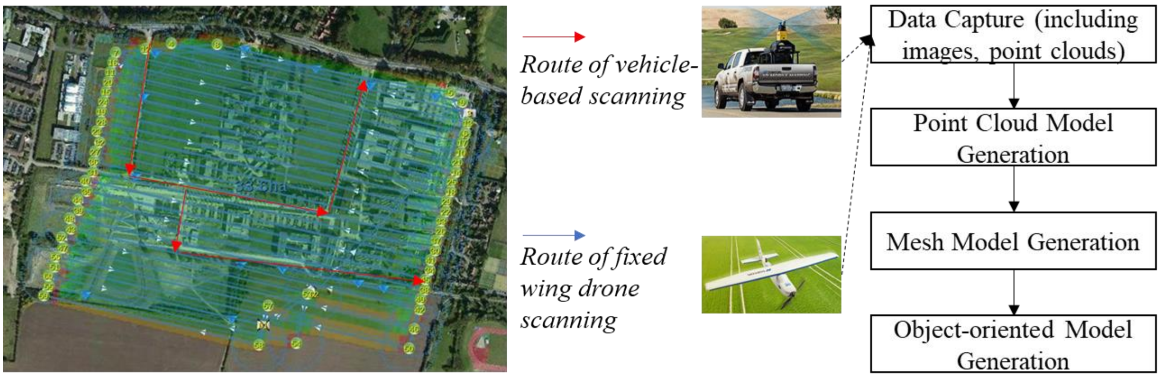


Figure 7. Digital model generation process and plan for west Cambridge site using fixed-wing drone and vehicle-based scanning

Data/Model integration layer

In addition to the data/model integration layer, the two developed DT instances have incorporated the proposed DT system architecture with the capabilities to store and analyse BIM object related data collected by heterogeneous data systems. These data include asset condition monitoring data (e.g., building's plant room assets, including boilers, heat circulating pumps, thermal extractors and energy readings from the HVAC system), asset historical records, environment monitoring data, utilisation monitoring data and energy consumption data (see Figure 8). In this project, two DT instances have been developed as shown in Fig.8: 1) the research-based instance was developed by our team for research purposes and 2) the commercial instance was developed through cooperating with Bentley Systems, Inc. for providing a mature product option in the future market. Research questions, objectives and whole processes related to new functions and services would be completed and evaluated in the research DT instance firstly. Then, when handing over these research results to Bentley system, similar functions/services in the 'commercial' DT instance would be added with stronger software robustness.

For the DT research platform, Autodesk Revit was used to develop the RVT model and then export to Industry Foundation Classes (IFC) files. This platform was developed based on AWS DynamoDB, Autodesk forge API and web-based program design (i.e., .Net) using C# and Java script. For the commercial one, since cooperating with Bentley Systems, Inc., AECOsim building designer was used to develop the DGN model and then export to IFC files. Bentley Systems, Inc. developed this platform based on their available commercial off-the-shelf application (i.e., Assetwise).

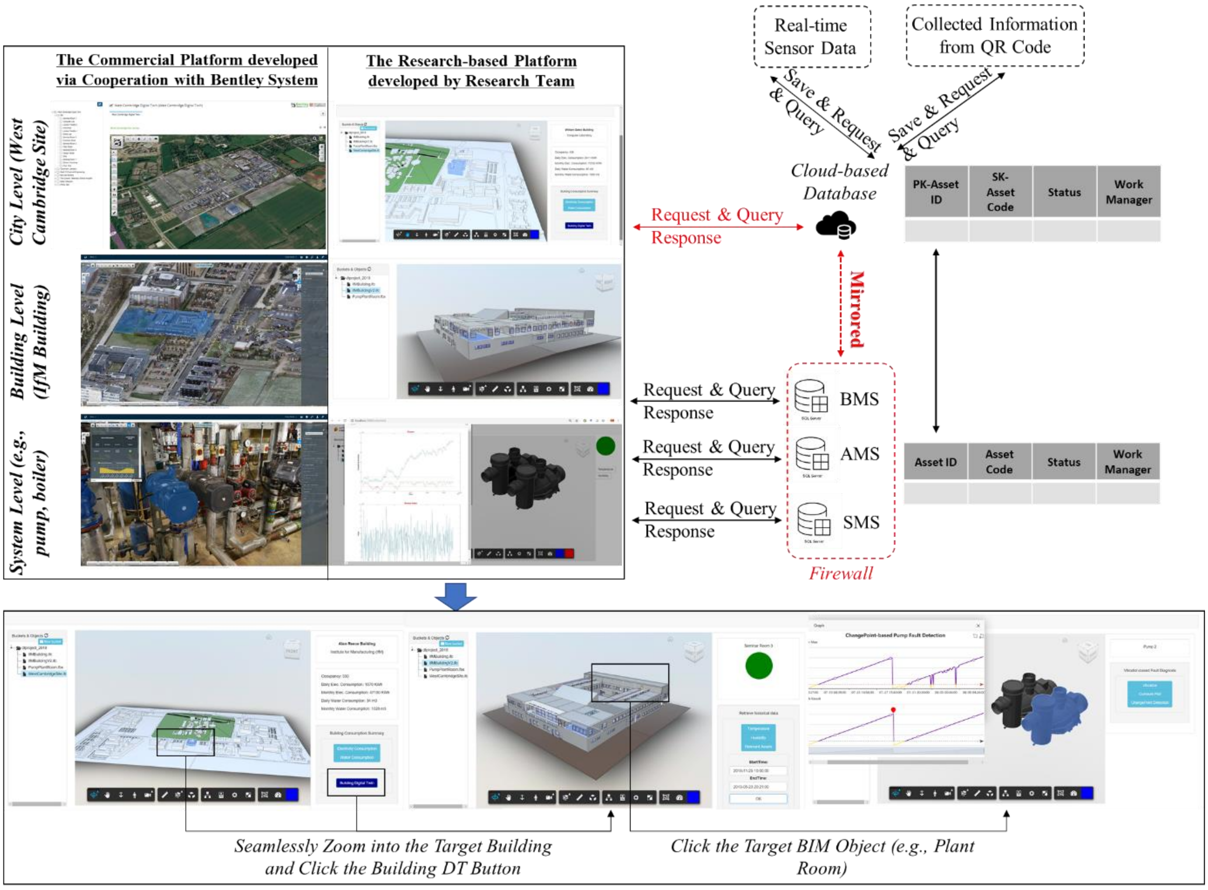


Figure 8. Diagram of the data/model integration layer and service layer

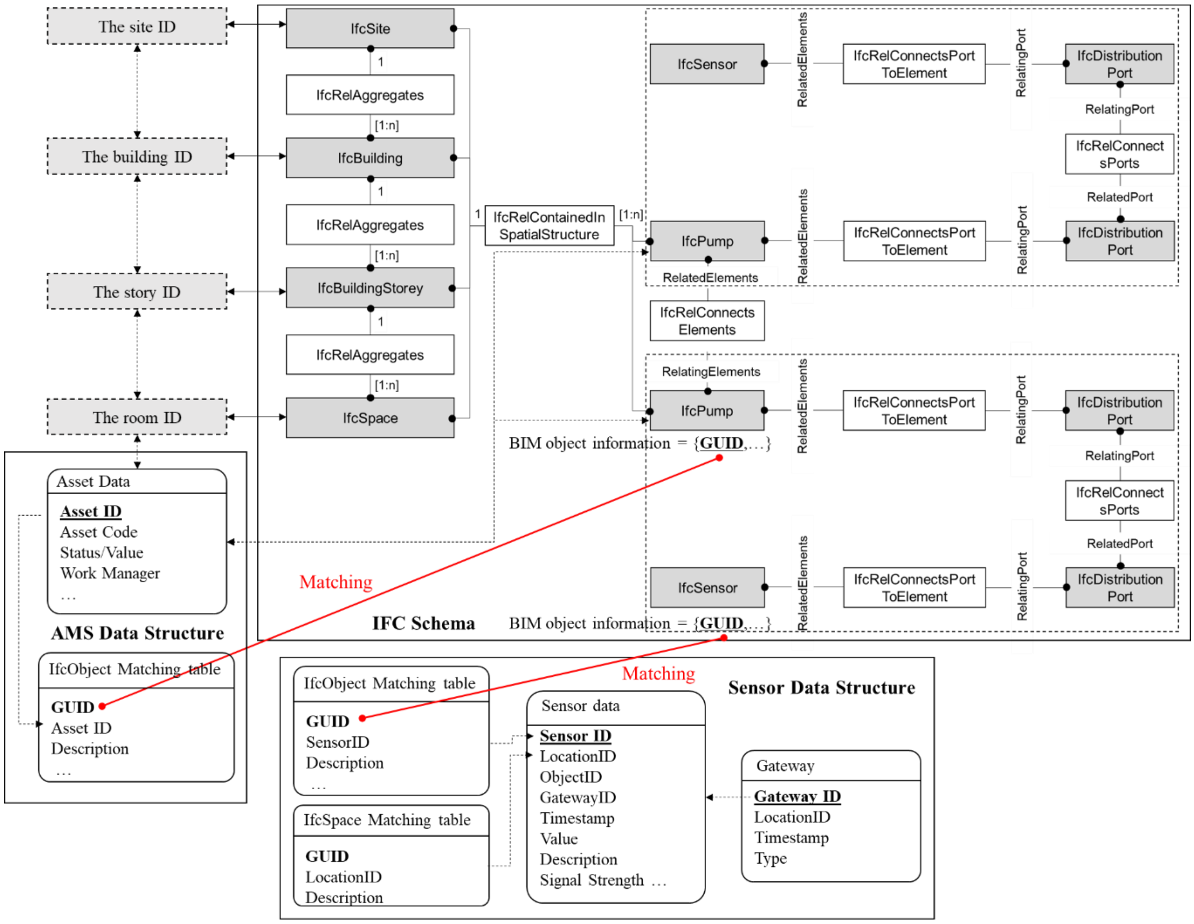


Figure 9. The IFC schema mapping with other data resources (using AMS as an example)

Due to the existence of University's security firewalls, the AMS, BMS, SMS and other datasets are not ubiquitously accessible beyond the scope of University local area network (LAN). To enhance the accessibility of these external data, a mirrored database is used, which basically stores all datasets stored in the protected AMS, BMS, SMS into DynamoDB NoSQL schema. Different from relational database management used in the AMS, BMS and SMS, a DynamoDB based non-relational database is adopted that is highly available, scalable, and optimized for high performance. Near-zero downtime migration could be realized using the AWS Database Migration Service (AWS DMS) (Balobaid and Debnath 2018), importing data from MySQL towards DynamoDB. After migration, the datasets stored in DynamoDB act as the primary data source for external asset-related information in this case. While real-time sensor data and QR code feedback information are stored and managed directly through DynamoDB. If there is no limited access (e.g., no firewall), it is suggested to query data from various databases based on the application requirements.

The AMS data is used as an example to explain the detailed data structure of data/model integration. To enable the IFC-based interoperability (Steel et al., 2012) between BIM and

AMS (which refers to the AMS data stored in DynamoDB), the data/model integration layer is designed to be capable of interchanging and interoperating external data related to each BIM object in the digital model on a semantic level (see Figure 9). IFC is a widely used standard data schema for BIM and is an object-oriented and semantic representation that includes components, attributes, properties, relationships, and linkages with other libraries or data resources (Romberg et al., 2004). Specifically, in this DT development, an IfcObject/IfcSpace matching table for AMS data integration is stored in DynamoDB, describing the relationship between the BIM object GUID and its corresponding asset ID from data resources (e.g., AMS). Shown as Figure 9, when asset data (saved in AMS) needs to be integrated or queried for some services in the upper layer, the IfcObject matching table provides a linking bridge between the targeted BIM object (GUID) and the corresponding asset ID in AMS. Through this matching approach, the matched asset ID is used as a primary key (PK) in the designed data schema (Figure 9) for searching the required data. Through the GUID in the IfcObject matching table and querying matched asset ID number, the required data would be searched automatically by their unique asset ID as primary key and further refined using sort key (SK). In this way, data resources could be kept in their original storage locations and saved in this distributed manner. This data integration method enables that IFC and other data sources (e.g., AMS) are independent from each other. To keep the consistency of the data, only the IfcObject/IfcSpace matching table needs to be maintained, which achieves CRUD (Create, Retrieve, Update, Delete). For instance, when a new BIM object is added to the IFC, a new linking pair would be added to the matching table via linking the GUID of the new object to the unique asset ID from the AMS database; when the asset ID number is changed, which would happen when assets are replaced, the asset ID that corresponds to the replaced object GUID in IFC should be updated, without modifying the IFC or the original database. Furthermore, the requested data would be visualised in the DT platform linked with the corresponding BIM objects (see Figure 8). Exchanging information across data source boundaries makes interoperability a primary issue, but IFC well solves this problem. Data processing and advanced functions (e.g., AI) are also designed in this layer, driving the KEs to understand the mechanisms behind assets, systems, buildings and cities. The supported services would be discussed in detail with their applications in the next layer.

Service layer

The DT pilot currently includes five services in building and city levels. Among five services, anomaly detection in pumps is described in detail, including the data resources used, functions implemented in the data/model integration layer, the proposed DT architecture demonstration and advantages of DT-supported decision-making processes. Other four services are expressed briefly as follows and will be extended in future publications.

- **Anomaly detection in pumps**. Given a set of vibration data that carries diagnostic information on the mechanical condition of pumps, this service is implemented in the proposed DT architecture, aiming at detecting change points in vibration data which indicates the occurrence of suspicious faults on pumps in the HVAC system (see Figure 11 (c)). Generally, BMS and real-time sensors keep track of the operating conditions, especially for principle assets, BIM provides additional information (e.g., geometry, location). Empowered by IFC schema implemented in demonstration, the data/model integration layer enables the intelligent extraction of pump relevant data. A typical change point detection method, cumulative sum charts (CUSUM), is adopted to analyse the extracted pump data and find those change points in an unsupervised manner where the underlying symptom parameters of vibration deviate from their normal values. A real case study is conducted to demonstrate the role that a building DT plays in the pump anomaly detection service. In the case, two identical pumps are installed in the plant room of IfM building. They work in parallel to pump return chilled water from the air handling units and fan coil units back to the chiller. For the convenience, the vibration frequency measured by sensors mounted on the pump casing (close to the bearing) is extracted using the established DT, as an indirect way of assessing the conditions of two pumps. Two scenarios are analysed, a scheduled operating condition change, and a pump failure event causing strong abnormal noises respectively. In the first scenario, the studied centrifugal pump undergoes a scheduled shutdown due to the Christmas holiday. The period of data starts from the 5th December of 2018 and lasts until 1st January of 2019 (4 weeks). Figure 10 (a) and (b) shows the recorded vibration frequency time series and CUSUM result within this given period. The shutdown can be seen to the naked eyes, and a rough judgement can be made that the studied pump stops working in the afternoon of 31st December of 2018. In the second scenario, one of the two pumps undergoes a highly suspicious anomaly causing a strong abnormal level of noise. Figure 10 (c) and (d) show the generated vibration frequency time series and CUSUM result within given period. It is relatively hard to distinguish the difference between the vibration of normal and faulty pumps by unaided eyes. But at least, the CUSUM based detector could locate the change point corresponding to the shutdown and anomaly scenario with a reasonable time delay. In alliance with the BMS, the found change points are matched against the recorded normal operation changes, so that change points caused by real faults can be uniquely identified. Comprehensively synthesizing the information from change point detection and cross over matching, the live knowledge engine (KE) for pump, realized in the data/model integration layer, can be established for modelling and updating the up-to-date status of pump. In summary, benefit from the DT, a centralised system that integrates heterogeneous available data sources is established, enabling the data interchange and interoperation. Supported by the strong data integration capability of DT, better-informed decisions can be made, including continuous condition monitoring and anomaly detection of pumps (Kaur et al., 2020; Costa et al., 2013).

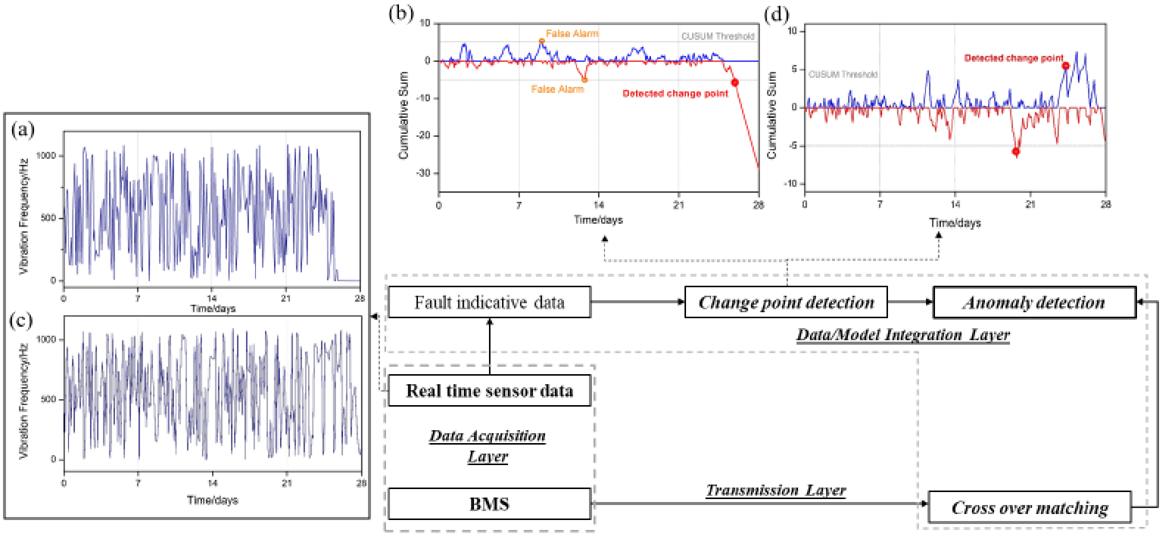
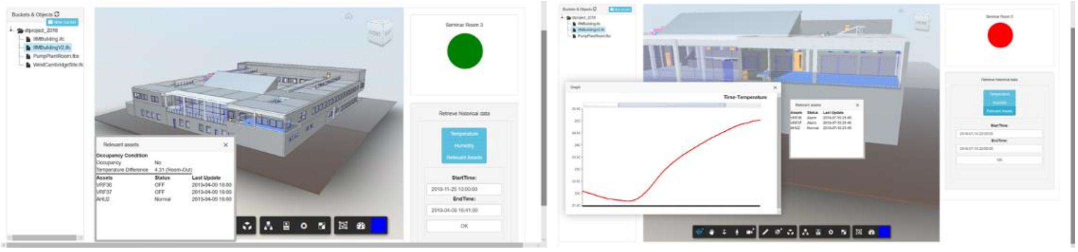


Figure 10. Pump anomaly detection implemented in the service layer of DTs

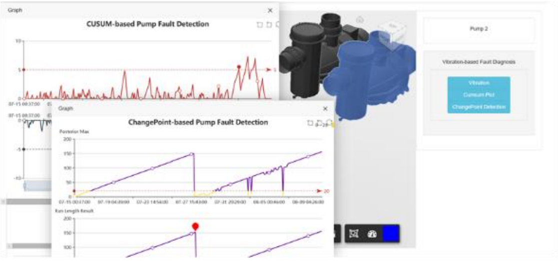
- **Ambient environment monitoring**. Ambient temperature and humidity monitoring are used to evaluate the comfort level of the working space. If the ambient condition is outside a pre-determined threshold of comfort, the DT platform will indicate this through a status indicator (coloured red for too hot, blue for too cold and green for comfortable status). The facilities manager can further check the real-time and historical temperature and humidity data to analyse faults in the system. This system will be further developed to enable the facilities manager to carry out effective root-cause analysis (Figure 11 (a)) and guiding the facilities manager to take appropriate corrective action if necessary.
- **Maintenance optimisation**. This application predicts unexpected temperature drops caused by the biomass boiler's malfunction by applying machine learning algorithms using the data collected from the building management systems and failure/maintenance logs.

Further, for assets that are not suitable for predictive maintenance, the application also includes a maintenance planning optimiser that develops the optimal maintenance/replacement interval based on the historical failure rates calculated using data from the maintenance/failure logs (see Figure 11 (b)).

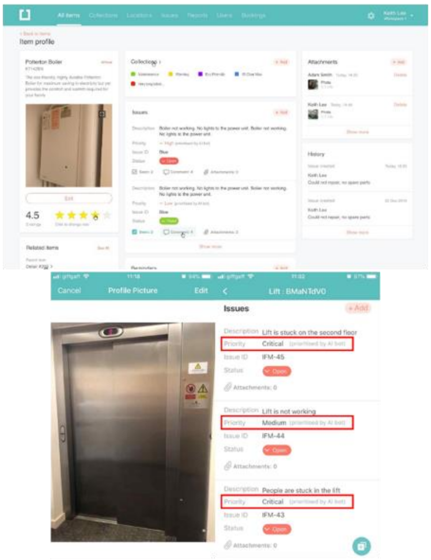
- **Maintenance/Repair prioritisation.** Maintenance task prioritization is essential for allocating resources. It is estimated that almost 1/3 of the maintenance cost is spent insufficiently (Mobley, 2002). Based on the developed DT, this application exploits the advances in mobile communications, social networking, and machine learning to address these shortcomings on a city scale. It also brings assets online using asset tags with an online 'asset digital profile'. Users of assets are able see the 'digital profiles' and enter 'comments' describing issues and problems by scanning these tags using a mobile phone app (Itemit, 2019). This feedback is the input of a machine learning based method (defined in the data/model integration layer) that infers the criticality of every asset defect reported. A prioritization label that indicates the response time is finally returned for each maintenance task in the West Cambridge site (see Figure 11 (d)).
- **Environmentally friendly urban energy planning**: Urban energy planning has moved beyond providing the necessities and societal needs to a stage of establishing an integrated methodology to solve environment and energy problems at the urban level in achieving low carbon intensity. In this application, the building DT exhibits a tight integration of sensing and computation capability, which estimates the characteristics of building energy demand patterns using sequence-to-sequence LSTM. Taking advantage of this information, quantitative energy demand figures and the spatial distribution of the forecasted energy can be acquired to decide the future need of the capacity of the energy supply facility and the energy production at the urban planning perspectives. In this way, a better energy demand pattern for urban space can be achieved by integrating the optimal amount of clean energy resources (see Figure 11 (e)).



(a) Ambient environment monitoring (b) Maintenance optimisation



(c) Anomaly detection in pumps



Issues

Description Lift is stuck on the second floor
Priority Critical (prioritised by AI bot)
Issue ID IFM-45
Status Open
Attachments: 0

Description Lift is not working
Priority Medium (prioritised by AI bot)
Issue ID IFM-44
Status Open
Attachments: 0

Description People are stuck in the lift
Priority Critical (prioritised by AI bot)
Issue ID IFM-43
Status Open
Attachments: 0

(e) Environmentally friendly urban energy planning (d) Maintenance/Repair prioritisation

Figure 11. Digital twin services

Analysis of the DTs Development from the Perspective of Data Management

Since DT is built on data, the pilot so far has revealed four key data management challenges that should be addressed in order to develop an effective DT at city and building levels.

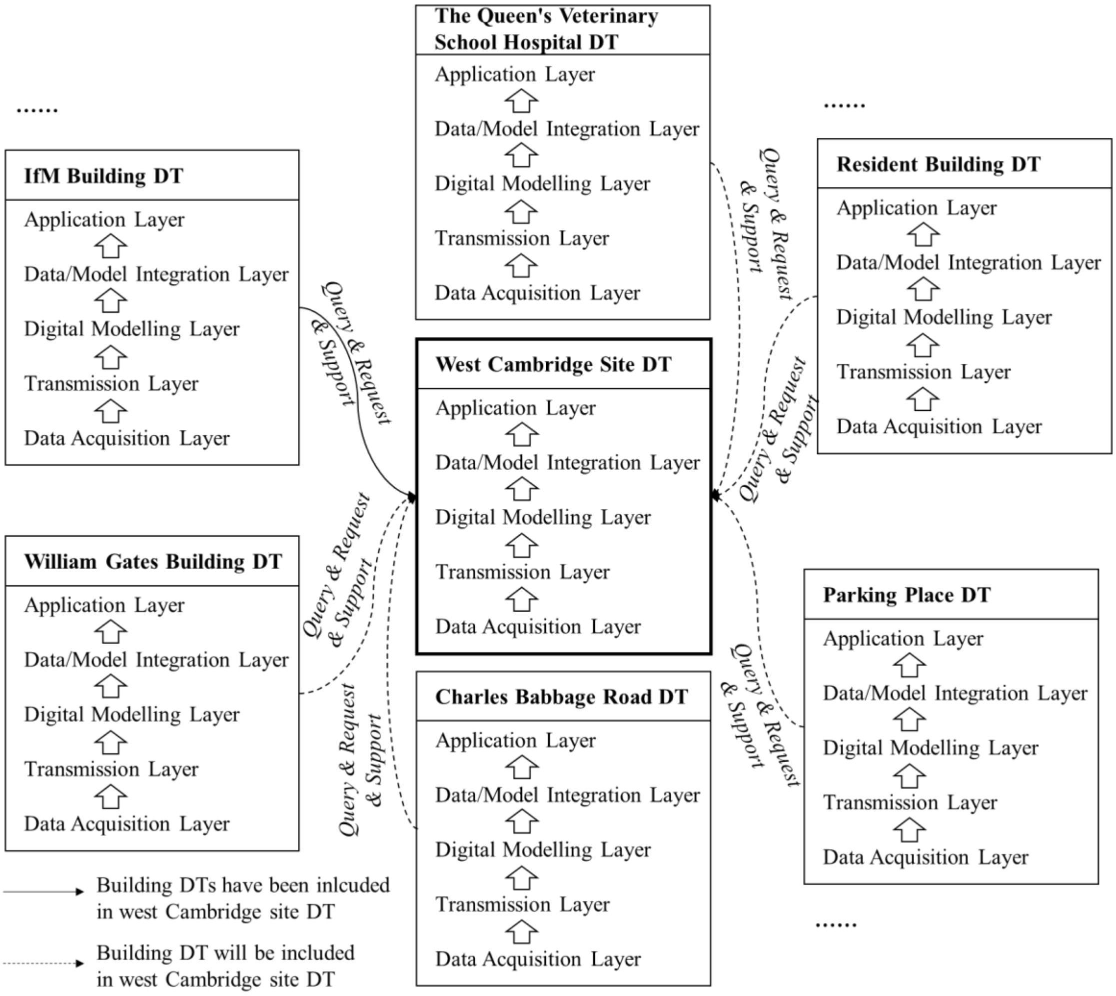


Figure 12. The data management among city DT and sub-DTs

- **Data integration**. To realise a DT poses various data management challenges; especially related to the integration of data from autonomous, disparate and heterogeneous sources. This is exemplified in this DT which integrates data from sources such as real-time sensors, BMS, cloud services, and AMS etc. From a technical point of view, there are many technologies available to support the integration of data, from ETL technologies that support the transfer of data between systems (Vassiliadis, 2009; Woodall et al. 2016), to Service-oriented Architectures that can expose data as a service (Budgen et al. 2007), to data virtualisation, to data warehouses and data lakes (Beyer et al. 2017). Generally, no one solution fits all problems and a mixture of these technologies is often deployed in organisational integration settings (Araújo et al. 2017). It is the foremost challenge of integrating different data resources and further linking various assets for DT development. Particularly, big data is an important part of a DT, which is characterized by high volume, high velocity, and high variety. Without big data, most of functions of digital twin would be the castle in the air. Semantic ETL workflow (Bansal and Kagemann 2015), as one of the potential solutions for DT data integration, could be investigated for integrating massive data from heterogeneous sources into a meaningful data model, which allows intelligent data querying and further creation of innovative applications. The semantic technologies are introduced in the transform phase of a traditional ETL process to find a semantic data model and then generate semantically linked data in the form of Resource Description Framework (RDF) triples to be stored in a data warehouse. While extract and load phases of the ETL process would remain the same as the traditional workflow.

- **Heterogeneity of source data systems**. The source systems containing the vital data needed as input for monitoring and prediction algorithms often reside in disparate systems running different software platforms and database systems. Efficient execution of queries to extract the data from these systems is non-trivial. For instance, a NoSQL engine used in DynamoDB is suitable not only for large-scale data storage and but also for massively-parallel data query across a large number of concurrent requests. This is especially important in DTs as there will often be a need for timely and up-to-date data. In extreme cases, a real-time stream of data would be needed, such as telemetry data. Moreover, city DTs need to query data from sub-DTs (see Figure 12).

The structure of the data models throughout systems often differs because there are many ways in which database designers can choose to store the same type of data. This manifests as differences in the choice of database tables, records, and attributes for data. One common problem is that without a globally unique identifier/standard for data records among data sources. It is difficult to know whether a data record in one system (e.g., a particular machine) is the same machine as referred to in another data record in another system. Various terms may be used in different systems, including entity linking, record linkage, entity resolution, data matching, and data de-duplication (Talburt 2011).

Also, how to reconcile the differences in the semantics and syntax of data is another challenge. For instance, the definition of a boiler in one data source may include the external pipework and in another system it may not cover. The area of Master Data Management (MDM) is a topic that deals with these issues to advise how to reach a consensus on the definitions of data and manage its changes and evolution over time (Loshin 2009; Otto 2012; Otto et al. 2012). Hence, there is also the need to reconcile the differences in specific values in databases to ensure that the nomenclature is consistent. For instance, one system may use degrees Celsius while another could use Fahrenheit.

- **Data synchronisation**. The demonstrator shown in the last section doesn't explicitly involve data synchronisation. It was basically done manually offline in the data/model integration layer. But a key problem in a practical DT is timing and frequency of synchronising different copies of data in order to provide up-to-date data to decision-makers. The problem is non-trivial because a trade-off exists between synchronisation costs and quality (staleness) of the data (Qu and Jiang, 2018). Synchronisation costs include the cost of resources used, such as Information Technology (IT) staff and computing resources etc. Computing resources can cause a considerable disruption cost to the business, because systems often need to be "locked" (Woodall et al. 2016) in order to access the data and any reduction in computer power can reduce the power available for critical business operations (Qu and Jiang, 2018). Organisations often resort to batch synchronisation of data which is attempted out of business hours (such as overnight). However, for DTs with a requirement to monitor engineering assets in real time, a continuous stream of data will be needed, which shifts the trade-off towards high synchronisation costs. For instance, if semantic ETL workflow is adopted, a mechanism must be integrated to make sure that the datasets are relatively consistent. Because heterogeneous data sources may have different timestamps, ETL workflow is required to be capable of holding back certain datasets until they are synchronized.
- **Data quality**. Data quality is defined as fitness for use (Wang and Strong, 1996), which captures the dual concepts of how the data is to be used and whether it meets the requirements of that use. The use of the data in the DTs must support various applications at once, such as enabling service decisions and predictions. However, in a DT, data may degrade causing it to be not fit for use for various reasons, including:

  a). the quality of the data extraction process from the data sources or sub-DTs (Figure 12);
  b). the inherent quality of the data in the underlying data sources;
  c). quality loss due to abstraction required by the integration of data;
  d). differences in the quality requirements from different data sources (repurposing).

  In this process, data quality can be lost when extracting data from source systems, for example, the query to extract the data is incorrectly formulated and gathers the incorrect records. Data can also be lost in this process if the transformations on the data (when performed in semantic) transform it incorrectly. Even if the data extraction process is perfect, if the data from the source systems contains errors, then these will propagate to the DT. There are, however, certain types of these errors that can be detected and corrected in the transformation process, such as incorrectly formatted data and invalid data (Woodall et al. 2014). DTs may utilise publicly available online data in the city level, such as using weather forecasts etc. However, the quality of online data can be questionable, and the use of this type of data could demand a different notion of data quality compared to traditional database systems (Lukyanenko et al. 2014).

In order to achieve data encapsulation and beneficial separation of concerns, each DT (from system-level to city-level) should be responsible for maintaining the quality of own data within its DT, and not offload it to another DT. Note that the vision for the high-level DT (West Cambridge site DT in this case) is not limited to a huge singular DT of the entire environment. Rather, suggested by Gemini Principles, it is envisaged as an ecosystem of sub-DTs joined together via securely shared data. Therefore, the high-level DT allows interdependencies across different sectors to be understood in a way that sub-DTs could hardly satisfy.

Summary and Discussion of the DTs Development

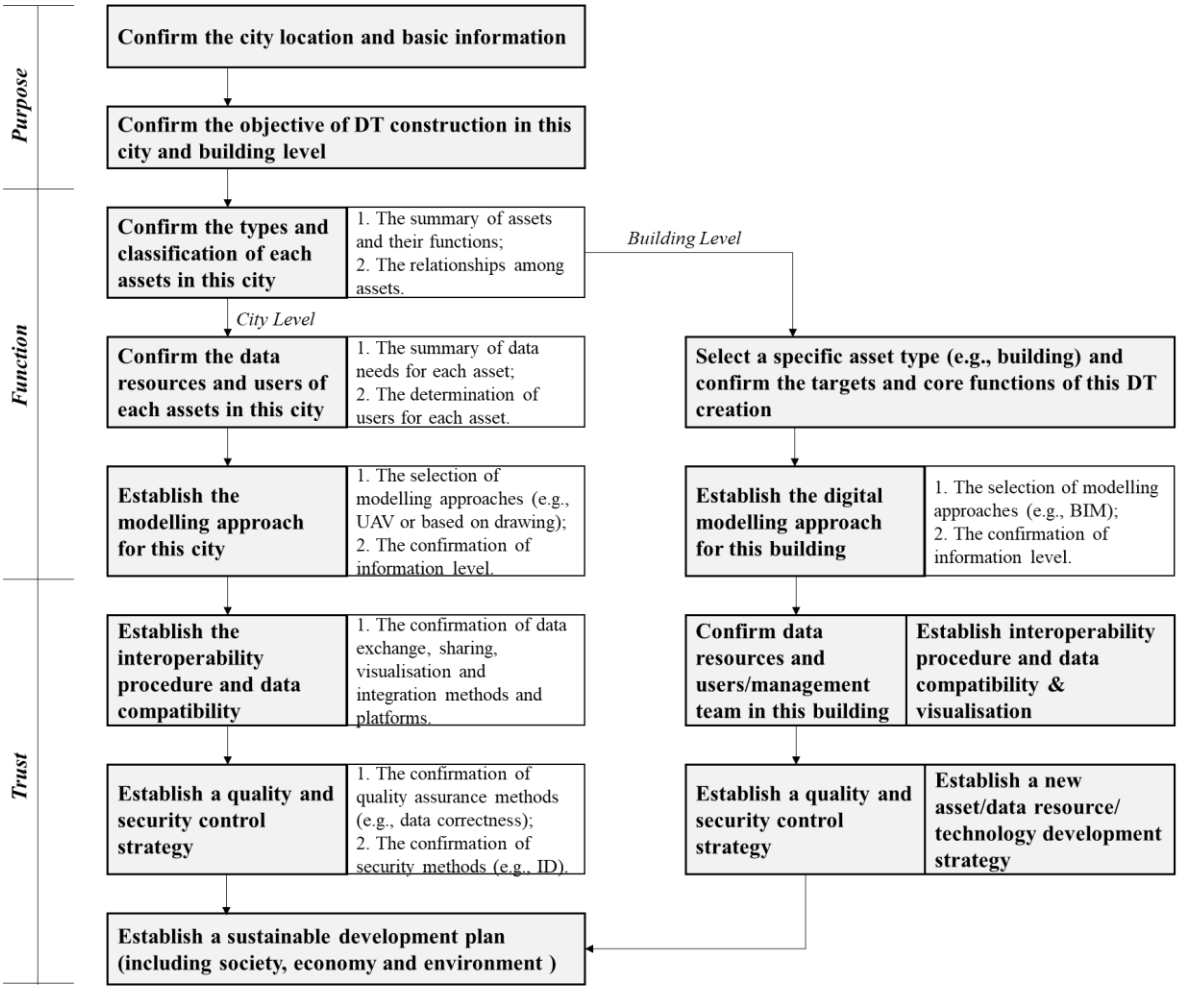


Figure 13 A road map for DT development in building and city levels

In-depth analysis was performed based on data management requirements (i.e., data integration, sources heterogeneity, data synchronisation and data quality) to highlight the key challenges of developing a DT in building and city levels. Thus, according to the analysis results from the perspective of data management and definitions of DTs (i.e., purposeful, trustworthy and functional) (Bolton et al., 2018) provided in the literature review, successful development of DT in building and city levels can be achieved with 1) a clear objective of DT construction (insight), 2) a clear definition of DT constitutes (value creation), 3) a well-designed and practical process of collecting, updating, transferring and integrating data/model throughout the life cycle (federation), 4) a well-executed and standardised interoperability procedure and data compatibility plan for curation and further possible evolution, which mean that the developed DT is able to adapt, develop and extend as technology advances (curation and evolution), and 5) a valid control strategy development, which guarantees the performances of DTs (security, openness and quality). This research examined a real-world dynamic DT development using the west Cambridge site, to determine the data required for DTs, to articulate the process of collecting, managing, and integrating various data resources, to test the seamless linkages among five layers and assets in different scales, to provide practical applications and functions, and to summarise challenges faced and lessons learned.

The major lessons learned on this DT developed based on the system architecture include the following: 1) organising a well-integrated project network and setting clear responsibilities, including representatives from the modelling, data collection, consulting, research as well as facility management team; 2) setting a clear objective, applications and functions development plan in advance; 3) confirming and classifying data resources according to different users aligning with their requirements; 4) choosing and creating central digital models (e.g., BIM), data schema (e.g., IFC) and authoring tools; 5) creating logical and reliable transmission networks, which allow efficient data transferring and communication between physical world and digital world; 6) designing intelligent and effective data processing and analysis functions according to predefined objective and applications; 7) conducting continuous data quality control and synchronisation assurance throughout the asset life cycle; 8) preparing a reasonable schedule and workflow process when developing DTs, since unexpected issues should be considered in real project.

Moreover, in order to visualise DTs and provide services for FM professionals, two DT instances were developed in this research project. A custom DT-specific instance was designed for research purposes and a commercial DT instance was developed by Bentley Systems, Inc. using their Assetwise platform. Both of them support for further development and evolution, and open for additional services and functions. These DT development processes provide two approaches to achieving DTs implementations in real practices.

These lessons learned are the unique contribution of this study and further can be widely generalized to DT development based on this system architecture. Some of the details presented in this pilot project (e.g., digital modelling, transmission network establishment) will be a solid reference for other projects with similar attributes and can further be applicable and extended to other areas. Future research is needed to consider different culture backgrounds (e.g., society, economy) and variations of DTs. The DTs in specific cities and further interacting with people must define and establish the appropriate data requirements, interoperability needs and cultures in that target areas (e.g., local policy, local BIM authoring tools and requirements) (Inyim et al.

2014). Hence, the system architecture and its details can be defined, revised and established accordingly. Based on the experience and lessons gained from this research, a road map is developed for DT development (see Figure 13). The proposed road map in Fig.13 provides a framework for future researchers to mention significant highlights and provide insight into the new field of DT development. These future proposed case studies can be then followed by a cross analysis of multiple cases to further enhance the existing architecture, and build the growing knowledge foundation of DT developments.

Conclusions

With the extensive attention to implementations of DTs and the expectations to take all the advantages of DTs into our daily lives, this study provided a comprehensive analysis from the definitions of DT and its applications in the AEC/FM sector firstly. In order to present the insight into the new field of dynamic DTs in building and city levels, this study provided a detailed description of the development of a DT. A system architecture informed the development of this DT pilot in building and city levels was also presented and explained. Following this developed architecture, a DT demonstrator of the West Cambridge site was developed, including a building DT (i.e., a sub-DT) using IfM building as a case study. In-depth analysis was conducted to highlight the challenges of developing DTs from the perspective of data management (including data integration, heterogeneity in source systems, data synchronisation and data quality). Lessons learned were discussed and a road map was provided for future researchers. Furthermore, it was clear that successful deployment and use of DTs face significant data management challenges and need well-organised guidance.

This research contributes to the body of knowledge by developing a novel system architecture for future researchers to systematically and strategically build the knowledge foundation on DTs development, developing one of the first few exploratory pilot projects on developing a DT in building and city levels, as well as proposing a road map for highlighting key perspectives for future research. The detailed implementation process and the lessons learned in this pilot project were discussed and presented in this paper, which provided valuable insights and future directions into the successful implementation of DTs in building and city levels. However, analysing value and usefulness of integrating city-level information are not discussed and studied enough in this study, which will be covered in future works.

In future work, we will collect occupant feedbacks and conduct performance evaluations through working with Estate Management department in this University, validate the proposed system architecture to broader city scale and investigate more practical applications of the DTs development in supporting the wider management activities and services. Moreover, we will also demonstrate the impact of digital modelling and analysis of infrastructure performance and use on organisational productivity and further provide the foundation to optimise city services such as power, waste, transport and understand the impact on wider social and economic outcomes.

Data Availability Statement

Some or all data, models, or code used during the study were provided by a third party (i.e., the cloud point of the west Cambridge site and IfM building; data resources including BMS, AMS and SMS). Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.

Acknowledgement

The research that contributed to this paper was funded by the EPSRC/Innovate UK Centre for Smart Infrastructure and Construction and the Centre for Digital Built Britain at the University of Cambridge.

References

Ahuja, V., Yang, J., and Shankar, R. (2009). “Benefits of collaborative ICT adoption for building project management.” *Construction Innovation*, 9(3), 323-340. https://doi.org/10.1108/14714170910973529.

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M., (2015). “Internet of things: A survey on enabling technologies, protocols, and applications.” *IEEE communications surveys & tutorials*, 17(4), 2347-2376.

Araújo, T. B., Cappiello, C., Kozievitch, N.P., Mestre, D.G., Pires, C.E.S., Vitali, M. (2017). “Towards Reliable Data Analyses for Smart Cities.” *Proceedings of the 21st International Database Engineering & Applications Symposium,* ACM, UK, 304-308, http://doi.org/10.1145/3105831.3105834.

Balobaid, A., and Debnath, D. (2018). Cloud Migration Tools: Overview and Comparison. *World Congress on Services*, Springer, Cham, 93-106.

Bansal, S. K., and Sebastian K. (2015). "Integrating big data: A semantic extract-transform-load framework." *Computer*, 48(3), 42-50.

Beyer, M., Thoo, E., Selvage, M., Zaidi, E., (2017). "Gartner Research Report: Magic Quadrant for Data Integration Tools." *Gartner research report*, No. G00314940.

Budgen, D., Rigby, M., Brereton, P., Turner, M., (2007). "A Data Integration Broker for Healthcare Systems." *Computer*, 40(4), 34–41.

Bolton, A., Enzer, M., Schooling, J., (2018). "*The Gemini Principles: Guiding values for the national digital twin and information management framework*" https://doi.org/10.17863/CAM.32260.

Cao, D., Li, H., Wang, G., Luo, X., Yang, X., Tan, D. (2016). "Dynamics of project-based collaborative networks for BIM implementation: Analysis based on stochastic actor-oriented models." *Journal of Management in Engineering*, 33(3), 04016055, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000503.

Chen, W., Chen, K., Cheng, J.C., Wang, Q., Gan, V.J., (2018). "BIM-based framework for automatic scheduling of facility maintenance work orders." *Automation in construction,* 91, 15-30, https://doi.org/10.1016/j.autcon.2018.03.007.

Costa, A., Keane, M.M., Torrens, J.I. and Corry, E., (2013). "Building operation and energy performance: Monitoring, analysis and optimisation toolkit". Applied Energy, 101, pp.310-316.

CRC Construction (2007). "Adopting BIM for facilities management: Solutions for managing the Sydney Opera House." Cooperative Research Centre for Construction Innovation, Brisbane, Australia, <http://www.construction-innovation.info/images/CRC_Dig_Model_Book_20070402_v2.pdf>.

Ding, L., Drogemuller, R., Akhurst, P., Hough, R., Bull, S., Linning, C., (2009). Towards sustainable facilities management, Peter Newton, Keith Hampson, Robin Drogemuller (Eds.), *Technology, Design and Process Innovation in the Built Environment*, Taylor & Francis, Oxon, Abingdon, http://eprints.qut.edu.au/20926/.

Fattah, S.M.M., Sung, N.M., Ahn, I.Y., Ryu, M., Yun, J., (2017). "Building IoT Services for

Aging in Place Using Standard-Based IoT Platforms and Heterogeneous IoT Products." *Sensors,* 17(10), 2311-2340, https://doi.org/doi:10.3390/s17102311.

Farghaly, K., Abanda, F. H., Vidalakis, C., Wood, G. (2018). "Taxonomy for BIM and Asset Management Semantic Interoperability." *Journal of Management in Engineering*, 34(4), https://doi.org/10.1061/(ASCE)ME.1943-5479.0000610.

France-Mensah, J., O'Brien, W. J. (2018). "Budget allocation models for pavement maintenance and rehabilitation: Comparative case study." *Journal of Management in Engineering*, 34(2), 05018002, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000599.

Gartner (2017). "Prepare for the Impact of Digital Twins." <https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins/> (Sep. 2017).

GE Digital (2017). "Digital Twins: The Bridge Between Industrial Assets and the Digital World." <https://www.ge.com/digital/blog/digital-twins-bridge-between-industrial-assets-and-digital-world> (Dec. 2017).

Ge, X., Tu, S., Mao, G., Wang, C.-X., Han, T., (2016). "5G ultra-dense cellular networks." *IEEE Wireless Communications*, 23, 72–79, https://doi.org/10.1109/MWC.2016.7422408.

Giel, B., Issa, R. R. (2015). "Framework for evaluating the BIM competencies of facility owners." *Journal of Management in Engineering*, 32(1), 04015024, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000378.

Gil, J., Almeida, J, Duarte, J. (2011). "The backbone of a City Information Model (CIM)", in Respecting Fragile Places: Education in Computer Aided Architectural Design in Europe (eCAADe) Ljubljana, 143-151, http://cumincad.scix.net/cgi-bin/works/Show?ecaade2011_104.

Gil, J., Beirão, J., Moutenegro, N., Dunantie, J. (2010). "Assessing computational tools for urban design: towards a "city information model"." 28th Conference on Future Cities, Switzerland: eCAADe, 316-324.

Glaessgen, E., Stargel, D., (2012). "The digital twin paradigm for future NASA and US Air Force vehicles." *Proceedings of 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences*, Honolulu, Hawaii, 1818-1832, https://doi.org/10.2514/6.2012-1818.

Grieves, M., Vickers, J., (2017). "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems." *Transdisciplinary perspectives on complex systems*, Springer, Cham, 85-113.

Hu, Z.Z., Tian, P.L., Li, S.W., Zhang, J.P., (2018). "BIM-based integrated delivery technologies for intelligent MEP management in the operation and maintenance phase." *Advances in engineering software*, 115, 1-16, https://doi.org/10.1016/j.advengsoft.2017.08.007.

Huang, J., Qian, F., Gerber, A., Mao, Z. M., Sen, S., Spatscheck, O., (2012). "A close examination of performance and power characteristics of 4G LTE networks." *Proceedings of the 10th international conference on Mobile systems, applications, and services*, ACM, 225–238.

High Value Manufacturing Catapult Visualisation and VR Forum (HVM Catapult) (2018). "Feasibility of an immersive digital twin: The definition of a digital twin and discussions around the benefit of immersion." <https://www.amrc.co.uk/files/document/219/1536919984_HVM_CATAPULT_DIGITAL_TWIN_DL.pdf> (Sep. 2018).

Inyim, P., Rivera, J., and Zhu, Y. (2014). "Integration of building information modeling and economic and environmental impact analysis to support sustainable building design." *Journal of Management in Engineering*, 31(1), A4014002. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000308.

Itemit (2019). "The Simple Asset Management Software." <https://itemit.com/> (Jan. 2019).

Lanzisera, S., Zats, D., Pister, K.S., (2011). "Radio frequency time-of-flight distance measurement for low-cost wireless sensor localization." *IEEE Sensors Journal*, 11(3), 837-845.

Lee, J., Jeong, Y., Oh, Y.S., Lee, J.C., Ahn, N., Lee, J., Yoon, S.H., (2013). "An integrated approach to intelligent urban facilities management for real-time emergency response." *Automation in construction*, 30, 256-264, https://doi.org/10.1016/j.autcon.2012.11.008.

Lee, J., Bagheri, B., Kao, H. A., (2015). "A cyber-physical systems architecture for industry 4.0-based manufacturing systems." *Manufacturing letters*, 3, 18-23, https://doi.org/10.1016/j.mfglet.2014.12.001.

Lee, E. A., (2008). "Cyber physical systems: Design challenges." *11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, IEEE, 363-369.

Lehr, W., McKnight, L.W., (2003). "Wireless internet access: 3G vs. WiFi?"

Telecommunications Policy, 27, 351–370, https://doi.org/10.1016/S0308-5961(03)00004-1.

Lewis, F.L., (2004). “Wireless sensor networks.” *Smart environments: technologies, protocols, and applications*, 11-46.

Lin, J., Zha, L., Xu, Z., (2013). “Consolidated cluster systems for data centers in the cloud age: A survey and analysis.” *Frontiers of Computer Science*, 7, 1–19, https://doi.org/10.1007/s11704-012-2086-y.

Lu, Q., Chen, L., Lee, S., Zhao, X., (2018a). “Activity theory-based analysis of BIM implementation in building O&M and first response.” *Automation in Construction*, 85, 317-332, https://doi.org/10.1016/j.autcon.2017.10.017.

Lu, Q., Lee, S., Chen, L., (2018b). “Image-driven fuzzy-based system to construct as-is IFC BIM objects.” *Automation in construction,* 92, 68-87, https://doi.org/10.1016/j.autcon.2018.03.034.

Lu, W., Fung, A., Peng, Y., Liang, C., and Rowlinson, S. (2015). “Demystifying construction project time–effort distribution curves: BIM and non-BIM comparison.” *Journal of Management in Engineering*, 31(6), 04015010. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000356.

Loshin, D., (2009). *Master Data Management*, Elsevier, https://doi.org/10.1016/B978-0-12-374225-4.X0001-X.

Lukyanenko, R., Parsons, J., Wiersma, Y.F., (2014). “The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content.” *Information Systems Research*, 25(4), 669–689, https://doi.org/10.1287/isre.2014.0537.

Kang, T.W., Hong, C.H., (2015). “A study on software architecture for effective BIM/GIS-based facility management data integration.” *Automation in construction,* 54, 25-38, https://doi.org/10.1016/j.autcon.2015.03.019.

Kaur, M.J., Mishra, V.P. and Maheshwari, P., (2020). “The Convergence of Digital Twin, IoT, and Machine Learning: Transforming Data into Action”. Digital Twin Technologies and Smart Cities. Springer, 3-17.

Knapp, G.L., Mukherjee, T., Zuback, J.S., Wei, H.L., Palmer, T.A., De, A., DebRoy, T., (2017). “Building blocks for a digital twin of additive manufacturing.” Acta Materialia, 135, 390-399, https://doi.org/10.1016/j.actamat.2017.06.039.

Kim, K., Cho, Y. K., Kim, K. (2018). "BIM-Based Decision-Making Framework for Scaffolding Planning." *Journal of Management in Engineering*, 34(6), 04018046, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000656.

Kleissl, J., Agarwal, Y., (2010). "Cyber-physical energy systems: Focus on smart buildings." *Design Automation Conference,* IEEE, 749-754.

Krylovskiy, A., Jahn, M., Patti, E., (2015). "Designing a smart city internet of things platform with microservice architecture." *3rd International Conference on Future Internet of Things and Cloud*, IEEE, 25-30.

Ma, X., Xiong, F., Olawumi, T. O., Dong, N., and Chan, A.P., (2018). "Conceptual Framework and Roadmap Approach for Integrating BIM into Lifecycle Project Management." *Journal of Management in Engineering*, 34(6), 05018011. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000647.

Ma, Y., Li, G., Xie, H., Zhang, H. (2018). "City Profile: Using Smart Data to Create Digital Urban Spaces." *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Delft, The Netherlands.

MiCAD (2019). <https://www.em.admin.cam.ac.uk/working-with-us/estate-management-it-systems/micad>. (Feb. 2019).

Mobley, R.K. (2002). "An Introduction to Predictive Maintenance." 2nd ed., Elsevier Inc. https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3.Monnit (2018a). "Monnit Wireless Sensors."

<https://www.monnit.com/ProductManagement/FilterResult/?productCategoryID=1&firmwareBaseID=1> (Oct. 15, 2018).

Monnit (2018b). "Monnit Wireless Sensor Datasheet."

<https://monnit.azureedge.net/content/documents/datasheets/sensors/MDS-001-Temperature-Sensor-Data-Sheet.pdf> (Oct.15 2018).

Monnit (2018c). "Monnit Ethernet Gateway."

<https://www.monnit.com/Products/Gateways/Ethernet-Gateway/Wireless-Ethernet-Gateways> (Oct.15 2018).

Motawa, I., Almarshad, A., (2013). "A knowledge-based BIM system for building maintenance." *Automation in construction,* 29, 173-182, http://doi.org/10.1016/j.autcon.2012.09.008.

Mohammadi, N., Taylor, J. E. (2017). “Smart city digital twins.” *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1-5.

National Infrastructure Commission (NIC) (2017). “Data for the public good.” <https://www.nic.org.uk/wp-content/uploads/Data-for-the-Public-Good-NIC-Report.pdf> (Nov. 2017).

Ohmura, N., Takase, E., Ogino, S., Okano, Y., Arai, S., (2013). “Material property of onmetal magnetic sheet attached on NFC/HF-RFID antenna and research of its proper pattern and size on, Antennas & Propagation (ISAP).” *Proceedings of the international symposium*, IEEE, 1158–1161.

Oliver, D., Adam, D., Hudson-Smith, A.P. (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, GIS Research UK (GISRUK).

Otto, B., (2012). “How to design the master data architecture: Findings from a case study at Bosch.” *International Journal of Information Management*, 32(4), 337–346.

Otto, B., Hüner, K.M., Österle, H., (2012). “Toward a functional reference model for master data quality management.” *Information Systems and E-Business Management*, 10(3), 395–425.

Pärn, E.A., Edwards, D.J., Sing, M.C.P., (2017). “The building information modelling trajectory in facilities management: A review.” *Automation in construction*, 75, 45-55, https://doi.org/10.1016/j.autcon.2016.12.003.

Peng, Y., Lin, J.R., Zhang, J.P., Hu, Z.Z. (2017). “A hybrid data mining approach on BIM-based building operation and maintenance.” *Building and environment,* 126, 483-495, https://doi.org/10.1016/j.buildenv.2017.09.030.

Planet (2019). <https://www.em.admin.cam.ac.uk/working-us/em-it-systems/planet> (Feb. 2019).

Qu, X., Jiang, Z., (2018). “A Time-based Dynamic Synchronization Policy for Consolidated Database Systems.” *Forthcoming in MIS Quarterly*.

Rajkumar, R., Lee, I., Sha, L., Stankovic, J. (2010). “Cyber-physical systems: the next computing revolution.” *Design Automation Conference*, IEEE, 731-736.

Rathore, M. M., Ahmad, A., Paul, A., Rho, S., (2016). “Urban planning and building smart cities based on the internet of things using big data analytics.” *Computer Networks*, 101, 63-80, https://doi.org/10.1016/j.comnet.2015.12.023.

Róka-Madarász, L., Mályusz, L., Tuczai, P., (2016). “Benchmarking facilities operation and maintenance management using cafm database: Data analysis and new results.” *Journal of building engineering*, 6, 184-195, https://doi.org/10.1016/j.jobe.2016.03.007.

Romberg, R., Niggl, A., van Treeck, C., Rank, E. (2004). “Structural analysis based on the product model standard IFC.” *Proceedings of International Conference on Computing in Civil and Building Engineering*, ICCCBE, Weimar, Germany.

Saini, H. (2012). “Li-Fi (Light Fidelity)-The future technology In Wireless communication.” *Journal of Computer Application*, 7(1), 13-15, https://www.akgec.in/journals/JCA/jan-Jun-2016.pdf#page=13.

Sackey, E., Tuuli, M., and Dainty, A. (2014). “Sociotechnical systems approach to BIM implementation in a multidisciplinary construction context.” *Journal of Management in Engineering*, 31(1), A4014005. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000303.Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., Wang, L. (2012). Modeling, Simulation, Information Technology & Processing Roadmap. Technology Area 11, 2012.

Shalabi, F., Turkan, Y., (2016). “IFC BIM-based facility management approach to optimize data collection for corrective maintenance.” *Journal of performance of constructed facilities*, 31(1), 04016081, https://doi.org/10.1061/(ASCE)CF.1943-5509.0000941.

Silva, B. N., Khan, M., Han, K. (2018). “Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities.” *Sustainable Cities and Society,* 38, 697-713, https://doi.org/10.1016/j.scs.2018.01.053.

Silva, B. N., Khan, M., Han, K., (2017). “Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management.” *Future Generation Computer Systems*, https://doi.org/10.1016/j.future.2017.06.024.

Son, H., Kim, C., and Kwon Cho, Y. (2017). “Automated schedule updates using as-built data and a 4D building information model.” *Journal of Management in Engineering*, 33(4), 04017012. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000528.

Song, J., Migliaccio, G. C., Wang, G., Lu, H. (2017). "Exploring the influence of system quality, information quality, and external service on BIM user satisfaction." Journal of Management in Engineering, 33(6), 04017036, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000549.

Steel, J., Drogemuller, R., Toth, B. (2012). "Model interoperability in building information modelling". Software & Systems Modeling, 11(1), 99-109, https://doi.org/10.1007/s10270-010-0178-4.

Suprabhas, K., Dib, H.N., (2017). "Integration of BIM and utility sensor data for facilities management." *Proceedings of ASCE International Workshop on Computing in Civil Engineering*, Seattle, Washington, USA, pp. 26-33, https://doi.org/10.1061/9780784480823.004.

Synapsys (2018). "SIP Interfaces."

< http://www.synapsys-solutions.com/products/interfaces/> (Nov. 12, 2018).

Talburt, J.R. (2011). *Entity Resolution and Information Quality*, Elsevier.

Taylor, J. E., and Bernstein, P.G., (2009). ″Paradigm trajectories of building information modeling practice in project networks.″ *Journal of Management in Engineering*, 25(2), 69-76. https://doi.org/10.1061/(ASCE)0742-597X(2009)25:2(69).

Vassiliadis, P., (2009). "A survey of Extract–transform–Load technology." *International Journal of Data Warehousing and Mining (IJDWM)*, 5(3), 1–27.

Wang, R.Y., Strong, D.M., (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, 12(4), 5–34.

Woodall, P., (2017). "The Data Repurposing Challenge: New Pressures from Data Analytics." *Journal of Data and Information Quality*, 8(3–4), 11:1–11:4.

Woodall, P., Borek, A., Oberhofer, M., Gao, J., (2016). "Data Quality Problems in ETL: The State of the Practice in Large Organisations." *Processing of International Conference on Information Quality (ICIQ)*, Ciudad Real, Spain.

Woodall, P., Oberhofer, M., Borek, A., (2014). "A Classification of Data Quality Assessment and Improvement Methods." *International Journal of Information Quality*, 3(4), 298–321.

Yick, J., Mukherjee, B., Ghosal, D., (2008). "Wireless sensor network survey." *Computer networks*, 52(12), 2292-2330, https://doi.org/10.1016/j.comnet.2008.04.002.