A message-driven architecture to integrate isolated building data into digital twins

by

Asad SHAHIN MOGHADAM

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER’S DEGREE
WITH THESIS IN CONSTRUCTION ENGINEERING
M.A.Sc.

MONTREAL, DECEMBER 6, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Ali Motamedi, Thesis supervisor
Department of Construction Engineering, École de Technologie Supérieure

Mr. Kim Khoa Nguyen, Chair, Board of Examiners
Department of Software Engineering and IT, École de Technologie Supérieure

Mr. Érik Poirier, Member of the Jury
Department of Construction Engineering, École de Technologie Supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON NOVEMBER 22, 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

"Let the future tell the truth, and evaluate each one according to his work and accomplishments. The present is theirs; the future, for which I have really worked, is mine." – Nikola Tesla.

ACKNOWLEDGEMENTS

First and foremost, I would like to show my sincere gratitude to my supervisor. Professor Ali Motamedi, thank you for the invaluable opportunities you offered, and all your support and guidance during this journey.

I wish to express my gratitude to *Division Ingénierie et soutien technique* of the *Ville de Québec* for their cooperation and data access opportunities.

I also would like to thank my uncle Akbar Shaygan Tabar, my friend Dr Farzaneh Golkhoo, and my cousin Mehrzad Shahin Moghadam, for their constructive comments on my text.

My beloved Sara (aka Queen Banoo!), there are no words that can thoroughly describe my feelings, and love for you as the best sister ever, friend, and co-supervisor. Thanks for always being there for me in every step of this thesis. I have always looked up to you and am genuinely grateful for having you.

This thesis is foremost dedicated to my beloved parents, Zari and Amir. I can never thank you enough for your unconditional love and support, for providing me with everything I needed to reach here and for your kindness. I love you!

Montreal, October 9, 2022
Asad Shahin Moghadam

Une architecture axée sur les messages pour intégrer les données de bâtiments isolés dans des jumeaux numériques

Asad SHAHIN MOGHADAM

RÉSUMÉ

Cette thèse vise à développer et à mettre en œuvre un cadre pour le jumeau numérique (DT) des bâtiments dont les systèmes se trouvent dans un environnement isolé en réseau, ce qui pose des problèmes d'intégration des données. Il existe plusieurs études qui se concentrent sur les modèles conceptuels et la mise en œuvre du DT dans l'architecture, l'ingénierie, la construction, l'exploitation et la gestion des installations (AECO-FM). Les DT peuvent être appliquées à la gestion des installations, à la maintenance préventive, à la surveillance et à la détection des anomalies, et améliorer l'efficacité des projets de construction et la sécurité de la main-d'œuvre. Cependant, en raison de certaines caractéristiques de cette industrie, ce secteur a été lent à adopter les DT. De plus, malgré le nombre croissant d'implémentations de DT, les défis liés à leur mise en œuvre ont fait l'objet d'une attention moindre. Les problèmes d'intégration des données, notamment l'accès, la collecte, le stockage et l'échange de données, figurent parmi les principaux défis identifiés par les chercheurs. Plus précisément, la majorité des études sur les DT ont supposé que les données étaient facilement accessibles pour la mise en œuvre des DT, alors que c'est rarement le cas dans la pratique.

Cette recherche se concentre sur le problème de l'intégration des données dans la mise en œuvre d'un DT dans les bâtiments où l'accès aux données des capteurs est limité pour diverses raisons telles que l'accessibilité des données et les problèmes de confidentialité, ou en raison de systèmes BMS/BAS existants. Un cadre méthodologique pour la mise en œuvre d'une DT dans de tels scénarios est présenté dans lequel les données sensorielles hors ligne basées sur des fichiers sont transformées en données en ligne basées sur des messages. En outre, une architecture distribuée axée sur les messages est conçue pour mettre en œuvre ce cadre méthodologique. Une étude de cas d'un centre communautaire est mise en œuvre pour évaluer l'applicabilité du cadre méthodologique proposé.

Cette thèse contribue au domaine en assouplissant l'hypothèse répandue dans la littérature sur la facilité d'accès aux données pour la mise en œuvre du DT. Le cadre méthodologique proposé et son architecture logicielle " cloud-native " peuvent être appliqués à des scénarios similaires dans lesquels les données sensorielles ne sont pas facilement accessibles. Le cadre proposé est réutilisable, évolutif et flexible dans les types de données qu'il peut traiter. De plus, les données sont essentielles aux collaborations au sein des projets AECO-FM. Par conséquent, cette étude est également bénéfique pour les praticiens en leur permettant d'utiliser le potentiel des données inexploitées qui sont constamment collectées dans les silos de données des systèmes existants des différentes parties prenantes, mais qui restent isolées.

Mots-clés: jumeau numérique, construction, intégration des données, gestion des installations, système de construction, capteurs isolés, DMZ, architecture axée sur les messages, informatique en nuage (cloud computing)

A message-driven architecture to integrate isolated building data into digital twins

Asad SHAHIN MOGHADAM

ABSTRACT

This thesis aims at developing and implementing a framework for the Digital Twin (DT) of buildings that their systems are inside a network-wised isolated environment, hence face data integration challenges. There are several studies focusing on conceptual models and implementation of DT in Architecture, Engineering, Construction, Operation and Facility Management (AECO-FM). DTs can be applied to facility management, preventive maintenance, monitoring and anomaly detection, and enhance the efficiency of constructions projects and workforce safety. However, due to certain features of this industry, this sector has been slow in adoption of DTs. Moreover, despite increasing number of DT implementations, less attention has been given to the challenges of implementing them. Data integration issues including data access, collection, storage and exchange are among the main challenges identified by scholars. Specifically, the majority of DT studies have assumed that data are readily accessible for implementing DTs, while it is rarely the case in practice.

This research focuses on the data integration problem in implementing a DT in buildings where accessing sensor data is restricted for various reasons such as data accessibility and privacy issues, or due to deployed legacy BMS/BAS systems. A methodological framework for implementing a DT in such scenarios is presented in which file-based offline sensory data is transformed to message-based online data. In addition, a distributed message-driven architecture is designed to implement this methodological framework. A case study of a community center is implemented to assess the applicability of the proposed methodological framework.

This thesis contributes to the field, by relaxing the widespread assumption of easy data accessibility for DT implementation in the literature. The proposed methodological framework and its cloud-native software architecture can be applied to similar scenarios in which sensory data is not readily accessible. The proposed framework is reusable, scalable and flexible in the data types that it can handle. Moreover, Data is critical to collaborations within AECO-FM projects. Therefore, this study is beneficial to practitioners as well allowing them to use the potentials of untapped data which is being constantly collected in data silos of legacy systems of various stakeholders but kept isolated.

Keywords: digital twin, construction, data integration, facility management, building system, isolated sensors, DMZ, message-driven architecture, cloud computing

TABLE OF CONTENTS

Page

INTRODUCTION 1

CHAPTER 1 LITERATURE REVIEW 11
1.1 Digital twin 11
1.2 Digital twin in construction 12
1.3 Digital twin application areas in construction 15
1.4 Challenges of adopting digital twins 19

CHAPTER 2 RESEARCH METHODOLOGY 25
2.1 Design science research methodology 25
2.1.1 Problem identification 26
2.1.2 Solution objectives 28
2.1.3 Design and development 29
2.1.4 Demonstration 30
2.1.5 Evaluation 31
2.1.6 Communication 31

CHAPTER 3 THE PROPOSED DT METHODOLOGICAL FRAMEWORK 33
3.1 DT methodological framework 33
3.1.1 Identification and analysis of sensor data 34
3.1.2 Simulating sensor data 34
3.1.3 Simulated IoT device provisioning 35
3.1.4 Storing time-series data 36
3.1.5 Visualization 36
3.2 DT software architecture 37

CHAPTER 4 CASE STUDY: BARDY COMMUNITY CENTER 41
4.1 Identification and analysis of sensor data from the enteliWeb (BAS/BMS software) 44
4.2 Simulating sensor data 45
4.2.1 Function A: Store attachments 45
4.2.2 Function B: Data preparation 48
4.2.3 Function C: Produce message 48
4.2.4 Function D: Consume message 51
4.3 Simulated IoT device provisioning 52
4.4 Storing time-series data 56
4.5 Visualization 56
4.5.1 Model viewer 57
4.5.2 Integration with collaborative tools 57
4.5.3 Integration with Azure IoT Hub 58

4.5.4 Mapping sensors 58
4.5.5 Integration with Azure Time Series Insight (TSI) 60
4.5.6 Visualization components 60
4.6 Evaluation of initial building temperature outputs 61

CHAPTER 5 DISCUSSIONS AND CONCLUSIONS 67
5.1 Discussions 67
5.2 Contributions 68
5.3 Reproducibility and reusability 70
5.4 Flexibility, scalability and availability 71
5.5 Multi-Sourcing capabilities 71
5.6 Limitations and future work 71

BIBLIOGRAPHY 75

LIST OF FIGURES

Page

Figure 2.1 DSRM process of the thesis ... 26

Figure 2.2 The conceptual representation of DT ... 29

Figure 3.1 The proposed methodological framework for implementing DT ... 33

Figure 3.2 The proposed software architecture for DT platform ... 38

Figure 4.1 The software architecture of the implemented DT for community center case study ... 43

Figure 4.2 The sample Excel file of sensor data ... 44

Figure 4.3 The overview of automated workflow designed in Logic App ... 45

Figure 4.4 The workflow that executes Function A ... 46

Figure 4.5 The workflow for storing the attached Excel files ... 47

Figure 4.6 The sample of the generated CSV file that contains sensor data ... 49

Figure 4.7 The execution count logs of Function B ... 50

Figure 4.8 The number of outgoing message logs of Azure Service Bus ... 50

Figure 4.9 The execution count logs of Function D ... 51

Figure 4.10 The Shared Access Policy that grants permissions to access IoT Hub ... 53

Figure 4.11 The Device Twin of a sensor ... 54

Figure 4.12 The number of connected devices to IoT Hub over 24 hours ... 54

Figure 4.13 The number of device-to-cloud messages over 24 hours ... 55

Figure 4.14 The total number of messages for all devices over 24 hours ... 55

Figure 4.15 An example of the sensor data visualization in TSI ... 57

Figure 4.16 The visualization of the Bardy Community Center on the Forge Viewer ... 58

Figure 4.17 The mapped sensors of 1st floor ... 59

Figure 4.18 The mapped sensors of 2nd floor 59

Figure 4.19 The sensor that is mapped inside a room 60

Figure 4.20 A selected sensor and its historical data on the right side panel 61

Figure 4.21 The planar heatmap to visualize the temperature measurements 62

Figure 4.22 The mechanical room of the building 63

Figure 4.23 The temperature data of mechanical room during a week in February and a week in July of 2022 64

Figure 4.24 The comparison of the daily temperature patterns between January 31 and July 4 of 2022 65

LIST OF ABBREVIATIONS

ETS	École de Technologie Supérieure
ASC	Agence Spatiale Canadienne
DSS	Decision Support Systems
DT	Digital Twin
IoT	Internet of Things
BIM	Building Information Modeling
GIS	Geographic Information System
IFC	Industry Foundation Classes
BMS	Building Management System
BAS	Building Automation System
O&M	Operations and Maintenance
SVM	Support Vector Machine
AECO-FM	Architectural, Engineering, Construction, Operation, and Facility Management
API	Application Programming Interface
DSRM	Design Science Research Methodology
DMZ	Demilitarized Zone
PaaS	Platform as a Service
TSI	Time Series Insight
HVAC	Heating, Ventilation, and Air Conditioning
SDK	Software Development Kit

INTRODUCTION

The fourth industrial revolution is characterized in terms of rapid and more cost-efficient development of technologies that contributes to increasing computing power, smart automation and inter-connectivity in various industries (Fuller, Fan, Day & Barlow, 2020). Cyber-physical systems, Internet of things (IoT), Artificial Intelligence and cloud computing are among the technological breakthroughs inducing this digital transformation. Digital Twin (DT) is one of the technological concept which is recognized as having great potentials in automation and enhancement of decision-making in various industrial sectors. It is a concept that integrates various components such as IoT, Simulations, Cyber-physical systems, leading to applications beyond what each of those components can offer. DT is a high-fidelity dynamic mapping between physical assets and their virtual representation that also deals with their bi-direction interaction (Delgado & Oyedele, 2021; Tao, Zhang, Liu & Nee, 2019).

DT was first introduced in a Product Lifecycle Management course at Michigan State University by Michael Grieves in 2002 (Grieves & Vickers, 2017). While recognizing that there exist a physical system and its mirror in virtual space, the emphasis was on its dynamic nature that linked those two systems over the life cycle of the product. In other words, the virtual should continuously adapt to provide a realistic view of the physical. Moreover, the discussions at that time were mainly manufacturing-industry-oriented. Since then, there has been a wide range of definitions of DT in the academic and industrial literature, some being generic and some being very context-specific. With the recent prompt growth in the number of research and applications of DT, there exist many recent review articles that have addressed the existing definitions of DT, their application areas, architectures and components (e.g., Barricelli, Casiraghi & Fogli, 2019; Negri, Fumagalli & Macchi, 2017; Semeraro, Lezoche, Panetto & Dassisti, 2021) . Most of the existing definitions share these three components: a physical asset, its virtual representation and the bi-directional dynamic connection between those two (Delgado & Oyedele, 2021).

Although originated in the manufacturing industry - and still the most implementations of DT are in this sector (Errandonea, Beltrán & Arrizabalaga, 2020)- many other application sectors for DTs quickly emerged; healthcare (e.g., Karakra, Lamine, Fontanili & Lamothe, 2020; Liu *et al.*, 2019), Oil and gas industry (e.g., Wanasinghe *et al.*, 2020) and Infrastructure and construction industry (e.g., Boje, Guerriero, Kubicki & Rezgui, 2020; Khajavi, Motlagh, Jaribion, Werner & Holmström, 2019; Shim, Dang, Lon & Jeon, 2019; Tchana, Ducellier & Remy, 2019). Within each of these sectors, DTs can have various functionality in different stages of the products' life cycle including, design (e.g., Martinelli *et al.*, 2019), construction/production and operations (e.g., Mannino, Moretti, Dejaco, Luciano & Re Cecconi, 2019) and maintenance (e.g., Errandonea *et al.*, 2020; Shim *et al.*, 2019). In addition, the ultimate goal of implementing DTs is prediction and enhanced decision-making for various stakeholders (Akanmu, Anumba & Ogunseiju, 2021; Boje *et al.*, 2020; Opoku, Perera, Osei-Kyei & Rashidi, 2021).

This thesis focuses on DTs within the construction industry and more specifically within Architecture, Engineering, Construction, Operation and Facility Management (AECO-FM). This sector is generally known as being slow in digital transformation and technology adoption and only recently there has been an increase in the number of conceptual publications as well as industrial applications of DTs in the built environment in general (Ozturk, 2021). Moreover, there are a number of factors that add to the complexities of this sector to which DTs may be able to contribute. The sector is inherently interdisciplinary and involves multiple heterogeneous stakeholders that need enhanced decision-making mechanisms and most importantly enhanced collaboration. Moreover, similar to many other sectors, construction is increasingly required to address the new societal, environmental and economic challenges which have implications for the whole lifecycle of built environments. In this context, the DT benefits arise from "[. . .] significantly better situational awareness that it can provide construction managers and workers at all levels, making construction management more proactive than reactive." (Sacks, Brilakis, Pikas, Xie & Girolami, 2020, P. 8). DT in construction goes beyond Building Information

Modeling (BIM) by integrating various technologies and information modeling approaches into a single system and promotes new services and value propositions.

There exist a number of DT conceptual models and implementations in the AECO-FM. Specifically, there are many successful reports of its application in asset management at the operations and maintenance phase of facilities and for monitoring, analysis and action (e.g., Akanmu *et al.*, 2021; Jiang, Ma, Broyd & Chen, 2021; Sacks *et al.*, 2020). DTs are discussed to be applicable in reducing workforce injuries through its monitoring and anomaly detection capabilities and (Mannino *et al.*, 2019; Khajavi *et al.*, 2019; Lu *et al.*, 2020), facility management(e.g., Xie, Lu, Parlikad & Schooling, 2020; Zaballos, Briones, Massa, Centelles & Caballero, 2020), and preventive maintenance (e.g., Pregnolato *et al.*, 2022; Shim *et al.*, 2019).

Although there have been efforts to learn from the DT knowledge already existing in more developed areas of manufacturing and aerospace, DT in the AECO-FM is in its infancy. Many conceptual and practical studies are still required in order to achieve a consensus on its definitions, principles and application areas (Delgado & Oyedele, 2021).

There exist various barriers and challenges in design and implementation of DTs, some of these challenges are generic to all industrial applications while some arise from the specific features of the AECO-FM sector. Data and information are the most critical elements of the DT paradigm. “The effectiveness of the AEC-FM industry activities heavily relies on continuous acquisition, share, store, and use of information via integration with available digital and cognitive technologies, and real-time data.” (Ozturk, 2021, P. 2).

This thesis focuses on challenges and barriers of DT implementation related to data access and data integration in the built environments. Data integration issues related to the accuracy of the collected data, its storage and exchange as well as privacy and security concerns are identified as one of the main challenges by Pregnolato *et al.* (2022). In spite of all these identified challenges, the majority of scholars have assumed that data are readily accessible for DT implementation, which is rarely the case. Data integration step is concerned with the physical and digital entities that need to interoperate effectively in a bi-directional way so that the data streams smoothly. The data that is produced in a physical environment (such as an IoT system) needs to be integrated into a virtual environment in real-time, these requirements for the DT platform bring complexity to the system.

Researchers have identified data integration as one of the main challenges in implementing DTs and argue that the field requires more empirical insights on it. Although critical, data integration challenges are not well-addressed in the context of DTs in general (Fan, Zhang, Yahja & Mostafavi, 2021; Feng, Chen & García de Soto, 2021; Möhring *et al.*, 2022; Tao & Zhang, 2017; Pregnolato *et al.*, 2022).

Möhring *et al.* (2022) identified three main categories of data integration challenges for DT implementation through a survey with experts within the field:

1. There is a lack of standardization of data structures and unclear semantics. Even when there exist standards in certain contexts, there is a lack of knowledge on how to use them.
2. There are various complexities involved in data integration in general. There are often heterogeneous data formats and structures coming from various sources (e.g., different information technology systems) which need to be integrated into a single DT.
3. There can be problems with various system interfaces. For example, in some legacy systems, there is no component designed for accommodating external data transactions.

There are a number of references to these concerns in the implementation of DTs in various sectors and for different purposes. For example, Morgado *et al.* (2020) M and Grieves & Vickers (2017) refer to the problems with “data silos” standing for large amounts of data being collected by various entities but being kept isolated in the Material Science sector and for their specific use case of mechanical testing of the materials.Voigt *et al.* (2021) refer to some data privacy concerns that hinder the application of DTs in healthcare.Singh *et al.* (2018) discuss challenges of internally and externally sharing data and integration of data in various formats and from different sources for the purpose of digital twinning as one of the main challenges of this technology within the high value manufacturing sector.

In the built environments, a number of researchers have identified similar barriers to the adoption of DTs (e.g., Borth, Verriet & Muller, 2019; Delgado & Oyedele, 2021; Tao & Zhang, 2017). This industry operates on the basis of collaboration among a wide range of stakeholders. For concepts such as DT to be feasible in this industry and to add value to this collaboration, data, its acquisition and quality are critical. However, "Silo problems" in data management in such fragmented and heterogeneous context can act as barriers to full access to data as well as timely and effective implementation of DTs.

Moreover, according to Borth *et al.* (2019) there exist various motives among stakeholders of built environment systems that may prevent data access and sharing. Considering that DTs are information-centric systems, data transfer and processing are critical to their performance. However, privacy concerns of the stakeholders and their data regulations can pose limitations, since refusing the data access is often considered a valid preventive approach for potential misuses and attacks. Therefore, privacy concerns of the involved stakeholders in AECO-FM projects need to be addressed so that they would not turn into barriers for the implementation of DTs and exploitation of their potentials.

This research adopts the DT definition offered by Center by Digital Built Britain: "a realistic digital representation of assets, processes or systems in the built or natural environment." (Bolton *et al.*, 2018, P. 6). The thesis focuses on the data integration problem in the implementation of DTs for buildings, in specific contexts in which access to data from sensor networks are hindered for various reasons. The sensor network that hosts the data may be isolated due to network infrastructure inherent limitations, such as the restrictions that are applied to the network because of data accessibility and privacy issues, or limitations corresponding to traditional BAS/BMS systems, e.g., data can only be transmitted in file-based format. IoT data is sensitive and access to it needs to be protected (Fuller *et al.*, 2020), for this reason, organizations are reluctant to share their data by giving access to raw data of sensor networks or BMS. Therefore, data is usually hosted and protected in a Demilitarized Zone (DMZ) which is not routed to the external networks. Current DT solutions offer limited role-based access solutions to the data. In addition, DT solutions may not be internally hosted (on-premise).

This research specifically addresses scenarios where a sensor network, BMS/BAS system, and on-premise data center are inside a network that is restricted to a private group of internal users (employees), and any public access through the internet/web is prohibited by the network administrator. Therefore, only the internal users (employees) can implement a program to push the data, from BAS/BMS (if API is already provided, or any other web integration method that is provided by the system), on-premise database, or directly from the sensor network, out of this isolated network. Considering the specific features of this industry mentioned above (existence of multiple stakeholders, data silos, slow technology adoption and privacy concerns), this specific data integration challenge is very common and widespread in practice.

Research Objectives and Research Questions

This thesis aims to develop a web-based production-ready Digital Twin (DT) for buildings with isolated systems that can integrate the time-series data with the BIM model and create visualization by utilizing cloud-native software architecture. These time-series data sources are assumed to be in an isolated network (DMZ) where access to the sensors and direct streaming of time-series data is not possible. More specifically this research objectives include:

- Identifying emerging technologies, tools, and current existing methods for digital twinning of buildings that are inside an isolated network.
- Presenting a general methodological framework for digital twinning of isolated building systems.
- Proposing a novel cloud-native software architecture for the implementation of the digital twin to overcome isolated network issues, by following the above methodological framework.
- Validating the applicability and effectiveness of the proposed methodological framework and the software architecture by implementing a case study.

The overall research question of the thesis is formulated as below:

> How can we implement a Digital Twin (DT) for built environments where data is hosted on an isolated network and the existing real-time data acquisition methods are not applicable, using emerging technologies and cloud services?

The main research question can be broken down into the following more specific questions:

- How can we access/gather real-time data inside an isolated environment?
- How to map sensor information to the BIM model as a foundation data model for DT?

To address the above objectives and questions, this research proposes a solution for accessing sensor data entrapped in an isolated network. Then, several serverless functions will be developed to simulate the streaming of sensor data. Next, certain processes are adopted for storing time-series data as well as mapping the sensors onto the corresponding object in the virtual environment (BIM model). Finally, the whole system will be implemented based on the proposed software architecture to achieve a web-based DT.

To validate the framework a real-world case study is implemented in a community center building which due to security measures, accessing the sensor network and BAS/BMS systems are not possible. Conventional methods for implementing a digital twin are not applicable in the case study. This DT platform enables access to historical/aggregated data and simulates the streaming of sensor data which is then integrated with the BIM model to visualize them through a web-based platform.

Delimitation

The thesis assumes that the sensor network already exists in the building. Moreover, the building is already equipped with a BAS/BMS System. The rooms and spaces should already be defined in the BIM model. It is assumed that all the devices are already installed within the building and data is already being collected by the BAS/BMS software. In addition, the BAS/BMS software API for the purpose of accessing sensory data is not provided.

The proposed solution is not meant to be generalized as a reference for implementing digital twins of buildings. The proposed architecture and all the developed components of this research can be reused for similar scenarios in which the user cannot access the data of building systems (including sensor network data) through existing methods, such as using APIs with streaming services, streaming channels/endpoint, direct communication with IP-based sensor network, etc.

Research Contributions

By addressing the above research questions, this study contributes to advancement of the knowledge and practice in the following areas:

- This study relaxes the widespread assumption within the field about the ease of direct access and collection of data from a sensor network within a DT platform. The author offers a solution for scenarios in which this assumption does not hold.
- The proposed solution is a framework including a cloud-native architecture, designed in a way that is reusable for other similar settings where a sensor network cannot be readily accessible.

- The proposed framework unfolds the opportunities that lie in untapped offline data, that is traditionally being collected in the construction industry but has rarely been taken advantage of due to outdated software and protocols. Therefore, practitioners can more readily benefit from DTs.

The Thesis Outline

In chapter 1, the relevant literature of the field are reviewed. This overview of the literature, includes the common existing definitions of Digital Twin adopted by scholars in various industries as well as in the construction industry. Various conceptual and empirical applications of digital twins specifically in the construction industry are referred to. Finally, the studies that have addressed the challenges of data integration in the DT context are discussed.

Chapter 2 discusses the methodological approach adopted in this thesis. Various steps of the Design Science Research Methodology (DSRM) are elaborated to illustrate the research process followed to design and implement the offered solution in this thesis.

Chapter 3 corresponds to the design and development phase of the DSRM where the author develops the artifacts of this research, namely a methodological framework and its associated software architecture for development of a DT that addresses data integration problems.

Chapter 4 is the demonstration and evaluation phase of the DSRM where the author illustrates the usability and performance of the proposed solution through a case study. In this case study, a DT is designed and implemented according to the author's proposed framework and software architecture for a community center building in Quebec City in Canada.

Finally, in chapter 5, the author discussed the results and implications of the proposed solution for the academia as well as practitioners within the field. The thesis concludes with a discussion of the research limitations. Thereafter, paths for future research are suggested.

CHAPTER 1

LITERATURE REVIEW

1.1 Digital twin

DT was first introduced in a Product Lifecycle Management course at Michigan State University by Michael Grieves in 2002 (Grieves & Vickers, 2017). While recognizing that there exists a physical system and its mirror in virtual space, the emphasis was on its dynamic nature that linked those two systems over the life cycle of the product. In other words, the virtual should continuously adapt to provide a realistic view of the physical. Moreover, the discussions at that time were mainly manufacturing-industry-oriented. Since then, there has been a wide range of definitions of DT in the academic and industrial literature, some being generic and some being very context-specific.

Although the concept and its applications have been discussed for the past 20 years, there is no one definition of DT that everyone adheres to and it may vary from one application context to another. There are many review articles available on the existing definitions of DTs (e.g., Barricelli *et al.*, 2019; Davari, Shahinmoghadam, Motamedi & Poirier, 2022; Delgado & Oyedele, 2021; Jiang *et al.*, 2021; Negri *et al.*, 2017; Semeraro *et al.*, 2021). The definitions may be slightly different depending on the main DT components or features that authors focus on. For example, many of the definitions emphasize the three components of DT: a physical asset, its virtual representation and the bi-directional dynamic connection between those two (Delgado & Oyedele, 2021). Definitions also exist that focus on five main elements of: physical part, virtual part, connection, data, and service that are expected from a DT implementation (Tao & Zhang, 2017; Semeraro *et al.*, 2021). Some of the most cited definitions of DT in industries are presented below:

- “A DT is an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin.” (NASA, 2012)
- “The Digital Twin is 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 physical manufactured product can be obtained from its Digital Twin." (Grieves & Vickers, 2017)

- "DTs are characterized by the seamless integration between the cyber and physical spaces." (Tao *et al.*, 2019)
- "a comprehensive physical and functional description of a component,product or system together with all available operational data." (Boschert & Rosen, 2016)

There are various DT application areas: Healthcare (e.g., Bruynseels, Santoni de Sio & van den Hoven, 2018; Karakra *et al.*, 2020; Liu *et al.*, 2019); Maritime and Shipping (e.g., Arrichiello & Gualeni, 2019) ; Manufacturing (e.g., Brenner & Hummel, 2017; Semeraro *et al.*, 2021; Tao *et al.*, 2019; Woitsch, Sumereder & Falcioni, 2022); and City Management (e.g., Fan *et al.*, 2021); Aerospace (e.g., Li, Mahadevan, Ling, Choze & Wang, 2017; Li, Aslam, Wileman & Perinpanayagam, 2021); and Construction (e.g., Ozturk, 2021).

1.2 Digital twin in construction

Construction sector and more specifically AECO-FM, has been slow in adoption of DTs, similar to any other digital transformation trends in the sector. Ozturk (2021) attributed this slow transition to some features of this sector such as "slow technology development, complex production, and procurement systems, chaotic supply chain operations, schedule pressure, and financial deprivation." (P8). Moreover, many of these challenges arise from the fact that this sector is highly interdisciplinary and involves various stakeholders at different stages of the projects which need to collaborate. Challenges in poor productivity and high rates of safety-related injuries are also often reported in the literature (e.g., Hou, Wu, Zhang, Tan & Wang, 2021; Opoku *et al.*, 2021).

On the other hand, data is critical to the functioning of this industry. Therefore, many potentials for DT applications and their benefits in terms of productivity, reduced costs, improved safety and most importantly enhanced collaboration are being discussed in the field (e.g., Boje *et al.*, 2020; Khajavi *et al.*, 2019; Opoku *et al.*, 2021).

The Center of Digital Built Britain defines the digital twin as:

> “a realistic digital representation of assets, processes or systems in the built or natural environment.” (Bolton *et al.*, 2018, P. 6)

DT consists of the following major components: physical, digital representation, integration, analytics and aggregation layer. The physical layer or data acquisition layer in particular includes all the built environment and its entities such as sensors, rooms, etc. (Lee & Lee, 2021). In the integration layers, the physical layer needs to be seamlessly integrated with all other virtual components in bi-directional way. This layer consists of the whole processes, patterns and technologies such as using web services, semantic technologies, modern message-based or event-driven patterns, etc. The digital representation includes any computer-graphics method to visualize the content of the corresponding model of the physical layer, e.g. using libraries such as Three.js that use WebGL for visualizing BIM on web (Autodesk Forge Viewer). Another major component of DT is to provide access to archive of historical data. The data aggregation layer consists of any technologies related to data warehousing such as Azure Data Explorer. The final component that plays a significant role in DT is the advanced analytics layer. Using AI algorithms and modern active learning approaches helps the industry to build a centralized intelligent system from a distributed source of data Botín-Sanabria *et al.*. The insight provided by this layer can be beneficial in several ways such as predictive maintenance, proactive decision making, etc.

The discussion of DTs by scholars within construction is often accompanied by a discussion of its similarities and differences with the BIM (e.g., Boje *et al.*, 2020; Delgado & Oyedele, 2021; Shahzad, Shafiq, Douglas & Kassem, 2022). Shahzad *et al.* (2022) through a review of literature within the field as well as a set of semi-structured interviews with industry experts showed that there existed no consensus about the similarities and differences of the two concepts. Douglas, Kelly & Kassem (2021) in another review of literature identified three dominant opinions on the relation between the two concepts among scholars; some consider DT as a development or continuation of BIM for the construction sector, some view the two concepts completely distinct with specific features and functionalities and some consider them complementary.

Clearly, an important goal of both is to offer a data-integrated ground that facilitates collaboration among stakeholders. However, certain differences exist too, e.g., in terms of the employed technologies they use and the stages of the life cycle of the projects that they focus on. Moreover, for DT the existence of a physical counterpart is essential (Jiang *et al.*, 2021) and that is probably the reason why Delgado & Oyedele (2021) and Khajavi *et al.* (2019) found out that BIM is often used in the design and construction stage, compared to DT which is mostly being applied for operations and facility management. BIM, if used in operations and maintenance, is merely a source of information that supports those activities (Lu *et al.*, 2020).

Relevant to that, is the importance of real-time or near-real-time data for DT while they are not required for the design of BIM. Moreover, DT focuses on creating a high-fidelity replica of the physical asset and addresses the features of the environment of the asset too, which are lacking in BIM (Tao *et al.*, 2019). Boje *et al.* (2020) in an extensive review of the literature, identified many of the shortcomings of the BIM and argued how the sector can address those shortcomings by moving towards DT, i.e., they viewed DT as a continuation of BIM or BIM as a component to DT. BIM is often known as a static representation of the physical asset while the dynamic two-way connection between the virtual and physical is essential to the functionality of DT (Delgado & Oyedele, 2021).

Although, according to Douglas *et al.* (2021) further understanding of the similarities and differences of BIM and DT and the boundaries between the two concepts is essential for the advancements of the sector, the discussion of the topic is out of the scope of this thesis. However, it is necessary to point out that in this thesis, the author views BIM as a component of DT. More specifically, BIM itself is a data source for DT, it contains many detailed information about the physical components. The information supports DT implementation in different ways e.g. geometric data of the BIM can be used in the digital representation layer or the relationship between building elements can be used in advanced analytics layer for inference and mapping purposes.

1.3 Digital twin application areas in construction

Although the adoption of DT within the construction industry has been slow compared to many other sectors, there has been an evident growth since 2018 (Opoku *et al.*, 2021; Ozturk, 2021). DTs are recognized to have high potential in the construction sector. DT is often proposed to be ideally applied in the management of the entire Life cycle. The potential arises from the high levels of data integration that lead to visibility over different stages of a construction project and consequently enhancing collaboration among multiple stakeholders (Akanmu *et al.*, 2021); a feature that is exceedingly needed in construction projects, considering their multidisciplinary and multi-stakeholder nature. DT helps connect various systems and data among multiple partners in the supply chain of the projects and supports managerial decision-making and negotiations happening along the project life cycle (Boje *et al.*, 2020). Moreover, various potential efficiency improvement benefits of DTs in different stages of the construction projects are being discussed by scholars within the field (e.g., Boje *et al.*, 2020; Khajavi *et al.*, 2019; Opoku *et al.*, 2021).

Most of the applications and discussions on DT within construction are related to the operations and facility management stage of the projects rather than the design and construction stages of the project's life cycle (Jiang *et al.*, 2021; Jones, Snider, Nassehi, Yon & Hicks, 2020; Opoku *et al.*, 2021). Alizadehsalehi & Yitmen (2021) integrated BIM with Reality Capture and XR (Extended Reality) technologies to design a DT-based framework for automating construction progress monitoring. The research study used a systems and software data modeling technique called Integration Definition (IDEF). The framework was validated through questionnaires distributed to construction project management professionals.

Jiang *et al.* (2021) developed a systematic framework to tackle the challenges corresponding to planning, scheduling, and execution of assembling on the construction site by collecting real-time data in the context of DT. It was stated that the framework could improve the decision-making process. Greif, Stein & Flath (2020) investigated the potential of integrating DT with decision support systems (DSS) in construction logistics and inventory management. The result of the study showed that this approach can decrease logistics costs. Lee and Lee 2021 simulated logistics operations of a modular construction project of a six-story apartment for which the modules were delivered from factories to the construction site by trucks. They developed and tested a DT framework for this purpose that uses IoT, BIM, and GIS. Considering the complexity of these projects and the importance of accurate logistics planning, they illustrate the benefits of DT as an "effective team coordination tool" for enhancing logistics performance in modular construction.

Sacks *et al.* (2020)proposed a workflow for digital twin in construction which included three core steps storing, processing and monitoring. It was also mentioned that DT should be considered as a closed-loop control system and not an additional tool for BIM that is equipped with sensor devices for analytics purposes.

Moreover, Jones *et al.* (2020) identified that there is a lack of DT application in the early stages of the projects as well as their last stage. Similarly, Opoku *et al.* (2021) in a review of articles within the construction industry with a specific focus on DTs applications, identified no application of DTs in the recovery or demolition stage of the built environments. The authors showed that DTs are applied in operations and maintenance of the buildings, building performance management, and optimization of their energy usage.

Khajavi *et al.* (2019) proposed a DT development method applicable to building facades. Their use case tested their DT and included the data collected from a large wireless sensor network (capturing light, ambient temperature, and humidity) installed on the facade of an office building at a university in Finland. As the authors mentioned, the proposed DT is limited and in its early stages of development. It is very much focused on the data collection steps (e.g., setup of their sensor networks) and some processing of the raw data collected. They discussed the technical challenges they faced in their experimentations (e.g., Bluetooth channel disruptions) and how they addressed them. Moreover, the advantages of integrating building life-cycle management into digital twins were discussed.

Mannino *et al.* (2019) focused on integrating the occupant live data with the BIM model for digital twining purposes and optimization of facility management-related operations in an office building.Lydon, Caranovic, Hischier & Schlueter (2019) presented a simulation method based on a DT approach for the thermal design of a heating and cooling system with a lightweight roof structure. Their approach was supposed to later use sensor data of the system in the context of a DT to contribute to monitoring and management of the operation stage of the building structure.

A digital twins' system architecture is proposed by Qiuchen Lu, Parlikad, Woodall, Ranasinghe & Heaton (2019) to integrate an IFC-based BIM model of a university campus with data gathered from multiple buildings' systems such as BMS, sensors network, etc. Forecasting future systems' failure and improving asset management are two other applications that are discussed in the study. Lu *et al.* (2020) further the previous study by Qiuchen Lu *et al.* (2019). They proposed a comprehensive system architecture for implementing a digital twin focusing on the O&M stage of the built lifecycle, which was validated on a university campus building. It was asserted that the proposed architecture can also be effective at the city level. Furthermore, the implemented digital twin was used by Xie *et al.* (2020) to design a framework for improving asset management by adapting anomaly detection methods and continuous monitoring in the O&M phase.

Angjeliu, Coronelli & Cardani (2020) argued for the high potential of DTs for structural study and maintenance operations in historical masonry buildings. They described a procedure for the development of a DT for studying the structural responses of historical buildings and for their maintenance operations. They specifically discussed how to create highly accurate geometric models of these structures which allow measurement and observation of damage in such structures. They further applied their proposed procedure in a use case on Milan Cathedral.

Schweigkofler *et al.* (2022) viewed DT as the integration of BIM and IoT with the benefits of provision of visualization and monitoring of real-time data specifically beneficial for operations and maintenance processes of buildings. They proposed a workflow for developing a cloud-based DT and tested it in a school building in Italy.

DTs can also contribute to monitoring activities, better management of assets and can enhance safety issues in construction projects; an area in which the construction industry is often criticized for. Still, there is a lack of practical application of this approach within the field (Boje *et al.*, 2020; Hoeft & Trask, 2022; Hou *et al.*, 2021). For example, Akanmu *et al.* (2021) proposed a DT framework that helped map the construction workers' body postures in a digital replica, assessing its ergonomics, and contributed to reducing injuries in workers. Liu *et al.* (2019) proposed a framework for indoor safety management by implementing a DT which integrates BIM with operational sensory data, combined with a supervised machine learning method (SVM) to classify and evaluate the level of hazardous events in a stadium complex in Beijing.

In the area of operations and maintenance of infrastructure, a theoretical research study on linear infrastructure such as roads, and railways by Tchana *et al.* (2019) indicated that a singular DT is required in the whole life cycle of such civil construction, especially in the operation and maintenance stage. Yu, Zhang, Hu & Wang (2020) proposed a DT-based method for forecasting the performance of the pavement of a highway tunnel in China, by integrating BIM with the machine learning algorithm and time-series data.

Shim *et al.* (2019) aimed to address the challenges corresponding to bridge maintenance management using a proactive computer vision-based approach by proposing a DT-based system that facilitates the inspection of the bridge's structural elements. In response to a call for AECO-FM sector-specific standards and procedures for developing DTs, Pregnolato *et al.* (2022) proposed a generic workflow for DT development for already existing built environments. They applied their proposed 5-step workflow in a use case on the Clifton Suspension Bridge in Bristol, UK, with no BIM data. Their implementation aimed at preventive maintenance of the bridge.

Lin, Xu, Lu, Guan & Li (2021) proposed a DT-based approach for analyzing the structural health of cable-stayed bridges under high-frequency vibration situations such as earthquakes. To verify the approach, it was tested on a Multi-Functional Shaking Table Laboratory of Tongji University.Yu *et al.* (2021) developed a DT-based framework for improving the O&M of the Wenyi Road Tunnel in Hangzhou, China by integrating an extended COBie standard and Semantic Web technologies. It was indicated that the proposed framework facilitates the federation of the data residing in the digital twin and thereafter reasoning of the federated data. The approach assisted operations and maintenance administrators with better failure analysis of the installed fans in the tunnel.

A case study of a railway bridge was implemented by Gürdür Broo, Bravo-Haro & Schooling (2022) to investigate the systematic and informational as well as organizational aspects of DT in structural health monitoring of smart infrastructure.

In conclusion, many researchers have investigated DTs in various capacities, with different functionalities, and for different lifecycle stages in the AECO-FM industry. However, less attention has been given to the challenges of adopting DT in this industry.

1.4 Challenges of adopting digital twins

Although there has been a fast-growing trend of discussions both within academia and industry about DTs, still their implementation and insights on their practical application are in their infancy specifically in the construction industry (e.g., Möhring *et al.*, 2022). This shortcoming may be attributed to the implementation challenges identified in many studies. For instance, Möhring *et al.* (2022) through a survey of experts in Europe identified three main challenges that hinder the adoption of DTs, namely organization, compliance and data integration challenges. Other studies within different sectors have come to similar conclusions.

Implementation of DTs within organizations is reported to require new or different skills and expertise among workforces (e.g., Delgado & Oyedele, 2021; Feng *et al.*, 2021; Möhring *et al.*, 2022; Pregnolato *et al.*, 2022; Shahzad *et al.*, 2022). The surveys conducted by Shahzad *et al.* (2022) and Möhring *et al.* (2022) both concluded that there is a lack of competencies among the organizations for the implementation of DTs. The interviews in these studies also pointed out that DTs are also not well-addressed in training and education.

Moreover, to capture the full potential of DTs in organizations, high levels of collaborative mindset are required. Scholars have identified cultural changes (e.g., Grieves & Vickers, 2017; Pregnolato *et al.*, 2022; Shahzad *et al.*, 2022) within organizations as well as the requirement of new business models and organizational structures as some of the reasons hindering the adoption of DTs in organizations (e.g., Delgado & Oyedele, 2021). Grieves & Vickers (2017) argued that cultural issues and fragmentation within businesses and within sectors are among contributing issues to organizational siloing, with each partner and unit holding the pieces of information to themselves. Similarly, Singh *et al.* (2018) pointed out that organizational cultures and policies often inhibit internal and external data sharing. Considering the importance of data sharing and integration for harvesting the benefits of DT, cultural changes in organizations are required and are said to be more difficult to handle than technological changes (Grieves & Vickers, 2017).

Another important reason for the lack of DT's application emphasized within the literature is that its benefits are not yet tangible enough to make organizations trust the concept and be confident about its adoption (Fuller *et al.*, 2020; Pregnolato *et al.*, 2022; Ríos, Staudter, Weber, Anderl & Bernard, 2020; Shahzad *et al.*, 2022). Tao & Zhang (2017) argued that organizational motivations behind the adoption of this concept are missing since sufficient analysis and discussion of DT's costs and benefits do not exist.

A number of technical challenges are also discussed by scholars that contribute to the slow practical adoption of DTs. For example, existing IT infrastructure may not be suitable for the implementation of DTs, and their adaptation may be costly and require valid investment justifications (Laborie, Røed, Engdahl & Camp, 2019). Fuller *et al.* (2020) argued that high-performance hardware and software are required to accommodate the unprecedented growth in AI. This challenge is probably more significant in sectors such as construction which is traditionally slow in digitalization (Ozturk, 2021). Although technologies such as sensors and IoT are being widely adopted in recent years, still, Boje *et al.* (2020) argue that their use is often taken for granted and the challenges of their applications in DTs are rarely discussed. Moreover,Möhring *et al.* (2022) stated that there is often a lack of interfaces with the legacy systems of organizations.

Ríos *et al.* (2020) indicated that most of the existing DT applications in the literature have medium fidelity levels. Then, the technical consideration with DTs is that their functionality and benefits are enhanced with high-fidelity models (Opoku *et al.*, 2021; Ozturk, 2021). On the other hand, Pregnolato *et al.* (2022) and Botín-Sanabria *et al.* (2022) explained that the complexities of creating such models and their computational demand (in data collection, virtual-physical synchronization, and data integration) are considerably high. These high levels of computational demand can hinder their widespread applications and are also often translated into higher costs for businesses (Feng *et al.*, 2021; Grieves & Vickers, 2017).

Related to the challenges reviewed above and more important to the purpose of this research are the challenges associated with data in DTs. Pregnolato *et al.* (2022) discussed challenges in data collection, specifically guaranteeing the accuracy and quality of the data, and privacy issues associated with data collection, specifically when client data is being collected. Moreover, considering the amount of data that is constantly being collected by DTs (Delgado & Oyedele, 2021) - to support their functionality, their storage, how it is done, and where they are located- are among the challenges that need to be further addressed by scholars (Akanmu *et al.*, 2021; Boje *et al.*, 2020).

Data integration is the most cited challenge of DT implementations in the literature. The value of DT arises from its capabilities in integrating various types of data. For a DT to work and to enhance information flow among stakeholders (Opoku *et al.*, 2021) various types of data from different resources and models, and collected by different technologies need to be integrated which is often a very challenging task, specifically, considering that multiple stakeholders and organizational units are involved (Möhring *et al.*, 2022; Shahzad *et al.*, 2022). This task as well as guaranteeing the interoperability of multiple systems within DT is quite challenging both because of technical limitations and organizational limitations. As mentioned earlier in this section, and specifically within construction projects, we are facing organizational siloing; organizations having their own data silos (both internally and externally) with limited potential for usage and sharing for example due to privacy reasons (Borth *et al.*, 2019; Grieves & Vickers, 2017; Shahzad *et al.*, 2022). Jones *et al.* (2020) argued that the issue becomes even more challenging when the interdisciplinary nature and multiple stakeholders of the project become bolder as we move towards implementing DTs for the entire life-cycle of the assets and not only for operations and facility management. Tao & Zhang (2017) and Singh *et al.* (2018) also emphasized the importance of this challenge within shop-floor manufacturing and high-value manufacturing sectors, respectively.Laborie *et al.* (2019) addressed the problem of data silos in the oil and gas industry. They mentioned that although large amounts of data are collected, still their usefulness is hampered due to poor technical infrastructure and as well as the lack of interoperability of the systems. They further investigated how these challenges were being addressed using an example from the largest European oil and gas company Aker BP. Shahzad *et al.* (2022) indicated the integration challenges as a key reason for the fact that within the construction industry managers are only using BMS/BAS facility management systems and have difficulties in integrating them within a DT. Lee & Lee (2021) also stated many DT integration problems are not yet addressed. For example, even though in some studies there exist some levels of interoperability among BIM, GIS, and data formats such as IFC, still information is lost when transforming data.

During all these data collection and data integration steps in DT implementation, a critical challenge that requires significant attention is assuring the security and privacy issues associated with that data. This challenge is widely mentioned within the DT literature in different industries (Botín-Sanabria *et al.*, 2022; Feng *et al.*, 2021; Shahzad *et al.*, 2022). However, as Alshammari, Beach & Rezgui (2021) argued, the issue of cybersecurity is rarely practically addressed. They identified only a few studies that had proposed frameworks for improving the security of IoT in the built environment. Again due to the high number of stakeholders and disciplines involved in the construction project, ensuring safe and secure data collection and transmission is of utmost importance for the enterprises involved. The Gemini principles for implementation of DTs put forth by the Centre for Digital Built Britain (Bolton *et al.*, 2018) within the National Digital Twin program in the UK, have also emphasized the trustworthiness of the DT and its security. A privacy-related issue is data ownership in DTs, which as mentioned above, is of critical importance if DT is collecting client-related data (Fuller *et al.*, 2020; Jones *et al.*, 2020).

Such discussions are currently ongoing specifically within the smart automobile and smart home industries as they have legal implications for the stakeholders involved (e.g., Singh *et al.*, 2018).

After all, with all the challenges of DT implementation that scholars have mentioned over years, there is a lack of empirical research in different industries to practically address these challenges and offer solutions. Specifically, as indicated by many scholars, DT implementation requires guidelines for data collection, integration, processing, standard workflows for the implementation processes as well as regulations and policies addressing the privacy and security issues. Among the few studies in this area is Delgado & Oyedele (2021), which through inspiration from a review of literature in the manufacturing sector, identified a set of standards for the representation of DTs in AECO-FM applications. Considering that there is a reported lack of all these standards and regulations (Botín-Sanabria *et al.*, 2022; Fuller *et al.*, 2020; Möhring *et al.*, 2022; Shahzad *et al.*, 2022; Singh *et al.*, 2018), and the risk concerns of organizations are not being addressed, adoption of DTs specifically within the construction industry is progressing at a slow pace, hence organizations do not see tangible benefits of their application.

As was also evident from the studies of DT within AECO-FM, not only the above data integration challenges are hardly addressed, but also accessibility of data and functionality of IoT are often taken for granted. This might be understandable considering the very limited existing practical applications of DT within AECO-FM.

This thesis focuses on the general challenge of data integration in DTs. More specifically, the author tries to practically address the data integration problem that arises from data silo issues in the built environment, the incompetency of legacy systems, and the technical IT infrastructure of organizations that can offer only poor levels of security. Moreover, the expertise level in construction is also quite variable with stakeholders from many different disciplines. The asset in construction can be located in the large spectrum of an old residential house to a modern facility building. Therefore, options for methods of data collection, access, transfer, and processing may vary a lot. Consequently, as is evident from the studies above, the application of DT is prevented due to security reasons or traditional organizational cultures that are hostile to data sharing.

And finally, widespread adoption of DTs is prohibited because the organizations do not get to implement the concept at all, yet alone perceive any benefits arising from it (Shahzad *et al.*, 2022). Therefore, this thesis focuses on the common assumption of the ongoing DT studies within construction about the easy accessibility of data, ignoring all the practical challenges that may hinder smooth data access and streaming. These challenges may have different sources as elaborated above, e.g., poor legacy systems and unaddressed security concerns as well as traditional organizational cultures that all act as barriers to smooth data access. In this study, this assumption is relaxed and a solution is offered for DT implementation under these circumstances.

A close work to this thesis is Pregnolato *et al.* (2022) which as explained above proposed a generic workflow for DT development for already existing built environments which was applied to a bridge management use case. However, they took for granted that sensor data was easily accessible. They indicated that the sensor network requires to be integrated with DT software using APIs.

Another close work to this thesis in terms of the DT framework and application area is Schweigkofler *et al.* (2022) which proposed a similar DT framework for building energy management. In their proposed framework, the access to IoT sensor network was assumed to be existing, similar to all the studies with use cases in the construction sectors which were mentioned in the previous section. Moreover, the authors claimed that they implemented the framework in a school building, however, the process is not well-presented.

CHAPTER 2

RESEARCH METHODOLOGY

In this chapter, different steps of the research process are explained in detail.

2.1 Design science research methodology

This thesis adopts the Design Science Research Methodology (DSRM) (Hevner, March, Park & Ram, 2004; Peffers, Tuunanen, Rothenberger & Chatterjee, 2007) to develop a framework for a digital twin specifically designed for buildings with isolated systems to address their challenge in data integration. Design Science paradigm is "[. . .] fundamentally a problem-solving paradigm" (Hevner *et al.*, 2004, P. 77). Moreover, the DSRM helps to make sure that the research and its outcomes are applicable to the actual problems that practitioners face in the digital transformation of the construction industry. Adopting this methodology helps to ensure that the artifact developed and implemented in this research follows the standards of the field and the research process which led to that artifact is communicable in a scientific way.

The design artifact in this thesis is both a Method and an Instantiation (Hevner *et al.*, 2004; March & Storey, 2008). It is a methodological framework which aims to provide a set of steps on how to solve a problem (data integration). The artifact here is also an Instantiation (in the form of a software architecture) that is derived from the proposed methodological framework (Method artifact) to illustrate the applicability of the Method artifact in real-world systems. In other words, both the design "process" and the designed "product" are evaluated.

In the following paragraphs, we adopt the specific Design Science Methodology developed by Peffers *et al.* (2007) to present the research process adopted to address the objectives of this research. The original methodology proposed in Peffers *et al.* (2007) includes six steps; problem identification, definition of the solution objectives, design, and development, demonstration, evaluation and communication. It is important to note that in line with many other Design Science researches (e.g., Hevner *et al.*, 2004; Markus, Majchrzak & Gasser, 2002), this research process was not a straightforward and linear one, and the solution was developed in an iterative and incremental way going back and forth between the demonstration and evaluation step and the development step. Figure 2.1 illustrates the Design Science Methodological steps adopted in this thesis.

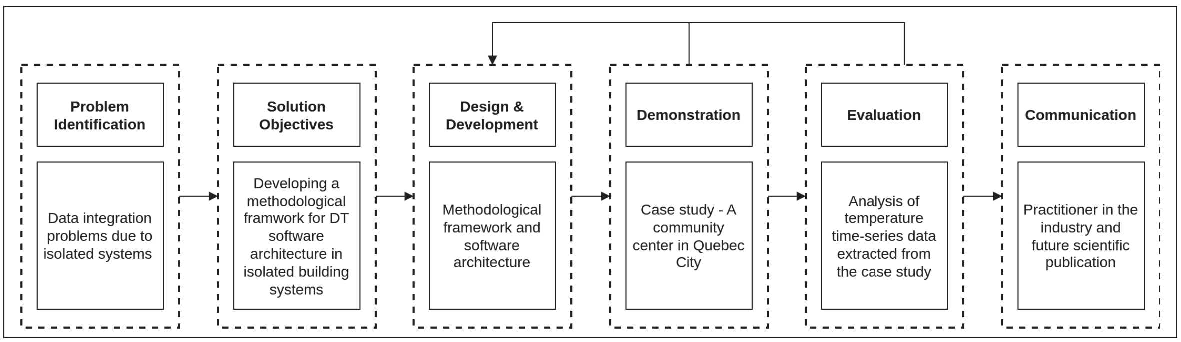


Figure 2.1 DSRM process of the thesis adopted from Peffers *et al.* (2007)

2.1.1 Problem identification

Considering that DTs are information-centric systems, data and its handling are critical to their performance. One of the main challenges related to data in implementation of DTs is data integration for the purpose of efficient interoperability of the physical assets and their digital replica (Fan *et al.*, 2021; Feng *et al.*, 2021; Möhring *et al.*, 2022). One of the contributing factors to data integration problems in DTs is data privacy and data sharing concerns, which can complicate and hinder the implementation of DTs in industries and consequently prevent benefiting from the full potential of DTs. Since adaptations in regulations and conventional approaches of the stakeholders often lag behind the technological innovations, privacy concerns can sometimes unnecessarily slow down the adoption of those technologies. DTs' inherent characteristics make them susceptible to these concerns. There are a number of references to these concerns in the implementation of DTs in various sectors and for different purposes (e.g., Morgado *et al.* (2020) in the Material Science sector; Voigt *et al.* (2021) in the healthcare sector; Singh *et al.* (2018) in the high-value manufacturing sector).

In the built environments, a number of researchers have identified similar barriers to the adoption of DTs (e.g., Borth *et al.*, 2019; Delgado & Oyedele, 2021; Tao & Zhang, 2017). However, privacy concerns of the stakeholders and their data regulations can pose limitations, since refusing the data access is often considered a valid preventive approach for potential misuses and attacks.

Although these concerns are recognized by digital twin scholars, still they are not empirically addressed in the field. In the majority of the DT research (whether theoretical or empirical), specifically the scholars focusing on DTs within AECO-FM fail to explicitly address any of these concerns. The assumption of unrestricted access to data from sensor networks is quite widespread.

In addition, the industry partner of this study (*Ville de Québec*) had a significant role in inspiring of the problem with the presence of a clear practical application. The process of identifying the problem was through having several discussions with the employees of *Ville de Québec*. As part of a larger project, *Ville de Québec* was aiming to implement a DT for a community center building located in Quebec City. The initial step was to collect the data from the building systems, which due to security concerns and technical limitations of the existing network infrastructure and BAS/BMS system, was not feasible. Therefore, the discussion helped the author to narrow down the data integration problem to a specific focus on data access challenges of DT implementations in this study.

This research aims to address this data integration problem by relaxing the common assumption of unrestricted access to sensor network data in a built environment. The data accessibility problem, and in the specific context focused here are highly relevant to the AECO-FM. Network infrastructure requirements may lead to the isolation of the sensor network and building systems that host the data with no public access. Moreover, releasing certain data may expose other parts of a system to threats. There could also be limitations related to the traditional BAS/BMS systems that cause the data to be transmitted only in file-based formats. In another word, the old legacy systems cannot accommodate the transaction of external data. Considering that DT solutions may not be hosted internally, this can be a barrier to its implementation and can prevent benefiting from its potential. The focus of this thesis on these data integration problems is also partially inspired by conversations that the author had with industry practitioners.

2.1.2 Solution objectives

The objective is to purposefully design an artifact that addresses the data integration and access problem in the above setting. There exist a number of common data streaming methods to access the sensory data, for instance, the data can be streamed to the cloud from a network of devices through common communication/messaging protocols such as gRPC, MQTT, WebSocket, etc. or it can be streamed out from a streaming channel provided on the BMS/BAS system API. In any case that the data streaming feature is not already provided/enabled on the BMS Software API or the sensor network, the previously mentioned methods are inapplicable. Moreover, If the systems (including BMS and sensor network) are hosted inside a Demilitarized Zone (DMZ), data cannot be streamed to the cloud using common approaches. These restrictions may exist in built environments due to security measures such as the existence of a firewall.

This research focuses on formulating a methodological framework for the development of a web-based production-ready digital twin for built environments with the data streaming restrictions mentioned above. More specifically, the objective of the solution proposed is to access time-series data sources that are hosted in an isolated network and consequently their direct streaming is not possible. Offline data acquisition methods such as scheduled emails for gathering data in file-based format are identified as suitable solutions in this context. Moreover, the proposed artifact needs to be modular to facilitate scalability and reusability and high availability for production use cases. Figure 2.2 is the conceptual representation of DT in which the Data Integration box illustrates where exactly the proposed solution is focusing on.

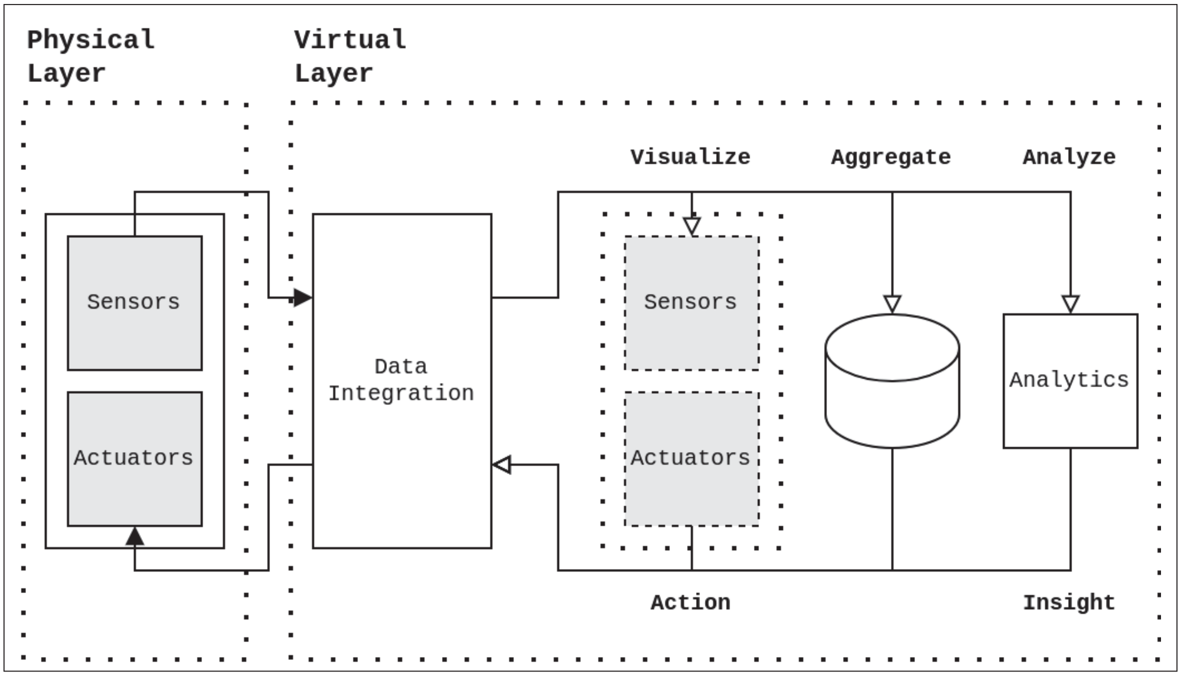


Figure 2.2 The conceptual representation of DT and the Data Integration element as the focus of this study

2.1.3 Design and development

A methodological framework is designed as a solution artifact for the digital twining of isolated building systems. It is assumed that the data collection devices (i.e., sensors) are already installed and configured in the building and a BAS/BMS software is collecting the data. However, external access to this data is prohibited due to privacy reasons and the BAS/BMS software API is not provided either. The suggested methodological framework comprises five following main steps and is described and discussed in detail in chapter 3:

1. Identification and analysis of sensor data
2. Simulation of sensor data
3. Simulated IoT device provisioning
4. Storing time-series data
5. Visualization

To transfer the offline data in the file-based format to messages a set of functions is implemented using a serverless design pattern. Each file is assumed as a data source and its corresponding transformed messages are stored in a database, which then is integrated with its twin (cyber counterpart) for visualization purposes.

This methodological framework artifact is then used to offer an instantiation artifact in the form of a novel cloud-native software architecture. The problem solving focus is mainly on the data integration layer of the software architecture. The rest of its components are included too because they are essential to the implementation process of the DTs. Both the proposed methodological framework and the software architecture cover the whole process and not just the data integration layer. The feedback connection of the DT to the physical world, e.g., in terms of automated analysis and reaction to the extracted time-series are not included in this research.

2.1.4 Demonstration

The suggested methodological framework of this research is demonstrated through a case study. The designed methodological framework is applied step by step to develop a software architecture for digital twining of a community center in Quebec City. The building is already equipped with more than 200 sensors (including around 50 temperature sensors), and sensory data are being collected through BMS Software. However, the building systems are inside an isolated environment. Due to the access issue to the data, we identified it as a suitable scenario to demonstrate the proposed solution of this study. The case study is presented in detail in chapter 4. Considering that the design process is an iterative process, there was constant feedback from this demonstration step to the development step of the design process, incrementally enhancing the performance and utility of the methodological framework. For example, the current configuration of the serverless functions used in the second stage of the methodological framework and its associated software architecture, is achieved after multiple tests and feedback to identify the most suitable setting.

2.1.5 Evaluation

To evaluate the applicability of the proposed methodological framework of this research and to verify its accuracy, the temperature time-series data extracted from the implemented software architecture of the community center were analyzed. The purpose was to verify that the extracted temperature data corresponding to the rooms of the building do follow the expected patterns. The problem addressed in this research is isolated sensor data which is a barrier to DT implementation. If the output time-series of the implemented DT match the temperature patterns that were expected to be observed for the built environment, it can be assured that the objectives of the proposed solution are achieved.

In addition, the solution was exposed to practitioners to validate the result of this study. The feedback received from the internal employees of *Ville de Québec*, and an external expert, confirmed the validity of the proposed solution. It is important to note that the focus of this research was addressing the data integration challenges in DT implementation specifically the data access issue. The aim was to make offline file-based data available for DT implementation. Therefore, the evaluation process was focused on assuring that the proposed methodological framework and software architecture achieved this purpose. Evaluating the general performance of the implemented DT was out of the scope of this study. This evaluation step of the process is described in the final section of chapter 4.

2.1.6 Communication

The author hopes that this thesis in itself and the future publications of the proposed solution can act as communication outlets of this research to the academia. Moreover, the proposed solution of this research is already communicated to the *Division Ingénierie et soutien technique* of the *Ville de Québec*, responsible for the Community Center. Further ongoing communication of the results to the GRIDD lab at the Construction Engineering Department of ETS University is also happening and they potentially can benefit from this solution since they are facing similar DT implementation problems due to data integration challenges.

CHAPTER 3

THE PROPOSED DT METHODOLOGICAL FRAMEWORK

To accomplish the objectives of this study, a general methodological framework is developed which can be reused in cases similar to the use case of this research, in which accessing building systems data is restricted e.g., due to data privacy and security issues or when only offline file-based data is available. This methodological framework and its software architectures are applied to a specific use case in this study where the BMS system and sensor network of the community center building are inside a DMZ and the data are only accessible in file-based spreadsheet format.

3.1 DT methodological framework

The required steps are illustrated in Figure 3.1. In this chapter, the major stages of implementing a Digital Twin (DT) for scenarios where sensor data were generated inside an isolated network are discussed.

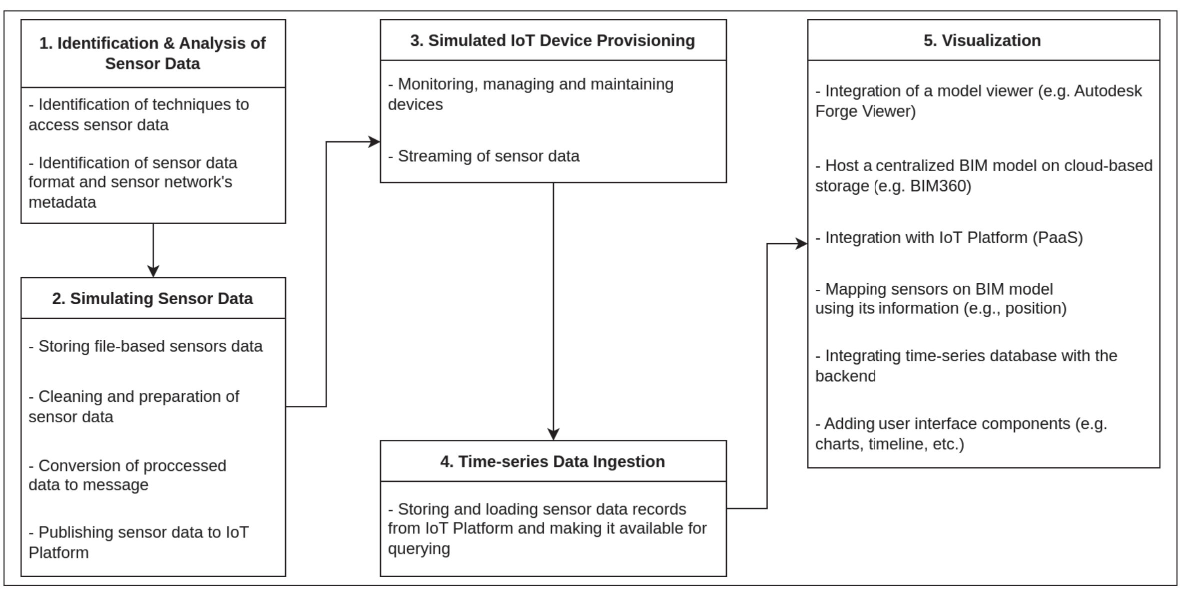


Figure 3.1 The proposed methodological framework for implementing DT, where building systems and sensors network are inside a DMZ

3.1.1 Identification and analysis of sensor data

An essential requirement for implementing a DT is to have access to constant real-time data from the physical environment which is usually collected through BAS/BMS software. Depending on the BAS/BMS software and network specifications, there are a couple of methods to acquire data from an isolated environment, such as scheduled emails, API without streaming channel, on-premise database, etc. Therefore, the sensor data format needs to be identified. For example, the data could be a response in JSON format from a request from a REST API or file-based, e.g. excel files from emails, or a result of a database query e.g., result-set of a SQL query in table format. Other metadata of the sensor devices of this specific network (e.g., the types or locations of the sensors) should be identified as well. The initial task is to identify the appropriate procedure to gain access to sensor data, based on the organization's BAS/BMS software and infrastructure. The focus of this study is on the situation that the sensor data can only be retrieved through daily basis scheduled emails. After identifying the sensor data and determining the technique to gain access to it (e.g., file-based format from emails), the next step is to simulate the sensor network based on the acquired data, which is discussed in the next section.

3.1.2 Simulating sensor data

As mentioned in the previous section, DT requires up-to-date and real-time information about the physical environment. Due to the type of data acquisition in such use cases, where direct access to the sensor data is not feasible and collected data are in batches or file-based format, it is required to simulate the streaming of sensor data from the records that are collected based on the selected methods (e.g., APIs or emails) in the previous step. The current study focuses on batches in the spreadsheet format. To simulate sensor data multiple serverless functions are developed to collect, clean, and process the data. The flow is initiated once the mail server receives an email and:

1. A serverless function (Function A) is triggered to remove the HTML from the email body and store the attached spreadsheet files;

2. Once the generated output (spreadsheet file) is downloaded to the storage, another function (Function B) is invoked to clean and split the file into multiple files for each sensor;
3. The output of the previous function triggers another function (Function C - Message Producer) to read the records, create a structured message, modify the timestamp and send messages to a queue service to be published based on the message timestamp. The term of Message is used when some data is sent directly to a specific component;
4. A function (Function D - Message Consumer) is immediately executed to consume the message, as soon as a message is published as an input of this function. The function is then responsible for registering the device and sending the message as its readings to the IoT Platform (PaaS), which is explained in the next section.

3.1.3 Simulated IoT device provisioning

A DT requires bi-directional communication with the physical device. To achieve this functionality, and facilitate the monitoring, managing, and maintaining of the devices (IoT device provisioning), a cloud-based IoT Platform (PaaS), called Azure IoT Hub, is utilized. The output of the previous step, which is the simulated devices, needs to be registered on the IoT Platform (PaaS). The aforementioned Function D registers the device based on the unique identifier inside the body of the message (If it was not already registered); sends the message, which includes the device measurement and other metadata; to the IoT Platform (PaaS). Utilizing serverless architecture provides the capabilities of parallel and auto-scalable execution of these functions, so that numerous devices can be simulated at the same time. Moreover, The IoT Platform (PaaS) provides a Device Twin to store the metadata (such as model ID, room name, coordinates, etc.) which can further be used in the visualization part. After the device provisioning is completed and streaming of the sensory data is initiated, in order to have access to the historical data, the data needs to be stored in a database. The process of storing the data is demonstrated in the next key step of this framework.

3.1.4 Storing time-series data

One of the necessary functionalities of a DT is to provide access to historical data of the built environment to the end-user. This aggregated data can be used for visual analytics and prediction depending on the type of information. Due to the time-series-based format of the sensory data, the produced data from the previous step is required to be stored in a time-series database. For the sake of rapid development and easier integration with the Azure IoT Hub (IoT PaaS), this study deploys a cloud-based time-series database called Time Series Insights (TSI), which is also one of the Microsoft Azure services.

3.1.5 Visualization

The visualization part of a DT brings all of its essential components together to provide a rich digital representation of the physical environment. To achieve this enriched visualization it is required to have access to the most recent BIM model of building, and integrate it with the real-time sensor data. The following tasks are conducted to accomplish this goal:

- To visualize the BIM model in a web environment, a web-based model viewer is required. In this case study Autodesk Forge Viewer is integrated into this DT platform as the model viewer;
- A centralized BIM model is needed, so that all stakeholders can access an up-to-date BIM model. Therefore, the DT platform is integrated with a cloud-based collaborative tool. In this case study, Autodesk BIM 360 (part of the Autodesk Construction Cloud platform) is deployed. Therefore, any future updates will be applied on a centralized model;
- The already deployed IoT Platform (PaaS) is integrated with the server-side of the visualization service;
- The registered sensors on the IoT Platform (PaaS) are automatically mapped to the BIM model using the provided device model ID and the sensors' location information, which is retrieved from metadata of devices (the aforementioned Device Twin in step 3 of this methodological framework contains this metadata);

- To provide access and visualize historical data on the BIM model, the previously deployed time-series database is integrated with the backend;
- Some visualization components such as charts, timeline, calendar, animated heatmap, etc. are implemented on the client side. The purpose is to facilitate better insights by integrating aggregated and streamed data with the BIM model to create a rich visualization for the end-user.

3.2 DT software architecture

An instantiation of the methodological framework above is a software architecture illustrated in Figure 3.2. The figure shows an abstract version of the architecture and in the next chapter, the components of it are explained in detail when the framework is applied to a use case.

The "Building System" on the left side shows the first step of the proposed framework, which refers to analyzing the building systems including BAS/BMS and sensors network to identify the the available techniques for acquiring data from the isolated environment. The focus of this research is on the scenarios that only scheduled emails is possible for data retrieval. Therefore, the "Mail Server" is added to the architecture for receiving the emails. The "Serverless Computing" is designed to achieve the second step of the framework "Simulating Sensors Data". This component consists of a sequence of serverless functions to perform the following tasks respectively: A) Storing file-based sensors data to a object-based storage on the "Data Layer"; B) Data preparation and cleansing, and storing them to the storage; C) Converting the data to message and publish them to a queue on the "Message Broker"; and D) Consuming the message as soon as it leaves the queue, then registering the simulated device and sending the message to the "IoT Platform". The "IoT Platform" is utilized in this architecture to facilitate device provisioning and to meet the requirements of the third step of the framework. According to the fourth step of the framework, the streaming time-series data is required to be stored for data aggregation purposes and further analysis. Therefore, a time-series database is added to the "Data Layer". The "User Interface" and "Web Apps Container" components are designed from the "Visualization" step of the framework. This step requires an integration with BIM model storage, for collaboration and visualization purposes. The “BIM model storage” is included on “Data Layer” component to meet these requirements. These components are also integrated with “IoT Platform”, and “Time-series Database” for visualizing streaming data and analytics dashboards.

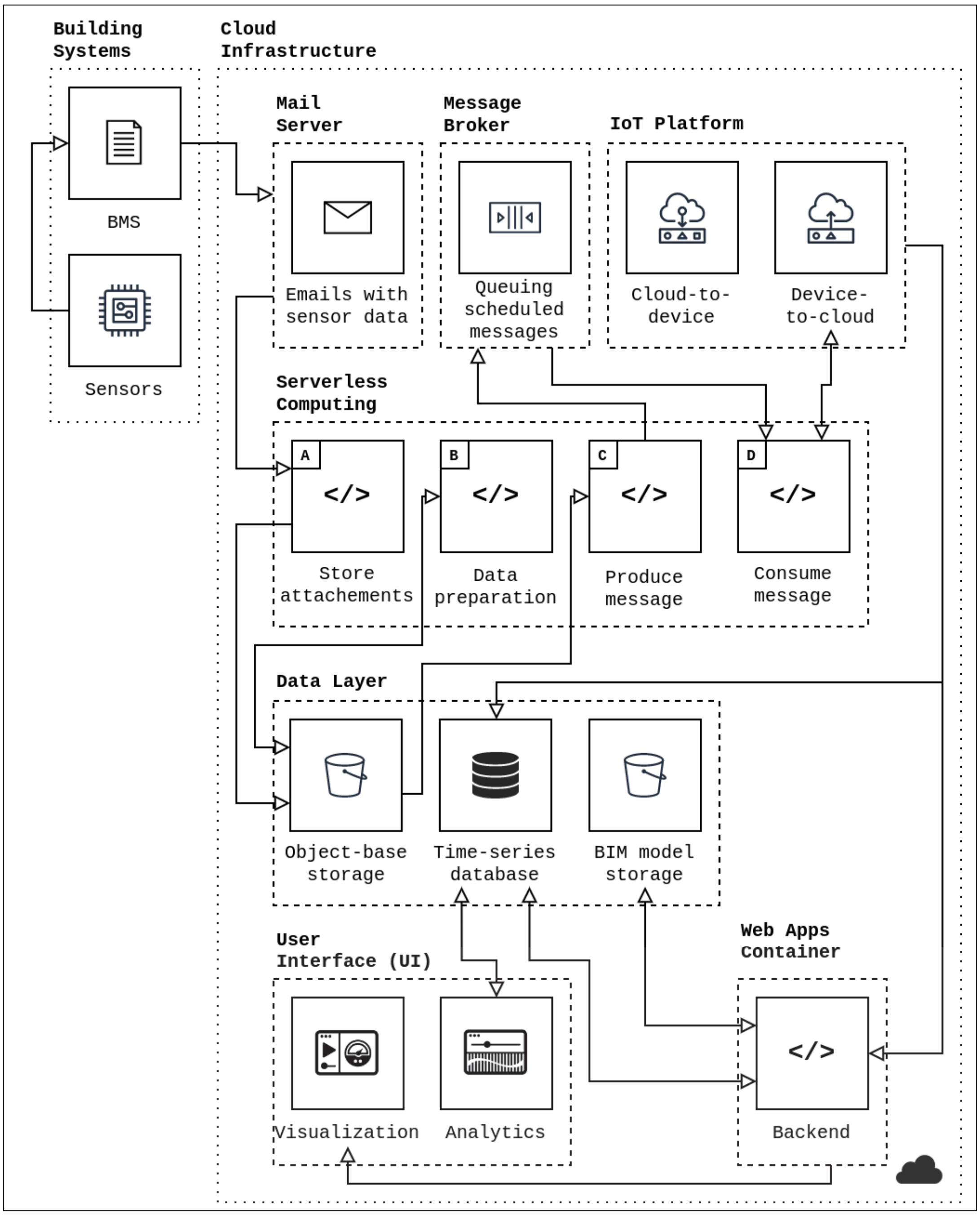


Figure 3.2 The proposed software architecture for DT platform

In the next chapter, in order to validate the designed methodological framework above, it is implemented in a case study of a community center in Quebec City.

CHAPTER 4

CASE STUDY: BARDY COMMUNITY CENTER

In this chapter, the proposed methodological framework and its associated software architecture which are developed as the artifacts of this research in the previous chapter, are demonstrated by implementing a real-world case study as a proof-of-concept process. Moreover, in the final section of the chapter, an initial visual analysis is done to time-series data extracted from the case study to evaluate the artifacts.

The case study is implemented on Bardy Community Center, which is one of the *Ville de Québec* (City of Quebec) buildings and located in Quebec City. The building is already equipped with more than 200 sensors, and the sensor data is being collected using the BAS/BMS software "enteliWeb" inside an isolated network. The goal is to implement a web-based Digital Twin (DT) for this community center that integrates indoor temperature data with the BIM model and creates a visualization including a heat map. The building systems are considered isolated because direct access to this system cannot be provided, due to security concerns, i.e, streaming data directly from the BAS/BMS system is not an option. Due to certain technical limitations imposed by the Community Center itself, the use of API is not feasible either.

In the following chapter the steps of the proposed methodological framework explained in chapter 3, and their results are described thoroughly: 1) Identification and analysis of sensor data; 2) Simulating sensor data; 3) Provisioning of simulated IoT device; 4) Time-series data ingestion; and 5) Visualization.

There are a number of assumptions that delimit the case here. The BAS/BMS contains a series of subsystems with a large number of sensors and actuators. The methodological framework implemented here uses the collected data and list of devices from the enteliWeb BAS software of Bardy Community Center. However, the process of setting up a BAS/BMS system, implementing a sensor network, and data collection are not within the scope of this thesis. The BIM model that is used for the visualization part is provided by Ville de Quebec (Municipality of Quebec) and all of the floors, rooms, and zones that were required for mapping the devices into the BIM model are already defined. The process of designing and improving the quality of the BIM model is out of the scope of this research.

The software architecture of the DT platform is illustrated in Figure 4.1. This architecture was adapted from the generic architecture (Figure 3.2) which was elaborated in chapter 3. The following steps show how the components were adapted from the proposed generic architecture to follow the proposed framework in this thesis: Step 1, on the “Building Systems” the offline file-based sensor data were exported automatically in Excel (.xls) and scheduled to be sent to an external mail server “Outlook”; Step 2, A workflow was designed and implemented in Microsoft Azure Logic Apps (“Logic Apps” component) to automate the process of storing the attached Excel files. A sequence of “Azure Functions” as the “Serverless Computing” to perform several specific tasks, “Azure Blob Storage” as the object-based storage, and “Azure Service Bus” as the “Message Broker” were included for this architecture to implement the case study; Step 3, “Azure IoT Hub” as the “IoT Platform” was chosen for device provisioning; Step 4, “Azure Time Series Insight (TSI)” was utilized to store the time-series date and it was integrated with “Azure Iot Hub”. TSI also provides analytics dashboard which the user can perform custom operations on data; and Step 5, “Autodesk Construction Cloud” was integrated as the BIM model storage to facilitate collaboration and model visualization. “User Interface” and “Web Apps Container” contain the custom developed backend, which was integrated with “Azure IoT Hub”, and “Time Series Insight” for visualizing streaming data and aggregated data on the client-side of the platform. The details of each step are explained in the following sections.

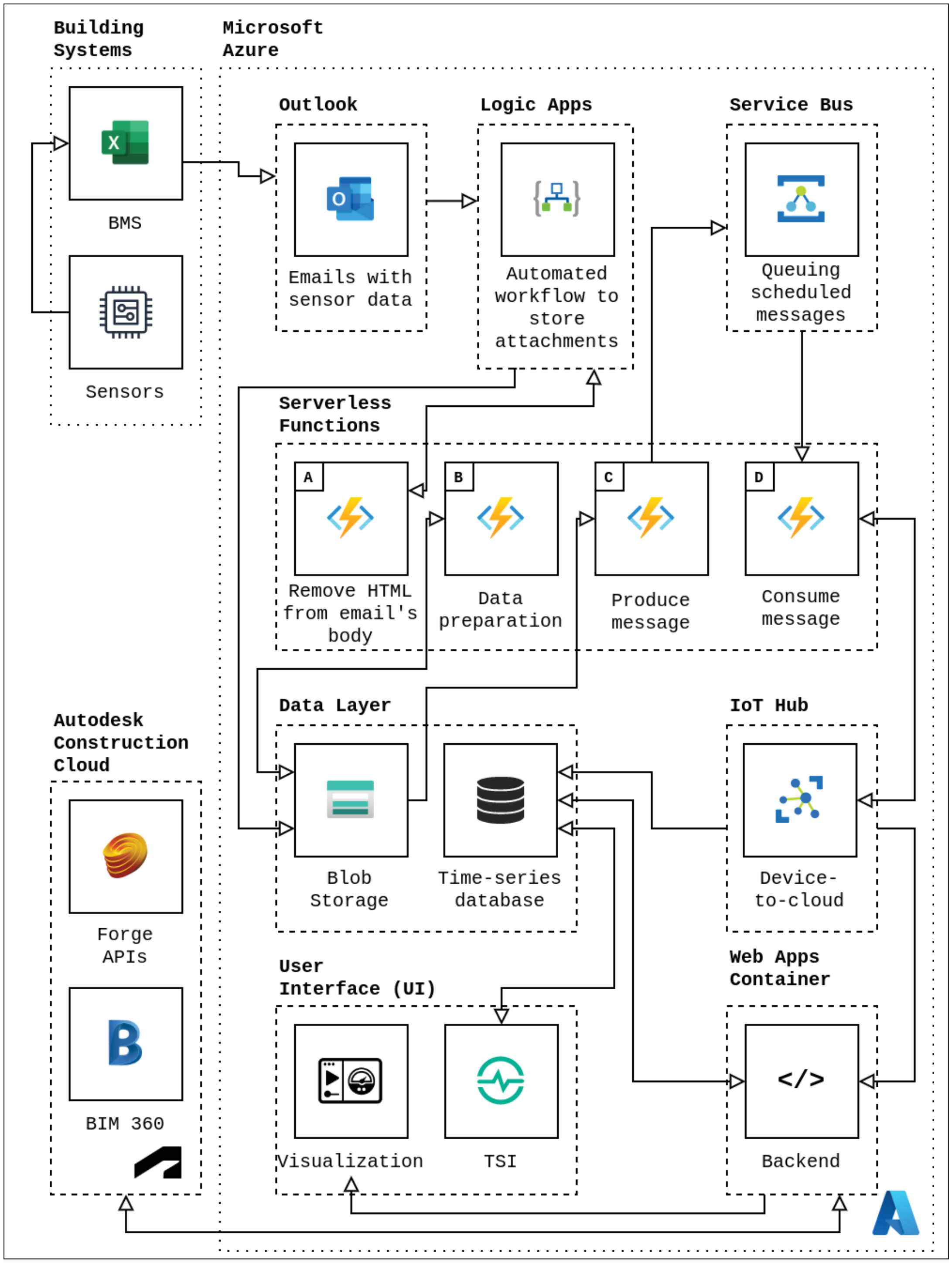


Figure 4.1 The software architecture of the implemented DT for community center case study

4.1 Identification and analysis of sensor data from the enteliWeb (BAS/BMS software)

Direct access to database was not allowed. The different existing options for acquiring data were investigated to provide access to enteliWeb data. The options were 1) through scheduled emails; 2) through developing a plug-in inside the enteliWeb to push data to an external database; 3) Through pulling the data from enteliWeb API (considering that the API had no streaming channel). Due to the sensor network data and software inaccessibility, the only option was via scheduled emails. This feature is already being provided by enteliWeb software. The list of devices and the BIM model of the building was provided by *Ville de Québec*. 54 temperature sensors were selected, and the application is configured to send the data readings of them via seven scheduled daily basis emails to an external mail server in the morning. Each email contains an attachment of an Excel file including the last 24 hours data of the selected temperature sensors with the frequency of 5 minutes. The sample excel file can be found in Figure 4.2.

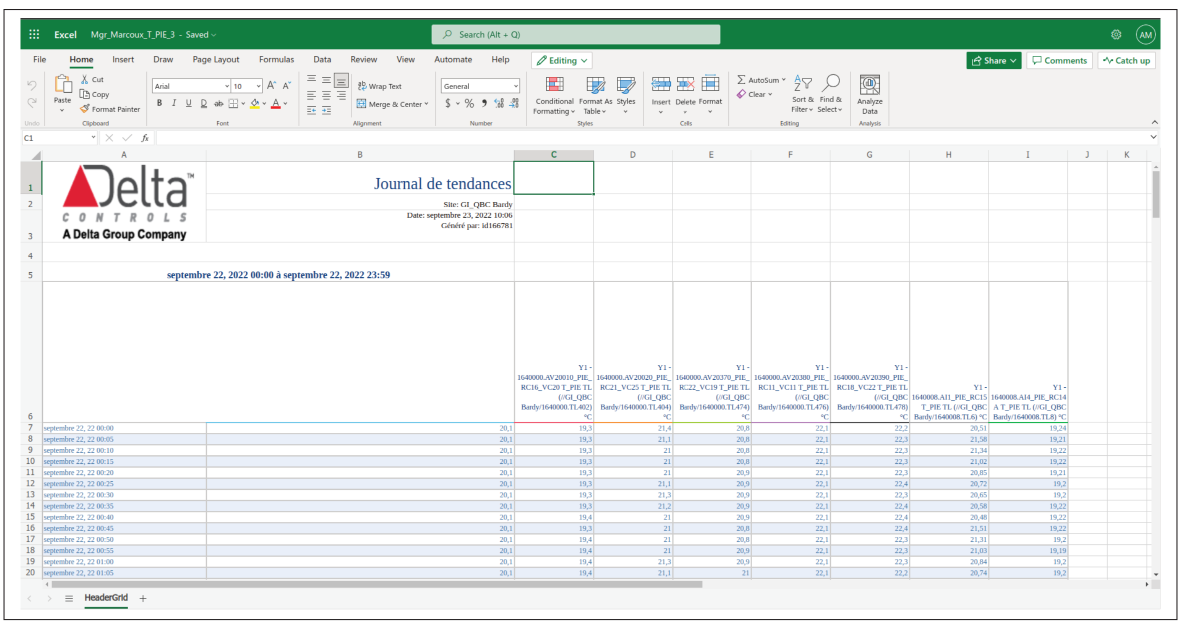


Figure 4.2 The sample Excel file of sensor data

4.2 Simulating sensor data

As mentioned previously, in this step the streaming sensor data needs to be simulated by designing a number of serverless functions that are responsible for the collection, cleaning, and processing of the data. The following flow of functions begins as soon as an email including the data in file-based batches is received:

4.2.1 Function A: Store attachments

In this step the daily-basis emails (scheduled to be sent by "enteliWeb" software) are processed to store the generated Excel files that are attached to the body of the emails. To do so, an algorithm is designed on Azure Logic Apps to reduce the development time. The overview of this algorithm can be found in Figure 4.3.

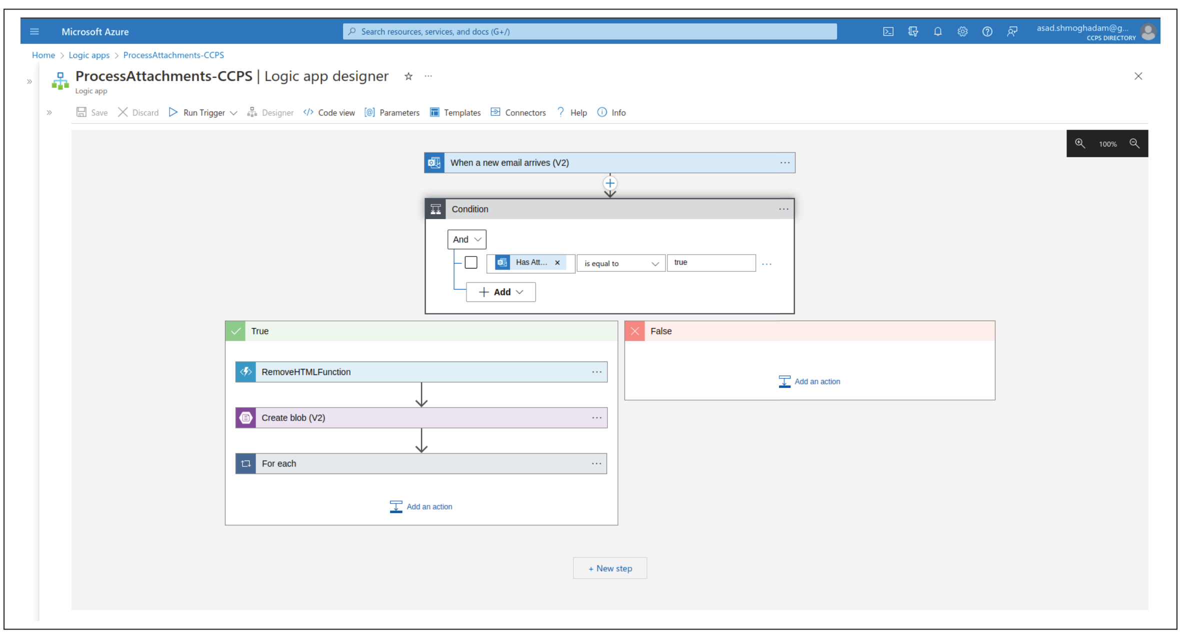


Figure 4.3 The overview of automated workflow designed in Logic App

Once the email is received from a specific domain, it triggers this algorithm. The next step is to set a condition to check if the email contains any attachments. If the condition is true (Figure

4.4), it executes Function A[1] with the email as its input. Function A removes the HTML and unnecessary characters from the body of the email. The output of Function A is a clean email, and it invokes a built-in Azure function to save the cleaned email as a Blob (Binary large object) in Azure Blob Storage (which is already configured).

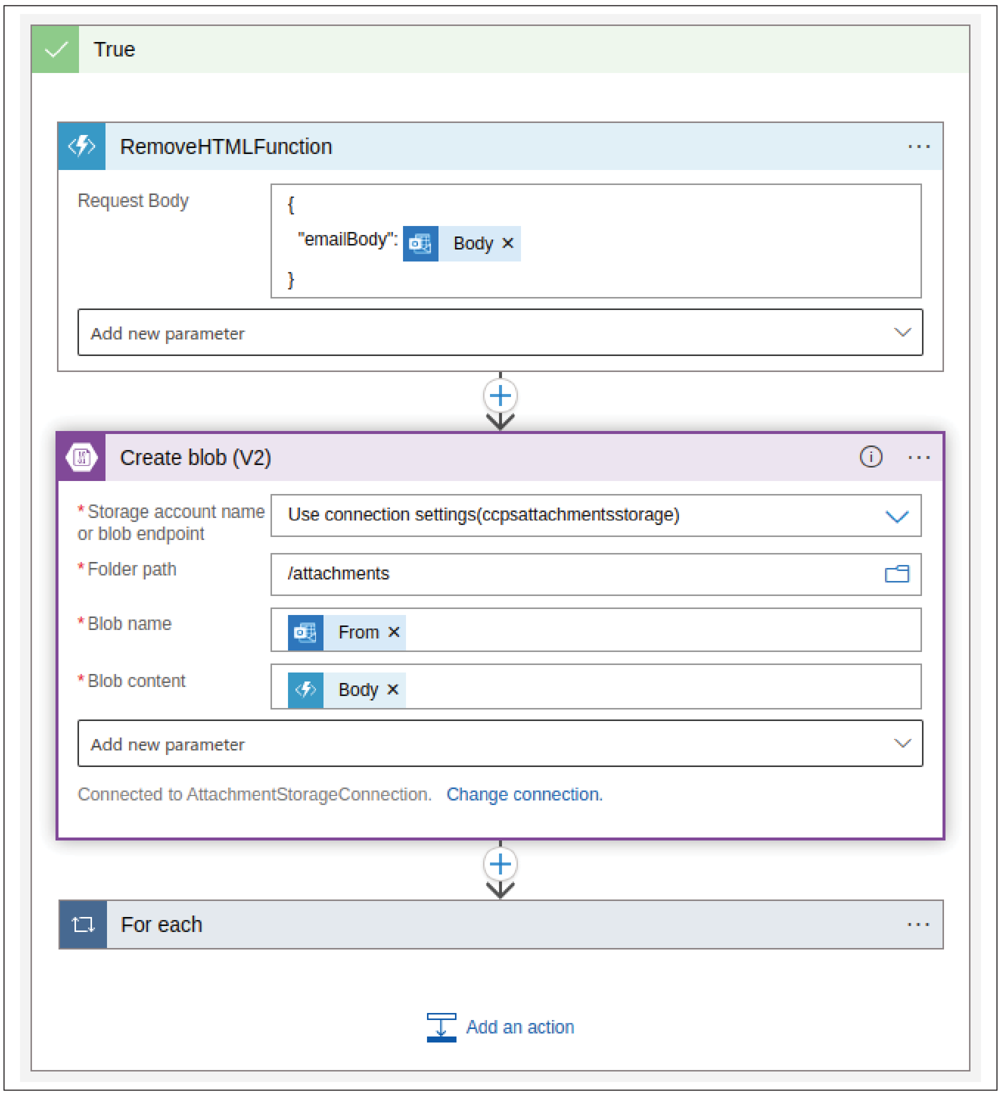


Figure 4.4 The workflow that executes Function A for removing HTML from the body of the email, and stores the cleaned email as a blob

[1] Function A - is developed with C# script and shared openly on the GitHub: https://github.com/asad-shmoghadam/az-func-a-remove-html.git

Once the email is saved as a Blob, the flow will run a for-loop over the attachments of the email. The next step is to check if the file is in the Excel format (.xls) to avoid storing unnecessary files, and if the attachment meets the condition, save it as a Blob in a directory called "/attachments" by utilizing the Azure built-in function “Create Blob”. A screenshot of this for loop can be found in Figure 4.5.

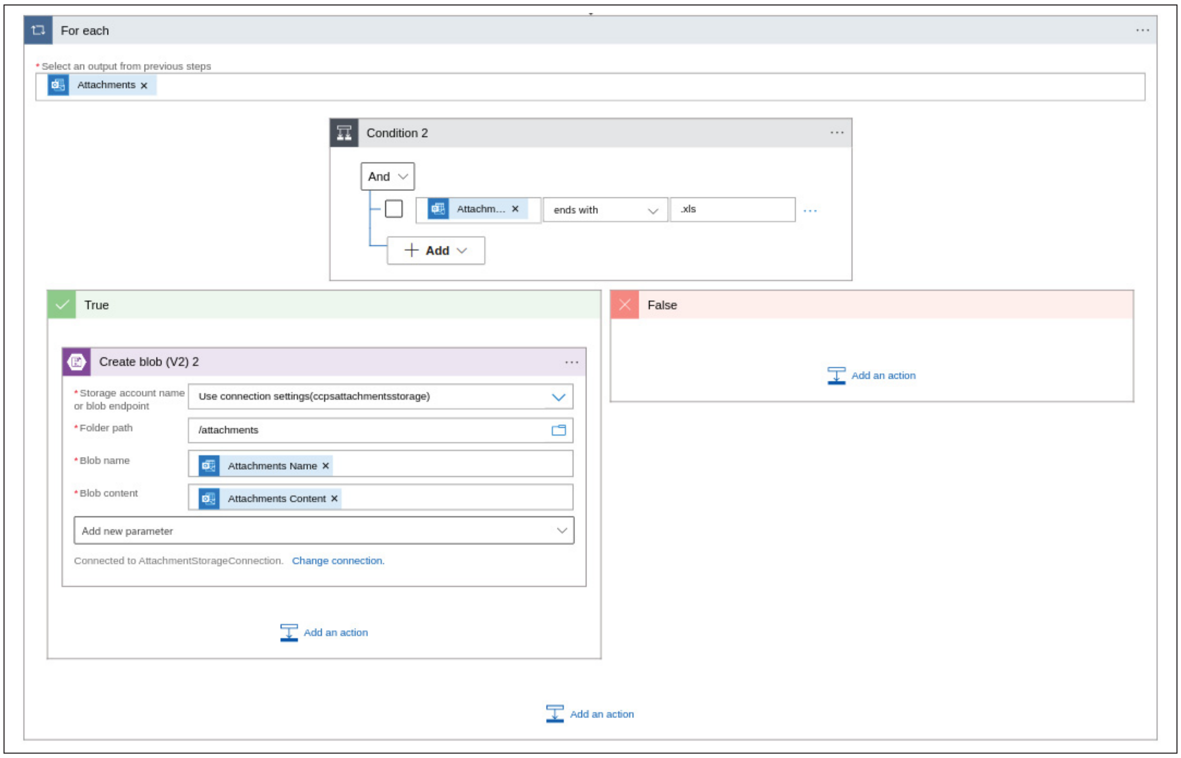


Figure 4.5 The workflow for storing the attached Excel files

Utilizing logic apps provides parallel computing, so all of the seven emails can be processed at the same time, and Logic Apps run the flow for each email separately. These attachments will be overwritten every day when the new emails arrive. The author has shared the designed workflow that is developed on Azure Logic App on Github[2]. The result of this step is seven Excel files, which are the inputs for the next function.

[2] https://github.com/asad-shmoghadam/logicapp-attachments-to-blob.git

4.2.2 Function B: Data preparation

This function, according to the proposed methodological framework, is for processing the Excel files, extracting the data of each sensor from them, and storing them in a separate CSV file. Function B is developed with Python and the code is shared by the author[3]. This function looks for the .xls file inside the attachments folder ("/attachments/{name}.xls"). Once an Excel file is downloaded by Logic Apps, in the attachments directory, Function B is triggered. Function B processes the Excel file, extracts the required information, and generates the CSV file for each sensor. The output of Function B, which is a CSV file will be saved in “/sensorsdata” directory (Azure Blob Storage). The sample generated CSV file is shown in Figure 4.6.

The sample Function B execution count logs are shown in Figure 4.7. Figure 4.7 shows that this function is executed seven times (number of stored excel files) each day. The outputs of this function are 54 CSV files containing the time-series data of 54 temperature sensors for the last 24 hours, which will be the inputs for the next function.

4.2.3 Function C: Produce message

In this stage the generated CSV files from the previous function will be processed. Function C is developed with C# and is shared on Github[4]. Function C watches the directory "/sensorsdata/{name}" and each time a CSV file is created this function is invoked, and starts to read the CSV file line by line and structures a message[5] (Figure 4.6), and adds 1 day to the original timestamp to compute the new timestamp of the message. Then, the message including all the metadata is sent to Azure Service Bus (Queue Service). Each message contains the following fields: sensor_id, sensor_name, sensor_type, org_timestamp (original timestamp). Azure Service Bus has a feature to schedule messages to be released and published at a specific time. Function C schedules the message based on the modified timestamp to be released and

[3] https://github.com/asad-shmoghadam/az-func-b-etl.git

[4] https://github.com/asad-shmoghadam/az-func-c-publish-message.git

[5] The term of Message is when some data is sent directly to a specific component.

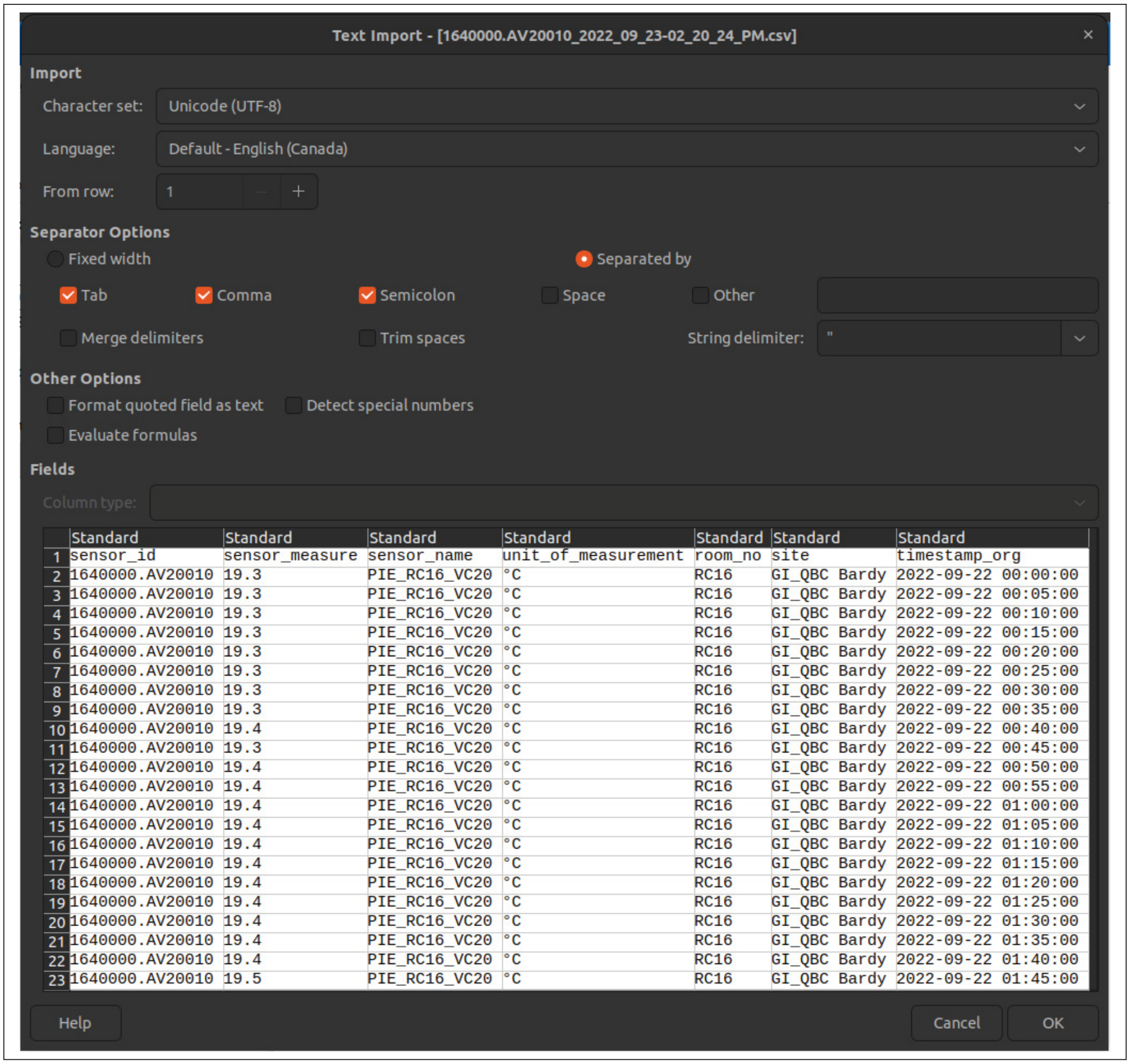


Figure 4.6 The sample of the generated CSV file that contains sensor data

published on the current day (the original timestamp incremented by 24 hours), and then adds it to the queue.

The sample outgoing message logs of one hour is shown in Figure 4.8. As it was mentioned before the serverless function provides parallel computing, so all CSV files can be processed at once. As it is shown in Figure 4.8, every 5 min, the quantity of published messages is equal to the number of temperature sensors. These messages are published to a queue that will be consumed by Function D.

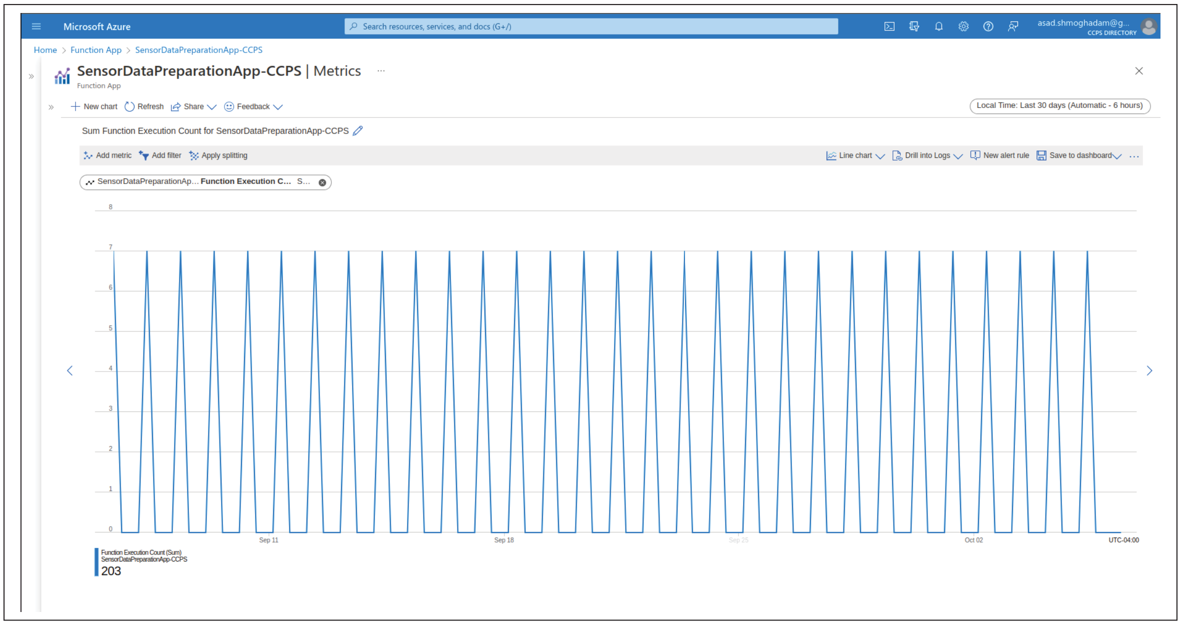


Figure 4.7 The execution count logs of Function B

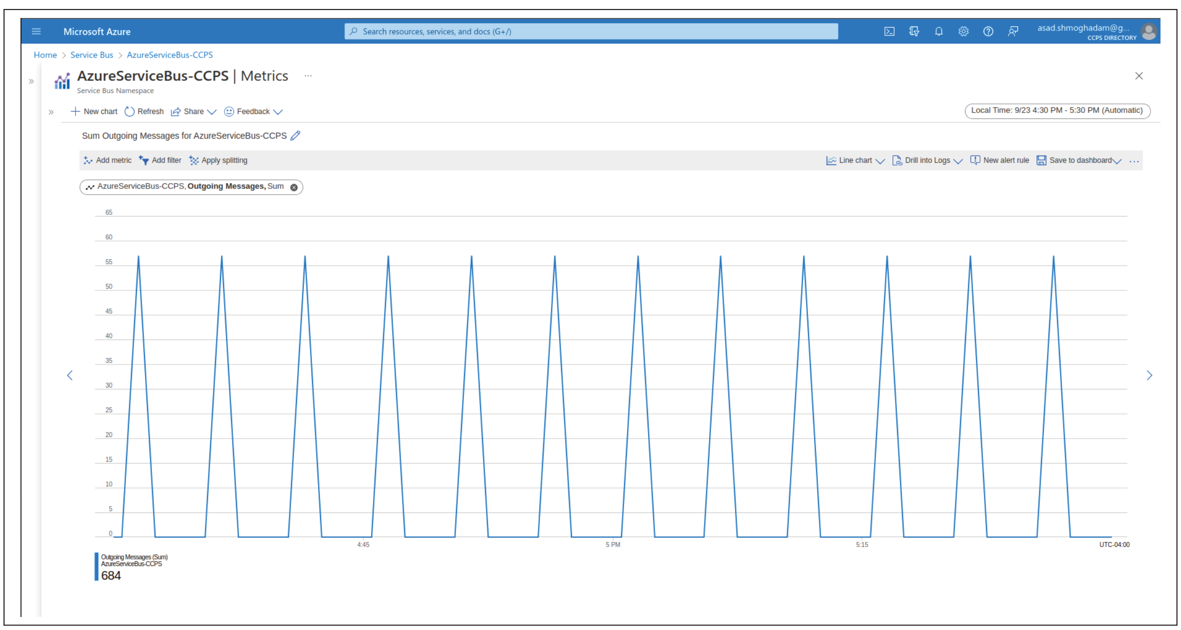


Figure 4.8 The number of outgoing message logs of Azure Service Bus

In the next step the function responsible for the processing of reading the messages, device registration, and publishing of these records to the Azure IoT hub (IoT Platform - PaaS) is discussed.

4.2.4 Function D: Consume message

The scheduled messages which were discussed previously, are now published to a queue, and Function D as its consumer triggers and receives the message as its input. Function D has been developed in C# and the codes are shared on GitHub[6]. Function D processes the message and checks if the device is already registered, and if not it registers it as a new device on Azure IoT Hub based on the provided "sensor_id" in the message. Thereafter, Function D sends the message as a sensor reading to Azure IoT Hub. Figure 4.9 shows the execution count of Function D, in a range of 1 hour and an interval of 5 minutes. The execution count is the same as the number of sensors that have been received in Excel files via emails.

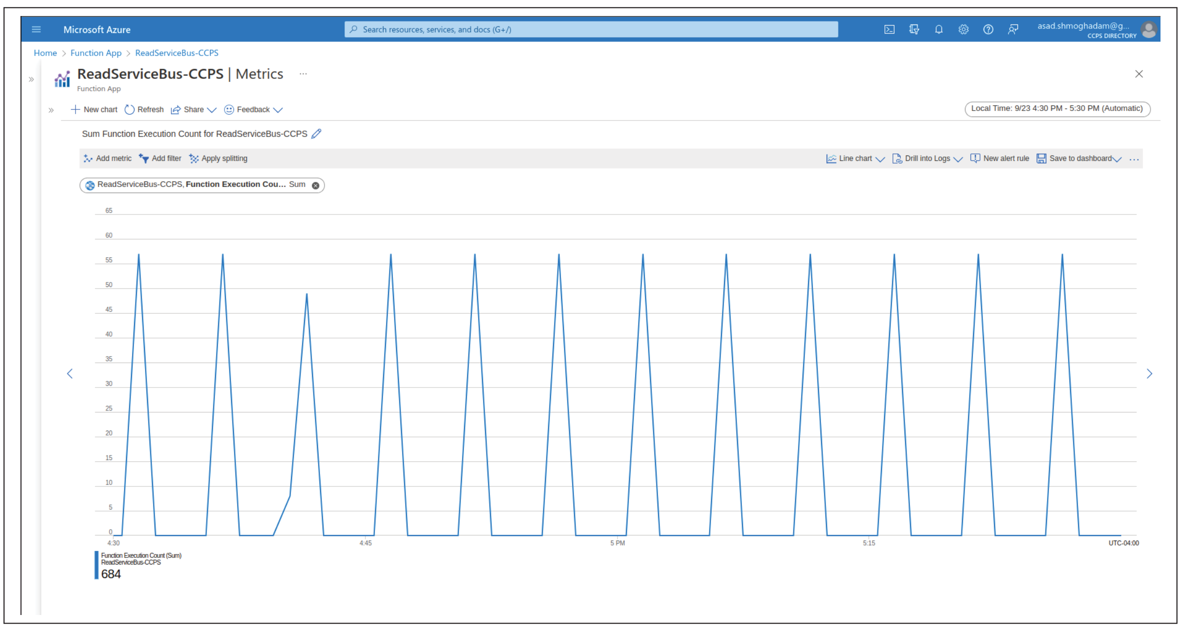


Figure 4.9 The execution count logs of Function D

In the next section, the Azure IoT Hub and how Function D is integrated with it, are explained in more detail.

---

6 https://github.com/asad-shmoghadam/az-func-d-sim-sensor.git

4.3 Simulated IoT device provisioning

Function D communicates with Azure IoT Hub using a defined "Shared access policies" (Figure 4.10) and connects to the Azure IoT hub through the SDK (Software Development Kit). Authorization happens via a connection string, so that Function D has an access to read and write data to Azure IoT Hub. Function D registers devices on the Azure IoT Hub.

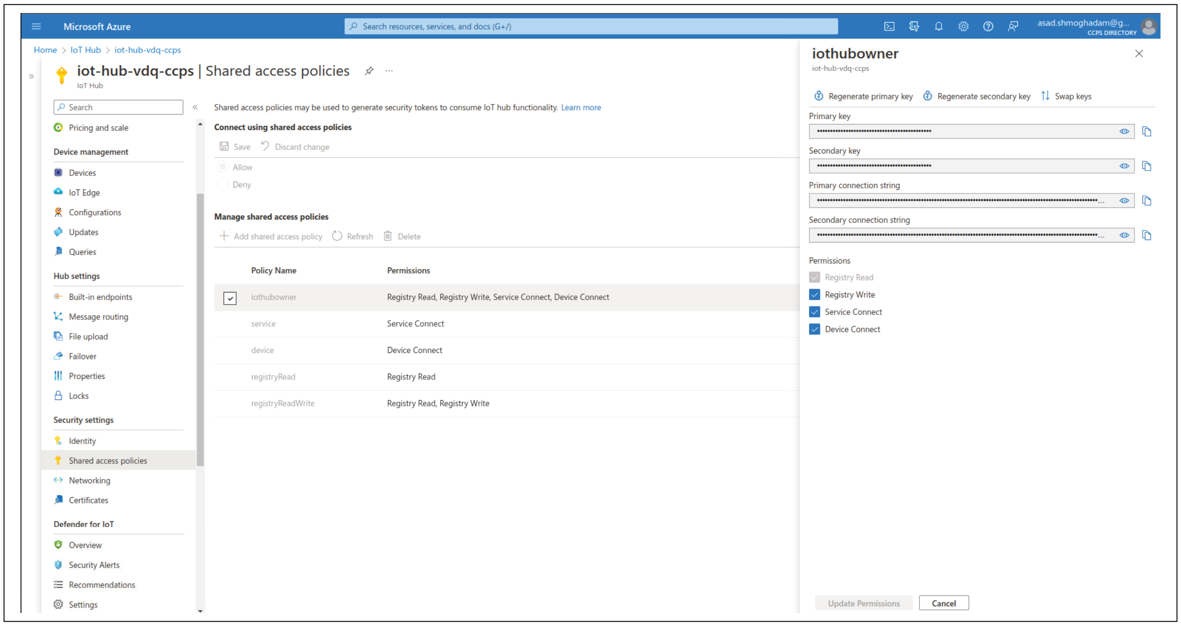


Figure 4.10 The Shared Access Policy that grants permissions to access IoT Hub

Azure IoT Hub gives the capability of bi-directional communication with the device. Also, it creates a "Device Twin" for each of the registered devices. The device twin contains metadata of the device and custom metadata can be added to it. Some custom tags such as model ID, position of the sensor (x, y, z coordinates), etc. are added to Device Twin in order to map the sensors in the visualization section. A sample Device Twin is shown in Figure 4.11.

Figure 4.12 shows the number of connected devices in 24 hours. Figure 4.13 shows the number of messages sent to Azure IoT Hub (device-to-cloud messages). The jump in chart in Figure 4.13 (approximately 1k messages) is during the morning when the emails are received. The messages that their timestamp has already passed, will be sent all at once. The Azure IoT Hub will continue to receive the messages as usual at the specified timestamp. The number of received messages is equal to the number of connected sensors. Figure 4.14 shows the total number of messages for all devices in 24 hours, which is approximately 17k messages per day.

Now that the simulated devices are connected to the IoT Platform (Azure IoT Hub) and the readings are streaming, in order to store these records, it is needed to integrate the IoT Platform (PaaS) with a database in order to have access to the historical data for further analysis. In the

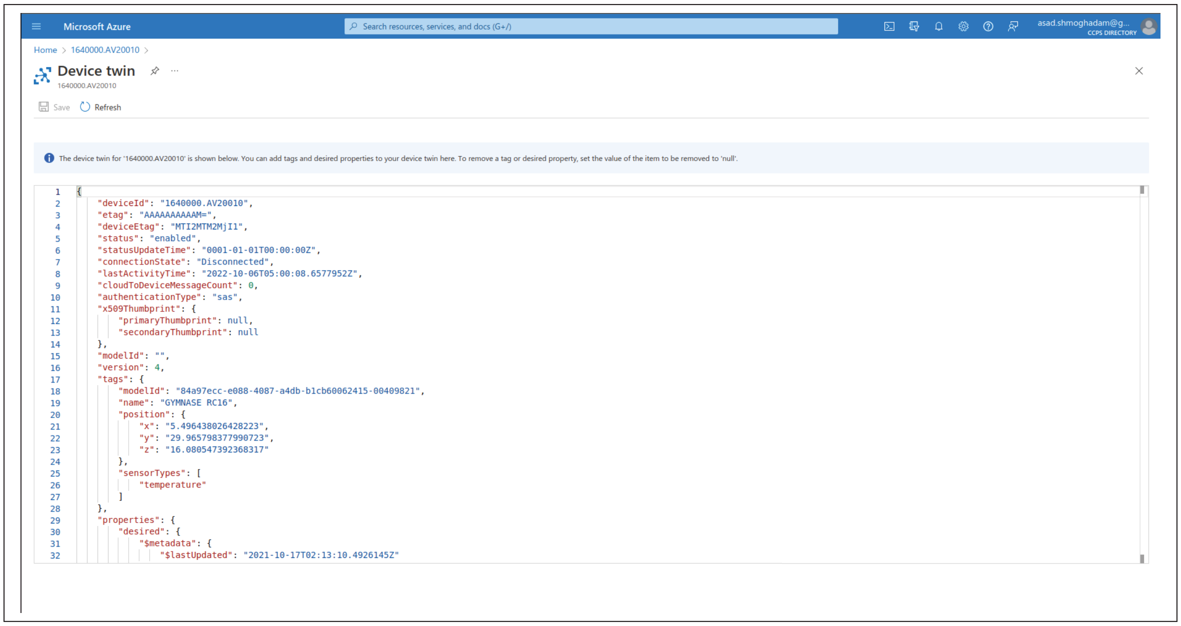


Figure 4.11 The Device Twin of a sensor

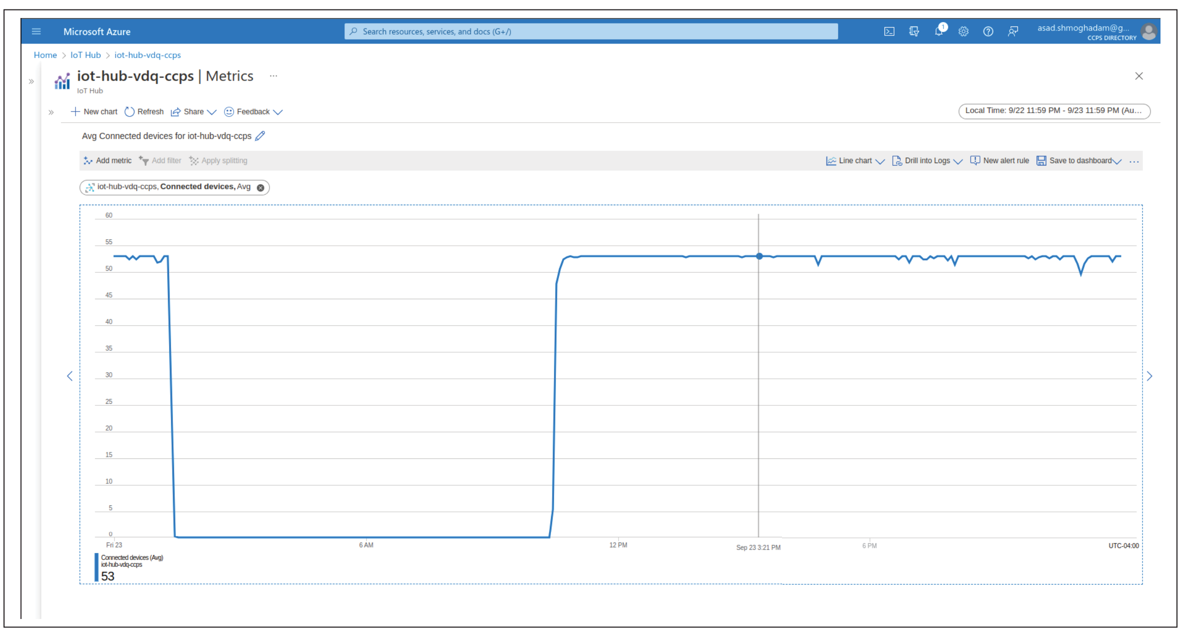


Figure 4.12 The number of connected devices to IoT Hub over 24 hours

next step, the process of integrating Azure IoT Hub with a database, and storing the messages are described.

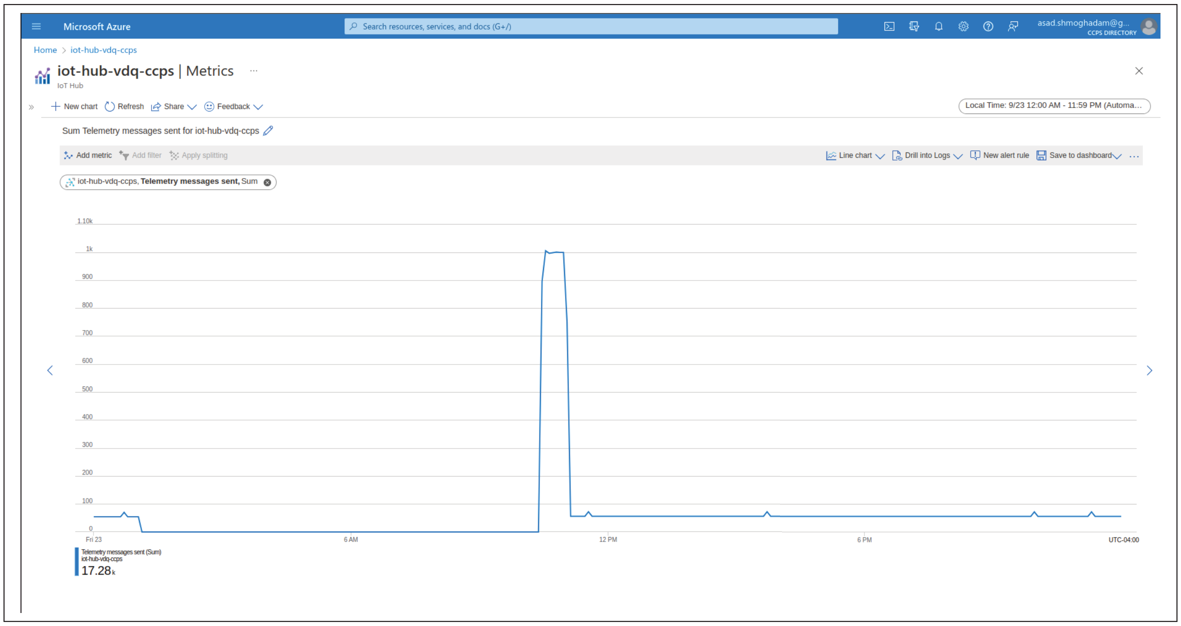


Figure 4.13 The number of device-to-cloud messages over 24 hours

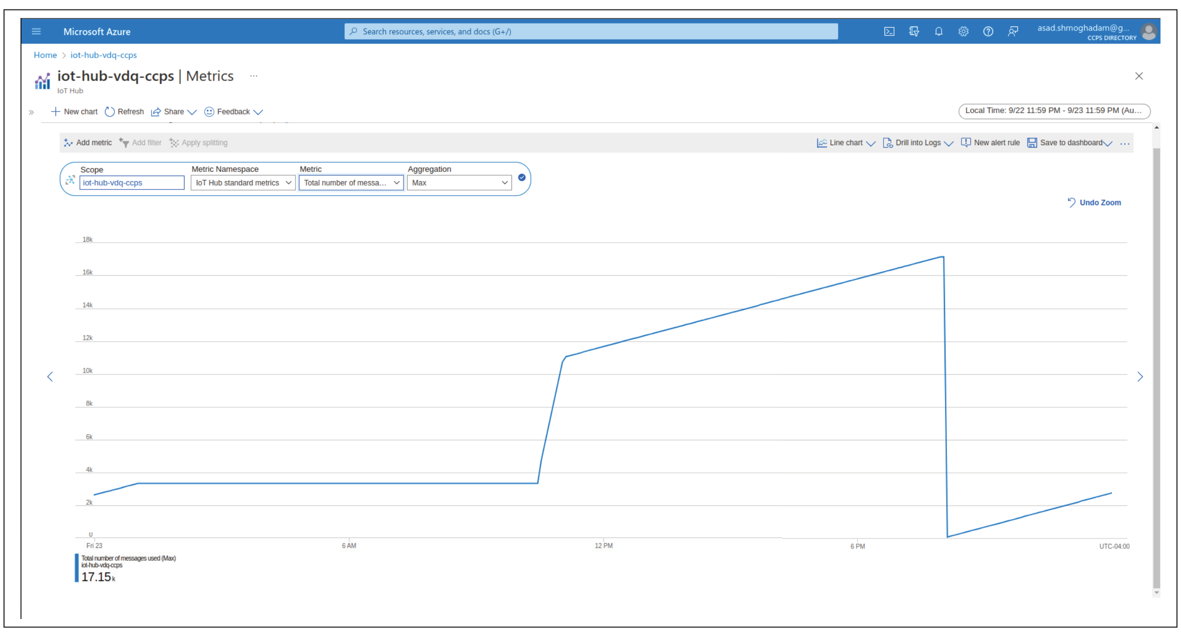


Figure 4.14 The total number of messages for all devices over 24 hours

4.4 Storing time-series data

To store the time-series data of sensors, the Azure Time Series Insight, which is the time-series database, is deployed. The reason that this service is utilized is that TSI can be easily integrated with Azure IoT Hub, and connect Azure IoT Hub endpoint as its data source, so the sensor data can be streamed directly to TSI without any coding. The source of the events (Azure IoT Hub) that TSI is subscribed to require to be configured in TSI panel (it calls Event Sources). Moreover, in the Azure portal, it is required to define the "Timestamp property name" field as "Timestamp" (the timestamp that was generated and assigned to the message in Function C), otherwise, the TSI will use the timestamp at which the event[7] was consumed by TSI, not the one that we customized. For example, the sensor data between 00:00 AM to the time that the emails are received (around 10:00 AM) in the morning, will be consumed all at once by TSI If the new Timestamp property is not configured. Basically, it tells the service to use that field for storing the data with that specific timestamp.

TSI provides an enriched visualization of the time-series data. Figure 4.15 shows an example of the collected data of two temperature sensors (X-axis is timestamp, and Y-axis is device measure).

In the next step the Azure IoT Hub and TSI will be integrated with the BIM model on the model viewer (Autodesk Forge Viewer) to map the sensors and visualize them in a single view.

4.5 Visualization

To visualize the collected data, and provide an interface for the end user, a web application is developed. The application consists of a backend developed with Nodejs, and the frontend developed with React framework[8]. The visualization steps that are also described in the proposed methodological framework in chapter 3 come next.

---

[7] The term of Event is when some data is emitted to any consumer that is listening.

[8] https://github.com/asad-shmoghadam/ccps-dt-platform.git

Figure 4.15 An example of the sensor data visualization in TSI

4.5.1 Model viewer

Autodesk Forge Viewer is utilized to visualize the BIM model of the Bardy Community center. Figure 4.16 shows the Bardy Community Center BIM model that is visualized with the Autodesk Forge Viewer. Next, the integration with the BIM 360 is discussed to provide a collaborative environment.

4.5.2 Integration with collaborative tools

DT requires an up-to-date BIM model. The BIM 360 is integrated into the application using Autodesk BIM 360 API to always have access to the latest versions of the model. Moreover, it provides a collaborative environment so that every stakeholder can work on the same model, and all modifications will be applied to a central model.

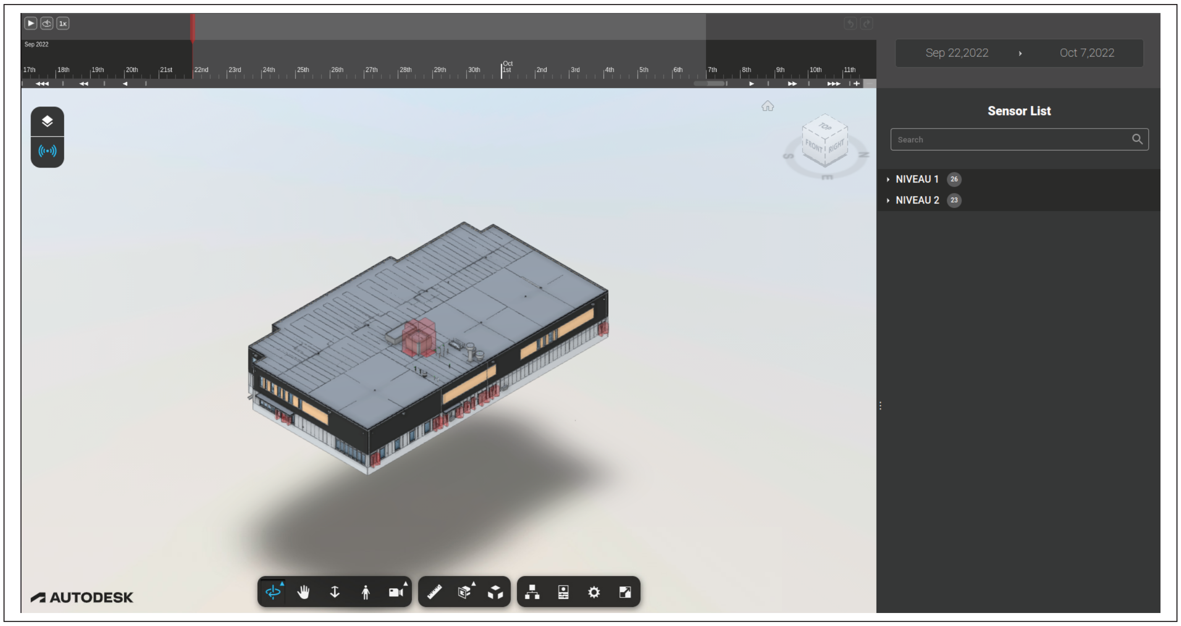


Figure 4.16 The visualization of the Bardy Community Center on the Forge Viewer

4.5.3 Integration with Azure IoT Hub

The web application is integrated with the Azure IoT Hub (developed using the provided SDK for Javascript) to have access to the latest readings from sensors and get the list of devices. It also has access to metadata such as model ID, name, and position of the devices that were added to the device twin (previously explained in section 3 of this chapter).

4.5.4 Mapping sensors

Autodesk Forge Viewer has the capability to map the sensors automatically in the room based on the device position in x, y, and z coordinates using Autodesk Forge SDKs. By providing the position information of the sensor in the device twin, the application will be able to access the coordinates of the device by querying metadata that exists in the device twin. The backend gets the list of devices with metadata (device twin), and checks if the Autodesk Forge Viewer can locate the device inside the room space coordinates based on the provided x, y, and z coordinates of the sensor. Figure 4.17 and Figure 4.18 show the mapped sensors on Floor 1 and Floor 2.

Figure 4.19 shows a sensor mapped inside the room space. If the Autodesk Forge Viewer does not find a room space that contains the coordinates, it will not map the sensor in the room.

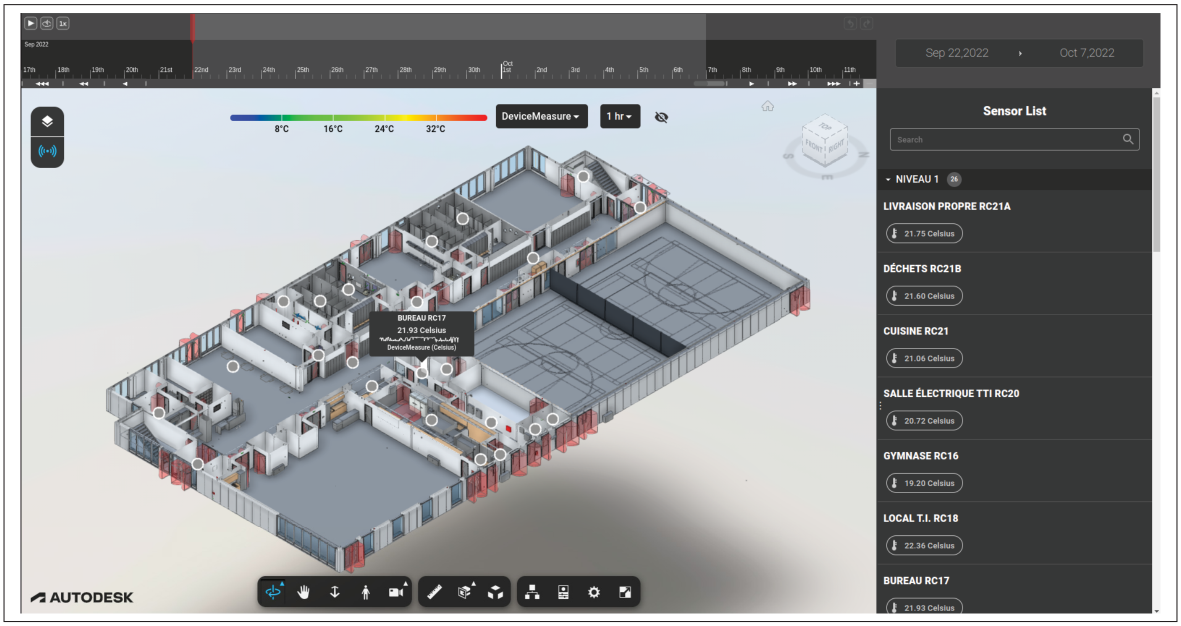


Figure 4.17 The mapped sensors of 1st floor

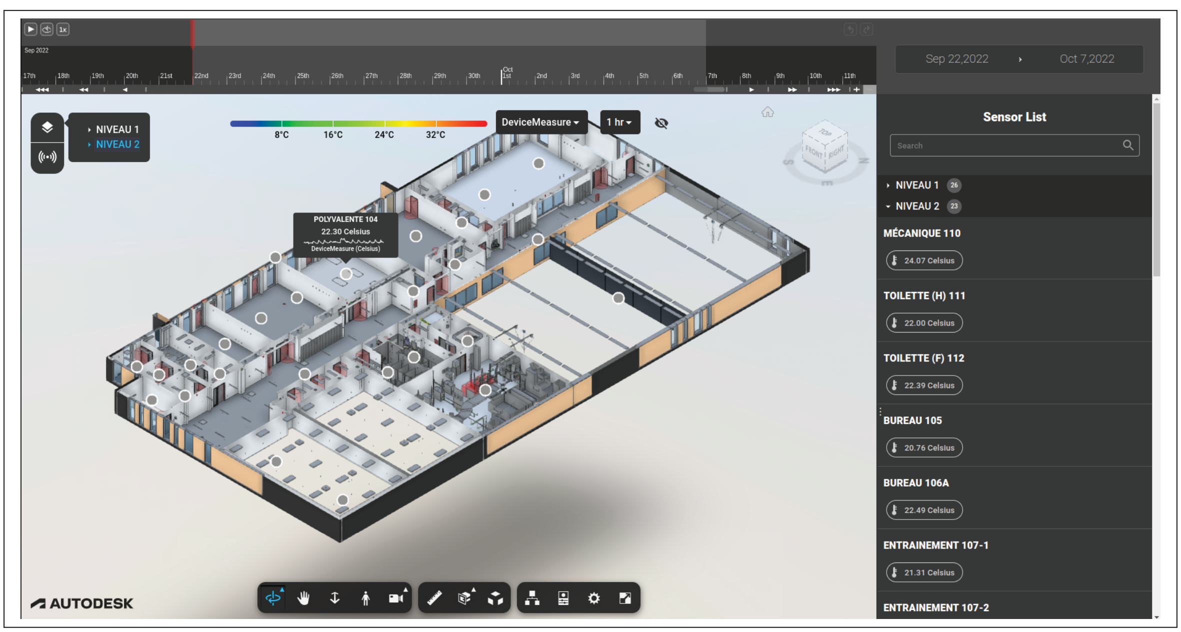


Figure 4.18 The mapped sensors of 2nd floor

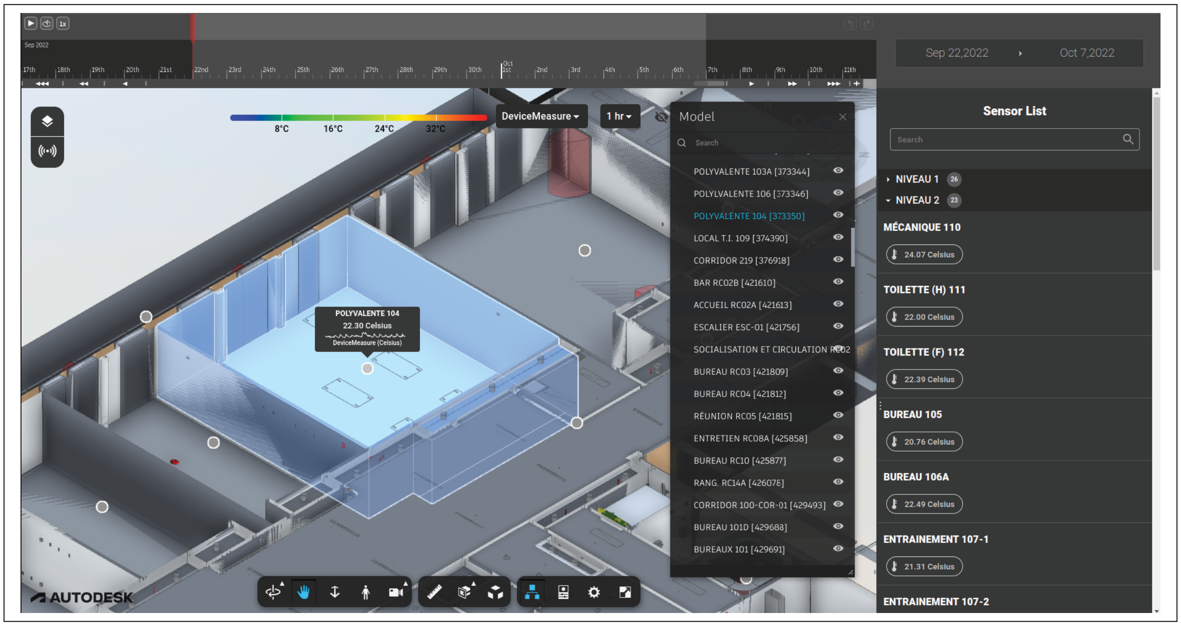


Figure 4.19   The sensor that is mapped inside a room

4.5.5 Integration with Azure Time Series Insight (TSI)

To provide historical data for the end-user, the application is integrated with TSI which is a time-series database. The backend sends a request to the TSI endpoint, which includes a range of dates and device IDs. The response contains the device measures of that range of data, which then will be visualized on a chart so that the end-user can monitor the historical data. Figure 4.20 shows the historical data of the sensor.

4.5.6 Visualization components

To provide a rich visualization for the end user, some components are implemented on the frontend. The Reactjs components (forge-dataviz-iot-react-components) are developed by Autodesk Forge to facilitate the frontend development.

The temperature data is visualized with heatmap on the model (Figure 4.21). The user can select a range of dates with a calendar, and see the heatmap of temperature data on the model. The end-user can also animate the heatmap based on the historical data through the provided

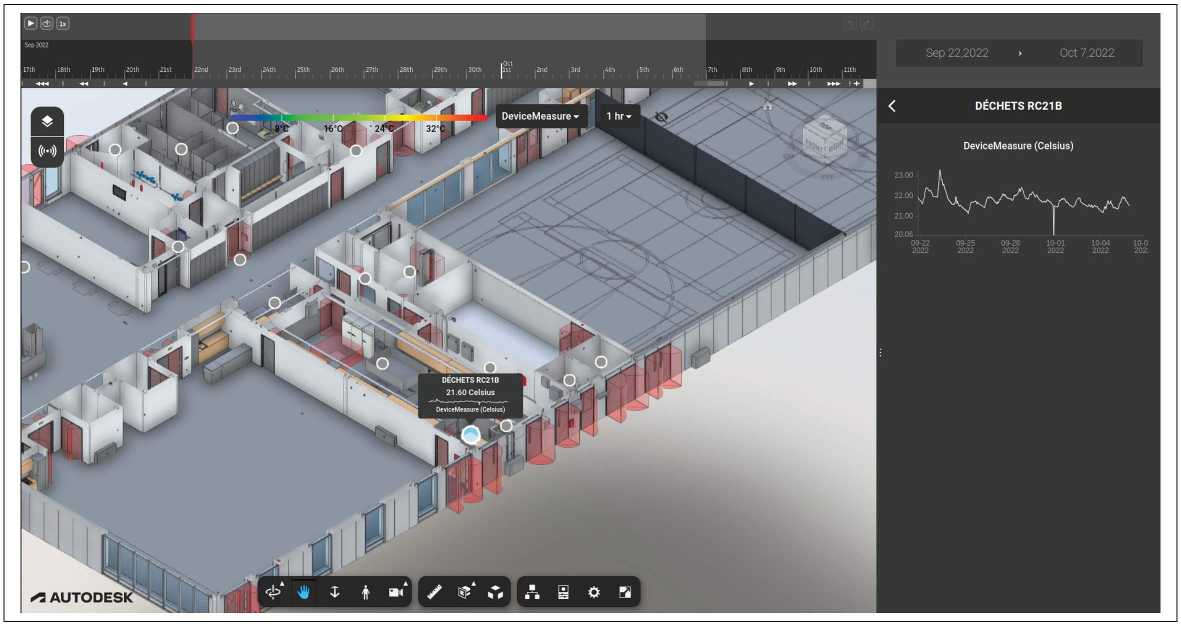


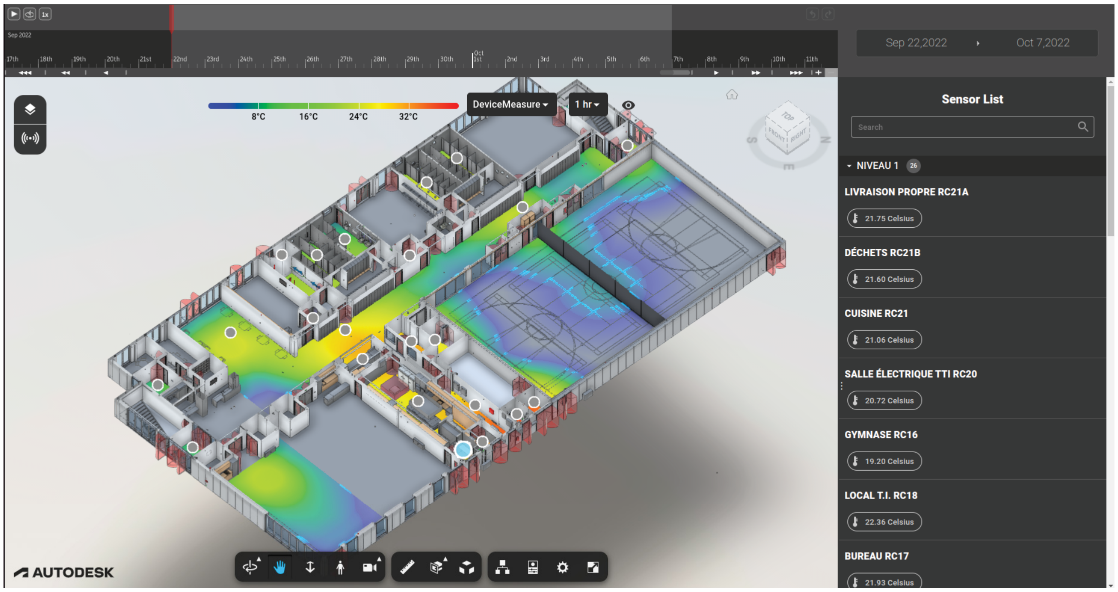
Figure 4.20 A selected sensor and its historical data on the right side panel

timeline on the top. This can be useful for visual analytics. The access to the DT platform can be granted upon request.

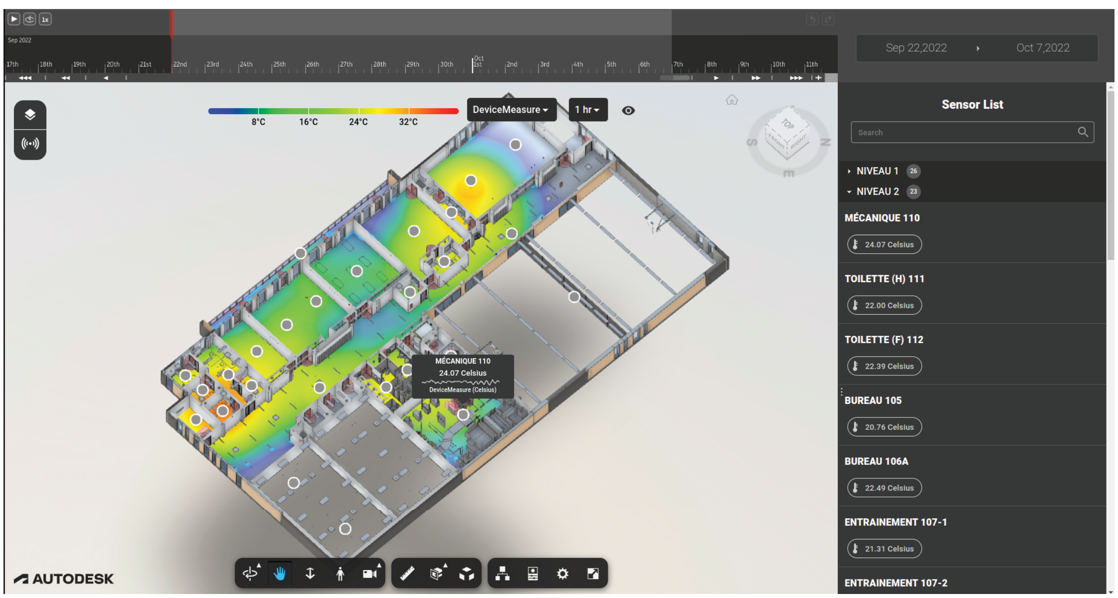
4.6 Evaluation of initial building temperature outputs

As an initial step to evaluate the proposed methodological framework and its output - the software architecture- the author decided to explore some visual analytics of the temperature time-series data, extracted from the Bardy Community Center building by the implemented DT. The purpose is to confirm that the implemented DT and more specifically the part which transforms the file-based offline data to message-based online data and is responsible for handling the data integration problem is actually working, i.e., it is extracting the right data from the isolated building system.

Considering that the building system is equipped with HVAC, the temperature is already being controlled automatically to match a certain average to assure comfort of the occupants. Therefore, extreme temperature cases are not expected. For instance, there should not be large differences



a) 1st floor



b) 2nd floor

Figure 4.21 The planar heatmap to visualize the temperature measurements, which can also be animated using the timeline on top

between the temperature of the rooms between summer months and winter months or within a day. One room in which *relatively* higher variations in temperatures are expected to be observed, is the mechanical room (called *MÉCANIQUE 110* in Figure 4.22) of the building. All boilers, heat pumps, tanks, pipes, and heat exchangers are located in this room.

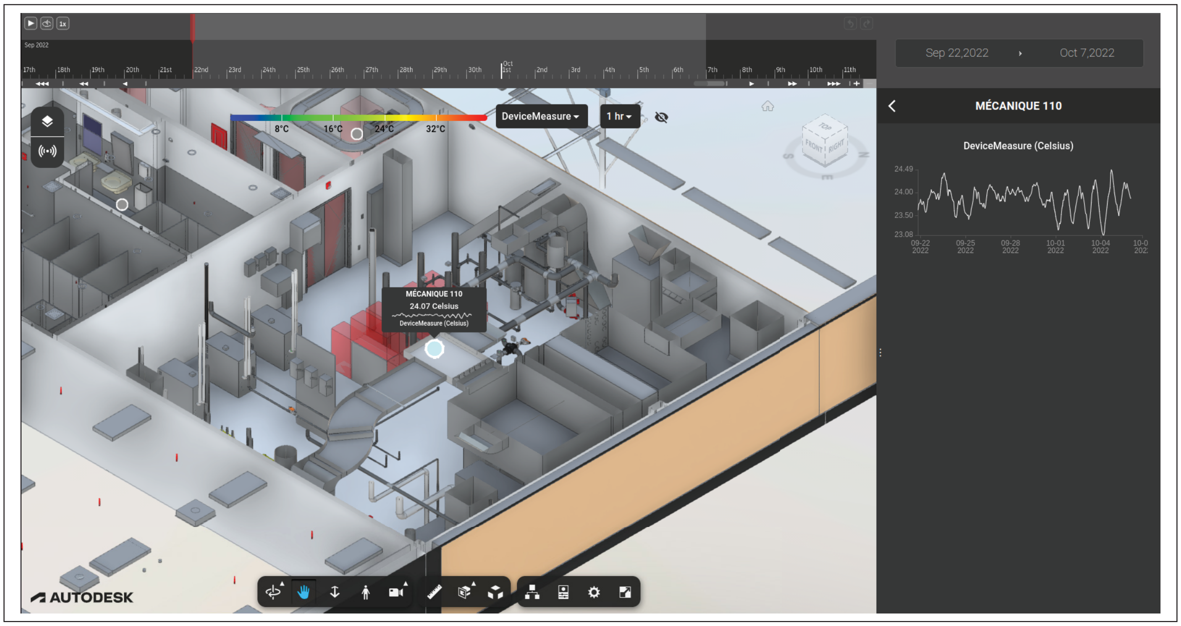


Figure 4.22 The mechanical room of the building

Figure 4.23 shows the temperature data of the mechanical room for first week of February 2022 and the second week of July 2022 (starting from Monday to Monday equivalent to 7 days). The data clearly shows the expected differences in the temperature of the room in February compared to July, i.e., on average the temperatures in February are lower than the temperatures in July.

Figure 4.23 also shows another pattern that can further verify that the data are being extracted in the right way. Note that, although the variation in temperature is not extreme, still one can observe an almost opposite daily pattern of temperature between February and July. For example, the temperature patterns of January 31 and July 4 are delimited in the boxes shown in Figure 4.23 and magnified in separate charts in Figure 4.24. On January 31 (the blue line) the temperature slightly decreased during the day until it reached a valley. However, during the same time of the day on July 4 (the red line) the temperature increased until it reached the peak. This is due to the fact that, in winter, heating equipment of HVAC are working more rigorously during the night to be able to maintain the temperature at acceptable levels considering the

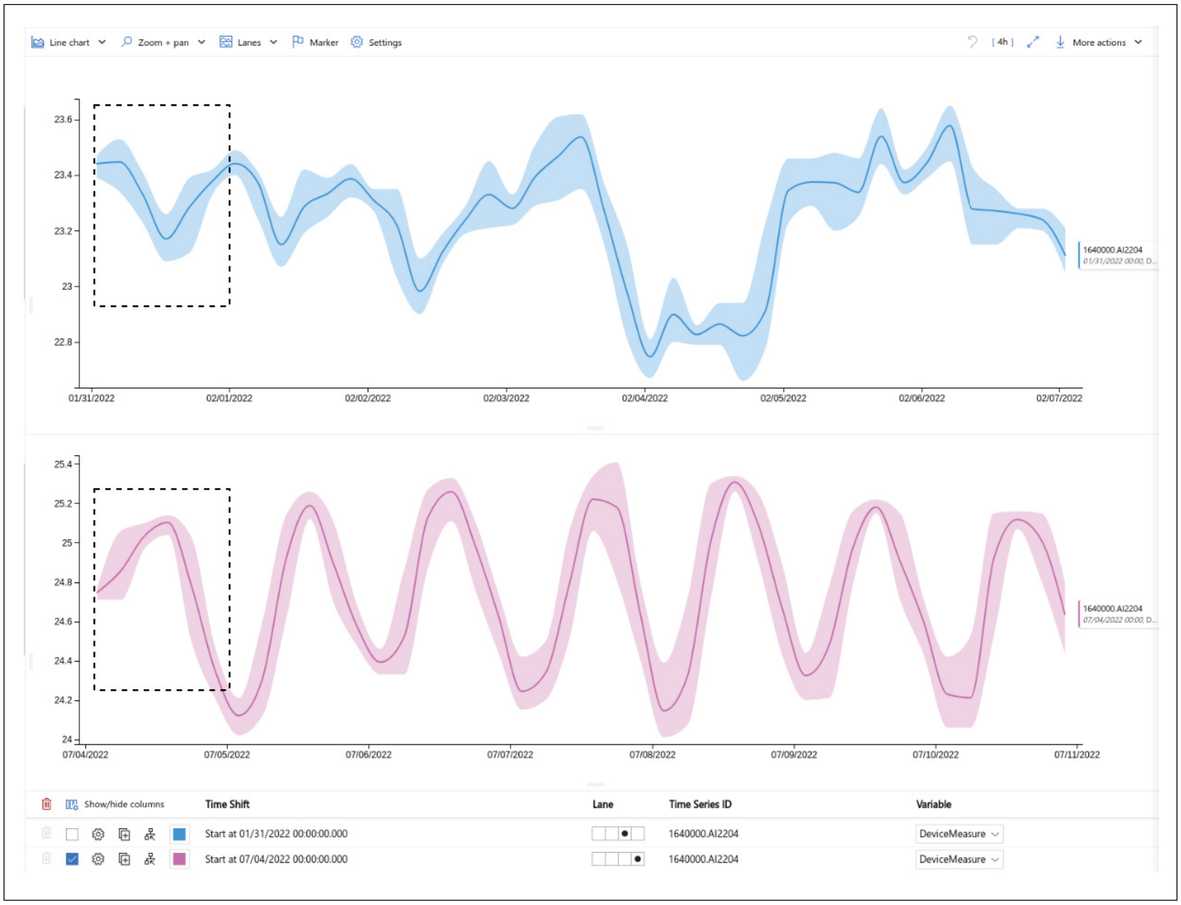


Figure 4.23 The temperature data of mechanical room during a week in February and a week in July of 2022

colder outside temperature of the night (compared to the day). Their operations cause the mechanical room temperature to be at relatively higher values. Then, at around sunrise, the power of heating equipment can be reduced since the rise of outside temperature can contribute to the target inside temperature, i.e., less HVAC-induced heat is required to match the target inside temperature. Then again, at around sunset and with the eventual decrease of the outside temperature, heating equipment's heat is needed during the night. Therefore, their power is raised again and consequently the mechanical room's temperature is increased.

During summer, the cooling equipment of HVAC is highly activated. Similar to the mechanism above, during the night, with the decrease in outside temperature, cooling equipment's activities are reduced. In other words, not much cooling equipment's power is required to keep the inside

Figure 4.24 The comparison of the daily temperature patterns between January 31 and July 4 of 2022

temperature of the building at the target level. At early stages of the day, when the outside temperature starts to rise, the power of cooling equipment is also increased to be able to control the rise of temperature during the day and keep it at the average target comfort levels. With more activity of the cooling equipment, the temperature inside the mechanical room increases as can be seen in Figure 4.24 and finally in later stages of the day, and with the sunset, again less power from the cooling equipment of HVAC is needed since the outside temperature is reducing as well.

Subsequent to the system verification, i.e., confirming that the extracted data from the DT designed and implemented in this thesis do match the temperature patterns that were expected to be observed in the mechanical room, system validation was carried out as follows: In order to evaluate the applicability and effectiveness of the proposed solution, a series of meetings were held with teams of experts from *Ville de Québec*. As the case study was advancing, the results of the project were progressively presented to the technical teams with members with backgrounds in both IT and facility management. At middle stages of the project's progress, the IT technicians confirmed that all the sensor data were able to be conveyed to the DT environment.

Subsequently, the experts with backgrounds in facility operations confirmed that the proposed DT system allows technicians to benefit from visualizations and data history exports. In particular, it was affirmed by the expert team that the rich spatio-temporal visualizations enabled by the implemented DT system can assist practitioners with gaining insights about the building's day-to-day operations, in a way that is significantly faster than looking at time-series records and matching them with the building spaces. The validation process was conducted based on the fact that this study was focused on a specific data access challenge and aimed to transform the offline file-based data into online data. The users who tested the developed solution agreed that the proposed framework and the software architecture have achieved the key objectives of the solution needed to overcome the data integration challenges and deliver effective visualizations of the spatio-temporal data. However, while testing the developed solution, issues were raised with regard to a more profound management of the information requirements, as well as organizational change management, to be able to actualize the transition of conventional workflows to a DT-driven paradigm. Since addressing the mentioned issues and validating the applicability of DTs for facility management, operation and maintenance fall outside the scope of the present study, a thorough investigation in this direction is recommended as a key direction for future research.

CHAPTER 5

DISCUSSIONS AND CONCLUSIONS

This study aimed at developing a message-driven system to transform offline sensory data into online data for digital twinning purposes of the buildings through implementing a cloud-native architecture. More specifically, the thesis focused on a solution applicable to network-wise isolated building automation systems.

5.1 Discussions

The developed methodological framework of the thesis addresses the following main research question:

> How can we implement a Digital Twin (DT) for built environments where data is hosted on an isolated network and the existing real-time data acquisition methods are not applicable, using emerging technologies and cloud services?

This study offers a solution to this question by the proposed framework elaborated in chapter 3, which involves a procedure to acquire the offline data and transform them into online data. This approach is implemented using a message-driven architecture. This architecture consists of the integration of several cloud services and serverless components specifically developed for this purpose. The proposed solution of this thesis shows that distributed cloud-native architectures are suitable for DT in buildings. The implementation details may vary depending on the context and applications. DT platforms of building, in general, need an architecture that allows high levels of scalability. Various sensors may be added to or removed from the model overtime. The adopted architecture must be able to accommodate these emergent changes without considerable effort. This thesis adds flexibility and improves the performance of the platform by enhancing its resilience through a granular architecture. Moreover, the proposed solution can be rapidly developed and is easily reproducible in similar DT projects. The real-world scenario described in chapter 4 verifies the viability of this solution by applying it for a DT of a community center building in Quebec City.

Furthermore, the proposed solution in this thesis addresses the following two specific sub-questions:

> How can we access/gather real-time data inside an isolated environment?

The common data streaming approaches (e.g., BMS Software APIs with streaming service, stream via direct access to a sensor network, etc.) fell outside the specific assumptions of the problem focused in this research about the network access difficulties to online data. In other words, the specific isolated features of the network, such as the existence of a firewall or limited access to the internet, automatically make the use of those common approaches irrelevant. For such scenarios, this research shows the applicability of offline data acquisition methods such as scheduled emails for gathering data.

> How to map sensor information to the BIM model as a foundation data model for DT?

This thesis adopts a mapping method that is about the sensor's coordinates and the bounding box of the rooms. If these coordinates are identified as lying within the bounding box of a specific room, then that sensor is acknowledged to be located in that room and hence is mapped to that room on the BIM model viewer. Coordinates of each sensor should be stored on the Device Twin in order to be retrieved at the time of mapping.

5.2 Contributions

The majority of research in the domain of digital twinning has assumed that there are no barriers for collecting data from the physical environment for digital twinning purposes. It is often assumed that a sensor network can be directly connected to an IoT Platform-as-a-Service (PaaS) that facilitates configuring, provisioning, and monitoring devices, and also enables bi-directional communication between IoT devices and the application. Therefore, the IoT PaaS can be easily integrated with a cloud-based DT platform. However, in certain scenarios, a sensor network cannot be readily accessible. For instance, data is protected in a DMZ due to security concerns, e.g., a sensor network can be behind a firewall, which prevents external access to the network; or there could be no internet connectivity.

This research proposes a methodological framework for developing a cloud-native software architecture that collects, processes, streams, and stores offline time-series data, which integrates sensory data with BIM model data and visualizes them in the DT platform. An important part of this framework addresses one of the data integration problems which concerns integrating the offline sensory file-based data that have limited accessibility. The proposed framework in this thesis relaxes the common assumption of having access to readily accessible sensory data, by designing a message-driven architecture comprising a set of functions and cloud services that simulate the behavior of a sensor network. Assuming that the offline data is already collected in spreadsheet format, the simulation consists of the transformation of these documents into events that can be stored or streamed for further processing in the digital twinning of buildings. Moreover, this methodological framework and its cloud-native architecture are developed in a systematic and abstract way that makes them reusable in other settings, specifically when sensory data are not readily accessible. These features will be elaborated later in this section.

Another contribution of this thesis relevant to the practitioners within the construction industry, relates to the unique features of the industry. Traditionally, AECO-FM is known to be a slow adopter in digital transformation. In addition, the field is inherently interdisciplinary and involves various stakeholders at different stages of the product's life cycle. The progress and success in any AECO-FM project is dependent on stakeholders' collaborations. A key element in the collaborations is access to integrated data. However, this access and data integration among various stakeholders is often hindered for different reasons. The fragmentation existing in the industry has contributed to emergence of "data silos" inside stakeholder organizations. Large amounts of data with various formats are being collected in these organizations but kept isolated often for privacy and regulation reasons or due to the structure of the legacy systems, which make the data integration from multiple sources a challenging task. Therefore, the solutions offered in this thesis help the AECO-FM sector to overcome these challenges and benefit from the potential of valuable untapped data for various purposes such as digital twinging, advanced data analytics, etc.

In addition to the contributions mentioned above, the current study is valuable in terms of its reusability, scalability, reproducibility and the different data types that it can handle.

5.3 Reproducibility and reusability

The outputs of the proposed framework applied to the specific case of this study are reproducible and the proposed cloud-native software architecture in chapter 3, can be reused in similar scenarios. It consists of five important components:

1. Azure Logic apps, the designed workflow in Figure 4.3 which automates the workflow and integrates the services in this architecture, with slight customization can be reused in future projects.
2. Azure Function A, this function is responsible for storing the attachments of the email and can be reused in other projects.
3. Azure Function B, applies some processes to clean up the sensory data and prepare it for the next step. The input is an excel file and the outputs are multiple CSV files of each sensor. The function is reusable for any excel file with slight modifications.
4. Azure Function C, generates messages based on the trends in a CSV file. The input is a CSV file of a sensor and the outputs are numerous messages (each message contains a reading of the sensor and other information about the sensor) which the function publishes to a message broker (Azure Service Bus). Each message is assigned a timestamp for its scheduled release and stays in the message queue up until then. The function can be reused in any scenario by applying some slight changes.
5. Azure Function D reads/consumes the message from the queue at the time of its release. Then based on the content of the message, it registers a sensor on the Azure IoT Hub (if the sensor does not exist on the Azure IoT Hub), then publishes the message to the Azure IoT Hub endpoint. The input and output of the function is a message. The function is reusable in similar cases.

The algorithm behind all of the previous functions can be reused to develop those functions for any cloud provider.

5.4 Flexibility, scalability and availability

All of the components of the proposed architecture automatically scale based on the number of requests made to the system. In developing the proposed solution, scalibility has been one of the main design criteria as it is essential for DT implementation. Moreover, for further experiments it is possible to increase/decrease the frequency of the offline data provision.

Additionally, The proposed framework provides a high degree of availability. This arises from it being cloud-native and most of the cloud providers (in the use case of this thesis, Microsoft Azure) guarantee very high uptime with a very low chance of failures.

These two specific features have important practical implications. The solution developed here is readily transferable to the production stage which makes it compelling for industrial applications.

5.5 Multi-Sourcing capabilities

The AECO-FM sector involves multiple heterogeneous stakeholders, which require collaboration. There are often different data formats and structures coming from various sources (e.g., different information technology systems or various legacy systems) which need to be integrated into a single DT. Therefore, one of the challenges of data integration in the construction industry is the disparity of the data sources. The proposed software architecture can easily be integrated with a physical sensor network, a legacy database, etc. as another data source, and the developed DT federates these data sources under a single platform.

5.6 Limitations and future work

This study has four main limitations:

- The proposed architecture in this research project is developed and deployed on Microsoft Azure infrastructure. This may lead to vendor lock however, as it is previously mentioned, all of the components can be redeveloped based on the algorithm and logic elaborated in section 4. They can be redeployed to similar services (e.g., AWS Lambda Function, Google Cloud Function, etc.), supplied by other cloud providers.
- The template of the input excel file associated with Function B that cleans the data must be exactly the same as the file presented in this study. However, this limitation can be resolved by applying some slight changes in the developed codes.
- Considering that this framework handles offline data (by adding 24 hours to their timestamps in order to simulate the previous day), the data are not inherently real-time. Akanmu *et al.* (2021) and Delgado & Oyedele (2021) mention the issue of data communication latency and that it is very context-dependent. In some applications, the actual near-real-time data are needed for quick response. However, there might be scenarios in which a more long-term perspective on learning from data patterns is required. Therefore, there is more flexibility in terms of communication latency, and also the frequency of offline data provision can be increased.

  The proposed solution of this thesis is not suitable for DT use case scenarios where it is critical to access and analyze near real-time data. However, the proposed approach in this study is applicable to use cases in which trends of data over time are the most important and are analyzed with the purpose of learning and prediction such as the energy performance of the building.
- Digital twins in their ultimate form must model a complete bi-directional interaction between the physical asset and its digital twin. They should be able to act upon the physical world by analyzing the collected data and automatic response. However, the majority of current existing DTs in industries lack this feature. Similarly, the implemented DT platform for the community center does not provide a component for end-user to interact with the physical environment, e.g. controlling the HVAC system via actuators directly on the user interface. However, the IoT Platform (PaaS) used in this platform has this built-in functionality for

bi-directional communication, therefore the DT Platform has the potential, and it can be a feature to be added to the system in future works.

There are a number of ways in which the current study and its architecture can be improved or further expanded in future projects.

A natural next step in the extension of the current study is introducing new use cases for the DT platform. For instance, occupancy detection can be used to adjust the temperature of the room accordingly to ensure occupants' comfort. Moreover, by observing the occupancy rate patterns over time, one can predict the future occupants' behavior and adjust the HVAC systems' settings autonomously through actuators in a way that the energy efficiency of the buildings is optimized. This way the DT platforms are operated in their entirety when there is bi-directional communication between the digital world and the physical world. Hence, the outcome of this study, can be used by other researchers for further development depending on the use case at hand. By adding further intuitive visual components to the user interface, the user can interact with the DT platform more effectively and have more control and insight over the physical entities.

This platform facilitates addressing the well-known challenge of disparate sources of data within the construction industry. With the offline data being handled already in this platform, one can add a typical online data source such as a sensor network to this platform and have multiple sources of data federated. All these sources of data can then be integrated with a BIM model and visualized in a model viewer. Additionally, another future contribution can be federating these disparate data sources through a GraphQL server and querying them under a single endpoint. In GraphQL we can design a graph-based schema to define the relationships between elements in different data sources, e.g. Sensor ID and Room ID to facilitate the mapping of the sensors to their corresponding object.

BIBLIOGRAPHY

Akanmu, A. A., Anumba, C. J. & Ogunseiju, O. O. (2021). Towards Next Generation Cyber-physical Systems and Digital Twins for Construction. *Journal of Information Technology in Construction*, 26, 505-525. doi: 10.36680/j.itcon.2021.027.

Alizadehsalehi, S. & Yitmen, I. (2021). Digital Twin-based Progress Monitoring Management Model Through Reality Capture to Extended Reality Technologies (DRX). *Smart and Sustainable Built Environment*, 37. doi: 10.1108/SASBE-01-2021-0016.

Alshammari, K., Beach, T. H. & Rezgui, Y. (2021). Cybersecurity for digital twins in the built environment: current research and future directions. *Journal of Information Technology in Construction*, 26, 159-173. doi: 10.36680/j.itcon.2021.010.

Angjeliu, G., Coronelli, D. & Cardani, G. (2020). Development of the simulation model for Digital Twin applications in historical masonry buildings: The integration between numerical and experimental reality. *Computers & Structures*, 238, 106282. doi: 10.1016/j.compstruc.2020.106282.

Arrichiello, V. & Gualeni, P. (2019). Systems Engineering and Digital Twin: a Vision for the Future of Cruise Ships Design, Production and Operations. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 14, 115-122.

Barricelli, B. R., Casiraghi, E. & Fogli, D. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access*, 7, 167653-167671. doi: 10.1109/ACCESS.2019.2953499.

Boje, C., Guerriero, A., Kubicki, S. & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction*, 114, 103179. doi: 10.1016/j.autcon.2020.103179.

Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A. & Luck, A. (2018). *Gemini Principles*. CDBB.

Borth, M., Verriet, J. & Muller, G. (2019). Digital Twin Strategies for SoS 4 Challenges and 4 Architecture Setups for Digital Twins of SoS. *2019 14th Annual Conference System of Systems Engineering (SoSE)*, pp. 164-169. doi: 10.1109/SYSOSE.2019.8753860.

Boschert, S. & Rosen, R. (2016). Digital Twin—The Simulation Aspect. In Hehenberger,Peterand Bradley, D. (Ed.), *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers* (pp. 59-74). Cham: Springer International Publishing. doi: 10.1007/978-3-319-32156-1_5.

Botín-Sanabria, D. M., Mihaita, A.-S., Peimbert-García, R. E., Ramírez-Moreno, M. A., Ramírez-Mendoza, R. A. & Lozoya-Santos, J. d. J. (2022). Digital Twin Technology Challenges and Applications: A Comprehensive Review. *Remote Sensing*, 14(6), 1335. doi: 10.3390/rs14061335.

Brenner, B. & Hummel, V. (2017). Digital Twin as Enabler for an Innovative Digital Shopfloor Management System in the ESB Logistics Learning Factory at Reutlingen - University. *Procedia Manufacturing*, 9, 198-205. doi: 10.1016/j.promfg.2017.04.039.

Bruynseels, K., Santoni de Sio, F. & van den Hoven, J. (2018). Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Frontiers in Genetics*, 9, 31. doi: 10.3389/fgene.2018.00031.

Davari, S., Shahinmoghadam, M., Motamedi, A. & Poirier, E. (2022, June). Demystifying the Definition of Digital Twin for Built Environment.

Delgado, J. M. D. & Oyedele, L. (2021). Digital Twins for the built environment: learning from conceptual and process models in manufacturing. *Advanced Engineering Informatics*, 49, 101332. doi: 10.1016/j.aei.2021.101332.

Douglas, D., Kelly, G. & Kassem, M. (2021). BIM, Digital Twin and Cyber-Physical Systems: crossing and blurring boundaries. doi: 10.48550/ARXIV.2106.11030.

Errandonea, I., Beltrán, S. & Arrizabalaga, S. (2020). Digital Twin for maintenance: A literature review. *Computers in Industry*, 123, 103316. doi: 10.1016/j.compind.2020.103316.

Fan, C., Zhang, C., Yahja, A. & Mostafavi, A. (2021). Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. *International Journal of Information Management*, 56, 102049. doi: 10.1016/j.ijinfomgt.2019.102049.

Feng, H., Chen, Q. & García de Soto, B. (2021, 11). Application of digital twin technologies in construction: an overview of opportunities and challenges. doi: 10.22260/ISARC2021/0132.

Fuller, A., Fan, Z., Day, C. & Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access*, 8, 108952-108971. doi: 10.1109/ACCESS.2020.2998358.

Glaessgen, E. H. & Stargel, D. (2012). The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles.

Greif, T., Stein, N. & Flath, C. M. (2020). Peeking into the void: Digital twins for construction site logistics. *Computers in Industry*, 121, 103264. doi: 10.1016/j.compind.2020.103264.

Grieves, M. & Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In Kahlen, F.-J., Flumerfelt, S. & Alves, A. (Eds.), *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (pp. 85-113). Springer International Publishing.

Gürdür Broo, D., Bravo-Haro, M. & Schooling, J. (2022). Design and Implementation of a Smart Infrastructure Digital Twin. *Automation in Construction*, 136, 104171. doi: 10.1016/j.autcon.2022.104171.

Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105. Retrieved from: http://www.jstor.org/stable/25148625.

Hoeft, M. & Trask, C. (2022). Safety Built Right in: Exploring the Occupational Health and Safety Potential of BIM-Based Platforms throughout the Building Lifecycle. *Sustainability*, 14(10), 6104. doi: 10.3390/su14106104.

Hou, L., Wu, S., Zhang, G. K., Tan, Y. & Wang, X. (2021). Literature Review of Digital Twins Applications in Construction Workforce Safety. *Applied Sciences*, 11(1), 339. doi: 10.3390/app11010339.

Jiang, F., Ma, L., Broyd, T. & Chen, K. (2021). Digital twin and its implementations in the civil engineering sector. *Automation in Construction*, 130, 103838. doi: 10.1016/j.autcon.2021.103838.

Jones, D., Snider, C., Nassehi, A., Yon, J. & Hicks, B. (2020). Characterising the Digital Twin: A Systematic Literature Review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36-52. doi: 10.1016/j.cirpj.2020.02.002.

Karakra, A., Lamine, E., Fontanili, F. & Lamothe, J. (2020). HospiT'Win: a digital twin framework for patients' pathways real-time monitoring and hospital organizational resilience capacity enhancement. *Proceedings of the 9th International Workshop on Innovative Simulation for Healthcare (IWISH 2020)*, pp. 62-71. doi: 10.46354/i3m.2020.iwish.012.

Khajavi, S. H., Motlagh, N. H., Jaribion, A., Werner, L. C. & Holmström, J. (2019). Digital Twin: Vision, Benefits, Boundaries, and Creation for Buildings. *IEEE Access*, 7, 147406-147419. doi: 10.1109/ACCESS.2019.2946515.

Laborie, F., Røed, O. C., Engdahl, G. & Camp, A. (2019). Extracting value from data using an industrial data platform to provide a foundational digital twin. *Offshore Technology Conference*.

Lee, D. & Lee, S. (2021). Digital Twin for Supply Chain Coordination in Modular Construction. *Applied Sciences*, 11(13), 5909. doi: 10.3390/app11135909.

Li, C., Mahadevan, S., Ling, Y., Choze, S. & Wang, L. (2017). Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin. *AIAA Journal*, 55(3), 930-941. doi: 10.2514/1.J055201.

Li, L., Aslam, S., Wileman, A. & Perinpanayagam, S. (2021). Digital Twin in Aerospace Industry: A Gentle Introduction. *IEEE Access*, PP, 1-1. doi: 10.1109/ACCESS.2021.3136458.

Lin, K., Xu, Y.-L., Lu, X., Guan, Z. & Li, J. (2021). Digital Twin-based Collapse Fragility Assessment of a Long-span Cable-stayed Bridge Under Strong Earthquakes. *Automation in Construction*, 123, 103547. doi: 10.1016/j.autcon.2020.103547.

Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., Liu, R., Pang, Z. & Deen, M. J. (2019). A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin. *IEEE Access*, 7, 49088-49101. doi: DOI:10.1109/ACCESS.2019.2909828.

Lu, Q., Parlikad, A. K., Woodall, P., Don Ranasinghe, G., Xie, X., Liang, Z., Konstantinou, E., Heaton, J. & 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. doi: 10.1061/(ASCE)ME.1943-5479.0000763.

Lydon, G., Caranovic, S., Hischier, I. & Schlueter, A. (2019). Coupled simulation of thermally active building systems to support a digital twin. *Energy and Buildings*, 202, 109298. doi: 10.1016/j.enbuild.2019.07.015.

Mannino, A., Moretti, N., Dejaco, M., Luciano, B. & Re Cecconi, F. (2019). Office building occupancy monitoring through image recognition sensors. 9, 371-380. doi: 10.2495/SAFE-V9-N4-371-380.

March, S. T. & Storey, V. C. (2008). Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research. *MIS Quarterly*, 32(4), 725-730. Retrieved from: http://www.jstor.org/stable/25148869.

Markus, M. L., Majchrzak, A. & Gasser, L. (2002). A Design Theory for Systems That Support Emergent Knowledge Processes. *MIS Quarterly*, 26(3), 179-212. Retrieved from: http://www.jstor.org/stable/4132330.

Martinelli, I., Campi, F., Checcacci, E., Presti, G. M. L., Pescatori, F., Pumo, A. & Germani, M. (2019). Cost Estimation Method for Gas Turbine in Conceptual Design Phase. *Procedia CIRP*, 84, 650-655. doi: 10.1016/j.procir.2019.04.311. 29th CIRP Design Conference 2019, 08-10 May 2019, Póvoa de Varzim, Portgal.

Möhring, M., Keller, B., Radowski, C.-F., Blessmann, S., Breimhorst, V. & Müthing, K. (2022). Empirical Insights into the Challenges of Implementing Digital Twins. *Human Centred Intelligent Systems*, pp. 229-239.

Morgado, J. F., Ghedini, E., Goldbeck, G., Hashibon, A., Schmitz, G. J., Friis, J. & de Baas, A. F. (2020). Mechanical Testing Ontology for Digital-Twins: a roadmap based on EMMO. *SeDiT@ ESWC*, 2615, 10.

Negri, E., Fumagalli, L. & Macchi, M. (2017). A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, 11, 939-948. doi: 10.1016/j.promfg.2017.07.198. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

Opoku, D.-G. J., Perera, S., Osei-Kyei, R. & Rashidi, M. (2021). Digital twin application in the construction industry: A literature review. *Journal of Building Engineering*, 40, 102726. doi: 10.1016/j.jobe.2021.102726.

Ozturk, G. B. (2021). Digital Twin Research in the AECO-FM Industry. *Journal of Building Engineering*, 40, 102730. doi: 10.1016/j.jobe.2021.102730.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77. doi: 10.2753/MIS0742-1222240302.

Pregnolato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J. & Taylor, C. (2022). Towards Civil Engineering 4.0: Concept, Workflow and Application of Digital Twins for Existing Infrastructure. *Automation in Construction*, 141, 104421. doi: 10.1016/j.autcon.2022.104421.

Qiuchen Lu, V., Parlikad, A. K., Woodall, P., Ranasinghe, G. D. & Heaton, J. (2019). Developing a dynamic digital twin at a building level: Using Cambridge campus as case study. *International conference on smart infrastructure and construction 2019 (ICSIC) driving data-informed decision-making*, pp. 67-75. doi: 10.1680/icsic.64669.067.

Ríos, J., Staudter, G., Weber, M., Anderl, R. & Bernard, A. (2020). Uncertainty of data and the digital twin: a review. *International Journal of Product Life cycle Management*, 12, 329-358. doi: 10.1504/IJPLM.2020.10035102.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S. & Girolami, M. (2020). Construction with digital twin information systems. *Data-Centric Engineering*, 1, e14. doi: 10.1017/dce.2020.16.

Schweigkofler, A., Braholli, O., Akro, S., Siegele, D., Penna, P., Marcher, C., Tagliabue, L. & Matt, D. (2022, Apr.). Digital Twin as energy management tool through IoT and BIM data integration. *CLIMA 2022 conference*. doi: 10.34641/clima.2022.46.

Semeraro, C., Lezoche, M., Panetto, H. & Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. *Computers in Industry*, 130, 103469. doi: 10.1016/j.compind.2021.103469.

Shahzad, M., Shafiq, M. T., Douglas, D. & Kassem, M. (2022). Digital Twins in Built Environments: An Investigation of the Characteristics, Applications, and Challenges. *Buildings*, 12(2), 120. doi: 10.3390/buildings12020120.

Shim, C.-S., Dang, N.-S., Lon, S. & Jeon, C.-H. (2019). Development of a bridge maintenance system for prestressed concrete bridges using 3D digital twin model. *Structure and Infrastructure Engineering*, 15(10), 1319-1332. doi: 10.1080/15732479.2019.1620789.

Singh, S., Shehab, E., Higgins, N., Fowler, K., Tomiyama, T. & Fowler, C. (2018). Challenges of Digital Twin in High Value Manufacturing. *Aerospace Systems and Technology Conference*.

Tao, F. & Zhang, M. (2017). Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing. *IEEE Access*, 5, 20418-20427. doi: 10.1109/ACCESS.2017.2756069.

Tao, F., Zhang, H., Liu, A. & Nee, A. Y. C. (2019). Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics*, 15(4), 2405-2415. doi: 10.1109/TII.2018.2873186.

Tchana, Y., Ducellier, G. & Remy, S. (2019). Designing a unique Digital Twin for linear infrastructures lifecycle management. *Procedia CIRP*, 84, 545-549. doi: 10.1016/j.procir.2019.04.176. 29th CIRP Design Conference 2019, 08-10 May 2019, Póvoa de Varzim, Portgal.

Voigt, I., Inojosa, H., Dillenseger, A., Haase, R., Akgün, K. & Ziemssen, T. (2021). Digital twins for multiple sclerosis. *Frontiers in Immunology*, 12, 669811. doi: 10.3389/fimmu.2021.669811.

Wanasinghe, T. R., Wroblewski, L., Petersen, B. K., Gosine, R. G., James, L. A., De Silva, O., Mann, G. K. I. & Warrian, P. J. (2020). Digital Twin for the Oil and Gas Industry: Overview, Research Trends, Opportunities, and Challenges. *IEEE Access*, 8, 104175-104197. doi: 10.1109/ACCESS.2020.2998723.

Woitsch, R., Sumereder, A. & Falcioni, D. (2022). Model-based data integration along the product & service life cycle supported by digital twinning. *Computers in Industry*, 140, 103648. doi: 10.1016/j.compind.2022.103648.

Xie, X., Lu, Q., Parlikad, A. K. & Schooling, J. M. (2020). Digital Twin Enabled Asset Anomaly Detection for Building Facility Management. *IFAC-PapersOnLine*, 53(3), 380-385. doi: 10.1016/j.ifacol.2020.11.061. 4th IFAC Workshop on Advanced Maintenance Engineering, Services and Technologies - AMEST 2020.

Yu, G., Zhang, S., Hu, M. & Wang, Y. K. (2020). Prediction of highway tunnel pavement performance based on digital twin and multiple time series stacking. *Advances in Civil Engineering*, 2020, 21. doi: 10.1155/2020/8824135.

Yu, G., Wang, Y., Mao, Z., Hu, M., Sugumaran, V. & Wang, Y. K. (2021). A digital Twin-based Decision Analysis Framework for Operation and Maintenance of Tunnels. *Tunnelling and Underground Space Technology*, 116, 104125. doi: 10.1016/j.tust.2021.104125.

Zaballos, A., Briones, A., Massa, A., Centelles, P. & Caballero, V. (2020). A Smart Campus' Digital Twin for Sustainable Comfort Monitoring. *Sustainability*, 12(21), 9196. Retrieved from: https://www.mdpi.com/2071-1050/12/21/9196.