Dataspaces in Urban Digital Twins: a Case Study in the Photovoltaics

Cristian Martella[1,*], Antonella Longo[1], Marco Zappatore[1] and Antonio Ficarella[1]

[1]*Department of Engineering for Innovation, University of Salento, via Monteroni - 73100 Lecce, Italy*

Abstract

The rise of the Internet of Things (IoT) fosters the massive deployment of ubiquitous devices in geographically distributed locations, which are reaching billions in number, covering disparate domains such as smart healthcare, transportation, logistics, energy management, industrial automation, and other services. The vast amount of data generated by these devices can be used to enhance the functionality and evolution of smart environments. In addition, digital twins (and especially urban digital twins) represent another scenario benefiting from a pervasive availability of sensor data. However, traditional cloud infrastructures are not designed to manage them efficiently. Fog computing and edge computing can complement cloud computing to improve latency, reliability, as well as to achieve faster response times. In this scenario, a Dataspace stands as a technology that can support the effective management of huge volumes of data generated by digital twins, also providing FAIRness-driven policies for data exchange. In this context, this work presents a case study focusing on a smart photovoltaic cyber-physical system and the development of its digital twin.

Keywords

Dataspace, Urban Digital Twin, Edge-Cloud Continuum

1. Introduction

The Internet of Things (IoT) facilitates connections between devices and the Internet, enabling more meaningful interactions between objects and people. This connection process typically involves integrating sensing, actuating, and control devices. IoT is gaining popularity in various domains such as smart healthcare, transport, logistics, retail, industrial automation, and more. The deployment of ubiquitous devices in geographically distributed locations is increasing rapidly, reaching billions in number. These smart devices generate a vast amount of data that needs to be transmitted through network infrastructures, which can pose challenges. This data can be utilized to enhance the functionality and evolution of smart environments.

However, in cloud infrastructures, the data is sent to cloud servers for processing and then returned to the devices, which presents limitations due to increased latency and potential outdated control information, as sensors and actuators are often located on the same physical device.

SEBD 2023: 31st Symposium on Advanced Database System, July 02–05, 2023, Galzignano Terme, Padua, Italy

*Corresponding author.

cristian.martella@unisalento.it (C. Martella); antonella.longo@unisalento.it (A. Longo); marcosalvatore.zappatore@unisalento.it (M. Zappatore); antonio.ficarella@unisalento.it (A. Ficarella)

0000-0001-9751-9367 (C. Martella); 0000-0002-6902-0160 (A. Longo); 0000-0002-8277-9390 (M. Zappatore); 0000-0003-3206-4212 (A. Ficarella)

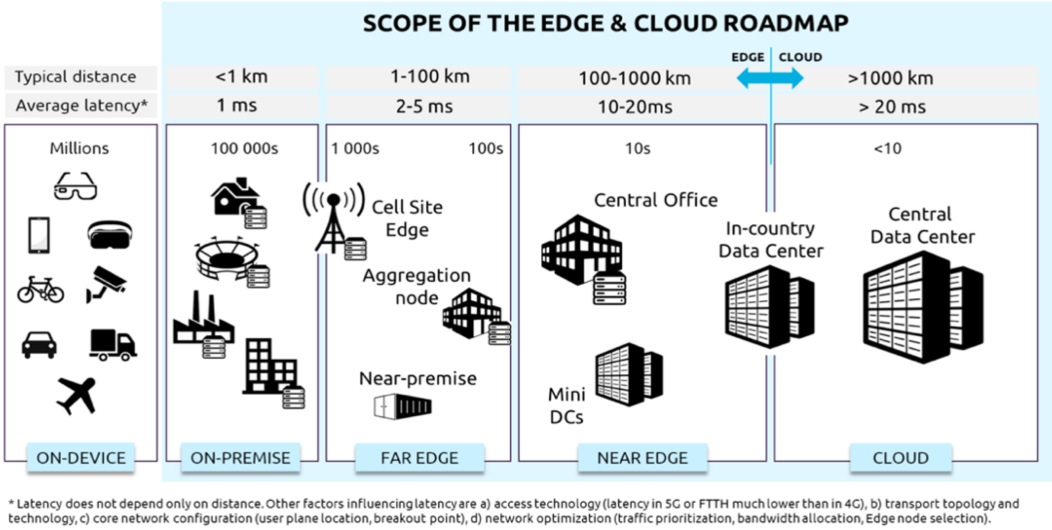

Figure 1: Scope of the European Industrial Roadmap in the cloud-edge continuum as presented in [1].

Fog computing and edge computing can complement cloud computing to address these limitations, providing improved latency, reliability, and faster response times. The geographically distributed nature of the fog layer and edge devices also enables location awareness.

By bringing intelligence away from the cloud, fog computing can process the IoT data in close proximity to the data sources. Afterward, it can use resources from the cloud (only if needed) more effectively than through individual devices.

As per the roadmap outlined by the European cloud-edge industry, the next generation edge-cloud continuum (ECC) framework aims to achieve the following objectives of attaining global recognition for technological leadership and competitiveness in areas such as cybersecurity and interoperability, developing energy and resource efficient infrastructures and deploying standardized accessible, scalable and compatible core services, making them accessible via APIs for seamless integration and portability across different cloud providers' environments.

The scope of the edge cloud roadmap, as illustrated in Figure 1 [1], encompasses the entire cloud-edge continuum, ranging from on-premise edge to "far edge" (within 100 kilometers of customers and premises) and "near edge" (typically hundreds of kilometers away from the customer), and even regional data centers [1].

In this scenario, a *Dataspace* represents a technology that can serve as an effective means of managing the vast amount of data generated by digital twins (DTs). By integrating data from diverse sources and facilitating the coexistence of heterogeneous data, dataspaces enables a wide range of applications in various fields. For instance, in the context of Industry 4.0, digital twin technology can be leveraged for predictive maintenance, quality control, optimization of production processes, and decision-making support [2].

Additionally, DTs find valuable applications in the realm of smart cities, where they can simulate and optimize the performance of complex systems such as energy management, transportation, and waterways management [3]. Through the creation of digital replicas of a

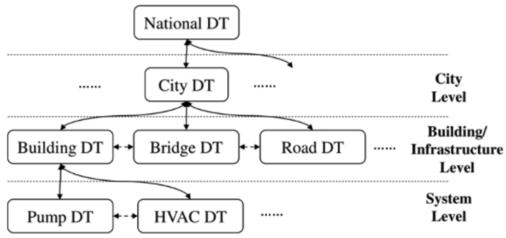

Figure 2: Digital twins connections and hierarchy levels.

city's infrastructure, DTs can continuously monitor and analyze real-time data from multiple sources, including sensors, cameras, and other IoT devices, to predict state changes in systems and forecast potential future actions. This enables the study of what-if scenarios, estimation of outcomes, and proposal of appropriate preventive actions, thereby assisting city planners in making informed decisions related to urban planning and management [4].

DTs can be deployed in hierarchies of levels that span across the edge-cloud continuum scope. Each level comes with its own policies and action ranges that can be managed through dataspaces. Concerning smart cities, for instance, every infrastructure provides its own digital twin instance which communicates and manages system level DTs, interacts with other infrastructure level DTs and with city level DTs as well, as shown in Figure 2 [5]. Based on this schema, complex urban digital twins can be thought as a federation of DTs that address city elements at finer grains of detail. Therefore, each element, from the water pump to the city itself, can be considered as a building block of a city DT ecosystem. This work focuses on the case study related to the creation of the DT of a smart photovoltaic panel, that is one of such building blocks. The approach adopted in this work allows it to be tailored to further case studies, investigating efficient ways to deal with the challenges that arise in the real-time bidirectional communication between multiple DTs.

This paper is organized as follows. After this introductory section, a literature overview discussing the current state-of-the-art is provided in Section 2. Section 3 describes the case study of a smart photovoltaic cyber-physical system and the development of its digital twin, while section 4 concludes the paper.

2. Background

Over the last decade, several international initiatives and projects focused on providing standards and implementations of data models and, more recently, on dataspaces for digital twins. In the next paragraphs, these concepts are introduced along with some significant related works.

A data model can be defined as a standardized representation of data to offer advanced capabilities for the description of context, providing a common language for describing data in various contexts, such as smart cities, agriculture, and energy management [6, 7]. FIWARE[1] provides a rich catalog of Smart Data Models, organized in contexts and compatible with NGSIv2/NGSI-LD APIs. In this way, developers can deliver their applications, avoiding interoperability concerns,

[1]https://www.fiware.org/smart-data-models/

as they are managed by the FIWARE's stack [8, 9]. As a matter of fact, interoperability between an arbitrary number of heterogeneous data sources is managed by a dataspace, whose goal is to obtain an upfront unifying schema across all such sources [7]. Other than the FIWARE's initiative above, a list of international initiatives related to the development of dataspaces is discussed in the following paragraphs.

First, the International Data Spaces Association (IDSA)[2] is a non-profit organization that promotes secure and standardized data interchange and connectivity among businesses. IDSA establishes a reference architecture and specifications for developing secure and trustworthy dataspaces that facilitate interoperability between organizations. Gaia-X, a European initiative, aims to establish a secure and reliable data infrastructure for Europe, with principles based on data sovereignty, interoperability, transparency, and openness. It aims to create a federated system of cloud services and dataspaces to enable secure data exchange among organizations.

The European interoperability framework [10] is a commonly agreed approach for the delivery of European public services in an interoperable manner, defining guidelines in the form of common principles, models, and recommendations.

Similarly, the Data Spaces Support Centre (DSSC)[3] investigates requirements of data space initiatives, defines common requirements, and establishes best practices to accelerate the formation of sovereign data spaces in digital transformation. DSSC has developed a resource inventory tool that provides an overview of state-of-the-art resources for organizations interested in creating or participating in data spaces.

More specifically, the dataspace concept originates from typical challenges in large-scale integration scenarios with thousands of data sources, where obtaining an upfront unifying schema across all sources is difficult and expensive from the following points of view [7]: (1) Data variety: many different types and formats of data need to be integrated in order to provide a comprehensive view or analysis; (2) Semantic heterogeneity: differences in meaning or interpretation of data between different information sources; (3) Initial effort: traditional approaches to data integration require significant initial effort to set up, including manual matching and mapping generation techniques; (4) Scalability: as the number of data sources increases, it becomes increasingly difficult to perform information integration on-the-fly; (5) Cost: the cost associated with integrating large amounts of data can be prohibitive, particularly for smaller organizations.

Consequently, it can be also seen as a novel data management approach that prioritizes integrating data "as needed" and deferring labor-intensive data integration tasks until they are required. Dataspaces leverage automatic matching and mapping techniques to minimize the initial effort needed for data integration, resulting in a loosely integrated collection of data sources. When more comprehensive semantic integration is needed, specific mappings between the required data sources can be incrementally employed. Each data source within the dataspace, such as databases, CSV files, or web services, is referred to as a participant. The dataspace can effectively model the relationships between data in different participants, thereby defining a dataspace as a collection of participants and their interrelations [1].

Dataspaces are well suited for managing context related data flows, as in the case of digital

[2]https://internationaldataspaces.org/ (last access: 12 April 2023)
[3]https://dssc.eu/ (last access: 12 April 2023)

twins. A digital twin (DT) can be defined as the coupling between the state of a physical asset and its live virtual counterpart, with a functional output [11]. According to this definition, the data flows bidirectionally between the physical asset and its digital replica, and the state is kept up to date in real-time. DTs can be adopted in a wide range of applications, fostering the proliferation of advanced analytics tools for decision support systems and what-if analysis and simulations in disparate scenarios. In the context of urban planning and city development, for instance, the concept of DT has been tailored to meet the ontology of the various use cases [12, 13, 14]. Examples of the state-of-the-art urban DT implementations are the TU-Delft project and the Digital Twin Cities Centre.

The Dutch Kadaster collaborates with the 3D Geoinformation research group at TU Delft [15, 16] to generate and disseminate a sustainable 3D city model covering the Netherlands. The project aims to the creation of a toolset to provide real-world data based detailed 3D representations of buildings and infrastructures in a Country level scope. The platform can be used to generate detailed simulations and visualizations of urban areas, to aid in urban planning and decision-making tasks.

The Digital Twin Cities Centre (DTCC)[4] research program aims to create sustainable and efficient cities using digital tools and methods. DTCC seeks to virtually develop and manage cities, to transform how urban places are planned, managed, and renewed. In particular, it enables the simulation of different scenarios to forecast the effects of changes on corresponding city's environment.

3. Case Study

The considerations discussed so far are applied in the photovoltaic topic thanks to the case study presented in this section. More specifically, the cyber-physical infrastructure developed in this scenario consists of a smart photovoltaic system (PV), an edge node, and a mini data center, as shown in Figure 3.

The PV system features both sensors and actuators. The sensors include: power meters, to collect information on the power output of the panel and the state of charge of the batteries; basic meteorological sensors (i.e., thermo-hygrometer and barometer) to assess the environmental conditions related to a given reading, a GPS sensor, a magnetometer and a gyroscopic sensor to assess the position and the orientation of the panel. In this way, it is possible to collect consistent data related to the power output of a PV system as well as the context information. Conversely, on the actuators side, the PV system mounts two servo motors that drive a gear work to orientate the panel as requested. Both manual and automatic sun tracking modes are supported.

Data is streamed to an on-premise edge node, that can also perform preliminary data cleaning tasks and format the information according to a predefined format, to foster interoperability between different domains and reduce the workload of upstream components. While the proximity edge node can be constituted by a Single Board Computer (SBC), a desktop data center made of a cluster of SBCs was chosen to model the far edge data center. This component

[4] https://dtcc.chalmers.se/ (last access: 12 April 2023)

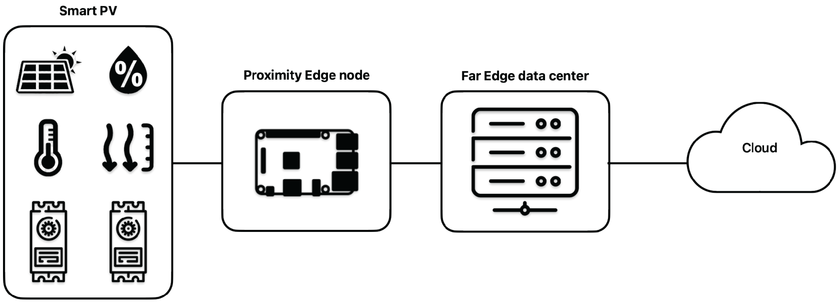

Figure 3: Architecture of the system.

is meant to collect data streamed by the on-premise edge node(s) and perform more advanced pipelines to provide insights and data analytics.

This infrastructure is suitable for the development of a DT of the PV system. In this context, a digital replica can be used to evaluate the current state of the system as well as to relate it either with the past evolution of the variables under examination or with their predicted values. Therefore, it is possible to run simulations and evaluate the feasibility and sustainability of possible scenarios, thus behaving as a suitable component for a decision-making support tool. Such decisions can actually result in actions on the real-world PV system, to change its state by driving the actuators (e.g. the servo gears) and make it assume a different orientation in order to optimize a set of variables of interest. This interaction can be leveraged to perform comparative analyses between real-world and simulated scenarios, which is a corner stone for deploying more realistic DTs by learning how to properly handle and fine tune a set of parameters [17]. This is where dataspaces come into play, integrating and facilitating the coexistence of data from various sources. In this way, it is possible to seamlessly integrate data generated by the PV with weather forecasts, for example, enabling more efficient use of energy resources and better management of energy consumption.

On the implementation side, a cluster of five Raspberry PI 4B is used as far edge data center. This device serves as Hadoop cluster, enabling distributed storage, processing and analytics of structured, unstructured and semi-structured data. The proximity edge node, instead, is implemented through a single Raspberry PI 4B, which runs a Mosquitto MQTT Broker to provide a message publisher/subscriber communication protocol between the system nodes.

A further addition to this setting consists of including a FIWARE context broker [18] on the far edge data center. FIWARE enables the deployment of context-aware smart solutions, integrating data from a variety of sources using standard Smart Data Models, and providing connectors to third-party components. In this way, it is possible to add a further layer atop the Hadoop Distributed File System (HDFS) to harness FIWARE context awareness functionalities. According to [8], FIWARE provides a platform to manage data spaces, by allowing the systems to share data using the API they prefer. The specification of each API can be published as a manifesto that can be dynamically processed by platform integrated components, in order to perform an automated adaptation from/to the APIs.

The system is scalable and allows the integration of multiple IoT devices with ease. One of the goals of its implementation is to optimize the energy management of PV panels, by tracking the delta between produced energy and consumed movement energy. In this way, it is possible to measure which is the best orientation update rate, also based on further context information. For instance, one can correlate energy production with atmospheric conditions, in the context of predictive analyses focused on a better power management. Furthermore, it can serve as a testbed to improve the accuracy of analytic models used to predict the behaviour of photovoltaic cells, by constantly updating a set of parameters used as input of such models, until the error between real-world model output and its digital counterpart's prediction is properly minimized.

4. Conclusion

The Internet of Things (IoT) has enabled more meaningful interactions between objects and people, but it also presents challenges in terms of integrating devices that have sensing, actuating, and controlling capabilities. Traditional cloud infrastructures are not designed to manage these data efficiently, which has led to the emergence of fog computing, edge computing, and dataspaces as potential solutions. The benefits of these technologies in improving latency, reliability, and offering faster response times are numerous. In particular, dataspaces can effectively manage the huge volumes of data generated by DTs and provide FAIRness-driven policies for data exchange.

These concepts can be applied to different scenarios, especially when DTs and urban DTs are involved, as demonstrated in the proposed case study which describes the development of a DT for a smart PV cyber-physical system. Yet, there are several potential extensions and future direction that are worth investigating further. For instance, future works can integrate different artificial intelligence approaches to data-driven predictive analysis and compare the outcomes precision, given the true reference set by the real-world counterpart of the digital replica.

Acknowledgements This work is partially supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU and by CETMA DIHSME.

References

[1] E. Commission, European industrial technology roadmap for the next generation cloud-edge offering, Technical Report, 2021. URL: https://ec.europa.eu/newsroom/repository/document/2021-18/European_CloudEdge_Technology_Investment_Roadmap_for_publication_pMdz85DSw6nqPppq8hE9S9RbB8_76223.pdf.

[2] Z. Wang, Digital twin technology, in: Industry 4.0 - Impact on Intelligent Logistics and Manufacturing, IntechOpen, 2020. URL: https://doi.org/10.5772/intechopen.80974. doi:10.5772/intechopen.80974.

[3] L. Deren, Y. Wenbo, S. Zhenfeng, Smart city based on digital twins, Computational Urban Science 1 (2021). URL: https://doi.org/10.1007/s43762-021-00005-y. doi:10.1007/s43762-021-00005-y.

[4] O. T. Eleftheriou, C.-N. Anagnostopoulos, Digital twins: A brief overview of applications,

challenges and enabling technologies in the last decade, Digital Twin 2 (2022) 2. URL: https://doi.org/10.12688/digitaltwin.17581.1. doi:10.12688/digitaltwin.17581.1.

[5] Q. Lu, A. K. Parlikad, P. Woodall, G. D. Ranasinghe, X. Xie, Z. Liang, E. Konstantinou, J. Heaton, J. Schooling, Developing a digital twin at building and city levels: Case study of west cambridge campus, Journal of Management in Engineering 36 (2020). doi:10.1061/(asce)me.1943-5479.0000763.

[6] A. J. Jara, Y. Bocchi, D. Fernandez, G. Molina, A. Gomez, An analysis of context-aware data models for smart cities: Towards fiware and etsi cim emerging data model, volume 42, International Society for Photogrammetry and Remote Sensing, 2017, pp. 43–50. doi:10.5194/isprs-archives-XLII-4-W3-43-2017.

[7] E. Curry, Dataspaces: Fundamentals, principles, and techniques, in: Real-time Linked Dataspaces, Springer International Publishing, 2019, pp. 45–62. URL: https://doi.org/10.1007/978-3-030-29665-0_3. doi:10.1007/978-3-030-29665-0_3.

[8] U. Ahle, J. J. Hierro, FIWARE for Data Spaces, Springer International Publishing, 2022. URL: https://link.springer.com/10.1007/978-3-030-93975-5_24. doi:10.1007/978-3-030-93975-5_24.

[9] F. Cirillo, G. Solmaz, E. L. Berz, M. Bauer, B. Cheng, E. Kovacs, A standard-based open source iot platform: Fiware, IEEE Internet of Things Magazine 2 (2020) 12–18. doi:10.1109/iotm.0001.1800022.

[10] E. C. D. G. for Informatics., New European interoperability framework: promoting seamless services and data flows for European public administrations., Publications Office, 2017. URL: https://data.europa.eu/doi/10.2799/78681. doi:10.2799/78681.

[11] H. Catapult, Untangling the requirements of a digital twin, Univ. Sheff. Adv. Manuf. Res. Cent. (AMRC) (2021) p7.

[12] G. Schrotter, C. Hürzeler, The digital twin of the city of zurich for urban planning, PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88 (2020) 99–112.

[13] F. Dembski, U. Wössner, M. Letzgus, M. Ruddat, C. Yamu, Urban digital twins for smart cities and citizens: The case study of herrenberg, germany, Sustainability 12 (2020) 2307.

[14] J. Ferré-Bigorra, M. Casals, M. Gangolells, The adoption of urban digital twins, Cities 131 (2022) 103905.

[15] B. Dukai, R. Peters, T. Wu, T. Commandeur, H. Ledoux, T. Baving, M. Post, V. van Altena, W. van Hinsbergh, J. Stoter, Generating, storing, updating and disseminating a countrywide 3d model, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIV-4/W1-2020 (2020) 27–32. URL: https://doi.org/10.5194/isprs-archives-xliv-4-w1-2020-27-2020. doi:10.5194/isprs-archives-xliv-4-w1-2020-27-2020.

[16] J. E. Stoter, A. Ohori, G. A. K. . Noardo, Digital twins a comprehensive solution or hopeful vision?, 2021. URL: www.3Dbag.nl.

[17] V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, K. Stefanidis, End-to-end entity resolution for big data: A survey, 2020. arXiv:1905.06397.

[18] B. Otto, M. ten Hompel, S. Wrobel (Eds.), Designing Data Spaces, Springer International Publishing, 2022. URL: https://link.springer.com/10.1007/978-3-030-93975-5. doi:10.1007/978-3-030-93975-5.