[DISCUSSION PAPER] Identifying key factors in designing data spaces for Urban Digital Twin Platforms: a data driven approach*

Cristian Martella[1,*,†], Angelo Martella[1,†], Amro Issam Hamed Attia Ramadan[1,†] and Antonella Longo[1]

[1]*University of Salento, Lecce, Italy*

Abstract

This study explores the design of data spaces for Urban Digital Twin Platforms using a data-driven approach. It interviews professionals from academia, urban design, engineering, research, software development, and other fields to identify 12 key factors organized in 4 categories: Data Support, Data Interoperability, Data Sovereignty, and Data Economy. The research aligns with European data space pillars and aims to guide future developments. The results emphasize the importance of managing diverse data formats for platform success and suggest that data interoperability and sovereignty will become more crucial as technologies mature. Future research should validate and expand on these characteristics to meet evolving data space requirements in smart cities.

Keywords

Urban Digital Twin, Data space, Urban data space, Data Interoperability, Data Sovereignty

1. Introduction

The rise of digital technology in smart cities is transforming hybrid spaces, requiring innovative data management approaches. Cities are evolving towards complex cyber-physical models, impacting digital systems and data supporting processes. Planning and designing public urban services are crucial for modeling and implementing the city's Digital Twin (DT), which is a crucial aspect of the integration of digital technologies in urban environments. Urban Digital Twins (UDTs) are being developed using various market solutions, but data management remains a significant challenge. Data spaces [1, 2], a common solution, offer efficient data sharing and exchange between different instances. These spaces aim to encompass the city-level to individual-level data ecosystem, facilitating the creation and development of UDTs in smart cities. The design and implementation of a data space within an Urban Data Twin (UDT) platform requires enabling key factors for efficient data sharing, ensuring it supports urban planning and decision-making capabilities. A survey was conducted to identify relevant factors for the design of a data space that must support and be integrated with a DT, including Urban DT. This research paper aims to match key factors proposed in the European context for data space design with the results of a survey conducted in collaboration with the Digital Twin Cities Centre (DTCC) research group at Chalmers University. The main European bodies involved in drafting data spaces development specifications include BDVA [3], IDSA [4], and Fiware [5]. The research focuses on key factors specialized professionals consider when designing a data space that supports and integrates with a Data Transport System (DT), including UDT. It is specific to academic, industrial, and research sectors and focuses on data management aspects within the design process. The study uses the Grounded Theory Methodology (GTM) to identify relevant characteristics in designing a UDT platform. The resulting data characteristics are organized into merit tiers, highlighting their influence on the decision-making

SEBD 2024: 32nd Symposium on Advanced Database Systems, June 23-26, 2024, Villasimius, Sardinia, Italy
*Corresponding author.
†These authors contributed equally.
cristian.martella@unisalento.it (C. Martella); angelo.martella@unisalento.it (A. Martella);
amroissam.ramadan@unisalento.it (A. I. H. A. Ramadan); antonella.longo@unisalento.it (A. Longo)
0000-0001-9751-9367 (C. Martella); 0000-0002-1082-7293 (A. Martella); 0000-0003-0042-5095 (A. I. H. A. Ramadan);
0000-0002-6902-0160 (A. Longo)

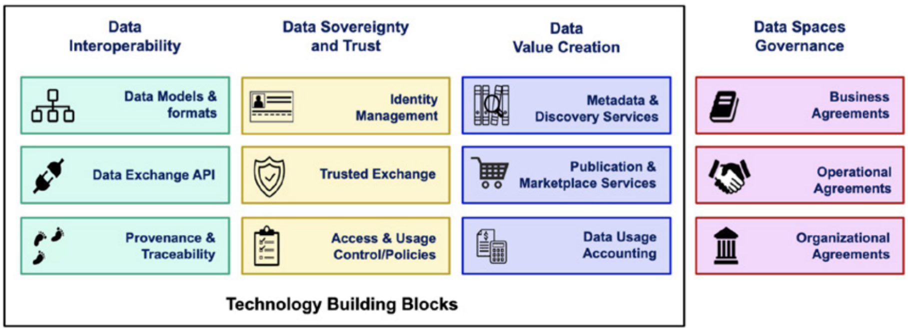

Figure 1: Building blocks of a Data Space [6].

process related to future UDT platform design. Four merit tiers have been identified, with essentials and supporting characteristics reported to ensure successful creation of UDTs that can natively integrate with data space instances. Data spaces require governance, which involves participants adhering to business, operational, and organizational agreements [6]. Business agreements regulate data exchange and legal constraints, operational agreements cover data spaces' policies, and organizational agreements ensure compliance with product specifications (as shown in Figure 1 [6]). This paper uses the technological pillars considered fundamental by European sector bodies as key concepts for identifying and organizing questionnaire sections for survey professionals and analyzing results.

The paper is structured into several sections, including background, methodology, results, discussion, and conclusions. Section 2 discusses the background of the study, whereas section 3 delves into the methodology used, and the definition of figures involved in the GTM-based survey. Section 4 provides a detailed description of the survey's results, while section 5 analyzes the survey results and provides final considerations. Section 6 presents potential future developments and conclusions.

2. Background and related work

This section introduces the fundamental concepts of data spaces, models, standards, and specifications in the paper dissertation. Several approaches exist, particularly in the European context. Basic concepts include DTs [7, 8, 9] and UDTs [10, 11, 12], which are well-documented in literature. A specific contribution is proposed in [13], providing a hypothesis of UDT architecture and a comparative analysis of license-free UDT platforms. This paper can be considered a sequel to the current research work.

2.1. Data Space

Data space is a technology that enables data management by creating a data ecosystem where stakeholders can exchange and share data. This seamless data sharing can provide value, especially when combined with data analytics. In a smart city, public transportation companies and local businesses can collaborate in a data space, benefiting from greater retail demand forecasts and optimizing traffic management. This data economy contributes to the creation of a smart city, as the Internet of Things has already addressed the concept within the context of smart cities [14].

2.2. Data Spaces in Europe

Gaia-X [15] and the International Data Spaces Association (IDSA) [16] are European initiatives focused on studying models, specifications, and standards for data spaces. Their synergies [17] have sparked interest in Europe [18, 19] and beyond [20]. Open ecosystems based on standards, like Fiware [21], can provide building blocks for data platforms, such as data brokering via standardized data models [22]. Data spaces involve data suppliers, consumers, and service providers, allowing them to exchange data and accommodate a wide range of data sources with varying models. Service providers can use the

Role	Professionals
Academics	P1, P3, P8, P11
Urban designer	P2, P6, P11
Engineers and architects	P2, P5, P6, P11, P12
Researcher	P2, P4, P5, P6, P12
Software architects	P3, P4, P7, P9, P10
CEO	P7, P8
CTO	P9
Data scientist	P13
Software developer	P6, P13

Table 1
Professional contacts and their occupation

shared data space for data analytics services, while data consumers can share analytics results. IDSA, Gaia-X, and Fiware are developing data connectivity building blocks. The Big Data Value Association (BDVA) is another significant European initiative [2, 23], focusing on data management principles and techniques, including data life cycle management, usage of data lakes and spaces, underlying data storage services, and connection to data sharing platforms.

2.3. Data Space and Digital Twin Integration

The literature discusses the integration of Digital Twins (DTs) and data spaces. A prototype [24] demonstrates that product carbon footprints can be stored in a DT's data space instance and shared with participants. Smart cities are shifting from technical to socio-technical perspectives [25], aiming to involve citizens in urban planning strategy decision-making. However, these DTs face challenges due to their limited variables and processes. A Living Lab is proposed in [26] to demonstrate the interaction between UDTs under the UN's Sustainable Development Goals of sustainable cities, health, well-being, industry, innovation, and infrastructure. The system processes real-time data using onboard sensing devices, edge computing paradigms, and Machine Learning algorithms. The authors highlight the constant conflict between what is real and ideal in urban-scale DTs.

3. Methodology

The study employs an exploratory methodology based on the principles of GTM [27], which involves two iterative steps: data collection and analysis. Data collection involves in-depth interviews, transcripts, questionnaires, and recordings to provide a comprehensive observation of a phenomenon. Data analysis extracts meaningful insights and shapes the theoretical framework of the phenomenon. In [28], the GTM is used to evaluate existing digital twins (DTs) within different application domains of Cyber-Physical Systems. The authors in [29] propose an analytical approach that compares DT instances and universally define and structure digital twins using a feature-based digital twin framework (FDTF). In this work, the GTM methodology is used to organize qualitative interviews with professionals involved in multidisciplinary projects related to DTs. The interviews involve professionals from various industries, including enterprises, startups, engineers, architects, technicians, and universities across Europe, as listed in Table 1. The questionnaire comprises sections about the degree of knowledge about DT development platforms and their support evaluation in terms of data management. Some interviews were conducted in presence, while the major parts were systematically recorded and subjected to Natural Language Processing-based analysis techniques. Interviews are organized according to the professional's role, with professionals with multiple roles participating in each round. Cross-references are made by matching the results obtained from the survey with models, standards, and specifications currently being studied at the European level. The methodology develops distinct and homogeneous characteristics that align with the existing body of literature related to fundamental pillars for DS design.

Category	Subcategory	Academics	Urban Designers	Engineers and Architects	Researchers	Software Architects	CEO	CTO	Data Scientists	Software Developers	Subcategory score	Category score
Data Support	**Input Data**	4	3	5	5	5	2	1	1	2	**28**	**116**
	Output Data	4	3	5	5	5	2	1	1	2	**28**	
	Object Representation	2	3	3	2	2	1	1	0	1	**15**	
	Simulation Output Data	4	2	4	5	4	2	1	1	2	**25**	
	Metadata Support	4	1	1	4	4	2	1	1	2	**20**	
Data Interoperability	**Standard and Open Data Formats**	4	1	1	4	4	2	1	1	2	**20**	**65**
	Data Provenance and Traceability	4	2	3	3	2	2	1	1	0	**18**	
	Data Exchange APIs	3	0	0	3	3	0	1	0	2	**12**	
	Data Integration	4	1	0	3	5	0	1	0	1	**15**	
Data Sovereignty	**Trusted Data Exchange**	3	0	0	4	4	2	1	0	1	**15**	**37**
	Control/Mechanisms on Data Access and Usage	3	2	4	3	4	2	1	1	2	**22**	
Data Economy		4	0	0	4	1	2	0	0	0	**11**	**11**

Table 2
Data types characteristics used for the creation of UDT

4. Results

The study identifies 12 key factors for designing data spaces for Urban Digital Twin Platforms based on professional interviews. These factors are organized into four categories: Data Support, Data Interoperability, Data Sovereignty, and Data Economy. These categories are based on European data space pillars (discussed in section 2) and will be further explored in future developments. The Data Support category includes input and output data, object representation, simulation output data, and metadata support subcategories. Data Interoperability includes standard and open data formats, data provenance and traceability, data exchange APIs, and data integration subcategories. Data Sovereignty includes trusted data exchange and control mechanisms on data access and usage. The study provides a detailed description of these 12 key factors and proposes further discussion into the four categories introduced.

4.1. Data Support

The Data Support category comprises key subcategories related to the requirements and features a UDT platform must offer. These include Input Data, Output Data, Object Representation, Simulation Output Data, and Metadata Support. Each subcategory is detailed, with a focus on data sources and sinks, identifying their main characteristics and importance in the context of UDT platforms.

Input Data Urban Digital Twin (UDT) platforms use aerial photography, LIDAR point clouds, and cadastral data to create accurate 3D models of urban environments [13]. Additional input data, such as live IoT sensor streams, historical collections of samples, and external web APIs, may also be required. IoT streams are encoded based on communication protocol requirements, with lightweight data formats like JSON and Ultralight 2.0 being popular [30, 31]. Historical collections provide timed behavioral representations and machine learning models for simulation. Popular storage options include tabular files like CSV and collections of JSON documents [32]. Web APIs are valuable sources of information for real-time simulations and decision-making [33, 34]. Standard formats are necessary to organize input data, fostering interoperability between systems [34], allowing policy application, and ensuring trusted data sharing between participants. This ensures interoperability, shared lingua, and trusted data sharing in the supporting data space for an UDT [6].

Output Data The assessment of a generic UDT platform involves considering various outputs related to various contexts, such as the 3D City Model, Solar irradiance, Solar shadow, citizen interaction, future city development, digital infrastructure for IoT services, and energy demand. These outputs are derived from commonly used scenarios in research. For example, solar irradiance and solar shadow can be traced back to simulation and analysis outputs, while the 3D city model and digital infrastructure for IoT services and energy demand are crucial outputs for planning a smart city [13].

Object Representation and Level of Details Object representation is crucial in choosing a Universal Design Tool (UDT) platform as it affects the accuracy and complexity of displayed models, but also requires higher rendering performance. The Level of Detail (LOD) metric helps identify the representation complexity of building geometries, from footprints to complex 3D geometries and textures. Other UDT platform characteristics include natural elements representation, building components representation, and representation of city infrastructures and service networks on surface and underground [13].

Simulation Output Data An UDT platform uses simulation output data to integrate tools and display phenomena using dashboards and interactive visualizations. It uses high-performance computing and AI-driven predictive tools to analyze what-if scenarios. The platform offers various solutions for analyzing geotechnical behavior, urban development, solar shadowing, sea level rise effects, weather conditions, air quality, and noise models, enabling accurate analyses based on what-if scenarios [13].

Metadata Support Digital twins rely on consistent data and information [35], and metadata can enhance the quality of data and simplify interoperability between systems [36]. Attribute metadata can provide a uniform shared context for data exchanges, reducing data mismatch and fostering accurate analyses. Attribute metadata integration can play a key role in selecting the most suitable UDT platform, enabling effective system integration and collaboration using a common lingua [34] (more on it in section 4.2). Additionally, metadata can include data structure information like keys, indexes, and columns, allowing search capabilities and interoperability among open data platforms [37].

4.2. Data Interoperability

In the data economy era, organizations often operate as isolated data silos, referred to by a few specific software systems or platforms. This presents a challenge in supporting data sharing and exchange. Data interoperability is a key challenge, and organizations must be prepared to comply with new technologies like data spaces that aim to break data silos and foster interoperability. To achieve interoperability, organizations must align their underlying data and data models. Standardization is crucial for achieving full data interoperability [38], but the heterogeneity of data sources and data is a significant challenge. Data access and models are key factors for data interoperability, as they must be agreed upon and shared for efficient and transparent data exchange. Initiatives like IDSA [39] and Gaia-X [40] aim to define technical specifications of data spaces, enabling concepts of data sovereignty and trust.

Standard and Open Data Formats Standard and open data formats are crucial for data sharing and exchanging in data spaces [6], as they enable a common lingua definition for seamless integration. However, the main challenge lies in mechanisms of data sharing/exchange that are independent of corresponding protocols and data formats. LinkedScales is a multiscale-based data space architecture that uses a graph-based integration process over a graph database [41], implementing an integration and enrichment pipeline to incrementally obtain ontology-like data structures from raw data representations. Linked Open Data provides high-quality retrievals in Exploratory Search Systems (ESSs) [42], while JSON-LD is a JSON-based format for serializing Linked Data [43], designed to be easily integrated into deployment environments that already use JSON as the reference data format. JSON-LD is a reference format for urban data management using emerging technologies [44], while Fiware enhances data quality [36] through metadata integration. CityGML is an international open standard for representing, exchanging, and storing 3D city models, supporting various applications like urban planning, environmental simulation, disaster management, and navigation in UDTs [45].

Data Provenance and Traceability Data provenance refers to the origin and chronology of data, including its inception, ownership, and utilization. Cryptographic measures are used to ensure data integrity and security, while data creation and storage procedures establish connections between current events and their preceding occurrences. This results in transparent and tamper-resistant documentation of data origins [46]. Data traceability involves monitoring and tracing data provenance throughout its life cycle, including preservation, processing, and access stages. Blockchain infrastructure can be used to achieve comprehensive data traceability through meticulous documentation and auditing of all data stages. This process ensures the reliability and credibility of produced data by capturing and documenting essential information at every stage of data generation [47]. Integrating data provenance and traceability within a digital twin framework enhances accountability, transparency, and credibility. Blockchain technology is used to establish a verified record of asset transactions, ensuring data authenticity and traceability [48].

Data Exchange APIs The data economy's limited expansion is due to heterogeneity in data access, which hinders the implementation of solutions relying on diverse sources. To ensure data interoperability within data spaces, consensus on technological interfacing and data modeling must be established. Current initiatives focused on establishing technical soft infrastructure for data spaces do not address data modeling, as their specifications only focus on metadata exchange rather than actual data. Additionally, there is a lack of standardization and harmonization in data distribution specifications across various data-providing platforms. The Next Generation Service Interfaces Linked Data (NGSI-LD) standard [49], developed by the European Telecommunications Standards Institute (ETSI), offers a comprehensive specification for managing context data. It enhances the accessibility of contextual information by establishing Application Programming Interfaces (API) and data models that stakeholders can use within a given data environment. Fiware offers a suite of open-source components that can be effectively used in building data platforms, including the Context Broker, which incorporates the NGSI-LD API [50]. The NGSI-LD API is based on an abstract information model centered around entities with various characteristics, types, properties, and relationships. Several ongoing initiatives are focused on developing a corpus of NGSI-LD-compatible data models, providing a standardized reference for semantically modeling data that will be exchanged within future data spaces. One notable program is the Smart Data Models program [51], which supports semantic interoperability of context information within data spaces.

Data Integration The rapid growth of data generation, processing, and storage is due to the recognition of data as a crucial resource for organizations, fostering innovation and value creation [52]. The integration of information technologies facilitates data utilization for economic purposes [53], and organizations engage in collaborative efforts within value-oriented socio-technical networks [54, 55] to share and transfer data [56, 57]. However, the establishment and sustainability of these ecosystems face challenges [57] such as digitization [58], diverse data sources integration [59], external data incorporation [60], and addressing organizations' hesitancy to share data [61]. The concept of data spaces offers a potential solution to these challenges, particularly technical complexities associated with integrating diverse datasets [1]. Information systems research has shown increasing attention towards data ecosystems [62, 63, 64, 65], with political bodies like the European Union supporting entities to promote data sovereignty, innovation, and organizational competitiveness [66]. However, the exact nature of the connection between data spaces and data ecosystems remains uncertain.

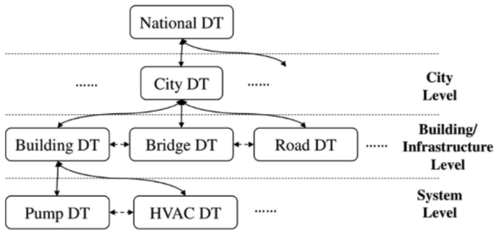

Figure 2: Digital twins connections and hierarchy levels

4.3. Data Sovereignty

Urban digital twins (DTs) are a family of digital assets that range from building assets to smart buildings and cities. UDT platforms aim to coordinate them in a federated fashion using schemas like the layered hierarchy proposed in [67] and also discussed in [68] (see Figure 2 [67]). Each DT is designed for high-performance bidirectional interconnections with multiple data sources, including other digital twins. However, data security and privacy issues are growing, and data sovereignty addresses these concerns. To achieve these goals, trusted data exchange paradigms and data access and usage control mechanisms are implemented. Both are discussed in the following paragraphs.

Trusted Data Exchange Trusted data security measures are crucial for the confidentiality, integrity, and availability of sensitive data exchanged between nodes in the smart city ecosystem [69]. These measures include sensitive data handling, data privileges, user activity logging, pseudonymization, and privacy-aware data interlinking services [70]. In UDTs, the data space layer should manage data exchanges effectively while providing Fair Access (FAIR) [68] policies and ensuring the authenticity and trustworthiness of data sources in the UDT's data space ecosystem [71].

Control Mechanisms on Data Access and Usage Data sovereignty is essential in data spaces, as it safeguards sensitive data from unauthorized access and malicious exploitation. Verifiable credentials, based on data ownership and selective disclosure, ensure transparency, provenance, and reliability. This allows secure data sharing between functional units in smart cities while maintaining control over access and purpose [71]. Implementing effective data governance and control mechanisms can be challenging, especially in inter-organizational data-sharing networks [72]. Decentralized access control, which turns a blockchain into an automated access manager, allows users to control their data without relying on third parties [73]. Knowledge-control regimes for exchanging samples and sequences can also control data access and use [74].

4.4. Data Economy

The data economy is the value generated by data collection, storage, and analysis within a data space [6]. International initiatives like Fiware, part of the Gaia-X European Association for Data and Cloud alliance, the Big Data Value Association (BDVA), and the International Data Spaces Association (IDSA) aim to accelerate business transformation in the data economy, enabling efficient near real-time data exchange [75]. Data markets and data exchange services are being addressed to enhance the data value chain, bridging the gap with traditional data value chains for consumable goods [76]. Emerging data-driven technologies, such as data spaces, are fostering interest in making data a new economic value by identifying and assigning new data properties, such as STREAM (Sovereign, Trusted, Reusable, Exchangeable, Actionable, and Measurable) data properties [76].

5. Discussion

The survey results discussed in section 4 and illustrated in Table 2 have identified four merit-based ranking tiers for data types characteristics (DT) in a data management solution. The **core aspects** (25-28 rank) include essential aspects such as input and output data, **fundamental aspects** (20-24 rank) include metadata support, standard and open data formats, and control mechanisms on data access and usage, **relevant aspects** (15-19 rank) include data provenance and traceability, data integration, trusted data exchange, and object representation, and **slightly relevant aspects** (11-14 rank) include data exchange APIs and data economy. These tiers are based on the scores achieved in the survey and are crucial for ensuring the development of DT solutions. The survey's findings provide valuable insights into the data management needs of DT developers. The evaluation framework for selecting an UDT platform focuses on data management subcategories. Data Support is the most important factor, with four out of six merit tiers belonging to this category. Data Interoperability and Data Sovereignty are considered on average significant, while the Data Economy category appears to be the least relevant. The analysis considers the requirements and maturity of reference technologies for UDT development platforms, such as managing and generating a wide range of data formats, supporting adequate data exchange and sharing mechanisms, and supporting the data value chain. The lack of precise protocols and standards in data spaces reduces the need for guaranteeing aspects related to Data Interoperability and Data Sovereignty, while the reluctance of companies to share their data and the perception of data as a source of value reduces the need for support for the Data Economy. The results underline that Data Support is the most important factor to consider when selecting an UDT development solution. Despite the limited availability of mature technologies, the Data Interoperability and Data Sovereignty categories still highlight professionals' strong perception of their potential relevance.

6. Conclusion

The study analyzed four main characteristics of data used for urban digital twin (UDT) creation and their 11 subcategories. 13 professionals rated each characteristic and subcategory's impact on choosing a UDT platform using a GTM methodology. The results showed that platform ability to manage diverse data formats is crucial for UDT platform success. Data interoperability and sovereignty were less influential, but their full potential will be unleashed as technologies mature. Future research should validate and refine identified characteristics, as data spaces and UDTs evolve with professionals' needs.

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 (PNRR-HPC, CUP:C83C22000560007). We are grateful to the Digital Twins Cities Centre for their invaluable collaboration and support, that made this survey a success.

References

[1] M. Franklin, A. Halevy, D. Maier, From databases to dataspaces: A new abstraction for information management, SIGMOD Rec. 34 (2005) 27–33. URL: https://doi.org/10.1145/1107499.1107502. doi:10.1145/1107499.1107502.

[2] E. Curry, S. Scerri, T. Tuikka, Data Spaces: Design, Deployment and Future Directions, Springer Nature, 2022.

[3] BDVA Data Sharing Practices Survey, Bdva data sharing practices survey, https://www.bdva.eu/-DataSharingPracticesSurvey, 2024. Last access 03-11-2023.

[4] The Data Spaces Support Centre conducts survey, The data spaces support centre conducts survey, https://internationaldataspaces.org/the-data-spaces-support-centre-conducts-survey/, 2024. Last access 03-11-2023.

[5] Smart Cities Survey, Smart cities survey, https://www.fiware.org/about-us/smart-cities/smart-cities-survey/, 2024. Last access 03-11-2023.
[6] U. Ahle, J. J. Hierro, FIWARE for Data Spaces, Springer International Publishing, 2022, pp. 395–417. URL: https://link.springer.com/10.1007/978-3-030-93975-5_24. doi:10.1007/978-3-030-93975-5_24.
[7] Q. Qi, F. Tao, Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison, Ieee Access 6 (2018) 3585–3593.
[8] D. Jones, C. Snider, A. Nassehi, J. Yon, B. Hicks, Characterising the digital twin: A systematic literature review, CIRP journal of manufacturing science and technology 29 (2020) 36–52.
[9] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: Enabling technologies, challenges and open research, IEEE access 8 (2020) 108952–108971.
[10] V. V. Lehtola, M. Koeva, S. O. Elberink, P. Raposo, J.-P. Virtanen, F. Vahdatikhaki, S. Borsci, Digital twin of a city: Review of technology serving city needs, International Journal of Applied Earth Observation and Geoinformation (2022) 102915.
[11] Y. Chang, I. Jang, Technology trends in digital twins for smart cities, Electronics and telecommunications trends 36 (2021) 99–108.
[12] M. Jafari, A. Kavousi-Fard, T. Chen, M. Karimi, A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future, IEEE Access (2023).
[13] A. Martella, A. I. H. A. Ramadan, C. Martella, M. Patano, A. Longo, State of the Art of Urban Digital Twin Platforms, 2023, pp. 299–317. URL: https://link.springer.com/10.1007/978-3-031-43401-3_20. doi:10.1007/978-3-031-43401-3_20.
[14] L. Sánchez, J. Lanza, L. Muñoz, From the internet of things to the social innovation and the economy of data, Wirel. Pers. Commun. 113 (2020) 1407–1421.
[15] G. Eggers, B. Fondermann, B. Maier, K. Ottradovetz, J. Pfrommer, R. Reinhardt, H. Rollin, A. Schmieg, S. Steinbuß, P. Trinius, et al., Gaia-x: Technical architecture, Federal Ministry for Economic Affairs and Energy (BMWi) Public Relations Division, Berlin 6 (2020).
[16] International data spaces association. 2021. international data spaces enabling data economy, IDSA Brochure (2021).
[17] International Data Spaces Association. 2021. Gaia-X and IDS. Position Paper, Version 1, 2021.
[18] S. Autolitano, A. Pawlowska, Europe's quest for digital sovereignty: Gaia-x as a case study, IAI papers 21 (2021) 1–22.
[19] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to european digital sovereignty with gaia-x and idsa, IEEE network 35 (2021) 4–5.
[20] A. Sakaino, International collaboration between data spaces and carrier networks, in: Designing Data Spaces: The Ecosystem Approach to Competitive Advantage, Springer International Publishing Cham, 2022, pp. 471–483.
[21] 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 (2019) 12–18.
[22] Scorpio broker documentation, https://scorpio.readthedocs.io/en/latest/., ???? Accessed: 2023-11-3.
[23] S. Scerri, T. Tuikka, I. Lopez de Vallejoan, Towards a european data sharing space, Report. Big Data Value Association (2020).
[24] F. Volz, G. Sutschet, L. Stojanovic, T. Usländer, On the role of digital twins in data spaces, Sensors 23 (2023) 7601.
[25] M. Charitonidou, Urban scale digital twins in data-driven society: Challenging digital universalism in urban planning decision-making, International Journal of Architectural Computing 20 (2022) 238–253.
[26] D. M. Botín-Sanabria, J. G. Lozoya-Reyes, R. C. Vargas-Maldonado, K. L. Rodríguez-Hernández, R. A. Ramírez-Mendoza, M. A. Ramírez-Moreno, J. d. J. Lozoya-Santos, Digital twin for urban spaces: An application, in: Proceedings of the International Conference on Industrial Engineering and Operations Management, Monterrey, Mexico, 2021, pp. 3–5.
[27] H. Noble, G. Mitchell, What is grounded theory?, Evidence-based nursing 19 (2016) 34–35.

[28] K. Josifovska, E. Yigitbas, G. Engels, Reference framework for digital twins within cyber-physical systems, in: 2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS), IEEE, 2019, pp. 25–31.
[29] J. Autiosalo, J. Vepsäläinen, R. Viitala, K. Tammi, A feature-based framework for structuring industrial digital twins, IEEE access 8 (2019) 1193–1208.
[30] T. Storek, J. Lohmöller, A. Kümpel, M. Baranski, D. Müller, Application of the open-source cloud platform fiware for future building energy management systems, volume 1343, Institute of Physics Publishing, 2019. doi:10.1088/1742-6596/1343/1/012063.
[31] J. Conde, A. Munoz-Arcentales, A. Alonso, S. López-Pernas, J. Salvachua, Modeling digital twin data and architecture: A building guide with fiware as enabling technology, IEEE Internet Computing 26 (2022) 7–14. doi:10.1109/MIC.2021.3056923.
[32] M. Fresta, A. Capello, F. Bellotti, L. Lazzaroni, M. Cossu, R. Berta, Supporting a .csv-based workflow in mongodb for data analysts, IEEE, 2023, pp. 1–4. doi:10.1109/ISIE51358.2023.10228044.
[33] R. Ala-Laurinaho, J. Autiosalo, A. Nikander, J. Mattila, K. Tammi, Data link for the creation of digital twins, IEEE Access 8 (2020) 228675–228684. doi:10.1109/ACCESS.2020.3045856.
[34] M. Jacoby, T. Usländer, Digital twin and internet of things-current standards landscape, 2020. doi:10.3390/APP10186519.
[35] B. Vogel-Heuser, F. Ocker, I. Weiß, R. Mieth, F. Mann, Potential for combining semantics and data analysis in the context of digital twins, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379 (2021) 20200368. doi:10.1098/rsta.2020.0368.
[36] J. Conde, A. Munoz-Arcentales, A. Alonso, G. Huecas, J. Salvachua, Collaboration of digital twins through linked open data: Architecture with fiware as enabling technology, IT Professional 24 (2022) 41–46. URL: https://ieeexplore.ieee.org/document/10017408/. doi:10.1109/MITP.2022.3224826.
[37] A. Ojo, L. Porwol, M. Waqar, A. Stasiewicz, E. Osagie, M. Hogan, O. Harney, F. A. Zeleti, Realizing the Innovation Potentials from Open Data: Stakeholders' Perspectives on the Desired Affordances of Open Data Environment, 2016, pp. 48–59. doi:10.1007/978-3-319-45390-3_5.
[38] S. Thirumuruganathan, N. Tang, M. Ouzzani, A. Doan, Data curation with deep learning., in: EDBT, 2020, pp. 277–286.
[39] International data spaces association. 2021. international data spaces enabling data economy, IDSA Brochure (2021).
[40] Gaia-X - Architecture Document - 22.04 release, 2022.
[41] M. S. Mota, F. L. Pantoja, J. C. dos Reis, A. Santanchè, Progressive data integration and semantic enrichment based on linkedscales and trails, in: Workshop on Semantic Web Applications and Tools for Life Sciences, 2016. URL: https://api.semanticscholar.org/CorpusID:18522320.
[42] K. Jacksi, N. Dimililer, S. R. M. Zeebaree, State of the art exploration systems for linked data: A review, International Journal of Advanced Computer Science and Applications 7 (2016). URL: https://api.semanticscholar.org/CorpusID:41500402.
[43] G. Kellogg, P.-A. Champin, D. Longley, Json-ld 1.1 – a json-based serialization for linked data, 2019. URL: https://api.semanticscholar.org/CorpusID:202784811.
[44] J. Dambruch, A. Stein, V. Ivanova, Innovative Approaches to Urban Data Management using Emerging Technologies, 2016. URL: https://api.semanticscholar.org/CorpusID:196072984.
[45] T. Kutzner, K. Chaturvedi, T. H. Kolbe, Citygml 3.0: New functions open up new applications, PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88 (2020) 43–61. doi:10.1007/s41064-020-00095-z.
[46] J. Xia, H. Li, Z. He, The effect of blockchain technology on supply chain collaboration: A case study of lenovo, Systems 11 (2023) 299. doi:10.3390/systems11060299.
[47] A. Pantazidis, A. Gazis, J. Soldatos, M. Touloupou, E. Kapassa, S. Karagiorgou, Trusted virtual reality environment for training security officers, in: 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), IEEE, 2023, pp. 518–524.
[48] A. Kalafatelis, K. Panagos, A. E. Giannopoulos, S. T. Spantideas, N. C. Kapsalis, M. Touloupou, E. Ka-

passa, L. Katelaris, P. Christodoulou, K. Christodoulou, et al., Island: An interlinked semantically-enriched blockchain data framework, in: Economics of Grids, Clouds, Systems, and Services: 18th International Conference, GECON 2021, Virtual Event, September 21–23, 2021, Proceedings 18, Springer, 2021, pp. 207–214.

[49] NGSI-LD API. Technical Report, Context information management (cim) etsi industry specifcation group (isg), www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.04.01-_60/gs_cim009v010401p.pdf, 2021. Last access 25-10-2023.

[50] Scorpio Broker, Fiware scorpio broker, https://scorpio.readthedocs.io/en/latest/, 2020. Last access 25-10-2023.

[51] H. Jilin, Smart data models, https://smartdatamodels.org/., ???? Accessed: 2023-10-31.

[52] B. Otto, R. Bärenfänger, S. Steinbuß, Digital Business Engineering: Methodological Foundations and First Experiences from the Field, 2015.

[53] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, F. Feldberg, Debating big data: A literature review on realizing value from big data, J. Strat. Inf. Syst. 26 (2017) 191–209.

[54] F. Ullah, L. Shen, S. H. H. Shah, Value co-creation in business-to-business context: A bibliometric analysis using histcite and vos viewer, Frontiers in Psychology 13 (2023) 1027775.

[55] M. I. S. Oliveira, B. F. Lóscio, What is a data ecosystem?, in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, dg.o '18, Association for Computing Machinery, New York, NY, USA, 2018. URL: https://doi.org/10.1145/3209281.3209335. doi:10.1145/3209281.3209335.

[56] F. Kitsios, N. Papachristos, M. Kamariotou, Business models for open data ecosystem: Challenges and motivations for entrepreneurship and innovation, in: 2017 IEEE 19th Conference on Business Informatics (CBI), IEEE, 2017. doi:https://doi.org/10.1109/CBI.2017.51.

[57] Y. Yoo, O. Henfridsson, K. Lyytinen, Research commentary—the new organizing logic of digital innovation: An agenda for information systems research, Inf. Syst. Res. 21 (2010) 724–735.

[58] F. de Prieelle, M. de Reuver, J. Rezaei, The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry, IEEE Trans. Eng. Manage. 69 (2022) 940–950. doi:https://doi.org/10.1109/TEM.2020.2966024.

[59] J. Lu, L. T. Yang, B. Guo, Q. Li, H. Su, G. Li, J. Tang, A sustainable solution for IoT semantic interoperability: Dataspaces model via distributed approaches, IEEE Internet Things J. 9 (2022) 7228–7242.

[60] R. Arnold, C. Hildebrandt, S. Taş, European Data Economy: Between Competition and Regulation, 2020.

[61] C. Kaiser, A. Stocker, M. Fellmann, Understanding Data-Driven service ecosystems in the automotive domain, in: 25th Americas Conference on Information Systems (AMCIS), Cancun, Mexiko, 2019.

[62] R. Abraham, J. Schneider, J. vom Brocke, Data governance: A conceptual framework, structured review, and research agenda, Int. J. Inf. Manage. 49 (2019) 424–438.

[63] J. Gelhaar, T. Gürpinar, M. Henke, B. Otto, Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems, 2021.

[64] S. U. Lee, L. Zhu, R. Jeffery, Data governance for platform ecosystems: Critical factors and the state of practice (2017). arXiv:1705.03509.

[65] M. Schreieck, M. Wiesche, H. Krcmar, Design and Governance of Platform Ecosystems-Key Concepts and Issues for Future Research, 2016.

[66] E. Curry, A. Metzger, S. Zillner, J.-C. Pazzaglia, The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem, Springer International Publishing, Cham, 2021. doi:https://doi.org/10.1007/978-3-030-68176-0.

[67] 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.

[68] C. Martella, A. Longo, M. Zappatore, A. Ficarella, Dataspaces in urban digital twins: a case study in the photovoltaics, volume 3478, 2023.
[69] S.-K. Gan, T.-L. Wong, C.-P. Goh, W.-P. Lee, Y.-M. Lim, A review on the development of dataspace connectors using microservices cross-company secured data exchange, 2021. URL: https://www.i-scoop.eu/industry-4-0/industrial-data-space/.
[70] J. Hernandez, L. McKenna, R. Brennan, Tikd: A trusted integrated knowledge dataspace for sensitive healthcare data sharing, IEEE, 2021, pp. 1855–1860. doi:10.1109/COMPSAC51774.2021.00280.
[71] Z. Pervez, Z. Khan, A. Ghafoor, K. Soomro, Signed: Smart city digital twin verifiable data framework, IEEE Access 11 (2023) 29430–29446. doi:10.1109/ACCESS.2023.3260621.
[72] M. Hellmeier, J. Pampus, H. Qarawlus, F. Howar, Implementing data sovereignty: Requirements & challenges from practice, in: Proceedings of the 18th International Conference on Availability, Reliability and Security, 2023, pp. 1–9.
[73] J. Ernstberger, J. Lauinger, F. Elsheimy, L. Zhou, S. Steinhorst, R. Canetti, A. Miller, A. Gervais, D. Song, Sok: Data sovereignty, Cryptology ePrint Archive (2023).
[74] L. Fearnley, Viral sovereignty or sequence etiquette? asian science, open data, and knowledge control in global virus surveillance, East Asian Science, Technology and Society: An International Journal 14 (2020) 479–505.
[75] Bdva, fiware, gaia-x and idsa launch an alliance to accelerate business transformation in the data economy, 2021. URL: https://design-principles-for-data-spaces.org/].
[76] Y. Demchenko, W. Los, C. D. Laat, Data as economic goods: Definitions, properties, challenges, enabling technologies for future data markets, 2018. URL: https://www.itu.int/en/journal/002/Pages/default.aspx.