Digital Twins in Healthcare: A forefront for knowledge representation techniques

Swarnendu Ghosh[1,2,†]

[1]IEM Centre of Excellence for Data Science, Kolkata, WB, India
[2]Institute of Engineering & Management (UEM), Kolkata, WB, India

Abstract
Digital twins have recently gathered significant interest in the healthcare community. This concept promises to unlock various previously unavailable services such as remote monitoring, advanced visualization, simulation of medical procedures, predictive analytics, demographic studies, and so on. At present research in this area is localized and conducted independently. Thus, effective deployment of digital twins in healthcare is still a work in progress due to inconsistent data representation and isolated innovation without effective integration at large scale. Knowledge representation plays a vital role in structuring, integrating, and reasoning over heterogeneous healthcare data sources such as electronic health records, genomics data, clinical guidelines, reports, medical literature, and more. The process of digitization is relevant not only to patients but also to healthcare professionals, infrastructure facilities, devices, insurance providers, and even historical records. This work proposes to thoroughly highlight this research gap and the current initiatives addressing these issues. It aims to review and consolidate existing efforts in standardizing data structures for healthcare digital twins, with a focus on interoperability, representation and integration across diverse healthcare domains.

Keywords
Digital Twins, Healthcare Informatics, Knowledge Representation.

1. Introduction

The uncivilized human species have solely depended on their biological immunity and behavioural traits for treating physical ailments[1]. The earliest civilizations started adopting unorthodox practices to aid in the general well-being and lifestyle improvement. There have been numerous records of herbal remedies, spiritual and yogic practices, and even primitive surgical methods in ancient civilizations like the Egyptian, Indian, Greeks and so on [2]. Around the 15th century, the Renaissance period introduced several aspects of modern medicine such as diagnosis, anatomical studies and various surgical tools. As time progressed each century brought us new concepts such as the Germ theory[3] and introduction of anesthesia[4] in the 19th century and other advancements like vaccines[5], antibiotics[6], radiology[7], ECG[8] and so on in the 20th century. This accelerated innovation has continued to increase the periphery of modern medicine. With the dawn of the 21st century, the age of automation took over. With

2nd Symposium on NLP for Social Good (NSG), 25-26 April, 2024, University of Liverpool, United Kingdom
[†]This work is done with infrastructure support from the Innovation & Entrepreneurship Development Cell, IEM Kolkata, India
drghosh90@gmail.com (S. Ghosh)
0000-0002-2220-1677 (S. Ghosh)

immense advancement in machine learning, internet of things, 5G, cloud computing platforms, healthcare informatics have taken a new turn[9]. Present efforts are being focused on the automation of the diagnosis, robot assisted surgeries, remote clinical processing and more. The complete mapping of the human genome[10] has led to a new era of precision medicine that combines genomic and proteomic studies for drug design. These advancements have created the hotbed for the emergence of the concept of digital twin.

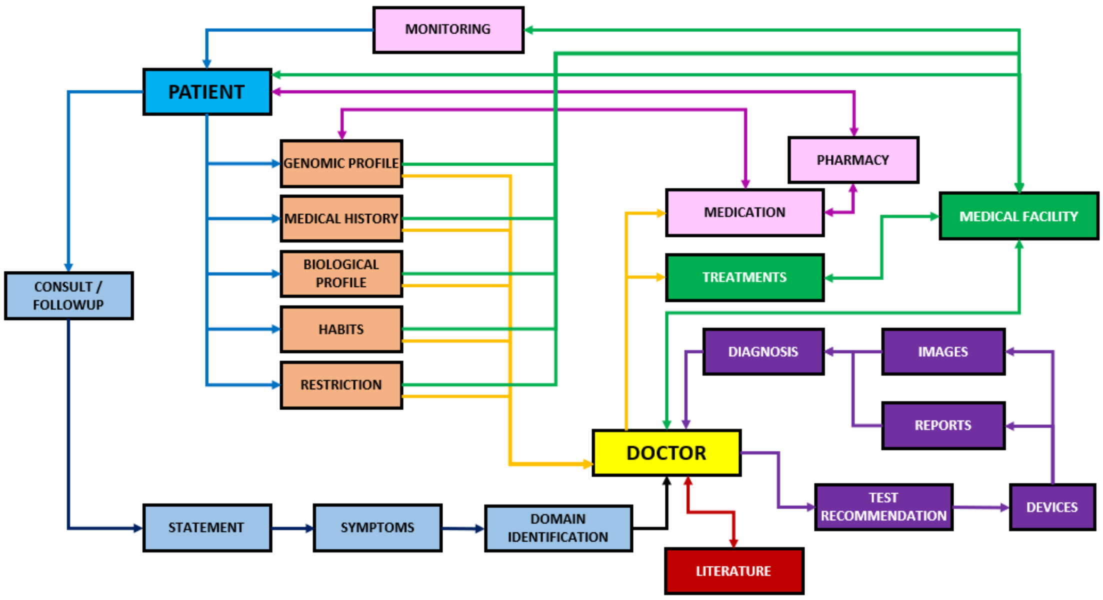

Figure 1: Hypothetical scenario in a healthcare ecosystem

Definition 1. *A **digital twin** is an accurate electronic representation or model of something that has physical existence created using real-world data and simulations to mimic its behavior, characteristics, and functionalities*

The concept of digital twins is being adopted by several industries[11] such as defence, transportation, manufacturing, urban planning, automobiles, e-commerce, environmental monitoring, and last but not the least healthcare.

Though several of previous studies have been conducted in this topic[12, 13, 14], we will discuss the possibilities of integrating the digital twin ecosystem with the current healthcare industry especially highlighting the role of knowledge representation. A hypothetical healthcare ecosystem is described in fig. 1.

2. Digital Twins in Healthcare: Scope, Target & Purpose

The healthcare industry has already started adopting the concept of digital twins in various different ways. Till now the innovation has mostly been localized. The objective of this paper is to propose a more global approach to make digital healthcare a reality. The applications of digital twins in healthcare can be classified into several categories based on the application scope, target areas and purpose that the digital twins would serve.

2.1. Scope

Digital twins is may be implemented at the level of a single patient[15], an entire organization[16] or even a geographical area[17]. Each level of implementation serves different use cases and application scenarios. A patient level twin may be developed by measuring physiological parameters, gene sequencing, radiology scans, medical histories, psychometric profiles and so on. At an organizational level, models of various organs, and biological processes can aid in running simulations or providing training. Molecular modelling is often necessary for drug designs and vaccine development. Furthermore various clinical infrastructures, services, equipment, etc. can also have virtual counterparts to enable process simulations and optimization of various clinical activities. Virtualization is not only limited to specific persons or organizations. They may also span over geographical regions to aid in various types of demographic surveys, and community based healthcare modelling.

2.2. Target

The next obvious aspect of digital twins is to figure out what components of the healthcare industry can be represented using a digital twin.

- **Multiomics:** Multiomics modelling[18] covers various aspects regarding the molecular dynamics of the human body. The genetic sequence and its electronic representation can be considered as a genetic twin of a person. This genetic profile may be used to prepare personalized treatment plans. Other than that models of various bio-molecules allow us to simulate molecular interactions. Prediction and 3D visualizations of protein structures is a big application area in this regard.
- **Drugs:** Molecular interactions can also be used for designing candidate drugs[19]. Digital twins of such molecules may be used to measure docking feasibility. This can be further extended to create precision medicine that takes genetic variations into considerations.
- **Diseases:** Modelling genetic variants of disease[20] causing microorganisms is a promising area to explore possible mutations and drug interactions or vaccine efficacy.
- **Epidemics:** Modelling mobility patterns of diseases[21] and various other environmental factors that may trigger healthcare concerns leads to a better understanding of epidemics and plan accordingly. Geographic twins of epidemic events can provide the platform for planning containment zones and vaccination drives.
- **Biological Systems:** Various organs, and respective biological processes like digestion, circulation, neural impulses can also have digital twins[22]. Generic models of these organs or processes can be used for training purpose. Even patient specific models can allow us to develop surgical or treatment plans by running simulations. Digital twins also help in designing personalized prosthetics.
- **Infrastructure:** Even in administration we can have twins of healthcare facilities[23], and personnel to optimize administration and services. Various clinical equipment and

devices can also have digital twins. These can be helpful in training and simulations. Electronic Health Records can provide decision support for defining health policies and insurance parameters.

2.3. Purpose

The final piece of the digital twin ecosystem is to define the purpose that is served by the twin. We have already discussed about the various scopes and targets for digitization. These twins would be useful in several scenarios.Digital twins may be used for training [24] of healthcare personnel on new equipment or clinical procedures may carried out. Surgical simulations [25] may be performed on patient specific twins of target organs. Demographic surveys allow us to plan region specific healthcare services[26] and also create decision support systems for epidemics and community healthcare. Past records may be used to forecast future outcomes[27] of treatment protocols, health camps, and also anticipate maintenance needs and provide supportive evidences for taking decisions. As discussed before, various vaccine[28] and drug design rely on digital twins for simulating molecular interactions. Personalized medicine[18] is also an application that consider genetic profile, medical history and lifestyle factors. Finally healthcare institutions can use digital twins to optimize their services, and reduce costs.

3. Holistic Healthcare Framework

The proposed holistic healthcare framework is a hypothetical model that ideologically establishes the potential of the digital twin industry. The objective of this study is to recognise the challenges of integrating digital twins into the existing healthcare industry and to propose essential steps for initiating a globally connected healthcare industry that can exploit modern technologies such as deep learning, internet of things(IoT), 5G/6G communications and cloud computing. But before we proceed we must understand that a digital twin is not a nuclear computational module working in isolation. It is an ecosystem consisting of various stakeholders such as patients, healthcare facilities, healthcare personnel, technology developers, device manufacturers, regulatory bodies, ethics committees, cybersecurity service providers, educational & research institutions, financial support providers, government, advocacy groups and more. It is built upon a versatile and robust technological stack [29]that interacts with various external data repositories and under the supervision of ethics committees and regulatory bodies.

3.1. Technological Stack

The necessary technological stack (as illustrated in fig. 2) for an ideal digital twin ecosystems would have multiple functional modules. Obviously the ecosystem would be built on top of the real world. This **physical layer** consists of the patients, healthcare professionals, medical facilities, equipment, and other relevant institutions. The virtual world would start with the **data layer** that would consist of the electronic data acquired from various sources such diagnostic equipment, patient records, reports, genomic profiles and so on. Once we have acquired the data the obvious next concerns are addressed by the **communication layer** and **storage layer**.

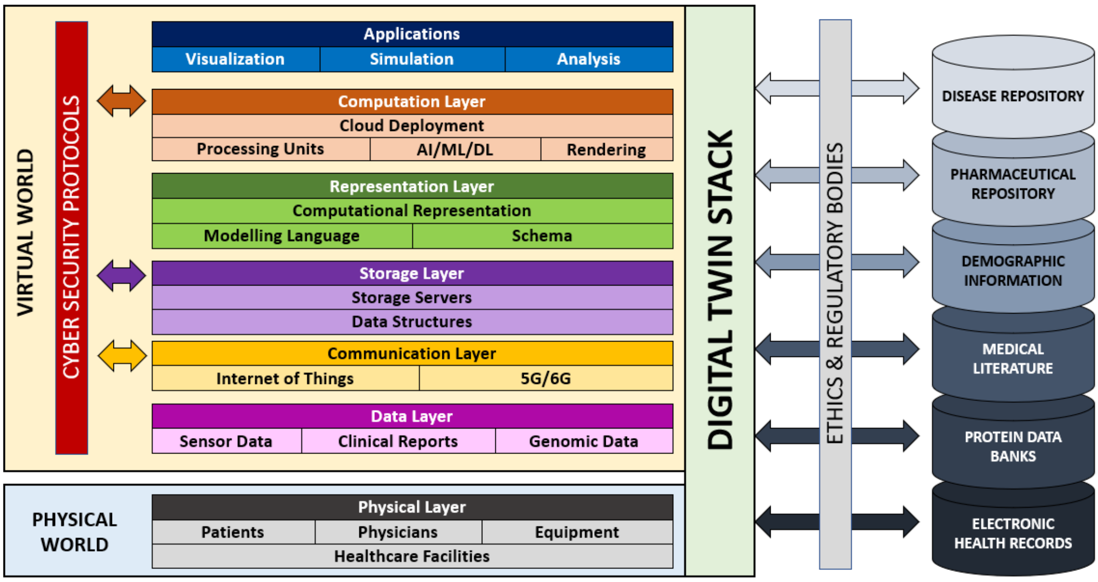

Figure 2: A schematic diagram for the technological stack for a digital twin ecosystem

Communication protocols may be defined as per implementation of the IoT infrastructure built upon 5G or 6G network backbone. The storage layer would deal with the necessary storage infrastructure. Blockchain techniques may be used for decentralization of information. The **representation layer** deals with the computational representation of the acquired data. The data would generally be acquired from different sources with different representation format. A common modelling language is needed for a consistent embedding. For a globally holistic healthcare framework common schema would be needed to assemble and integrate multi-modal information in a consistent data structure. The **computation layer** consists of the high performance computing infrastructure along with the advanced machine learning algorithms, image processing and language processing toolkit and models along with rendering engines. Finally the **application layer** would provide the necessary interface for visualization, simulation, and analysis. This can be carried out through standard modes of human computer interaction or even through augmented and virtual reality platform. Other than this there is a obvious needs of communication with external repositories such as electronic health records, government databases, insurance records, demographic information, pharmaceutical and disease repository, multiomics data banks and more. Additionally there is a need for cyber-security protocols across the stack and supervision from ethics and regulatory bodies.

3.2. One Species One Schema

For a digital twin ecosystem we need to operate with data acquired from various sources such as patients, equipment, healthcare facilities, and external data banks. All these data have vastly different formats and structure. The biggest challenge of this digital twin ecosystem is to progress towards the one species one schema concept. This refers to a single <ENTITY-

ATTRIBUTE-RELATION> model that can represent various elements of the healthcare industries along with their relevant parameters and contexts. This would allow us to create the knowledge graphs for healthcare digitalization[30]. The need of a common data structure is graphically summarized in fig. 3.

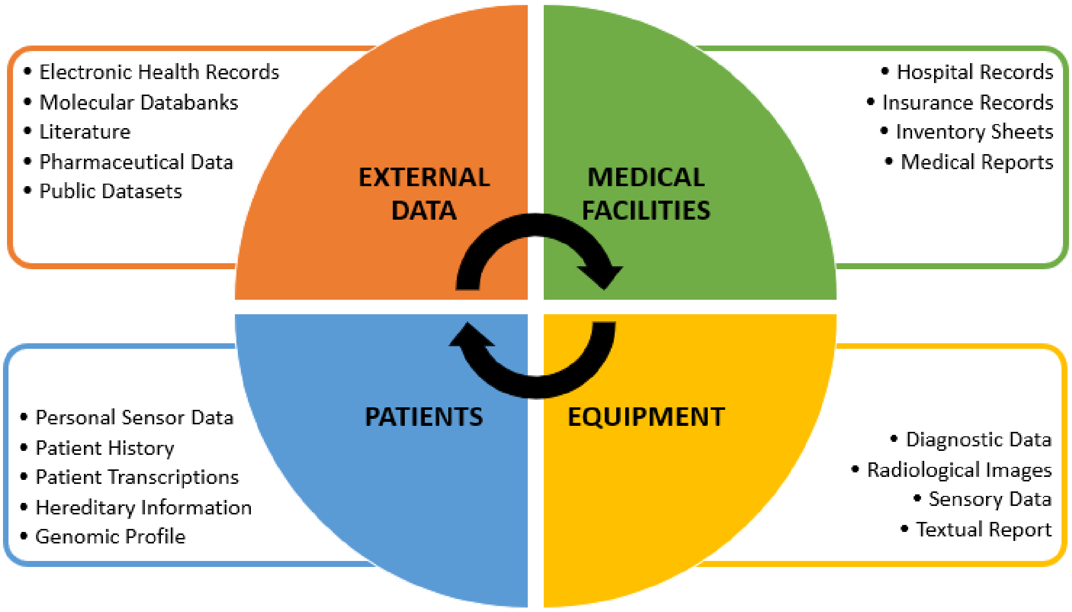

Figure 3: The variety among data sources and the need of a common data structure.

This schema for a global knowledge base must be built with some specific properties that ensures its sustainability. These properties, here abbreviated as M.U.S.C.L.E., refers to the aspects of

- **Multi-modality:** Data may come from various sources but must be mapped to the same schema
- **Uniformity:** Standardization of data representation is mandatory
- **Scalability:** The schema should be able to grow with the growing complexities of the data
- **Consistency:** The updates in the schema should be non-disruptive in nature
- **Longevity:** The schema must be able to adapt and grow with time to remain relevant with the fast moving pace of medical research as it can be a costly affair to revamp the entire schema
- **Ethics:** Each transactions in the knowledge based must be traceable to the responsible healthcare personnel or facility and must be explainable to avoid ethical or legal issues.

3.2.1. Knowledge Base Transactions

The knowledge base that is built upon the schema should be accessed and altered through serialized transactions. Transactions to update the knowledge base may be at three different levels:

1. **Content** : refers to nuclear updates that deals with specific parameters. These updates do not disrupt other nodes or edges. e.g. Updating the height and weight of a patient
2. **Context** : refers to updates that have effect in its immediate neighbourhood. These updates tends to display contextual significance. Updating diagnosis based on a test report result. Test report parameters will be updated which will trigger update in the diagnosis parameters
3. **Concept**: refers to the updates in the knowledge base that leads to addition, alteration or deletion of entire conceptual branches. A proposed treatment by a physician opens up a new branch for the healthcare facility that requires addition of several nodes and edges such as hospital beds, OT reservations, insurance parameters and so on.

3.3. Unique Biological Identifier (UBID)

For the realization of a global knowledge repository, the first and most important step is to develop a unique biological identifier. A unique globally standardized biological identifier must be created. This may be similar to other identification documents like passports, social security numbers, taxpayer identification numbers, Aadhar Card, VoterCard etc. However, a biological identifier must be associated to a biological signature of a being such as the fingerprints, retina scans or even genome sequence[31]. This must be standardized by the respective government as per uniform global standards. Most importantly it should be mandatory to associate with all healthcare related procedures similar to a taxpayer identification being connected with all financial transactions.

3.4. Globalization

Besides the UBID being necessary to unify healthcare transactions, a digital twin system would also require innovation tracking [32] and development of global policies, and standardization. We must understand that isolated innovation is the source of inconsistencies. Existing predictive models must be standardized and existing literature must digitized to create a viable ontology. Additionally an effort must be made to ensure every clinic, hospital, pharmacy, doctor, patients, equipment is connected to the common Healthcare Framework.

4. Conclusion

It is evident that revolution in global healthcare requires a migration to the digital twin ecosystem. However, several significant actions must be taken. From development of unique biological identifiers to moving towards a unified knowledge repository built on robust backbone schema. The proposed work tries to establish the need of "One Species One Schema" and the role that knowledge representation plays in the transformation of modern healthcare. The schema would need to be carefully designed to ensure long-term sustainability. The ultimate goal would be an unified knowledge graph that connects patients, healthcare facilities, professionals and all other stakeholder to ensure quality of medical services.

References

[1] L. Abi-Rached, M. J. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, et al., The shaping of modern human immune systems by multiregional admixture with archaic humans, Science 334 (2011) 89–94.

[2] B. V. Subbarayappa, The roots of ancient medicine: an historical outline, JOURNAL OF BIOSCIENCES-BANGALORE- 26 (2001) 135–143.

[3] L. Pasteur, On the germ theory, Science (1881) 420–422.

[4] D. H. Robinson, A. H. Toledo, Historical development of modern anesthesia, Journal of Investigative Surgery 25 (2012) 141–149.

[5] C. Mark, J. G. Rigau-Pérez, The world's first immunization campaign: the spanish smallpox vaccine expedition, 1803–1813, Bulletin of the History of Medicine (2009) 63–94.

[6] A. Fleming, Penicillin, British medical journal 2 (1941) 386.

[7] A. M. Thomas, A. K. Banerjee, The history of radiology, OUP Oxford, 2013.

[8] W. B. Fye, A history of the origin, evolution, and impact of electrocardiography, The American journal of cardiology 73 (1994) 937–949.

[9] S. D. Pruitt, J. E. Epping-Jordan, Preparing the 21st century global healthcare workforce, Bmj 330 (2005) 637–639.

[10] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al., The sequence of the human genome, science 291 (2001) 1304–1351.

[11] 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.

[12] T. Erol, A. F. Mendi, D. Doğan, The digital twin revolution in healthcare, in: 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT), IEEE, 2020, pp. 1–7.

[13] T. Sun, X. He, Z. Li, Digital twin in healthcare: Recent updates and challenges, Digital Health 9 (2023) 20552076221149651.

[14] H. Elayan, M. Aloqaily, M. Guizani, Digital twin for intelligent context-aware iot healthcare systems, IEEE Internet of Things Journal 8 (2021) 16749–16757.

[15] P. Barbiero, R. Vinas Torne, P. Lió, Graph representation forecasting of patient's medical conditions: Toward a digital twin, Frontiers in genetics 12 (2021) 652907.

[16] A. Karakra, F. Fontanili, E. Lamine, J. Lamothe, Hospit'win: a predictive simulation-based digital twin for patients pathways in hospital, in: 2019 IEEE EMBS international conference on biomedical & health informatics (BHI), IEEE, 2019, pp. 1–4.

[17] A. Perzynski, K. Berg, J. Dalton, Digital twin neighborhoods for precision population health: Initial results of community conversations, Innovation in Aging 7 (2023) 53–54.

[18] M. Cellina, M. Cè, M. Alì, G. Irmici, S. Ibba, E. Caloro, D. Fazzini, G. Oliva, S. Papa, Digital twins: The new frontier for personalized medicine?, Applied Sciences 13 (2023) 7940.

[19] K. Subramanian, Digital twin for drug discovery and development—the virtual liver, Journal of the Indian Institute of Science 100 (2020) 653–662.

[20] R. Laubenbacher, J. P. Sluka, J. A. Glazier, Using digital twins in viral infection, Science 371 (2021) 1105–1106.

[21] S. Alrashed, N. Min-Allah, I. Ali, R. Mehmood, Covid-19 outbreak and the role of digital twin, Multimedia Tools and Applications 81 (2022) 26857–26871.

[22] X. Zhong, F. Babaie Sarijaloo, A. Prakash, J. Park, C. Huang, A. Barwise, V. Herasevich, O. Gajic, B. Pickering, Y. Dong, A multidisciplinary approach to the development of digital twin models of critical care delivery in intensive care units, International Journal of Production Research 60 (2022) 4197–4213.

[23] Y. Peng, M. Zhang, F. Yu, J. Xu, S. Gao, Digital twin hospital buildings: an exemplary case study through continuous lifecycle integration, Advances in Civil Engineering 2020 (2020) 1–13.

[24] M. W. Zackoff, D. Davis, M. Rios, R. D. Sahay, B. Zhang, I. Anderson, M. NeCamp, I. Rogue, S. Boyd, A. Gardner, et al., Tolerability and acceptability of autonomous immersive virtual reality incorporating digital twin technology for mass training in healthcare, Technical Report, LWW, 2023.

[25] K. Hagmann, A. Hellings-Kuß, J. Klodmann, R. Richter, F. Stulp, D. Leidner, A digital twin approach for contextual assistance for surgeons during surgical robotics training, Frontiers in Robotics and AI 8 (2021) 735566.

[26] F. Pilati, R. Tronconi, G. Nollo, S. S. Heragu, F. Zerzer, Digital twin of covid-19 mass vaccination centers, Sustainability 13 (2021) 7396.

[27] A. Allen, A. Siefkas, E. Pellegrini, H. Burdick, G. Barnes, J. Calvert, Q. Mao, R. Das, A digital twins machine learning model for forecasting disease progression in stroke patients, Applied Sciences 11 (2021) 5576.

[28] T. Zohdi, Machine-learning and digital-twins for rapid evaluation and design of injected vaccine immune-system responses, Computer Methods in Applied Mechanics and Engineering 401 (2022) 115315.

[29] S. N. G. Gourisetti, S. Bhadra, D. J. Sebastian-Cardenas, M. Touhiduzzaman, O. Ahmed, A theoretical open architecture framework and technology stack for digital twins in energy sector applications, Energies 16 (2023) 4853.

[30] A. Carbonaro, A. Marfoglia, F. Nardini, S. Mellone, Connected: leveraging digital twins and personal knowledge graphs in healthcare digitalization, Frontiers in Digital Health 5 (2023) 1322428.

[31] C. Lippert, R. Sabatini, M. C. Maher, E. Y. Kang, S. Lee, O. Arikan, A. Harley, A. Bernal, P. Garst, V. Lavrenko, et al., Identification of individuals by trait prediction using whole-genome sequencing data, Proceedings of the National Academy of Sciences 114 (2017) 10166–10171.

[32] J. Barlow, Managing innovation in healthcare, World scientific publishing company, 2016.