Network Digital Twins for 6G: Defining Data Taxonomy and Data Models

Ayat Zaki-Hindi, Jean-Sébastien Sottet, Ion Turcanu, and Sébastien Faye
Luxembourg Institute of Science and Technology (LIST), Luxembourg
{ ayat.zaki-hindi, jean-sebastien.sottet, ion.turcanu, sebastien.faye }@list.lu

***Abstract*—As the telecommunications industry moves towards 6G, Network Digital Twins (NDTs) have emerged as a useful means for real-time monitoring, automation, and performance optimization in next-generation networks. This paper contributes to the ongoing research on NDTs within the framework of the 6G-TWIN project by presenting a comprehensive vision for NDT architecture and establishing baseline definitions for NDT data models, i.e., basic and functional models. We propose a structured taxonomy for the data required by these models to ensure a shared semantics consistent with existing standards. Furthermore, we introduce a graph-based modeling approach that provides a foundation for constructing basic models within NDTs, representing the physical network. Lastly, we discuss the challenges related to constructing basic models and propose future directions to address them.**

I. Introduction

The roadmap from 5G to 6G envisions not only improved performance, but also significant societal, business, and policy impacts. 6G will natively integrate AI, merge the digital and physical worlds, and address challenges such as sustainability, interoperability, and global coverage [1]. As 6G networks aim to provide even more diverse and demanding applications, the complexity of managing and optimizing these networks increases drastically. This has led to the emergence of Network Digital Twins (NDTs) [2], which offer real-time monitoring, automation, and advanced performance enhancement.

NDTs serve as virtual replicas of physical network infrastructures, providing real-time insights and predictive capabilities [3]. NDTs are driven by the need to optimize and manage the increasingly complex network infrastructures, providing significant advantages in terms of operational efficiency, predictive maintenance, and strategic planning.

The recognized recommendations set forth by the International Telecommunication Union (ITU-T) [2] outline key aspects essential for NDTs, summarized by data, mapping, modeling, and interface. **Data** serves as the foundation, providing a unified repository for accurate and up-to-date information. Real-time interactive **mapping** is what distinguishes NDTs from traditional network simulations, relying on the real-time data exchange between physical and virtual systems. **Models** within the virtual network reflect the key *basic* and *functional* features of the twinned physical entities. Standardized **interfaces** ensure compatibility and scalability, with southbound interfaces linking physical and virtual networks, and northbound interfaces facilitating information exchange between virtual networks and network applications. The European Telecommunications Standards Institute (ETSI) [4] and the Internet Engineering Task Force (IETF) [5] provide a similar view of the NDT architecture to the ITU-T recommendations.

Several survey papers benchmark the NDT research so far [6]–[8], providing insights into NDT requirements, use cases, and challenges. Other works focus on Radio Access Network (RAN) NDTs [9]–[11], since RAN accounts for the most complex and costly part of 6G networks.

While these works provide valuable insights for building NDTs, a clear gap remains in the formal definition of architectural components, data types, and NDT data models specific to 6G networks.

Within the framework of the 6G-TWIN project[1], this paper proposes a structured approach to defining a data taxonomy and data models for 6G NDTs. The main contributions of this paper are summarized as follows:

- Present the 6G-TWIN NDT architecture and provide formal definitions of the NDT data models.
- Propose a taxonomy for the data required in NDTs, aligned with existing standards.
- Introduce a graph-based approach representing the basis for NDT basic models.

The paper is organized as follows: Section II introduces the NDT architecture adopted by the 6G-TWIN project and our interpretation of data models. Section III presents a novel approach for constructing a 6G NDT data taxonomy. Section IV develops a graph-based network based on existing standards and perspectives on future research directions. Finally, Section V concludes the paper.

II. NDT Architecture

Building upon the reference architecture designed by ITU-T [2], we propose a conceptual architecture composed of three distinct layers: the physical network layer, the NDT layer, and the application layer. Figure 1 illustrates the different layers, their components, and the interaction between them via the double closed loop provided by the NDT management entity.

In the following, we describe the role of the different layers and elaborate on the building blocks of the physical network layer and the NDT layer, which serve the purpose of our paper.

[1] https://6g-twin.eu/

IEEE Conference on Standards for Communications and Networking (CSCN 2024)

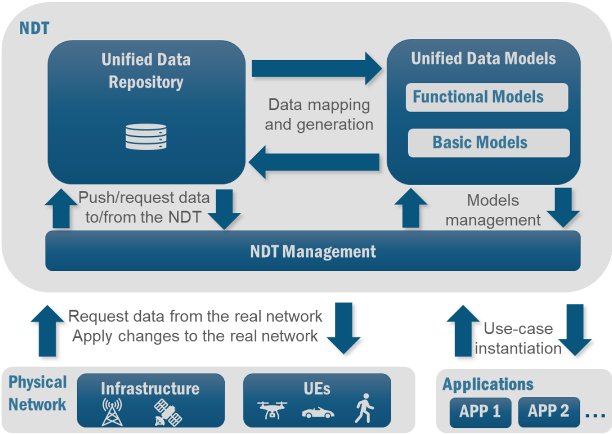

Figure 1. Network Digital Twin Architecture

A. Physical Network Layer

The physical network layer is also referred to as the infrastructure layer in the 6G end-to-end architecture, as proposed by 5G PPP [12]. The infrastructure layer hosts physical and virtual resources like RAN, Core Network (CN), and data networks, supporting the service for the User Equipment (UE). In order for the NDT to represent the 6G network infrastructure, we distinguish here the main domains to be twinned as follows: RAN, CN, and UE.

The RAN provides connectivity between UEs and the CN, represented by the Next Generation Node B (gNB) in 5G. The CN includes Virtual Network Functions (VNFs) responsible for collecting and managing data in the network, ensuring service continuity, and providing access to the data network, i.e., operator services, internet access, or third-party services. Finally, the UEs are the connected devices that access the network, such as smartphones, Unmanned Aerial Vehicles (UAVs), and other terminals, requesting services such as internet access and edge computing capabilities.

In addition, the RAN and CN network entities are actually a collection of sub-components, referred to as Network Elements (NEs). NEs can refer to intelligence, information, hardware, or software of a telecommunications network, and describing each NE ensures a standardized approach to efficiently managing these diverse components.

The 3GPP standards have effectively categorized NEs into classes, referred to as Information Object Classes (IOCs) [13], where each class is associated with specific attributes. This class structure defines the management aspects of network resources by specifying the information that can be exchanged through management interfaces in a technology-agnostic manner.

For example, RAN gNB has three main components: Radio Unit (RU), Distributed Unit (DU), and Central Unit (CU). The RU is the direct interface between the UE and the network, and it can be further decomposed into several sub-components, such as: *Beam* which represents the antenna beam and its properties, and *BWP* (Bandwidth Part) which manages the spectral resources available at the gNB.

Similarly, CN is the collection of VNFs, such as the *Access and Mobility Management Function (AMF)* and the *Session Management Function (SMF)*. Each VNF operates independently and is defined by its specific role in handling tasks like mobility, session management, and other essential network services. This modular approach enables greater flexibility, scalability, and easier management of the network functions, aligning with the broader goals of softwarization and automation in 6G networks.

These definitions will be further exploited in this paper to build the NDT data taxonomy and models.

What concerns UEs, it is harder to categorize and define common attributes since UEs can have a wide range of capabilities and are mostly vendor-specific. However, information about the UE can be depicted from their feedback data and behavior detected by the network, which will be the approach adopted in this work.

B. NDT Layer

The NDT layer is responsible for replicating the status and operational conditions of the components within the physical network. In addition to its core responsibility of creating and managing the Digital Twin (DT) representation, it handles data collection, communication with other layers, and the lifecycle management of NDT instances. To execute these functions efficiently, the NDT layer is composed of the following key building blocks:

1) Unified Data Repository (UDR): This component serves as a central hub for both historical and real-time data, facilitating seamless access and management of information crucial for the NDT accurate modeling. The UDR contains data from different heterogeneous sources such as network infrastructure, sensors, and contextual data, which requires defining adequate mechanisms for data collection, harmonization, and storage.

2) Unified Data Models (UDM): Within this layer, NDT data models are classified into two main types: (1) **basic models**, which provide real-time descriptions of the network's physical state, including configuration, environment, and topology, helping verify and emulate control changes before implementation, and (2) **functional models**, which leverage insights and data from basic models to optimize and predict the behavior of NEs. These models typically implement AI/ML algorithms aimed at a specific objective, i.e., optimization, anomaly detection, etc.

In the context of the 6G-TWIN project, one interpretation regarding the classification of models is that behavioral models, specifically algorithms that help emulate basic network behavior, should be categorized as basic models. This classification is not commonly addressed in the existing literature.

For instance, a physical layer resource allocation algorithm can be viewed as a basic model since it outlines the essential behavioral dynamics of the network. This algorithm is responsible for managing how resources, such as bandwidth and power, are allocated to different users and services, thereby directly simulating the normal network's performance. On the other hand, functional models are designed to enhance or optimize existing algorithms for specific purposes or scenarios. Taking the aforementioned resource allocation algorithm, a functional model can modify it to prioritize certain data flows over others.

Therefore, we propose the following formal definitions of basic and functional models from the 6G-TWIN perspective:

> A **basic model of a network element** is the collection of data describing its **properties**, **configurations**, and **operational status**, along with any **associated algorithms or protocols** used to emulate its dynamics and evolution with time.

> A **basic model of a network** is the aggregation of basic models of network elements, including their **physical and logical relationships** and the **interactions** that occur between them.

> A **functional model** of a network **builds upon basic models**, applying advanced processing techniques, often through **AI/ML algorithms**, under varying operational scenarios. These models are **designed for specific objectives** such as performance optimization, anomaly detection, or predictive maintenance.

In summary, basic models are associated to the physical network's layer, while functional models are associated with the application's layer, respectively.

3) NDT Management: This component oversees the creation and maintenance of NDT instances, ensuring that the application layer's requirements are met. It also manages the interactions between the UDR and the UDM, facilitating the mapping of relevant data entries and ensuring that the models reflect the current state of the network. The management entity is also responsible for data collection and implementing algorithms to the real network.

C. Application Layer

The application layer dictates to the NDT management entity the creation of NDT instances based on Key Performance Indicators (KPIs). The KPIs motivate the selection of appropriate basic and functional models related to the application under consideration. It also receives updates from the management entity regarding the results of the functional models to notify the user of actions taken or planned to be taken over the physical network.

III. Data Taxonomy for 6G NDTs

Data is the foundation of any DT, especially for complex systems like telecommunication networks. The NDT relies on an accurate and comprehensive representation of the network's physical and operational state, which is only possible through the collection, processing, and analysis of vast amounts of data. Data analysis allows capturing the dynamic, stochastic nature of the environment, unlike traditional methods which rely on mathematical modeling or simulations/emulations.

While it is important to study the data flow aspects from the physical to the digital system (data collection, storage and harmonization), our focus in this paper is on defining the content of UDR, which serves as the repository of all relevant data for NDTs.

In the case of 5G and 6G networks, the challenge lies in identifying the types of data that are most relevant for constructing such digital replicas. To address this, we aim to develop a taxonomy, i.e., a dictionary or classification of the data that needs to be handled by an NDT. This taxonomy forms the foundation for building basic models of NEs, ensuring that the NDT is represented with accurate and actionable data.

Our approach is grounded in existing standards for telecommunications networks, specifically those defined by the 3GPP. These standards provide a well-established framework for defining and managing network components, data flows, and performance metrics.

In mobile networks, data is categorized into User Plane (UP) and Control Plane (CP). UP handles the actual data transfer between the UE and the network, including data streams such as web browsing, video streaming, and application data. On the other hand, CP manages the signaling required to establish, maintain, and manage network connections, including session management, mobility management, and connection setup.

UP data itself is typically confidential and has limited relevance for network evaluation. Instead, the User Plane Function (UPF) within the CN manages UP operations and provides the needed information, for example, for modeling and optimizing content caching and delivery, through its CP data. Thus, only CP data is needed for building the NDT data taxonomy, which can be divided into attributes and measurements as follows.

1) Network Attributes: Attributes define various properties of the NEs including network topology, configuration aspects, and data exchanged to control the performance of the network. These attributes can support operations for network management services and event notifications. Key attributes include: radio resource management parameters, Quality of Service (QoS) parameters, security configurations, and network slicing.

2) Performance Measurements: Performance management [14] aims to collect data from NEs and VNFs, to verify network configurations, monitor traffic levels, assess resource access and availability, and ensure QoS requirements. Performance measurements cover several aspects, including user and signaling traffic, network configuration effectiveness, resource access, and QoS parameters such as call setup delays and packet throughput. Measurements are usually represented by statistics (e.g., mean, variance), aggregated over a time interval.

The proposed approach to build a data taxonomy prioritizes data generated directly from the physical network but does not comprehensively account for contextual data such as application requirements, sensors, and building environment, needed to create precise NDT models. These contextual elements are often covered by other well-established standards and consortia and can be integrated into the NDT. However, measurements can partially capture environmental factors, which are closely

Table I
EXAMPLE OF RAN IOCS, RELATED ATTRIBUTES, AND MEASUREMENTS.

	Attributes	Measurements
NRCellDU	cellLocalId; operationalState; administrativeState; cellState; pLMNInfoList; nRPCI; nRTAC; arfcnDL; arfcnUL; arfcnSUL; bSChannelBwDL; ssbFrequency; ssbPeriodicity; ssbSubCarrierSpacing; ssbOffset; ssbDuration; bSChannelBwUL; bSChannelBwSUL;	Packet Delay; Radio resource utilization; UE throughput; Number of active UEs; CQI (Channel Quality Indicator) related measurements; Transmit power utilization measurements; Received Random Access Preambles; power headroom measurement;
BWP	bwpContext; isInitialBwp; subCarrierSpacing; cyclicPrefix; startRB; numberOfRBs;	-
Beam	beamIndex; beamType; beamAzimuth; beamTilt; beamHorizWidth; beamVertWidth;	Intra-NRCell SSB Beam switch Measurement; RSRP Measurement; SSB beam related Measurement;

intertwined with network configurations, offering valuable insights into NDT models.

As mentioned in Section II-A, the network has three main domains to be twinned: UE, RAN, and CN. Each element is further represented by a collection of classes, i.e., IOCs. The existing approach to modeling these data structures uses YANG models, which are adopted by 3GPP[2]. YANG models define RAN and CN IOCs, their attributes, and their relations with other IOCs. However, a shortcoming of this modeling approach is that only attributes are modeled, and for a global network representation, measurements and performance statistics of the IOCs need to be incorporated.

Our proposal is to relate static (attributes) and dynamic (measurements) data of the same class in the same data model, in order to facilitate the network representation in a modular manner. An example of our proposed data taxonomy is illustrated in Table I, where we preview a set of RAN IOCs and their associated attributes and measurements, derived from [15] and [16], respectively. We note that not all IOCs necessarily have attributes or measurements, such as for *BWP*, for instance.

A similar table can be built for CN and UE data, but these are not shown here due space limitations.

IV. NDT BASIC MODELS

The backbone of the basic models for NEs is the taxonomy of network data defined in Section III. In addition to the proposed data taxonomy, these models also include all the functions that depict the standard behavior of the network, such as the 3GPP protocol stacks and procedures [17], e.g., Medium Access Control (MAC) protocols. However, basic models of the network require the modeling of the overall network topology and the relationships among NEs. For this purpose, we focus in this section on modeling basic models for networks as defined in Section II-B.

The comprehensive survey [18] offers a global perspective on how entities and interactions in Complex Networked Systems (CNS) can be represented, focusing on methods that preserve the complexity and heterogeneity of networked systems. The presented modeling paradigms include: basic graph-based models, probabilistic graph-based models, and network embedding-based models. Each of these approaches increases in complexity but also in capability, enabling richer and more accurate network representations as the complexity of the system grows.

Several research efforts have applied graph-based models to networking. For example, [19] proposes an NDT reference architecture from a software engineering perspective, identifying key modeling elements from standard network architecture documentation (e.g., ETSI) and network simulators (e.g., ns-3 and OMNET++). This research highlights the need for robust metamodels that define NDT elements, attributes, and relationships.

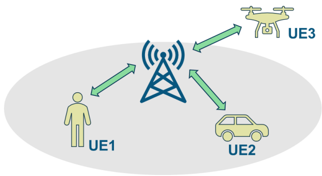

Figure 2. Scenario of a network.

Building a single NDT that includes all network aspects and scenarios is impractical due to the high complexity and processing power required. Therefore, the development of an NDT is inherently use-case-driven, with its data and models chosen precisely for the scenario it serves. For the sake of simplicity, we propose the scenario illustrated in Figure 2, where we have a set of UEs connected to the network via the RU.

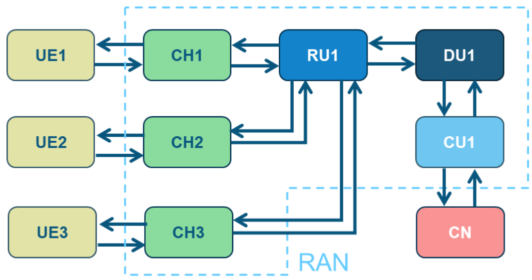

Figure 3. A graph representation of the scenario.

Building on the idea of graph-based modeling, we propose a preliminary graph representation of the scenario in Figure 3, illustrating the main entities and their logical or physical relationships. We treat the channel model, denoted by CH in Figure 3, as a distinct node due to its inherent complexity and stochastic nature. Complex nodes like the channel model can be captured through mathematical models, emulations, or data analysis. Furthermore, the channel model node contains all of the environmental and contextual factors that could not be listed in the data taxonomy.

In fact, with a "higher-granularity" level of representation, each node of the graph presented in Figure 3 represents a graph itself, following the decomposition into IOCs that we follow in this paper. Therefore, we show in Figure 4 an example of the

[2]https://forge.3gpp.org/rep/sa5/MnS

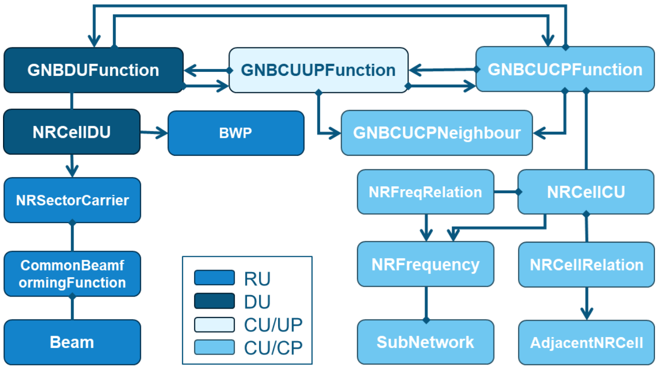

Figure 4. Several RAN IOCs and their relations.

different IOCs that represent the RAN part. These relations are mainly in line with the defined 3GPP standards [15]. A similar graph-based representation can be extended to the other nodes: CN, UEs, and channels.

Challenges and Future Directions

While the straightforward graph-based approach offers a structured and intuitive way to model simple network scenarios, it faces significant challenges in terms of scalability and flexibility. A major limitation is that it provides only a static snapshot of the network at a specific moment in time, making it inadequate for capturing the dynamic behavior of NEs. This restricts its ability to model ongoing changes, predict future states, or implement real-time adjustments.

To address these limitations, more advanced tools are required that can incorporate both temporal dynamics and uncertainties. Probabilistic graph models, such as Markov chains and Bayesian networks, offer a way to represent the inherent uncertainty in the relationships between network nodes.

For handling high-dimensional graphs more efficiently, graph embedding techniques provide a powerful alternative. Graph embedding reduces the complexity of graph representations by embedding node information into a lower-dimensional space, which allows for better graph construction and inference. This technique combines an encoder to extract node features and a decoder to reconstruct the graph in its simplified form. Common methods for graph embedding include Principal Component Analysis (PCA), deep learning techniques, and more advanced approaches like Graph Neural Networks (GNNs).

V. Conclusion

In this paper, we have made several key contributions to advance the conceptualization and implementation of Network Digital Twins (NDTs) for 6G networks. First, we have reviewed the NDT architecture with its foundational components and provided formal definitions for the basic and functional models of the NDT, which are related to the physical network layer and the application layer, respectively.

We also introduced a novel data taxonomy that sets the foundation for basic models, aligning them with existing standards. Our approach integrates both attributes and measurements into the same Information Object Classes (IOCs), addressing a limitation of traditional methods that only focus on attributes.

Lastly, we proposed a graph-based approach to model networks, providing a simple and structured way to represent the relations between the different network elements, and we identified existing challenges and future directions for NDT research, including the need for more advanced tools to capture network dynamics and the interaction between basic and functional models.

Acknowledgments

The 6G-TWIN project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation program under Grant Agreement No 101136314.

References

[1] ITU-R, "Future technology trends of terrestrial International Mobile Telecommunications systems towards 2030 and beyond," Tech. Rep. M.2516-0, Nov. 2022.

[2] ITU-T, "Digital twin network - Requirements and architecture," Recommendation Y.3090, Feb. 2022.

[3] S. Faye et al., "Integrating Network Digital Twinning into Future AI-Based 6G Systems: The 6G-TWIN Vision," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, IEEE, 2024, pp. 883–888.

[4] ETSI GR ZSM 015, "Zero-touch network and Service Management (ZSM); Network Digital Twin," Group Report V1.1.1, Feb. 2024.

[5] C. Zhou et al., "Network Digital Twin: Concepts and Reference Architecture," *Internet Engineering Task Force, Internet-Draft*, 2022.

[6] X. Lin et al., "6G Digital Twin Networks: From Theory to Practice," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 72–78, 2023.

[7] N. P. Kuruvatti et al., "Empowering 6G communication systems with digital twin technology: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112 158–112 186, 2022.

[8] A. Masaracchia et al., "Digital Twin for 6G: Taxonomy, Research Challenges, and the Road Ahead," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2137–2150, 2022.

[9] M. Polese et al., "Colosseum: The Open RAN Digital Twin," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 5452–5466, 2024.

[10] I. Vilà, O. Sallent, and J. Pérez-Romero, "On the design of a network digital twin for the radio access network in 5g and beyond," *Sensors*, vol. 23, no. 3, p. 1197, 2023.

[11] H. Gao, P. Kyösti, X. Zhang, and W. Fan, "Digital Twin Enabled 6G Radio Testing: Concepts, Challenges and Solutions," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 88–94, 2023.

[12] M. K. Bahare et al., "The 6G Architecture Landscape-European perspective," 2023.

[13] 3GPP TS 28.622, "Generic Network Resource Model (NRM) Integration Reference Point (IRP) (Release 18)," Mar. 2024.

[14] 3GPP TS 32.401, "Performance Management (PM); Concept and requirements (Release 17)," V17.0.0, Apr. 2022.

[15] 3GPP TS 28.541, "5G Network Resource Model (NRM); Stage 2 and stage 3 (Release 16)," V16.7.0, Dec. 2020.

[16] 3GPP TS 28.552, "Management and orchestration; 5G performance measurements (Release 18)," V18.5.0, Dec. 2023.

[17] 3GPP TS 24.501, "Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3; (Release 19)," V19.0.0, Sep. 2024.

[18] J. Wen, B. Gabrys, and K. Musial, "Toward Digital Twin Oriented Modeling of Complex Networked Systems and Their Dynamics: A Comprehensive Survey," *IEEE Access*, vol. 10, pp. 66 886–66 923, 2022.

[19] R. Verdecchia, L. Scommegna, E. Vicario, and T. Pecorella, "Designing a Future-Proof Reference Architecture for Network Digital Twins," in *European Conference on Software Architecture*, Springer, 2023, pp. 287–306.