Documentation of the Task G-1-1

**Definition and establishment of structure and dataflows in the digital twin framework**

**and**

Documentation of the Task G-1-3

**Phases and processes of a digital twin framework**

By

Atefeh Gooran Orimi

Task Area GOLO, NFDI4Ing, Leibniz University Hannover

Summer 2023

Contents

[1. Introduction 3](#_Toc147708932)

[2. Digital Twin Framework 4](#_Toc147708933)

[2.1. DT Definition 4](#_Toc147708934)

[2.2. Benefits of DT frameworks 7](#_Toc147708935)

[2.3. Types of DTs 8](#_Toc147708936)

[2.4. Difference from previously used technologies (ability of DTs) 9](#_Toc147708937)

[2.5. Applications 10](#_Toc147708938)

[3. Constructing a Digital Twin Framework 12](#_Toc147708939)

[3.1. Literature review 12](#_Toc147708940)

[3.2. DT characteristics 16](#_Toc147708941)

[3.3. Building a functional DT 18](#_Toc147708942)

[3.4. Dataflow in DTs 21](#_Toc147708943)

[3.5. DT and product lifecycle management 23](#_Toc147708944)

[3.6. Tools and software 26](#_Toc147708945)

[3.7. Current challenges 28](#_Toc147708946)

[4. General Structure of the Proposed Digital Twin for the Task Area GOLO 30](#_Toc147708947)

[4.1. Field data and DT for field-related distributed systems 30](#_Toc147708948)

[4.2. Digital master 32](#_Toc147708949)

[4.3. Digital shadows 35](#_Toc147708950)

[4.4. Communication 36](#_Toc147708951)

[4.5. Dataflow in the proposed DT framework 37](#_Toc147708952)

[4.6. The LUH use-case 38](#_Toc147708953)

[4.7. The DFKI use-case 38](#_Toc147708954)

[4.8. The TUD use-case 38](#_Toc147708955)

[5. Further notes 39](#_Toc147708956)

[6. Conclusion and Final Remarks 40](#_Toc147708957)

[Refrences 41](#_Toc147708958)

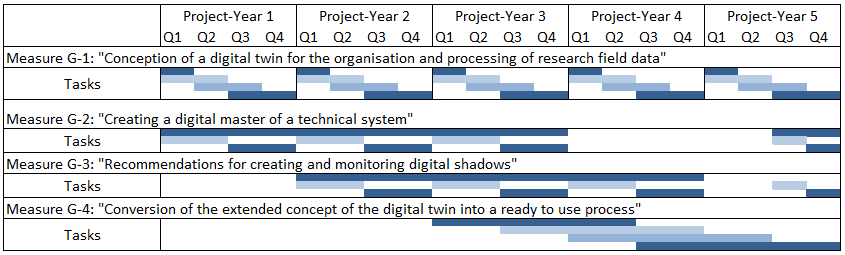
# Introduction

Digital Twin (DT) can be interpreted as a technology that has emerged in the fourth evolution of the industrial world (Industry 4.0). The application of such advanced technologies has been exponentially increased specifically in recent years mainly due to the advancement of the industrial and data-related technologies (including computer powers, availability of Big data, availability of huge storage platforms, advanced Machine Learning and Artificial Intelligence techniques, etc). Consequently, nowadays DTs have a variety of applications in different areas of science such as Engineering, Medicine, Aerospace and many more.

In this document, we discuss the different aspects of DTs and explore their specific advantages and disadvantages as well as their applications in different fields of study. We further investigate the details of a DT framework and discuss what characteristics and components a reliable DT framework should possess. We elaborate the details of these components and present a step-by-step method for creating an applicable DT framework. To this, we make use of the previously established frameworks and construct a thorough literature review, mainly in the field of engineering. We then summarize the prominent studies that have been carried out during the last years and provide an overview of the foundations of the DT frameworks and how to establish such frameworks.

This document then serves as the experimental results and documentation of the Measures G-1-1 and G-1-3 of the Task Area GOLO of the NFDI4Ing consortium. These measures can be realized in combination with other tasks and measures of the TA GOLO that are related to the DT framework and the establishment of a DT framework for field data and distributed systems. The Gantt-Chart of the TA GOLO and the different work packages (measures) throughout the whole project duration is then depicted in the following (Tab. 1):

Table 1: Gantt-Chart of the Task Area Golo presented in the NFDI4Ing proposal.



# Digital Twin Framework

The DT concept has been exponentially growth during the last years where a variety of researches have been developed (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) (Pantelidakis, Mykoniatis, Liu, & Harris, 2022) (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) (Liu, Fang, & Dong, 2021) (Bärring, Johansson, & Shao, 2020) (Barricelli, Casiraghi, & Fogli, 2019) (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) (Tao F. a.-Y., 2019) (Boyes & Watson, 2022) (Wang, et al., 2022). The development of such a concept has been mainly driven by the advancement of Industry 4.0 technologies such as the Internet of Things (IoT), cloud computing, sensor technology, additive manufacturing, artificial intelligence and machine learning, modeling, and simulation techniques, etc (Pantelidakis, Mykoniatis, Liu, & Harris, 2022).

In general, we are concerned with the specific systems that are of interest in the real-world physical space and attempt to simulate/analyze specific properties of these systems in a virtual space. Data and information exchanged between the two spaces are consequently of interest to our investigations. Specifically, advancing data-related technologies have a vital role in DT frameworks by collecting data from the physical space and further storing and sharing them with the digital space. This connection between the two spaces is formally known as the *twinning* which can be seen as a bridge between the two spaces (between the physical and digital spaces). Consequently, the term digital twin originated from this act of twinning. The investigated system (or process) in the physical space is commonly referred to as the *physical entity* or *physical twin*, and the twinned digital copy of it (in the virtual space) is commonly referred to as the *digital twin* or *cyber twin*.

Intuitively, the concept of DT has been developed in different fields of study. As a consequence, its definition does vary from one field to another and any researcher may use specific terminologies that are related to his/her specific task area. Nonetheless, there is a common understanding among these different points of view which can serve as a general definition for the DT concept.

## DT Definition

We follow the work of (Boyes & Watson, 2022) and (Catapult, 2021) and define digital twin as a “live digital coupling of the state of a physical asset or process to a virtual representation with a functional output”. This definition serves as a basis for our investigations where we seek an online coupling between the physical and virtual (digital) spaces, as well as a specific output that we desire (usually a functional and a real-time output). This specific coupling is, as mentioned above, known as the twinning between the two spaces that should be established. Other studies also follow a (somehow) similar definition of the DT and mainly seek how such a coupling/twinning should be constructed. We discuss a few points of these different studies regarding the DT concept in the following:

According to (Wang, et al., 2022), “DT is a digital replica of a living or non-living physical entity which has attracted extensive attention from different industries during the past decade”. We note the difference between the terms living and non-living. It should be mentioned that in many studies, the fact of having a real-time procedure for defining a DT is not justified, and any form of coupling between the investigated physical and virtual spaces would be satisfactory. In another word, a real-time exchange of data between the two spaces may be advantageous, but it is not required.

Furthermore, Tao et al. (Tao F. a.-Y., 2019) investigated the concept of DTs in connection to the different aspects of product lifecycle. They interpreted DTs as a “real mapping of all components in the product lifecycle using physical data, virtual data and interaction data between them”. This way of defining a DT is specifically important for companies (and also researchers) as it provides a complete digital footprint of the product throughout its lifecycle; and enables companies/researchers to track the process of a specific product from design and development till the end of its product lifecycle (see (Tao F. a.-Y., 2019) for more details).

In a newly established study, Rathore et al. (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) present an elaborate definition of the DT and the concept of twinning between the physical and virtual spaces. They discuss that “digital twinning is a process that involves the construction of 1) a cyber twin that digitally projects a living or non-living physical entity or a process (a system); and 2) a physical connection between cyber and physical twins to share data (and information) between them aimed at dynamic optimization, real-time monitoring, fault diagnostics and early prediction, or health monitoring of the physical counterpart”. They specifically consider the physical twin as a process that may include a human, a place, a device, or any other object that might be of interest. Such physical entities should be then replicated in the digital world as either a partial twin with limited functionalities, or a complete twin that incorporates the full behavior of its physical counterpart (see (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) for the details).

In general, we can argue that there are many different ways to define DT frameworks in industry and academia. Nonetheless, it is generally accepted among the researchers that (see, (Glaessgen and Stargel 2012)) a “DT, developed by NASA, is an integrated multi-physics and multi-scale probabilistic ultra-realistic simulation of systems or products which can mirror the life of its corresponding twin using available physical models, history data, real-time data, etc.”

Definitions above provide an intuition of the DT concept and illustrate how different points of views to the different aspects of this concept have led to a variety of interpretations and definitions of DTs. Of course, all these definitions do contain certain similar points which clarify the basic idea of a DT concept. In such previous studies, authors have even provided a list of different definitions that exist for a DT concept. These contributions will then provide thorough and complete information about DTs which can be also seen as our main references here (and in general for the documentation of the measure G-1-1 in TA GOLO). Tab. 2 presents a comparative list of definitions which is discussed and detailed in (Barricelli, Casiraghi, & Fogli, 2019); we here present the exact table from this reference and refer the reader to this study and also to the other studies mentioned above and references therein for further detailed information about DTs.

Table 2: *Different definitions of DTs in the literature; table is sourced from (Barricelli, Casiraghi, & Fogli, 2019).*

|  |  |
| --- | --- |
| **Key points** | **Definitions** |
| Integrated system | Integrated multi-physics, multiscale, and probabilistic simulation composed of physical product, virtual product, data, services and connections between them. |
|  | An ultra-realistic integrated multi-physics, multiscale, probabilistic simulation of a system. |
|  | A big collection of digital artifacts that has a structure, all elements are connected and there exists meta-information as well as semantics. |
|  | Comprehensive physical and functional description of a component, produc or system together with all available operational data. |
|  | A systematic approach consisting of sensing, storage, Synchronization, synthesis and service. |
|  |  |
| Clone, counterpart | Computerized clones of physical assets. |
|  | The virtual and computerized counterpart of a physical system. |
|  | Functional system formed by the cooperation of physical production lines with a digital copy. |
|  |  |
| Tines, links | Connections of data information that ties the virtual and the real product together. |
|  | New mechanisms to manage IoT devices and IoT systems-of-systems. |
|  | Technology that links the real and the digital worlds. |
|  |  |
| Description, construct, information | Comprehensive physical and functional description of a component, product or system. |
|  | Reengineering computational model of structural life prediction and management. |
|  | A simulation based on expert knowledge and real data collected from the existing system. |
|  | Virtual models for physical objects to simulate their behaviors. |
|  |  |
| Virtual, mirror, replica | A virtual representation of the system. |
|  | Digital mirror model of ourselves. |
|  | Virtual representation of a real product. |
|  | A digital copy of a physical system. |
|  | Virtual model of a physical asset. |
|  | A replication of real physical production system. |
|  | Cyber copy of a physical system. |
|  | A dynamic digital representation of a physical system. |
|  | A virtual model of physical object. |

## Benefits of DT frameworks

The concept of digital twin has gained significant attention and recognition across various applications due to its numerous advantages. In fact, the benefits of using DT frameworks in different projects ranging from scientific studies to industrial production and to novel technological advancements have led to their exponential growth in the past few years. Regarding the advantages of DT frameworks, several studies have identified their key benefits in different that we will list and discuss in the following (see, e.g., (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) (Pantelidakis, Mykoniatis, Liu, & Harris, 2022) (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) (Bärring, Johansson, & Shao, 2020) (Wang, et al., 2022) for some details):

* **Cost Efficiency:** The main advantages of DTs in industrial projects is that they can save costs, and time by, e.g., eliminating the need for physical prototypes. It is particularly beneficial in scenarios where developing physical prototypes is expensive or lab conditions cannot replicate actual scenarios accurately.
* **Fault Prediction:** Another important application of DTs is their power to predict and find the product flaws by assessing the virtual version of the products. Subsequently, DTs can predict impending dangers in time to mitigate them and thus help increase the product quality and lifespan.
* **Design Cycle Acceleration:** DTs can shorten product development time by enabling the assessment of virtual products and thus avoiding lengthy testing processes.
* **Real-Time Monitoring of Resources:** DTs can effectively monitor and therefore manage and save resources (including time and other feasible resources). In this regard, we are not required to build a physical prototype to perform some tests and instead, we can make use of the digital (virtual) prototype to perform the tests and predict timely risks in an effective and accurate manner. It should be noted that real-time monitoring and mitigation plans are important factors in some areas such as health systems and supply chain where we deal with a so-called ‘emergent behavior’. In such systems, the goal is to constantly monitor the physical asset and predict imminent problems can be both efficient and effective. This enables us to make an informed and fast decision and prevent potential losses.
* **Applications on Impossible/Impractical Projects:** DT technology offers major benefits in scenarios requiring extreme testing, real-time monitoring, and dealing with emergent behavior. They can be specifically used in projects where establishing a real-world physical prototype is expensive, time-consuming, or even impossible (like aerospace, supply chain, and manufacturing), allowing to perform/simulate different experiments on the digital prototypes.
* **Ability to Model Complex Problems:** DTs have been frequently used to model complex physical entities (e.g., *aggregated* DTs are commonly used).
* **Ability to Exploit Advanced Tools and Technologies:** DTs can be effectively used in conjunction with advanced technologies and tackle a variety of problems. DTs based on artificial intelligence (a.k.a. AI-based DTs), for e.g., have been widely used in recent years.
* **Revolutionizing Businesses:** DT frameworks provide a common environment with decision-making tools across different levels; and subsequently, revolutionize business operations, support decision-making, drive complete digitalization, and contribute to the advancement of the manufacturing industry. It spans across the enterprise, and supply chain.
* **Several DTs for one Entity:** Besides the advantages mentioned above, we can consider different DTs for the same physical entity and perform different examinations simultaneously. For instance, we can test different parameters settings for a prototype using multiple DTs with different experimental settings at the same time. This can be specifically beneficial in products with multiple parameters where we can optimize the model parameters jointly and/or complex cases where reducing the time of the product development is necessary.

The above points provide a list of commonly known benefits for using DT frameworks. In addition to these points, each study further develops the technology in different directions, investigating new applications in different areas. These recent studies then discuss new advantages, e.g., for one specific production. We refer the readers to (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021) (Pantelidakis, Mykoniatis, Liu, & Harris, 2022) (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) (Bärring, Johansson, & Shao, 2020) (Wang, et al., 2022) and references therein for example studies in this direction. We will also discuss such studies in detail in Sec. 3 further below.Formularende

## Types of DTs

After discussing the different definitions of a DT concept and its benefits in different fields, we now discuss what types of DTs have been established in previous studies. Before that, let us first discuss a specific scenario that has been considered in some previous works as a general interpretation of the DT concept. In such studies, one can assume a Digital Model as a framework in which the data will be manually exchanged between the virtual and physical spaces, and hence, a real-time state of the model is not considered/analyzed in the virtual space. This corresponds to, e.g., simple simulations of a physical system using the desired software/tools. In the next level of modelling, there exists the digital shadow which can be seen as a saved data copy of the physical entity/state, with one-way dataflow from physical object to the digital object. In this modelling, the connection between the two spaces is established, but as mentioned, this connection is only a one-way exchange of data. Finally, and most importantly, we have the Digital Twin technology which considers a fully integrated dataflow between the two spaces such that it properly and consistently reflects the actual state of the physical object. Consequently, a DT does replicate the physical entity (or a part of the physical entity) in an active and two-way manner (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022).

This way of explaining the DT concept, describes the different levels of data exchange that exist in manufacturing and engineering. A digital model which is a digital representation of a physical part or assembly without data exchange, a digital shadow that enables a one-way data flow between the physical and virtual representations of an object, and specifically a digital twin that allows for a bi-directional automatic data exchange between the physical and virtual representation. In fact, the DT concept encompasses specific properties that should exist / be established within the framework such that we can call it a proper DT framework. These properties and features then distinguish the DT concept from the previously used technologies for analyzing and investigating a physical system. In this regard, different reference models for DTs have been proposed, considering important properties, such as model scalability, interoperability, expansibility, and fidelity (Wärmefjord, Söderberg, Schleich, & Wang, 2020). We will further discuss these properties in the upcoming sections.

Besides the level of data exchange, another important concept is how complex a DT can be seen. Based on Grieves and Vickers (the pioneers of the digital twinning), the digital twin can be any of the following three types (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021):

* **Digital twin prototype (DTP):** DTP is a constructed digital model of an object that has not yet been created in the physical world, e.g., 3D modeling of a component. The primary purpose of a DTP is to build an ideal product, covering all the important requirements of the physical world.
* **Digital twin instance (DTI):** DTI is a virtual twin of an already existing object, focusing on only one of its aspects.
* **Digital twin aggregate (DTA):** DTA is an aggregate of multiple DTIs that may be an exact digital copy of the physical twin. For example, the digital twins of a spacecraft structure and a spacecraft engine are considered DTIs that may be aggregated into a DTA.

Generally, a DT might be that of a product or a product lifecycle component. However, one might also find it more feasible and easier to break down the product or product life-cycle’s components to sub-components, create several DTs, and establish connections between them. For example, for a car, one might not want a DT for the entire car but only the engine, brakes and gearbox to understand the functionality of these components interacting with each other; or if we take the example of the car’s product life-cycle, one could create different DTs for raw materials procurement, manufacturing, production and supply, for reasons like avoiding data sharing, if different components are handled by different vendors. The mechanism through which a DT interacts with other DTs is dependent on the extent of data sharing allowed and the IoT devices used. The dynamic property of synchronization again comes into play here, during regular updates among different DTs. Strong security protocols may form an important part of this communication (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022).

## Difference from previously used technologies (ability of DTs)

As mentioned above, the DT concept is sometimes inadequately considered as a high-fidelity modeling and simulation environment by many studies, since none of these studies fulfilled all functionalities of the Digital Twin concept (storage, modeling, learning, simulation, and prediction) by building a cloud–edge architecture (Wang, et al., 2022). A DT should be capable of self-healing and predictions (Bärring, Johansson, & Shao, 2020), and should possess certain properties and may differ substantially from the previously used technologies. Example of the dataflow above illustrates how a DT concept can be a generalization and at the same time, a much more sophisticated technology in comparison to the other existing ones. The following table then details the differences between a DT framework and other technologies that are mentioned and can be used either for data analysis or for the analysis of a physical system in the virtual space (and consequently, performing certain tasks). Each of these technologies can be discussed in detail, such that the difference become more evident, but this of course requires a thorough investigation which is not the subject of our study here and we refer the readers to from (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) for more information.

The major difference between a Digital Twin framework with an iteration framework, or a model-based design framework is that a DT framework always maintains synchronized versions of the physical system and its digital counterpart. In the proposed MDT framework of our study, this is guaranteed by the communication plane between the physical and digital spaces (Wang, et al., 2022). Also, the work by Sharma et al. (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) discuss the difference between a DT technology and other similar methods in detail, which is depicted in the following table (Tab. 3):

Table 3: *How DT differs from existing technologies; table is sourced from* (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022)*.*

|  |  |
| --- | --- |
| Technology | How the technology differs from DT |
| Simulation | No real-time twinning |
| Machine Learning | No twinning |
| Digital Prototype | No IoT components necessarily |
| Optimization | No simulation and real-time tests |
| Autonomous Systems | No self-learning (learning from its past outcomes) necessarily |
| Agent-based modelling | No real-time twinning |

## Applications

Typically, there are two distinct perspectives when it comes to the utilization of DT frameworks. One involves researchers investigating the application of DT within various scientific domains, aiming to uncover its adaptability across different fields. In this case, the emphasis is on comprehending how DT can be put into practice and its potential influence in specific domains. On the other hand, there exists another viewpoint where researchers explore the intrinsic capabilities of DT itself, transcending across various domains. In this regard, researchers explore the broader potential of DTs and their capacity to revolutionize industries and sectors. These two perspectives (one concerning the academic-based application of DTs, and the other corresponding to the application of DTs in industries and companies) are examined and researched independently, each with its own distinct focus and research objectives.

Digital twins contribute to optimization efforts across various industries, enabling better resource allocation and performance improvement. The work by Barricelli et al. (Barricelli, Casiraghi, & Fogli, 2019) discusses the application of DTs in three areas of Manufacturing, Aviation and Healthcare. The application of DTs in each of these different areas can be explained as follows (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021):

* **Manufacturing:** Digital twin application in manufacturing focuses on optimizing production processes and improving overall efficiency. These applications facilitate real-time monitoring, diagnostics, and predictive analytics for products, thereby assisting manufacturers in elevating product quality and reducing production costs. In fact, by employing DTs, manufacturers can predict product behavior, detect potential defects, and enhance product performance even before physical prototypes are produced.
* **Aviation:** Digital twin technology finds applications in the aviation industry, enabling virtual testing and simulations to optimize aircraft design and performance. It plays a crucial role in predictive maintenance by continuously monitoring the condition of aircraft components and identifying potential issues before they lead to failures.
* **Healthcare:** In the healthcare sector, digital twin applications are used for medical prognostics, health management, and enhancing patient monitoring and treatment planning. In fact, by creating virtual representations of patients or medical equipment, healthcare professionals gain the ability to make well-informed decisions and tailor treatments to individual needs.

In another study, Tao et al. in (Tao F. a., 2018) also explored various industrial applications of digital twin models across different categories. Based on their work, DTs find application in product lifecycle management including design, production, and prognostics, and in health management. Moreover, they discuss the superiority of the DTs over traditional methods in these areas, as DTs offer enhanced efficiency and accuracy. Other studies also add on the aforementioned points and discuss the general applications of DTs in different fields and how they can be applied. Examples of other industrial applications are in transportations, power and energy, business and so forth where DTs can be likewise employed. Another domain is, for e.g., smart cities, where DT applications contribute to their development, enabling better urban planning, infrastructure management, and resource optimization.

In addition to the points mentioned above regarding the industrial applications of the DT frameworks, we can further discuss the applications of the DTs in research. This is mainly concerned with the structure and the different components of DT frameworks (how such components interact with each other and how their potential can be extended/generalized to also subsume other complex cases). In TA GOLO, our focus is on exploiting DT frameworks for distributed systems in order to, e.g., increase the re-usability of the field data. Although we here focus on the fundamental structure and details of such an established framework, we would also be interested in its real applications in the field and how it can be implemented/employed in practice.

# Constructing a Digital Twin Framework

## Literature review

In the following, we discuss the related work on DT frameworks published during the last years. The first paper is the work by Wang et. al. (Wang, et al., 2022) that focused on a concept termed Mobility Digital Twin (MDT) in order to represent the digital format of various mobility entities together. The proposed MDT framework addresses limitations of traditional mobility systems (such as computing power, accessibility to Big data, and easiness of deployments and modifications) by providing powerful, shareable, manageable, and extendable functionalities. It allows rapid adjustments of cloud resources, offloading and retrieving bulk data, and facilitates over-the-air updates. The following contributions are then discussed for the MDT framework which can be seen as an addition to the benefits of DTs mentioned above:

Although a few Digital Twin studies have been conducted in the transportation domain very recently, there is no systematic research with a holistic framework connecting various mobility entities together. In this study, a mobility digital twin (MDT) framework is developed, which is defined as an artificial intelligence (AI)-based data-driven cloud–edge–device framework for mobility services. This MDT consists of three building blocks in the physical space (namely, Human, Vehicle, and Traffic), and their associated Digital Twins in the digital space. An example cloud–edge architecture is built with Amazon Web Services (AWS) to accommodate the proposed MDT framework and to fulfill its digital functionalities of storage, modeling, learning, simulation, and prediction. A case study of the personalized adaptive cruise control (P-ACC) system is conducted, which integrates the key microservices of all three digital building blocks of the MDT framework: 1) the Human Digital Twin with user management and driver type classification; 2) the Vehicle Digital Twin with cloud-based advanced driver-assistance systems (ADAS); and 3) the Traffic Digital Twin with traffic flow monitoring and variable speed limit. Future challenges of the proposed MDT framework are discussed toward the end of the article, including standardization, AI for computing, public or private cloud service, and network heterogeneity (Wang, et al., 2022).

1) **Powerful:** The MDT framework allows users to rapidly adjust cloud resources to meet fluctuating/unpredictable demands, providing high computing power at certain periods of peak demand.

2) **Shareable:** Bulk data generated by an end user is offloaded and stored on the cloud (and/or edge), which can be retrieved and utilized by the same user at a later time frame, or shared with other end users for microservices on demand.

3) **Manageable:** The MDT framework allows users to get their microservices up and running faster on the cloud platform, with more manageability and less maintenance. Over-the-air (OTA) updates are also available to the MDT framework.

4) **Extendable:** Arbitrary mobility microservices can be easily implemented to the MDT framework with minimal changes on the cloud–edge architecture and data structure.

The work by Wang et. al. (Wang, et al., 2022) is one of the few studies that discusses DT in a field-related system system and application. Although the concept may be a bit different from our work in Task Area Golo, it provides an intuition of the DT framework and its characteristics that are relevant to our goal in task G-1-3 (and also to our use-cases). Importantly, we put the emphasis on the concepts such as the “end-to-end” framework, and also to the “communication” between the two spaces of physical and digital. We will later propose the relevant sketch of our DT and specifically make use of the information we discuss in the following (see (Wang, et al., 2022) for more details):

The MDT framework is built on top of three different planes: 1) the physical space that has human beings, vehicles, and traffic infrastructures; 2) the digital space that has the digital replicas of aforementioned physical entities; and 3) the communication plane between these two spaces. … The communication plane plays a crucial role in this framework to allow real-time and nonreal-time data streaming for both upstream and downstream. …. the Human Digital Twin with user management and driver type classification, the Vehicle Digital Twin with cloud-based advanced driver-assistance systems (ADAS), and the Traffic Digital Twin with traffic flow monitoring and variable speed limit.

Physical space: If we consider this MDT framework as an end-to-end framework, then the physical space is in charge of both ends of this framework, namely, sampling and actuation. We assume no (or only minimal) computing work needs to be conducted in the physical space, since all (or majority) of that is offloaded to the digital space through the communication plane. For sampling, sensors in the physical space detect the dynamic status, operating process, or event occurrences, and then aggregate these measurements under various resolutions for their transmission to the digital space. On the other hand, once the processed results are received from the digital space, actuation can be made by physical entities to fulfill this end-to-end framework. Generally, the physical space is defined on a world coordinate, which may contain all the transportation related physical entities, and can be classified into three building blocks: Human, Vehicle, and Traffic (Wang, et al., 2022).

For description of the digital space and communication part see (Wang, et al., 2022). Another paper is (Wärmefjord, Söderberg, Schleich, & Wang, 2020), that discusses Digital twin framework for geometry assurance stated as following steps:

With the labeling digital model, digital shadow, and digital twin, a digital twin for geometry assurance can be interpreted as (in fact the following three pillars constitute together a digital twin):

1. a digital model, containing all nominal information about parts and assembly and joining processes. The model must also provide simulation capabilities (Digital Twin -- The Digital Model).
2. input to the digital model about deviations from nominal values on the part and process levels (Digital Twin -- The input data).
3. output from the digital model to adjust the assembly and joining processes (Digital Twin -- The output data).

The next paper we discuss here is (Liu, Fang, & Dong, 2021). Liu et. al. Discussed the key technologies for DT in three different categories: 1) Data related technology, 2) High fidelity modeling technology, and 3) Model based simulation technology. They described the details of each technology in their work and sum up them in the following figure (Fig. 1):

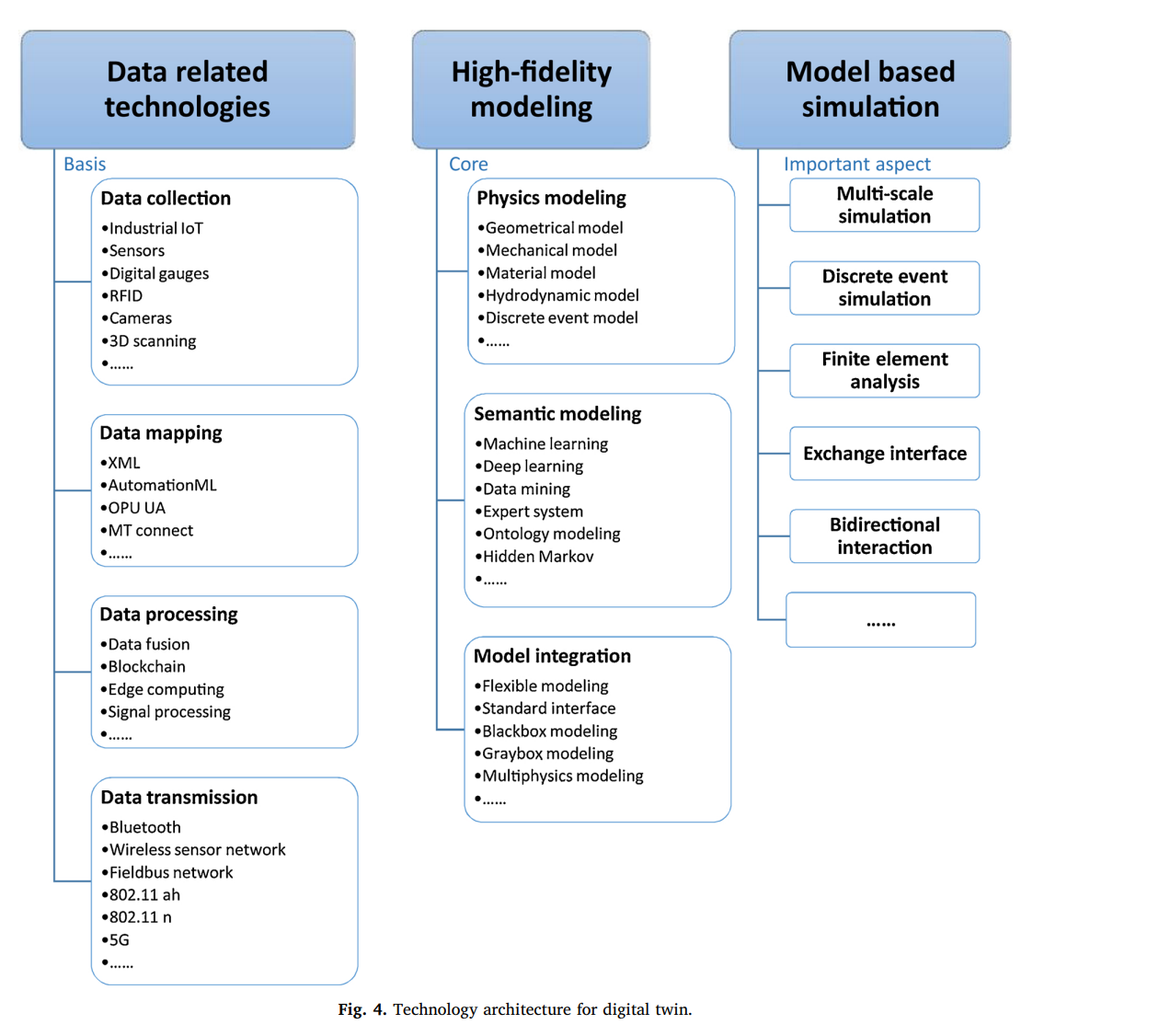


Figure 1: *Technology architecture for digital twin; figure is sourced from (Liu, Fang, & Dong, 2021).*

Another paper is (Tao F. a., 2018) which investigates the theoretical foundation of DTs in four parts given by:

1. DT modelling, simulation, and validation and accreditation.
2. Data fusion.
3. Interaction and collaboration.
4. Service.

These four different parts are coming from different science areas like, information science, production engineering, data science, and computer science. Each part is then discussed in detail in the paper.

We also review the work by (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) which investigates the components of a DT as follows (they are also listed in the corresponding table that is demonstrated further below):

1. **Elementary components:** The elementary components are those without which a DT cannot exist:
   1. Physical Asset (could be either a product or a product lifecycle)
   2. Digital Asset (the virtual component)
   3. Information flow between the physical and digital asset (this could be 1-way or 2-way/bijective)
2. **Imperative components:** The imperative components add to the properties of DT, to make it an all-encompassing tool of simulation, real-time monitoring, and analytics. Without these, the uniqueness of DT ceases to exist. The existence of each of these components depends majorly on the domain and application of DT.
   1. IoT devices — to collect sensors’ information from different sub-components of the physical asset and edge devices.

Requires: High-fidelity connection between IoT devices, for accurate and timely flow of information.

* 1. Data — gathered from different IoT components and software; it is required to monitor the system, guarantee correct behavior and provide input to the machine learning system.

Requires: Big data analysis and storage tools for extracting useful information from data.

* 1. Machine learning — for predictions and feedback, as well as to identify effective mitigation strategies, in exceptional circumstances.

Requires: A joint optimization feature for the subcomponents of the DT.

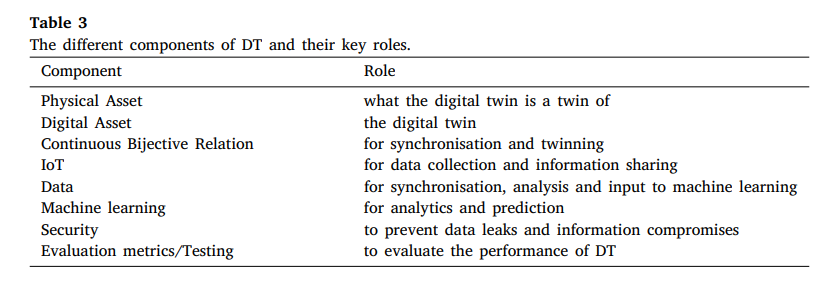
* 1. Security of data and information flow among various components involved in the DT.

Requires: Security protocols for information sharing and authentication, and authorization mechanisms.

* 1. DT performance evaluation.

Requires: Evaluation metrics (e.g., accuracy, resilience, robustness, costs), and evaluation method.

Table 4: *The different components of DT and their key roles; table is sourced from (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022).*



Based on (Glaessgen and Stargel, 2012), DT consists of three parts: **physical product**, **virtual product,** and **the linkage between physical and virtual product**. It serves as a bridge between the physical world and the digital world. On one hand, the physical product can be made more ‘intelligent’ to actively adjust its real-time behavior according to the ‘recommendations’ made by the virtual product. On the other hand, the virtual product can be made more ‘factual’ to accurately reflect the real-world state of the physical product. Also see Tab. 4 that represents the details of DT components and their key roles.

## DT characteristics

As discussed above, the DT framework replicates the physical entity, which we also called the physical twin, in a virtual environment. Such a sophisticated framework enables the researchers to, for e.g., perform different experimental analysis and/or examine the properties of the physical entity in a virtual environment.

Subsequently, a well-established DT framework should be able to interact between (A) the physical and virtual twins, (B) the DT and different DTs in the surrounding environment, and (C) the DT and the domain experts (i.e., the users or agents operating on the DTs) through proper accessible interfaces. This then raises the question of what specific characteristics a DT framework has, and based on what measurements we may assess the performance of a DT framework. Answering these questions strongly depends on the domain that we are working on, of course, but several important characteristics can be named that are commonly shared among different DT frameworks. Especially for our work in TA GOLO, we should find the specific characteristics that are related to the distributed systems and field-related tasks. We will discuss such important DT characteristics in the following.

Intuitively, data plays a substantial role in the DT framework, providing a well-established and active twinning between the physical and virtual spaces. As a consequence, there are a variety of data-related characteristics that a DT framework should poses:

First, note that all the collected (and later exchanged) data must be stored in a data storage that is accessible (specially by the DT, but also by the agents and users) through the whole lifecycle. For these storing processes, the researcher should then be aware of the data structures, details, and even properties to establish a suitable storage platform. For instance, data could be collected in real-time from the physical entity, and therefore require real-time storage and processing capabilities. Together with such a dynamic data, we also collect (A) historical static data, which corresponds to, e.g., the memory and information of the physical entity that does not change over the time (this historical information can be provided by human expertise or by past actions); and (B) descriptive static data, which describe important characteristics of the physical twin (e.g. its requirements and constraints, in the case of a product or device) (Barricelli, Casiraghi, & Fogli, 2019).

Besides data storage, the DT must be able to treat high-dimensional data and must therefore be equipped with effective high-dimensional data-(de)coding and analysis techniques, as well as data fusion algorithms for integrating the multiple data sources and produce more consistent, accurate, and useful information than that provided by any individual data source. This is a vital task since with the current status of the technologies, very Big data is available for the researchers that should be collected, stored and further processed, usually in a real-time procedure. In the case of distributed systems and field data that we are concerned with in TA GOLO, such a Big and various data are expected to be available to the DT framework.

We further refer to the abilities of the DT framework for analyzing the collected data and exploiting a variety of statistical and mathematical methods as well as advanced Artificial Intelligence (AI) and Machine Learning (ML) approaches. In fact, the DT characterizes, understands, clusters, and classifies the input data from the physical twin and/or the surrounding IoT environment, thanks to pattern recognition, unsupervised/supervised learning, and statistical applications. This allows to detect changes and identify important patterns and trends by analyzing data. We thus denote such characteristics of the DT framework as the ML-related characteristics which contain a variety of algorithms and techniques that can be used in the framework.

Among ML and AI algorithms, those algorithms, and models capable of dealing with the specific characteristics of the collected field data are of interest. For instance, algorithms for feature selection and/or feature extraction tasks have been frequently investigated in previous studies. As an important application, we can make use of such algorithms in order to reduce the data dimensionality of the collected data while keeping the most informative features; i.e., without (much) reducing the applicability of the used algorithms and the informative structure of the data. For analyzing and processing Big data this task can be seen as more distinct that reduces costs of storage and data analysis (computational costs) drastically.

Another important characteristic of the DT framework that we discuss here is the **self-improvements** and **closed-loop optimization characteristics**. The DT should have a self-adaptation and self-parametrization capabilities, which allow to resemble the physical twin during its whole lifecycle. This task might be easily and quickly accomplished, by developing a highly modular and parameterized DT. Modularity guarantees that changes in one module do not affect other modules. Parameterization guarantees an easy modification of DT status. Evolutionary algorithms, or stochastic optimization may be used for choosing the parameter settings producing the best fit between the DT and its twin.

In other words, the DT should have a closed-loop optimization of its components during the lifecycle. Closed-loop optimization refers to the continuous exchange of data between the cyber and physical worlds in order to continuously optimize the physical side. The DT is a virtual model of the physical object with the potential of understanding changes in the status of the physical entity through sensing data, to analyze, predict, estimate, and optimize changes. The physical entity should respond to the changes according to the optimized scheme received from the DT, should continuously send real-time data describing novel statuses, and then be ready to respond to novel optimizing ‘‘commands’’ received from the DT. Through such cyber-physical closed-loop optimization, DT technology could enable the performance improvement of the whole manufacturing process. In fact, extracting the salient value and information associated with the continuously acquired and exchanged big data is important to enable effective real-time cyber physical synchronization and such a closed-loop optimization.

One aspect of these continuously optimizing the DT components and self-improvement, corresponds to the investigated ML and AI algorithms, whose predictive capability is refined/optimized through the lifecycle as they process the continuously received sensed data from the physical twin and the surrounding environment. This virtual (cognitive) brain applies descriptive, predictive, and prescriptive algorithms thus allowing to perform a set of tasks as an intelligent product.

Another important characteristic of the DT that we concern with is the **decision-making characteristics**. As mentioned before, one of the benefits of using DT frameworks is their ability to exploit predictive analytics and consequently, predict future statuses and important changes (such as failures) in the product lifecycle.

The DT uses the result of descriptive and predictive techniques as input of prescriptive analytics to make decisions relevant to its own destiny, by computationally determining a set of high-value alternative actions or decisions given a complex set of objectives, requirements, and constraints (described by the historical and static data). It eventually applies (stochastic) optimization algorithms to achieve the best outcome, while addressing uncertainty in the data.

Beyond applying predictive and prescriptive algorithms, the DT codes the computed prescriptions and optimization schema by exploiting proper ontologies and high dimensional data-coding techniques. This allows sending feedback to both the physical twin and to other DTs in the whole environment. On the other side, end users may exploit interaction interfaces to access the computed information and to view the DT status. Finally, the DT provides modeling and simulation applications for representing, in a realistic and natural way, both the current status of the physical twin, and different ‘‘what-if’’ scenarios. (Barricelli, Casiraghi, & Fogli, 2019).

We finally consider the **integration level** or **Interoperability characteristics** of a DT framework. DTs are also characterized by their integration and interoperability capabilities. Integration is based on the aggregation of several digital twins and can be expressed by integration levels. Integration level can include for example the functional or self-sufficient state of a product in the field, the systematic interaction of the product with its environment, the presentation of a manufacturing component, a manufacturing machine, a production location. Interoperability means that the individual digital twins are computable with each other and communicate with each other. In addition, interaction of digital twin is the subject of current and future research activities (Stark, 2020).

## Building a functional DT

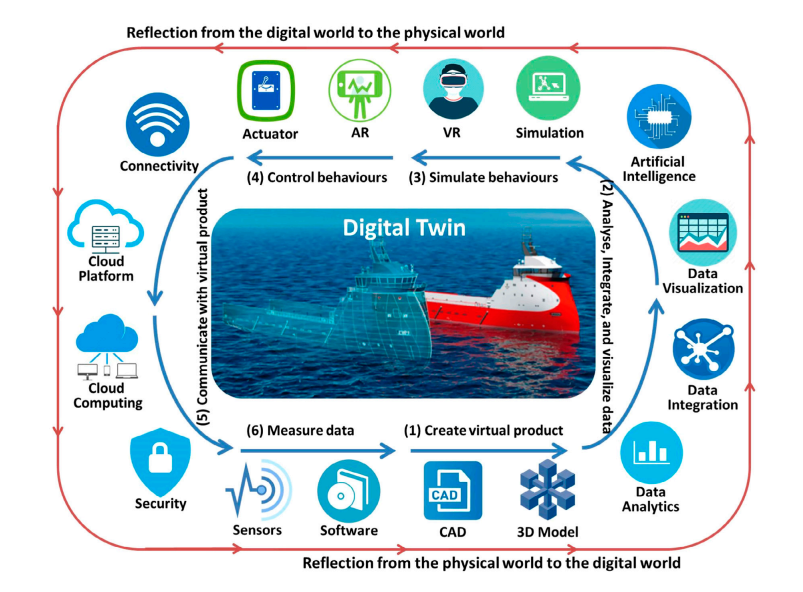
We now discuss the subsequent steps that are required to establish a DT framework in practice. As before, we make use of the previously established frameworks and discuss the different (substantial) points that have been considered in the literature (either for simple manufacturing tasks or for the case of complex physical systems).

The general digital twin mode for a product consists of three parts, which are the physical entities in physical space, the virtual models in virtual space, and the connected data that tie physical and virtual worlds. The physical entities are the real product that can be operated by the users. They are manufactured from raw materials or parts, through machining, assembly, and other processes. The physical entities have different characteristics, behaviors and performance in the course of manufacturing, use, Maintenance, Repair & Overhaul (MRO), disposal, and other operations, and a lot of data are generated. The virtual models are the mirror images and mapping of the physical products in the virtual space. They could reflect the whole lifecycle process, as well as simulate, monitor, diagnose, predict, and control the state and behaviors of the corresponding physical entities. The virtual models include not only the geo-metric models, but also all rules and behaviors, such as material properties, mechanical analysis, and health monitoring. The connected data include the subsets of physical data and virtual data, as well as some ‘new data’ that are acquired after the integration, fusion and analysis of physical data and virtual data. In the process of design and production, the parameters of the virtual models are passed to the production line and the virtual models are processed into real physical products. Through digital detection or measurement, the product attributes, operating status and other data are fed back to the virtual models, achieving a two-way data transmission process. By constructing the product models in the virtual space, as well as the feedback of the digital models to the physical space, the digital twin mode achieves a closed-loop process (end to end framework). The digital twin mode for a product can collect and accumulate the data continuously and knowledge on the entire lifecycle process, such as design, manufacturing, quality inspection, MRO; and these data and knowledge could continue to be reused and improved (Tao F. a.-Y., 2019).

The DT architecture investigated in (Wang, et al., 2022), known as MDT, leverages the concepts of cloud-edge computing to construct the dataflows throughout the framework. In detail, they make use of cloud computing and edge computing, enabling both real-time and bulk-batch ingestion, processing, and analytics of mobility data. They specifically divide the MDT architecture into four layers of 1) the cloud layer, which is built on AWS and its Virtual Private Cloud (VPC); 2) the edge layer, which has a computing component, a communication component, and a storage component; 3) the device layer, which generates data and consumes guidance; and 4) the API (Application Programming Interface) layer, which hooks up the cloud layer with external APIs (Wang, et al., 2022).

Another study that details the different steps of constructing a DT framework is (Tao F. a.-Y., 2019), which discusses this procedure in six subsequent steps. Although their work is mainly based on a DT product design, the discussed steps can be also used in general for other types of DT frameworks that aim at a more complex physical system (such as the distributed systems or field-related systems that are of interest for our work in TA GOLO). We discuss these six steps that may be carried out for creating a functional DT in the following; also Fig. 2 presents the details of this steps in a schematic way.

Step 1: Build a virtual representation of the physical product: The enabling technologies of this step are computer-aided design (CAD) and 3D modelling. Both are commonly used technologies in product design. The virtual product includes three aspects: elements, behaviors, and rules (Tao, Chenget al., 2017). At the level of elements, the virtual product model mainly includes the geometric model and physical model of the product, user and environment, etc. At the level of behaviors, the authors not only analyze the behavior of products and users, but also focus on the analysis of the product and user

Figure 2: The six different steps of creating a DT framework for manufacturing based on (Tao F. a.-Y., 2019).

interaction generated by the behavior and modelling. At the rules level, it mainly includes the evaluation, optimization and forecasting models established following the law of product operation.

Step 2: Process data to facilitate decision-making: Data collected from different sources (i.e., mainly from the physical product, and also from the Internet) are analyzed, integrated and visualized. Firstly, data analytics is necessary to convert data into more concrete information that can be directly queried by designers for decision-making. Secondly, since product data are collected from diverse sources, data integration is useful for discovering the hidden patterns that cannot be uncovered based on a single data source. Thirdly, data visualization technologies are incorporated to present data in a more explicit fashion. Finally, advanced artificial intelligence techniques can be incorporated to enhance a DT’s cognitive ability (e.g. reasoning, problem solving and knowledge representation), so that certain relatively simple recommendations can be made automatically. In this part we might consider some tasks like classification, clustering etc. for the sake of data processing or data analyzing.

Step 3: Simulate product behaviors in the virtual environment: The enabling technologies of this step include simulation and virtual reality (VR). The former is used to simulate key functions and behaviors of the physical product in the virtual world. In the past, simulation technologies are widely used in product design. On the other hand, virtual reality (VR) technologies play the role of involving designers and even users to ‘directly’ interact with the virtual product in the simulated environment. Recently, VR technologies are increasingly employed to support virtual prototyping and product design (Stark, Israel, and Wöhler2010). Many readily available VR hardware devices can be directly adopted for digital twins.

Step 4: Command the physical product to perform recommended behaviors: Based on the recommendations of DT, the physical product is equipped with a capability, by means of various actuators, to adaptively adjust its function, behavior and structure in the physical world. Sensors and actuators are the two technological backbones of a digital twin. The former plays the role in sensing the external world, whereas the latter plays the role in executing the desirable adjustments requested by DT. In practice, the commonly used actuators that are suitable for consumer products include, for example, hydraulic, pneumatic, electric, and mechanical actuators. In addition, augmented reality (AR) technologies can be used to reflect some parts of the virtual product back to the physical world. For example, AR enables end users to view the real-time state of their products. Recently, AR technologies are increasingly applied in the factory domain production engineering (Nee and Ong2013).

Step 5: Establish real-time, two-way, and secure connection between physical and virtual product: The connections are enabled using a number of technologies, such as network communication, cloud computing and network security. Firstly, networking technologies enable the product to send its ongoing data to the ‘cloud’ to power the virtual product. The feasible networking technologies for consumer products include, for example, Bluetooth, QR code, barcode, Wi-Fi, Z-Wave, etc. Secondly, cloud computing enables the virtual product to be developed, deployed, and maintained completely in the ‘cloud’, so that it can be conveniently accessed by both designers and users from anywhere with an Internet access. Lastly, since product data are directly and indirectly concerning user-product interactions, it is critical to guarantee the security of connections. In light of the Internet of Things, much effort has been devoted to connecting the physical and virtual product, which can be adapted for the DT research.

Step 6: Collect all kinds of product-related data from different sources: Generally speaking, there are three types of product-related data that should be processed by DT. For ordinary products, physical product data is usually divided into product data, environmental data, customer data and interactive data. Product data contains customer comments, viewing and download records. Interactive data consist of user-product-environment interaction, such as stress, vibration, etc. Using the sensor technology and IoT technology can collect some of the above data in real time, and analyze from the product manual, web page customer browsing records, download records, evaluation feedback, etc., can obtain the rest of the data. The collected data are fed to Step (1) in order to close the loop towards building more functional virtual products.

## Dataflow in DTs

The most important aspect of a DT framework is how dataflow is defined between the two physical and virtual spaces. Such an exchange of information and data has been commonly termed communication or communication phases that exist not only between the two spaces but also between the different components of a physical or virtual space. In the following, we discuss how such a communication phase is discussed in different studies.

In (Wang, et al., 2022), a specific communication plane between the physical and digital spaces is defined in the context of the MDT framework, which provides seamless connections between these two spaces. In fact, without such a communication plane, data cannot be transmitted between these two spaces to enable their interactions and synchronizations, hence Digital Twins cannot be formed. In detail, the MDT framework’s end-to-end process starts from sampling data in the physical space, where all or part of the data is transmitted *upstream* to the digital space via the communication plane. Those data will go through one or multiple processes in the digital space internally, including storage, modeling, learning, simulation, and prediction, and the resulting data are transmitted *downstream* to the physical space via the communication plane. Those data, upon receiving, is applied by the actuators of the physical space to fulfill the end-to-end process.

Since cloud computing is leveraged in this MDT framework, the digital space of the framework is deployed fully or partially on the commercial and/or private cloud. Therefore, the communication module needs to provide access to the cloud for the physical space, which is either direct access or indirect access (via edges). The MDT framework does not necessarily require any specific wireless communication technology (dedicated short-range communication (DSRC), C-V2X, or something else in the future) to be served as the communication plane, as long as it can be applied to transmit data between the physical space and the digital space (Wang, et al., 2022).

One of the biggest strengths of the MDT framework over traditional mobility system frameworks is the data lake, which is a centralized repository that allows structured or unstructured data at any scale to be stored. Traditionally, mobility data measured by a physical entity is only saved in its onboard data storage due to the lack of communication capability. Such data are only used for the physical entity itself without being shared with other entities and will be wiped out once the maximum size limit of the onboard data storage is met. However, with the proposed MDT framework, mobility data measured by Human, Vehicle, and Traffic blocks in the physical space can be transmitted to the digital space through the communication plane and stored in the data lakes of associated DTs for future use. Such data can be used for the microservices not only in the original mobility block but also in other blocks (e.g., traffic signal data measured by the Traffic block can be used for both the “real-time monitoring” microservice in the Traffic DT and the “cooperative control” microservice in the Vehicle DT) (Wang, et al., 2022).

**The physical-to-virtual connection** is established with a technology that allows the transfer of information from the physical environment to its virtual twin, including web services, cellular technology, WiFi, etc. The virtual twin is adjusted gradually with the functioning of the physical twin by continuously collecting the differences between the two environments. These connections allow the monitoring of responses to both conditions and interventions. The conditions mainly occur in the physical environment, whereas the interventions take place within the virtual twin. Thus, a digital twin holds a real-time status of the physical counterpart.

**The virtual-to-physical connections** represent the information circulating from the virtual to the physical environment. This information may change the state of the physical twin by displaying some data or changing the system’s parameters (for optimization, diagnostics, or prognostics). Although virtual-to-physical connections are very helpful in DT modeling, they are not always included in the description. Instead, it is common to consider a one-way connection, i.e., physical-to-virtual. Finally, the data and the information from both physical and virtual worlds are stored and analyzed at a centralized server—or a cloud computing platform— where the final decisions related to optimization, diagnostics, or prognostics, are made. (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021).

Besides the MDT discussed above and based on (Tao F. a.-Y., 2019), from the perspective of data science, digital twin can be regarded as an information filtering and integration system. On one hand, DT functions to filter the massive amount of data towards useful information that can be directly queried for design operations. On the other hand, DT functions to integrate different types of data to discover hidden patterns and cross-check analysis results. As illustrated in Figure 4, a typical data lifecycle includes data collection, data transmission, data storage, data integration, data processing, data cleaning, data analysis and data mining. Through such a lifecycle, raw data are converted to useful information that can be directly queried by designers to support their design decision-making. Some steps relevant to DTPD are illustrated as follows:

1. **Data collection:** Data collection is the first step of the whole process. In DTPD, the physical product data are collected by sensors from both consumers and products or extracted through interviews, downloads from online sources or reading documentation (Jin et al., 2015).
2. **Data integration:** According to the definition of DT, data from different sources, formats and characteristics must be integrated for the next step. Data integration involves combining data that reside in different sources and providing users with a unified view (Lenzerini, 2002).
3. **Data cleaning:** Data cleaning means the process of identifying, removing, and correcting various kinds of errors included in the dataset of DTPD. Some common tasks of data cleaning include record matching, identification of inaccurate data, evaluation of overall quality, deduplication, and column segmentation (Baheti and Gill, 2011).
4. **Data mining:** The goal of the data mining process in DTPD is extracting information from a data-set and trans-forming it into a visualization structure for further use (Keivanpour and Ait Kadi, 2017). Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online up dating (Chakrabarti et al.2006).

## DT and product lifecycle management

The Product Lifecycle Management (PLM) strategy has been present in the industry for many years and is considered as the most effective way of managing the components, products and systems of a company all the way across their lifecycles from the first idea of the product to its disposal. On the other hand, the digital twin can be defined as a set of models, linked with each other as well as with the physical product enabling data storage and real-time processing. In contrast to a digital twin, the PLM strategy provides a framework, which serves as a single source of truth connecting the partial models that describe the physical product. The models can receive the data stored in a product data management system (PDM) (Adamenko, Kunnen, & Nagarajah, 2020).

The by (Barricelli, Casiraghi, & Fogli, 2019) led us to describe two possible lifecycles for DTs, from their design to their dismissal. The former refers to a case where the object that has to be twinned still does not exist and, in this case, the design process simultaneously conceives both the object and its DT. The latter is about an object that already exists but has no DT in place; in this case, the design process focuses on the extension of the objects to make it connected. Both lifecycles share the same timeline: a first Design phase, followed by a Development phase, an Operational phase, and finally a Dismissal phase. For describing these two lifecycles, we use a running example of a medical device (the object) – i.e. a computer tomography scanner. In this first case, the DT starts living before the physical object as a Prototype (Prototype DT Object), which is then used by designers during the Design phase of the Prototype Object. During the initial part of the Design phase, the Prototype DT Object is used, as if it was the real prototype, to simulate, test, change, and eventually validate design choices, until the best solution is found. During this part of the design phase, designers exploit:

1) Historical data the Prototype DT Object acquires from any other already existing DTs linked to similar devices.

2) Static data (e.g., data describing the product requirements, customer preferences, bill of materials).

3) The results of simulations performed by the Prototype DT Object, the result of predictions computed by the Prototype DT Object, and its suggestions and optimization schema.

When the design of Prototype DT Object is completed, the process moves to the Design of the prototype Object, during which the Prototype DT Object is eventually modified to address technical constraints that may arise during the prototyping of the physical Object. During the Development Phase, the Prototype DT Object evolves becoming a Development DT Object, which must interact with the production machines to follow and optimize the assembly/construction of the Object, i.e. its physical twin. When the Object is finally built, the Development DT Object starts being a Product DT Object, and this moves the process into the Operational phase. The Product DT Object fully resembles the Object: it has the AI acquired by the preceding stages of its life, and is therefore ready to follow and mirror the medical device (Object) while it is being used. During its existence, the intelligence of the DT Object grows and self-adapts to the Object (in the case of the medical equipment, for example, it might start learning the most requested examinations, and the days when more or less examinations are performed).

When the Object stops being used (due to obsolescence or any other reason) it must be disassembled, and the Dismissal phase begins, first for the Object and then for the DT Object. The stored historical data of the Product DT Object are backed-up and made available to other DT Object as well as to domain experts; in this way, designers, or any other domain expert, will be able to use the collected information to optimize the production of future devices. The second lifecycle is shown in Fig. 4. The difference between this lifecycle and the previous one (Fig. 3), is that the Object is already in place and in use, but it does not have a connected DT yet. In this case, the Design phase regards the development of a novel Prototype DT Object (which is tested, changed and finally validated), the Development phase regards the development of connections between the existing Object and the DT Object (which is called Development DT Object in this phase), while the Operational phase regards the operational life of the two twins, the Connected Object and Product DT Object, which live in concert until their dismantle in the Dismissal phase. To sum up, during their lifecycle, the (physical and digital) twins base each step of their existence on a synergic and continuous interaction, which allow monitoring, predicting, and optimizing all their functionalities. The continuous interaction hides the differences among them, and they can act as a whole (‘‘Synergy is the creation of a whole that is greater than the sum of its parts’’) (Barricelli, Casiraghi, & Fogli, 2019):

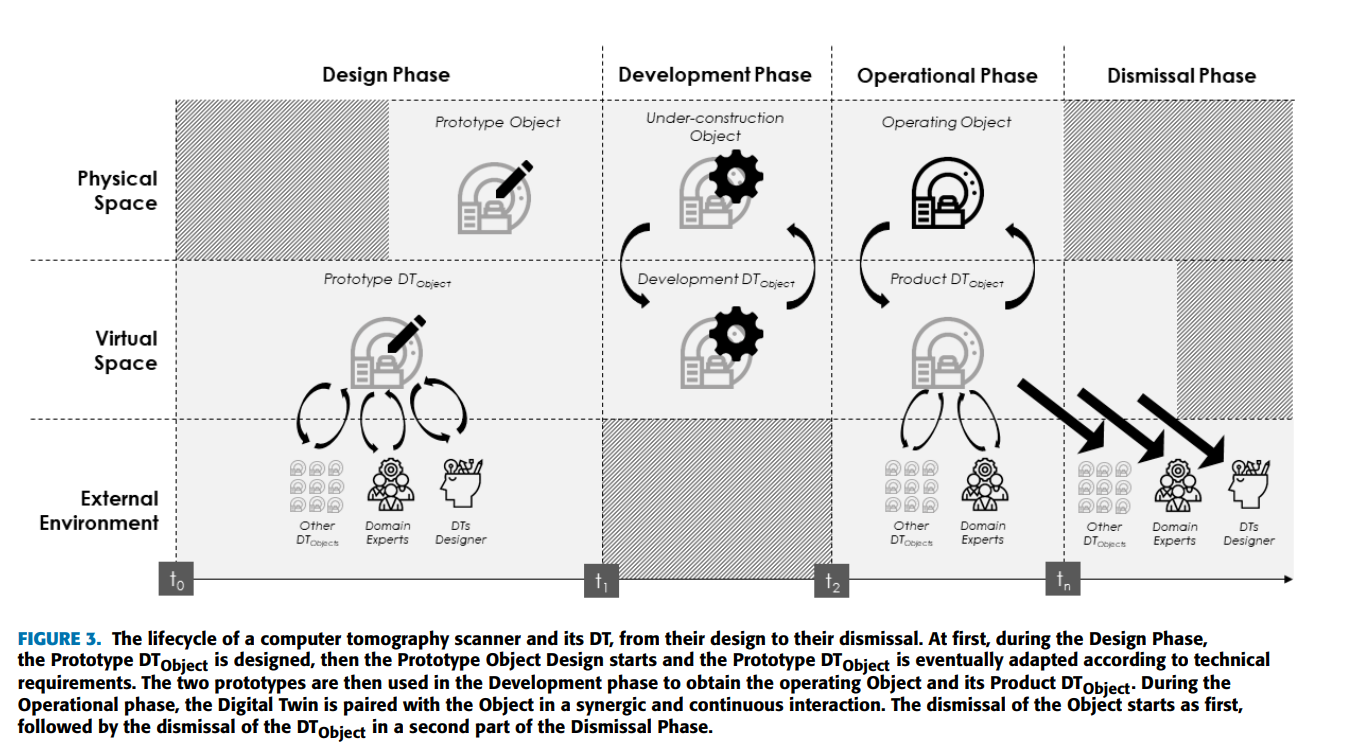


Figure 3: *The lifecycle of a computer tomography scanner and its DT, from their design to their dismissal. Figure is sourced from (Barricelli, Casiraghi, & Fogli, 2019).*

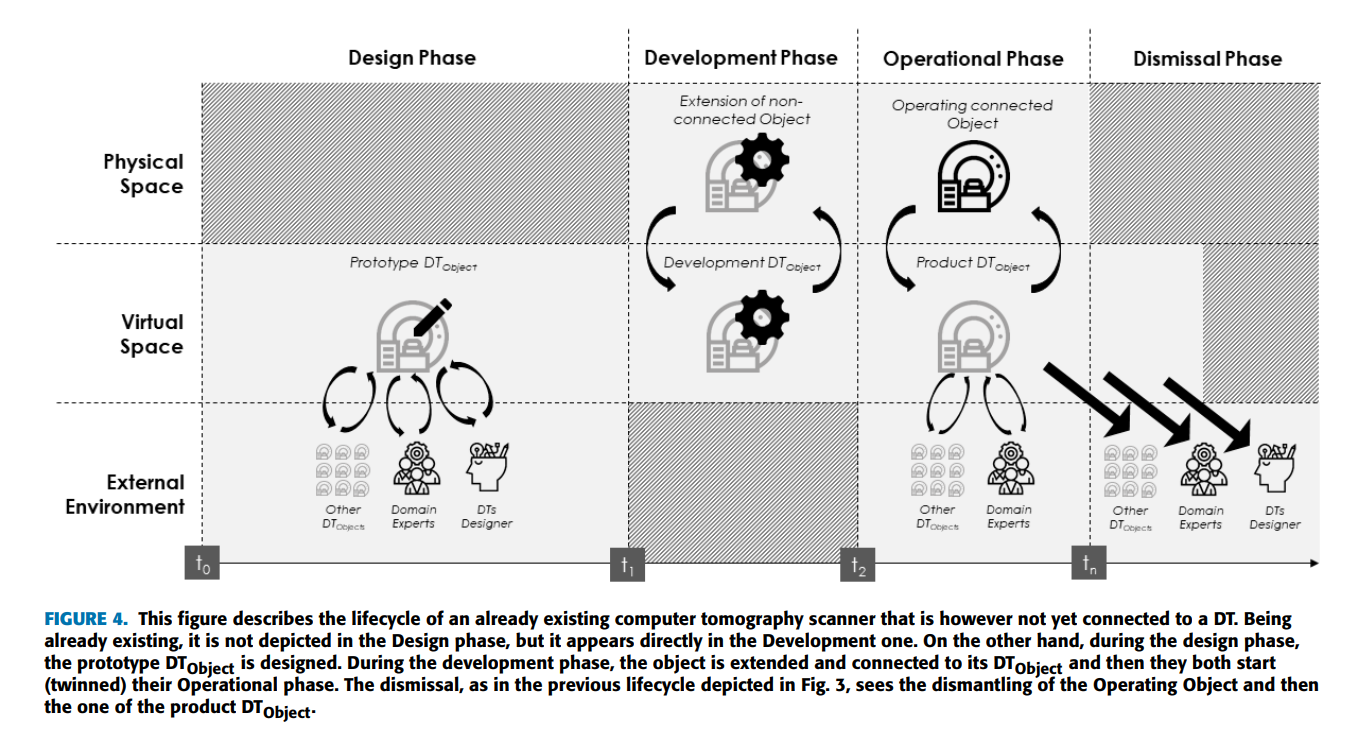


Figure 4: *The lifecycle of an already existing computer tomography scanner that is not connected to a DT yet. Figure is sourced from (Barricelli, Casiraghi, & Fogli, 2019).*

In addition to the discussion above, (Wärmefjord, Söderberg, Schleich, & Wang, 2020) states that a digital model, which will be the core of a digital twin, should be developed, reused, and updated in the different phases of the product development process in order to provide as much value as possible. Therefore, the respondents of the interviews were asked about the current situation regarding different gates and goals related to geometry assurance in the different phases. The results from the interviews reflecting the present situation for the three main product realization phases are presented.

In another similar study, (Liu, Fang, & Dong, 2021) investigated the industrial application of DT in different lifecycle phases and discussed in detail. However, the summary of this research study are shown in following (Fig. 5):

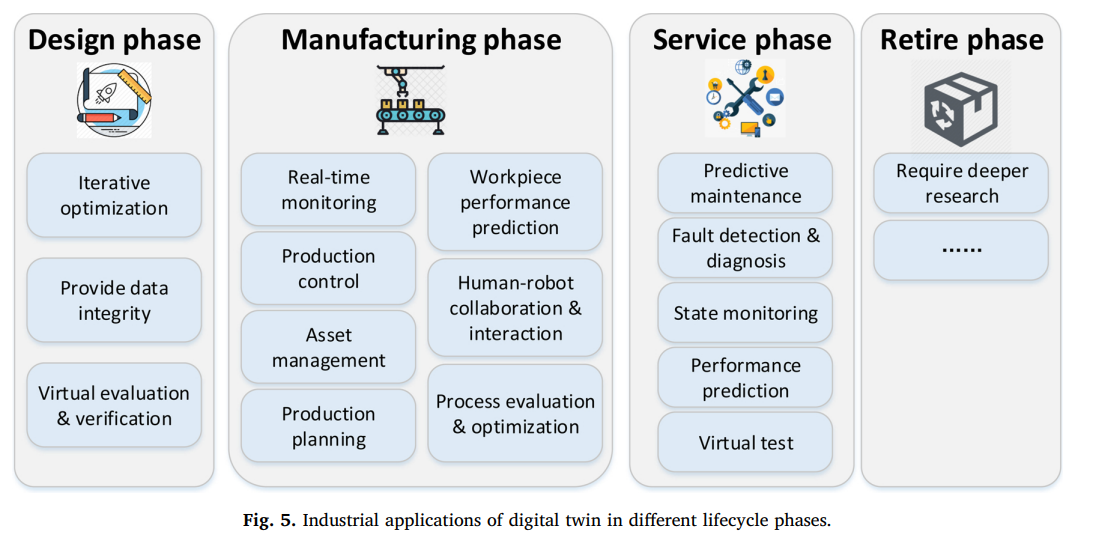
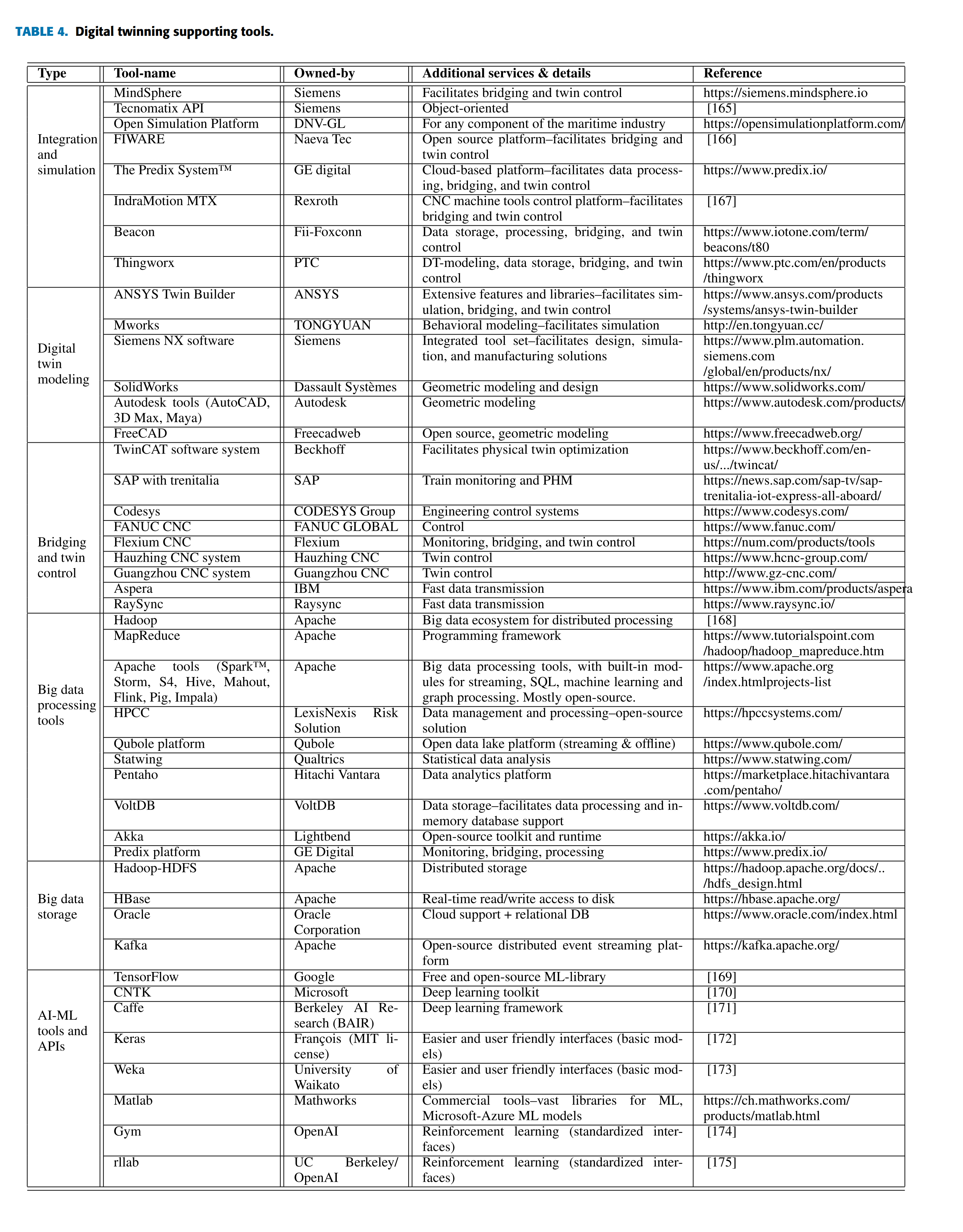


Figure 5: *Industrial applications of DT in different lifecycle phases; figure is sourced from (Liu, Fang, & Dong, 2021).*

## Tools and software

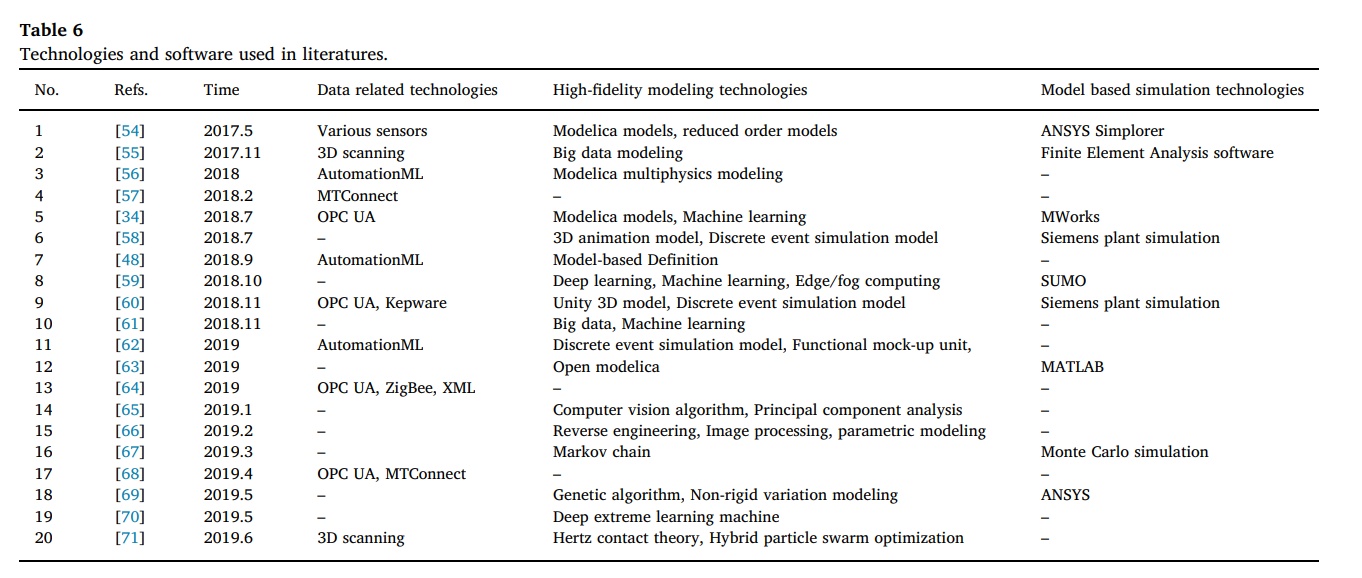
There is no standalone technology for DT implementation, rather, there is an integration of multiple technologies, including big data, AI-ML, IoT, CPS, edge computing, cloud computing, communication technologies, etc. Every technological component can be implemented with a variety of tools. Here, we only focus on the tools that facilitate component integration, digital twin simulation, twins bridging, physical twin control, data storage and processing, and machine learning. The following table (Tab. 5) presents the summary of widely used tools that may provide support in different stages of digital twinning (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021).

Table 5: *Digital twinning supporting tools; table is sourced from* *(Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021).*

****

Currently, commercial digital twin solution providers such as SAP, Visual Components, and AnyLogic are helping their customers to implement case-specific digital twins (Bärring, Johansson, & Shao, 2020). Plenty of technical challenges remain to be addressed to implement digital twin applications. To better grasp the development of digital twin key technologies, Tab. 6 lists technologies that academic publications used to build digital twins, ordered by paper accepted time to illustrate technology development (Liu, Fang, & Dong, 2021).

Table 6: *Technologies and software used in literatures; table is sourced from* *(Liu, Fang, & Dong, 2021).*



## Current challenges

Challenges related to the implementation of a DT are discussed in (Wärmefjord, Söderberg, Schleich, & Wang, 2020) and are listed below. They can be divided into four different categories: system level, simulation working process, management issues, and education.

In (Bärring, Johansson, & Shao, 2020) building a DT requires a homogenous perspective of the information that can persist across the organizational functional borders. Implementing such a perspective, requires overcoming five, existing obstacles that are making it difficult for manufacturers to effectively implement a digital twin:

1. The information associated with organizational functions (such as design, engineering, and manufacturing) is stored in silos. This significantly limits information sharing between functions.
2. There is a lack of detailed knowledge about the physical world. That knowledge is needed to understand and simulate the natural phenomena as it becomes virtually represented.
3. There is uncertainty about the possible states that a system can take and the time it will be in that state.
4. It is difficult to include human involvement, which happens in most manufacturing and logistics decisions, in the digital twin.
5. The needed digital data to represent the actual physical system is largely unavailable. (Bärring, Johansson, & Shao, 2020).

The rapidly increasing DT popularity and scope, as well as the involvement of IoT, big data, and AI technologies, broaden the research challenges of digital twinning. These challenges are categorized in the following five areas.

1. **Data collection:** IoT facilitates data harvesting from a physical twin (using sensors), data integration, and data sharing with the corresponding virtual twins.
2. **Big data challenges:** The explosive growth of IoT technologies in the industrial sector has led to the generation of large amounts of monitoring (sensor) data.
3. **Data analysis:** AI-algorithms for data analytics played a major role in DT for decision-making, as discussed in the literature. However, the selection of a particular model among hundreds of ML-models with customized configuration is challenging.
4. **DT standardization challenges:** Currently, there is no single standard that solely focuses on digital twinning.
5. **Security and privacy issues:** Some DT systems, such as human-DTs, product PHM, or defense-related DTs, are considered critical and may require stringent security and privacy guarantees. (Rathore, Shah, Shukla, Bentafat, & Bakiras, 2021).

Sharma et. al. (Sharma, Kosasih, Zhang, Brintrup, & Calinescu, 2022) discussed the following challenges for DT models. They also noted that these challenges vary in importance depending on the domain dependencies.

High-fidelity 2-way synchronization is especially hard for large scale industries, requires resources and high-stream IoT connection:

1. **Interoperability with existing software being used in a production lifecycle:** Industries use various software for tasks such as inventory, product management, operations. The compatibility of DT with these is a challenging issue, tackling which might lead to delay in implementations.
2. **Cybersecurity concerns, IoT security, cross industrial partners security**: With the digital twin operating across multiple industrial partners and inventory sites, the security concerns are inevitable. Not only the cross-industry security concerns but also the leak of real-time monitoring data can be hazardous to a firm.
3. **Add-Ons:** Using DT entails certain add-ons like cost, resources, and research. Since implementing DT and profiting from it is a timely process, DT can be costly if the life and span of a project are short. Building a software for DT also demands a team of programmers, developers, and domain experts to test the suitability of the software for the particular task. Moreover, like any technology, DT also needs to be updated according to the recent developments in the technologies it relies on (IoT, big data, machine learning). Industries with long-term DT use will therefore need to continuously invest in this research, which might lead to added cost. As DT requires interoperability among various components, real-time tools, formulating a joint optimization problem, and big data resources, putting these together can be time-consuming for an industry, and may lead to unwanted distractions.

# General Structure of the Proposed Digital Twin for the Task Area GOLO

In this section, we ask how the concepts of DTs can be generalized to the distributed systems and to the more intricate physical entities that encompass field and field-related data. This corresponds to the distributed systems investigated in TA GOLO, and as a specifical example, to the use-cases defined in the project. Despite the considerable benefits of DTs in the fields of industry and engineering, existing DTs are primarily domain-dependent and cannot be directly generalized to more complex scenarios. This is particularly evident when the physical system corresponds to a field-dependent entity within distributed systems; and hinders a straightforward implementation of DTs by researchers.

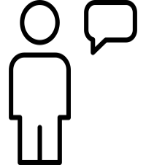
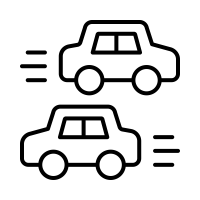
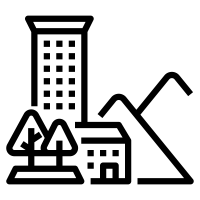
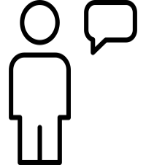
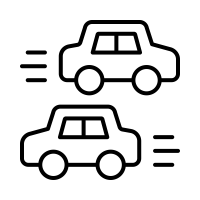
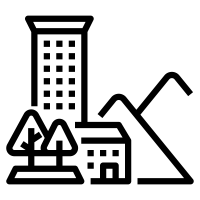
In this regard, we propose a generalized DT framework for distributed systems within both physical and virtual spaces, highlighting the communication between them as a primary focus. Subsequently, we define the necessary components and elaborate on their specific characteristics. This task is particularly challenging since the investigated distributed systems contain different moving and non-moving physical entities and thus the exact definition of the DT components becomes intractable. To address this, we consider three components in line with our research objectives: (A) a physical space which encompasses physical entities that will be modeled and analyzed, (B) a virtual space containing DTs that replicate the considered physical entities and provides the basis for the analysis and investigations, and (C) communication which links the physical and virtual spaces. As discussed before, the latter plays a vital role in DT technologies which distinguishes them from other simple simulation techniques. Similar to previous studies, here communication refers to the exchange of data and information between the two spaces forming an end-to-end model given the proposed framework. Specifically a real-time exchange of information allows to model the physical environmental conditions.

We will furthermore describe the specific characteristics of the proposed DT in connection to the investigated physical entity. The functionality of such DT components will be then specified and detailed. Within the virtual space, we define various digital shadows, each investigating a specific aspect of the considered physical entity. Additionally, we introduce a digital master that stores metadata and requirements of digital shadows. The prototype will be constructed in digital master which instantiates the phases and processes of digital shadows. Importantly, we later describe how dataflow (as a form of communication) between these different spaces and entities can be established.

## Field data and DT for field-related distributed systems

Our investigation in TA GOLO focuses on field data of distributed systems. We collect data from the physical space and transform it to the virtual space. Such field data, however, are commonly subject to a variety of difficulties corresponding to the environmental conditions, and one should carefully explore the different properties of the data, as well as the environmental surroundings in order to establish an effective data flow throughout the framework. To this, we make use of previous models and define a novel interaction between the physical and virtual spaces.

Fig. 6 presents a general overview of the investigated framework in TA GOLO. Field data ….. SOME NOTES ON FIELD AND DISTRIBUTED SYSTEMS



Data

Information

Physical Space

Virtual Space

Digital Thread

*Figure 6: A DT framework and the communication between physical and virtual spaces.*

Let us investigate the specifics of the proposed DT framework for distributed systems and explore the characteristics that such a framework should possess. As mentioned earlier, our focus encompasses three key components: the physical space, virtual space, and the communication that occurs between and within these two spaces. Within the virtual space, we precisely define our DT framework, comprising two integral elements: Digital Master which encapsulates metadata, requirements of digital shadows, tools for instantiation, and the necessary infrastructure, as well as Digital Shadows to store, analyze, predict, monitor, and simulate the behavior and performance of its physical counterpart by processing data from sensors, devices, and other sources. Moreover, we explore the dataflows between the digital master and digital shadows, as well as between the two spaces and entities. This process serves as a way of communication and enables effective information exchange and is particularly important since implementing a DT for field data and distributed systems requires tailored approaches based on data and dataflow to successfully address all the complexity of the domain. These tailored methods are necessary to handle the complex challenges present in this domain. Fig. 7 illustrates a visual representation of the proposed DT framework and the connection between the different components of the physical and virtual spaces. Data and information will be first transferred to the digital master and further analyses will be carried out in the digital shadows. In the initial loop of the figure, evaluation and outcomes refines the digital master. That is, if the chosen data analysis methods, for e.g., do not align with our data attributes or objectives, adjustments will be demanded. Similarly, alterations in the selected tools for instantiation might be necessary. In the second loop, enhancements can be made to the physical instantiation directly, leading to optimization, decision-making, or active monitoring. We will discuss each of these points in the following. We first begin by defining both the digital master and digital shadows, and then discuss the dataflow across the framework.

A diagram of a diagram

Description automatically generated

*Figure 7: Details of the proposed DT framework and the communication between physical and virtual spaces.*

## Digital master

The digital master encompasses metadata and all the necessary requirements, including information and specific settings, for both the digital shadow and physical instantiation. Essentially, the entire DT is primarily built upon its digital master, and the research objective aims to enhance the master by implementing and incorporating digital shadows. Note that for the sake of brevity, we hereafter use master to refer to a digital master, and likewise use shadow to refer to a digital shadow. Moreover, based on the objective of our study, we identify five key components within the master (as illustrated in Fig. 8), each containing distinct details for a successful implementation of the corresponding shadows:

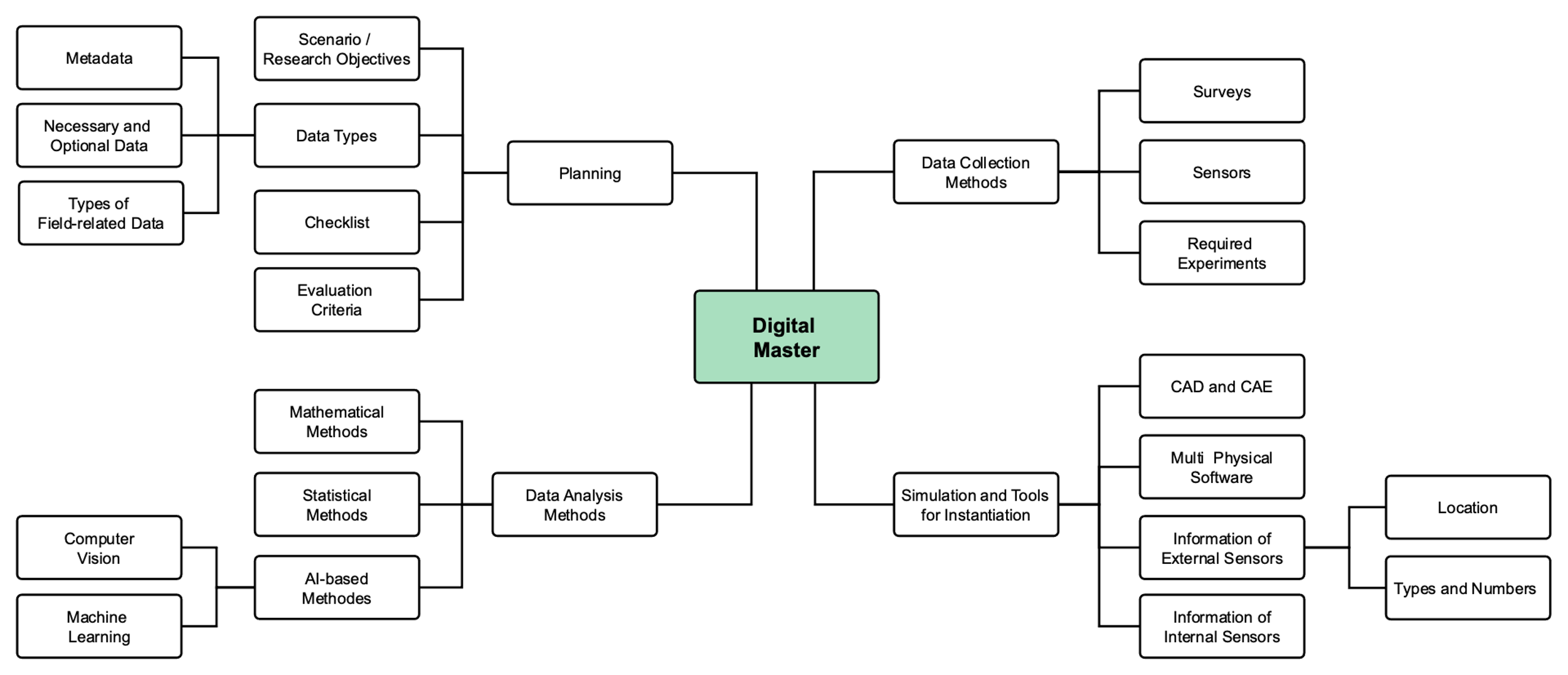
**Planning.** The initial component is ”planning”, following the fundamental objective questions and checklists of our research. This phase serves to provide clarity on the overarching goal of the established DT framework. Following that, we proceed to define the specific categories of data, encompassing both our metadata and the optional and essential features of data that are necessary in this context. Among the components within the master, metadata directly aligns with the FAIR principles, establishing a structured and coherent foundation for the entire process. Notably, metadata requires comprehensive exploration and refinement throughout the majority of DT phases. We update this information subsequent to the analysis of data within shadows and continuously refine the framework through an iterative optimization.

**Data Collection Methods.** Data collection is the systematic process of gathering important information from various sources. In the proposed DT framework, this step primarily occurs in the physical space, particularly after creating actual instances. In simpler terms, the digital master sets the initial requirements for creating our physical instantiation, and once that is done, we can start collecting data. Different methods of data collection lead to different types and when considering distributed systems in the field, we can classify data types into two main categories: static data and real-time data. Real-time data refers to information that is captured, processed, and transmitted instantly as events happen in the physical environment. It offers immediate insights and enables quick decision-making, setting DT apart from other technologies. Static data refers to information that remains unchanged over a certain period of time, usually until it is updated or replaced manually.

In the subsequent step, we focus on specifying the methods for data collection and transfer. We specifically consider techniques that enable us to efficiently collect data from physical entities and exchange them within both virtual and physical spaces. The level of IoT integration and technologies utilized in this step demonstrates the varying degrees of automation achieved by DTs in Industry 4.0. In addition, a structured and comprehensive metadata framework, coupled with effective classification, remains essential for ensuring a database of high quality.

We assume data to be obtained from different sources. That is, in addition to leveraging modern technologies for collecting sensor data, we also integrate alternative methods during this phase. For instance, the surveys and questionnaires given to pilot users can be used to provide insights about the field and further enriches our understanding. Consequently, such data can be used in conjunction with the collected sensor data for our analysis. Moreover, we expect to add experimental data when extra tests, typically done in real-world conditions, are needed.

Besides, this stage involves managing (and also specifying) both optional and necessary data in the direction of our research objectives. Although our foremost focus here is on facilitating the collection of necessary data that are crucial for achieving our research goals, it is equally important to also adopt a broader perspective and consider capturing data and details that might be valuable for the future investigations. The latter point can be specifically considered to enhance data reusability. To illustrate, assume the environment temperature might not be directly important for our objectives, but it may be beneficial to acquire it if the cost is also minimal. By doing so, we attain two key benefits: First, in the future development of our physical entity, temperature data might unexpectedly become relevant, and having historical data already available is advantageous; and second, other researchers pursuing different goals in the same domain could utilize our temperature data, exemplifying data reusability —a principle that fosters collaboration and shared knowledge.

****

*Figure 8: Different components of the digital master investigated here.*

**Data Analysis Methods.** We further gather crucial information about the methods that we use for processing and analyzing the collected data. This task can be seen as one of the important steps for finding valuable insights and improving the technical system, which depends directly on the data types and the techniques we use for data collection. Nowadays, a variety of different analysis methods are available for researchers that can be utilized in different tasks within the context of a DT framework. This can help researchers get more accurate results, but it can also make it difficult to choose the right analysis methods. In the master we sorted the methods, with a special focus on mathematical and statistical techniques, as well as advanced artificial intelligence and machine learning approaches. These methods are used for tasks like patterns recognition, objects detection, optimization, classification, and in general getting knowledge from the data collected in the field.

**Simulation and Tools for Instantiation.** In this step, our focus lies on investigating and identifying the essential tools and information required for both the instantiation phase and the pilot phase (TRUE?!). This enables us to obtain the necessary data and information for the creation of our digital shadow(s) which further leads to perform the desired analyses. Given our direct engagement with the distributed system in the field, certain sensors will be primarily utilized such as LiDAR, cameras, and Inertial Measurement Unit (IMU) that form our framework for the data collection, the instantiation process and, by extension, the establishment of our digital shadows.

Every tool we incorporate into our framework, whether attached to physical entities such as humans, vehicles, or the environment, contributes to this crucial step in our digital master strategy. These tools actively collect and provide us with information and data of the associated physical entities, which further constructs the foundation of our digital shadows. Particularly, the relation between the tools and the physical allows for what we refer to as instantiation —a comprehensive representation of the real-world elements in our digital domain.

## Digital shadows

The required information that are collected from the field will be transferred to the digital shadow(s) where the primary focus is on processing the collected data. The core function of the digital shadow lies in the conversion of raw data obtained from the physical space into meaningful insights. These insights directly contribute to monitoring and controlling the physical object, and also to enhancing the digital master based on the objective of our research. This will construct an iterative procedure that continues until predefined objectives are achieved.

Digital shadows assume responsibilities encompassing data processing (including data pre-processing, quality assurance, analysis, etc.), data integration, visualization, and simulation. Given that a crucial aspect of the shadows involves interacting with data, a portion of these responsibilities is executed within the context of data management within the DT framework. The outputs and information derived from shadows can lead to improvements in two distinct ways. Firstly, evaluations may reveal issues originating from the master. In such cases, enhancements should be made to the Digital Master itself, such as refining instantiation requirements, tools, methods, planning or requirements related to model fidelity. Secondly, the evaluation can have an impact on the physical entities, such as online decision making and monitoring. In our investigations here, we consider four essential tasks for each shadow that we elaborate in the following. Fig. 9 also illustrates the structure of a digital shadow and gives and overview of these four main tasks.

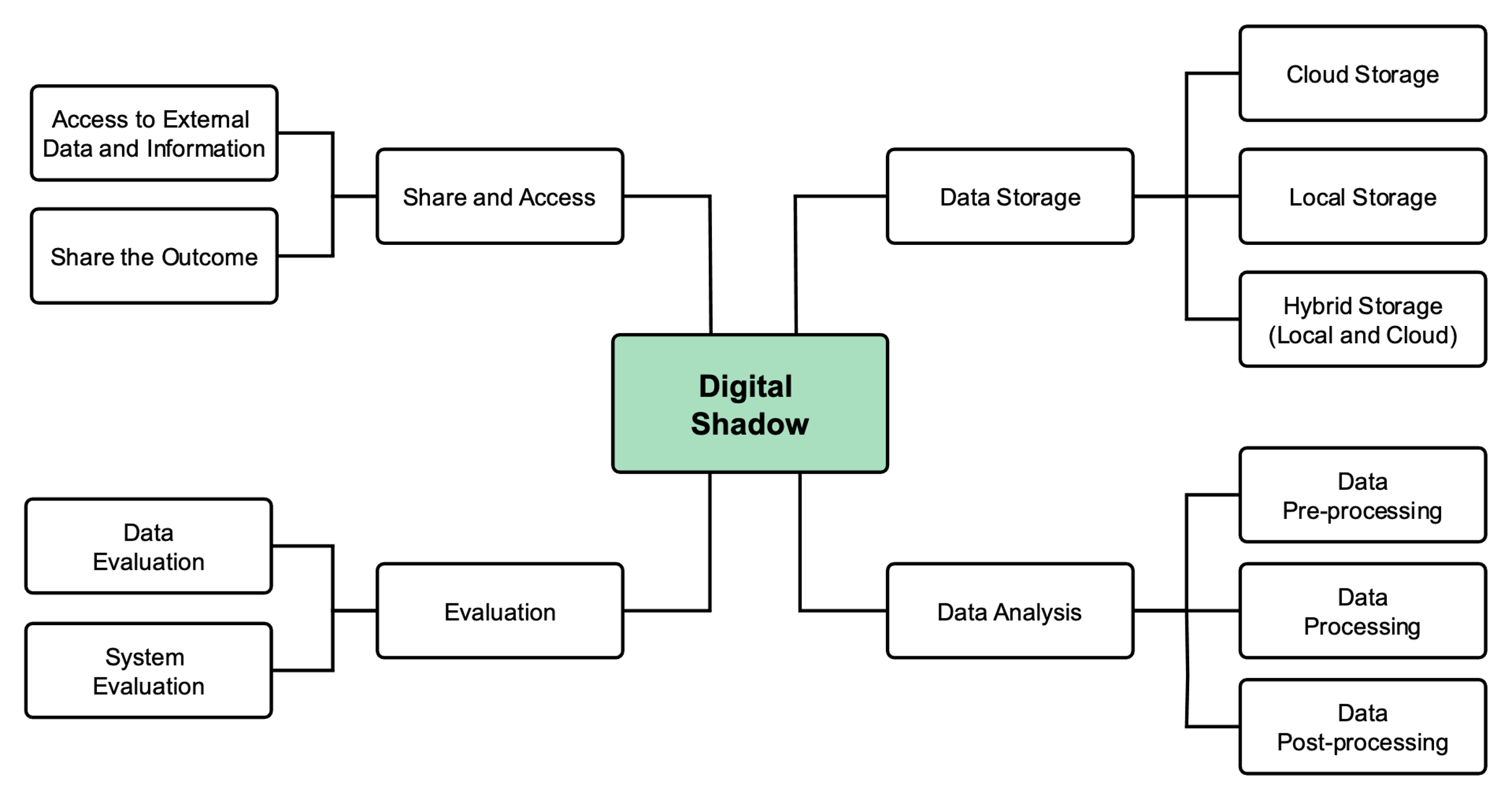
**Data Storage.** The digital shadow serves as the initial repository for the raw data and information gathered from the physical space. The acquired and analyzed data has the option to be stored through two main avenues: Cloud storage which offers a real-time accessibility, and local storage within a predefined memory space. Moreover, a hybrid approach that combines elements of both storage methods holds a particular significance. Cloud storage proves particularly advantageous in scenarios involving cloud computing and real-time data storage. This configuration facilitates online monitoring and active decision-making, as it grants quick and efficient access to the latest data updates. Meanwhile, local storage excels in safeguarding historical data, including both static and dynamic datasets. By retaining historical information within a local repository, the digital shadow enhances optimization efforts and supports more thorough evaluations. This dual approach to data storage—leveraging the strengths of both cloud and local storage—reinforces the comprehensive functionality of the digital shadow across diverse tasks and requirements.

**Data Analysis.** In the subsequent step, the digital shadow undertakes comprehensive data analysis encompassing various stages. These stages include data pre-processing, which involves ensuring data quality assurance, implementing data transformation, and integrating disparate data sources. Subsequently, the step involves data processing, where analytical methods such as statistical analysis, mathematical modeling, and artificial intelligence techniques are employed. Lastly, data post-processing takes place, further enriching the data based on the project’s objectives and requirements.

**Share and Access.** The extracted information from the collected raw data will be (usually) visualized and simulated in the procedure. Collectively, such simulations contribute to enhancing insights from the data and assists the decision-making, and overall operational efficiency within the DT framework. We will then share the decisions and all the related information with other parts of the virtual space. Furthermore, we may need to allow for access to other parts of the system, or to other external sources in order to collect and import extra data and information required for our analyses.

**Evaluation.** The last step involves assessing the outcome based on the specific performance standards and measures established in the master plan. This makes sure that the system’s outcome is reliable and can be later used to enhance both components, concluding the two loops displayed in Fig. 7.

Overall, digital shadows assume a range of responsibilities that revolve around various data management phases. These responsibilities that mentioned above can directly correspond to each of the field data managements discussed before and subsequently, construct an improved framework. They continuously update the master and also the physical entities based on received optimal outcomes till the desired model performance is reached. As mentioned before, nonetheless, data and the established dataflow plays the most crucial role in this framework which we discuss in the next section.

*Figure 9: Different components of the digital shadow investigated here.*

## Communication

We make use of previous models and define a novel interaction between the physical and virtual spaces. Creating this connection between the physical entity and its digital counterpart, involves establishing a bidirectional flow of information that allows real-time data from the physical world to be mirrored and used by the DT. This connection enables monitoring, analysis, simulation, and control that happens in virtual space. We subsequently consider four different types of communication that connects the different components of our framework:

**Physical-to-physical communication:** In the considered distributed systems, multiple physical entities (e.g., devices, field and environment, human, etc.) are interconnected which makes the investigations or operations more challenging. What is crucial in this context is the process occurring here and the communication between these physical entities. We define the relations and services between these entities using a physical-to-physical communication. These connections are mainly supervised/established by the experts beforehand. That is, the communication among these entities should be specified and digitally modeled. As an example, in our use-case here, the objective is to gather data that represents the interactions between a vehicle, the surrounding environment, and humans. Consequently, the data that humans, or more specifically, driver, gather through their visual observations, should be accurately reflected within the digital space.

**Physical-to-virtual communication:** This communication corresponds to the transfer of data from the physical space to the virtual space. Intuitively, the sampled data from the physical space will be first transmitted to the virtual space, using either sensor deployments or other communication methods such as wired connections, wireless protocols (WiFi, Bluetooth, etc.), or cellular networks to send data from sensors to a central processing point (Cloud storage, HDD storage, memory storage, etc). This transfer occurs in real-time and enables the DT to effectively represent the physical counterpart. Sensors can measure parameters like temperature, pressure, vibration, location, and more, depending on the entity’s characteristics and purpose; and the collected field data will be further processed and analyzed in the virtual space (the analysis are parts of the digital shadows discussed above).

**Virtual-to-virtual communication:** The considered DT components are in fact connected with each other to form a network for information and output exchange. As mentioned above, we here propose a digital master in combination with different digital shadows that each is responsible for a certain task. The interaction between these different components will be then defined internally within the DT framework using a virtual-to-virtual communication. Overall, the communications and interactions between Digital Master and Digital Shadows not only facilitate data-driven improvements but also enable more effective monitoring, decision-making, and enhancements in the digital and physical realms. Besides, we here discuss how different DTs can be combined to establish an aggregated system (referred to as aggregated DTs).

**Virtual-to-physical communication:** Finally, the output will be transferred from the virtual space to the physical space, where it proposes specific changes in the parameters of the system to optimize the framework. The output does involve monitoring or decision-making processes that can be implemented directly on the physical entities themselves, utilizing a combination of both direct control mechanisms and feedback loops. In fact, this final step, referred to as virtual-to-physical communication, further completes the end-to-end framework. In this way, the monitoring and decision-making processes involve a dynamic interplay between real-time data collection, digital analysis, decision generation, and physical action. This seamless integration of digital and physical spaces enables the optimization of various processes, enhances operational efficiency, and provides an agile approach to managing real-world systems. The communications above ensure an active data and information flow throughout the proposed DT framework.

## Dataflow in the proposed DT framework

The role of data in DT is discussed in different works, where data management is one of the most relevant topics for the creation of a DT framework. Our investigation here focuses on field data of distributed systems and undoubtedly, data plays a fundamental role in establishing both digital master and shadow. Different applications of a DT framework, however, lead to distinct representations of data in the DT. For instance, in manufacturing, we commonly employ CAD simulations during the digital instantiation and design phase of DT. This enables virtual modeling and testing of physical products, significantly reducing time and costs. On the other hand, such a concept lacks relevance in DT usage for distributed systems in the field, where the focus is on real-time monitoring and management of complex systems. Consequently, the data representation and dataflow differ significantly from manufacturing.

We here collect data from a physical space and transform it to the virtual space. Such field data is commonly subject to a variety of difficulties corresponding to the environmental conditions, and one should carefully explore the different properties of the data, as well as the environmental surroundings in order to establish an effective dataflow throughout the framework. The aforementioned communications ensure such an effective exchange of data and information between the different components of the investigated DT framework. The main remaining point, nonetheless, is how data is collected and later stored and processed in the framework. For this point, we discuss the details given any of the investigated use-cases.

## The LUH use-case

## The DFKI use-case

## The TUD use-case

# Further notes

We emphasize that there are in general different definitions and understandings for a DT and, usually, there is no unified structure for such a concept. DT is a practical and well-known technology which is mostly used for product manufacturing, and here we attempted to collect all the information abouts DTs that may be relevant for distributed systems, especially the field-related systems. However, there is not such a broad literature that we could exploit for applying the technology for field data and distributed systems in the field.

OTHER RELATED POINTS REGARDING THE USE-CASES AND THE COLLECTED FIELD DATA SHOULD BE DISCUSSED!

# Conclusion and Final Remarks

In this document, we investigated the details of DT frameworks and proposed a framework that can be used for distributed systems in the field. We specifically discussed the communication between the different spaces, namely the physical and virtual spaces, and also within the two spaces for an efficient dataflow and exchange of data and information. We further attempted to establish a first version of our framework for the considered use-cases and collect data directly from the field. Besides the details and efficiency of the proposed framework, we here emphasize that such a framework is shown to generalize the previously investigated models to the more intricate cases and enable us to perform a variety of tasks in the field. Details of the data analyses, and the considered use-cases are then discussed, and we refer the reader to the documentation of the other measures in TA GOLO or to other sources of materials (e.g., the published papers) for an elaborate discussion. This section then finalizes the details of this document and the corresponding studies of the measures G-1-1 and G-1-3 of the TA GOLO.

# Refrences

Bärring, M., Johansson, B., & Shao, G. (2020). Digital twin for smart manufacturing: the practitioner’s perspective. *ASME International Mechanical Engineering Congress and Exposition*, (p. V02BT02A015).

Rathore, M. M., Shah, S. A., Shukla, D., Bentafat, E., & Bakiras, S. (2021). The role of ai, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE*, 32030--32052.

Tao, F. a.-Y. (2019). Digital twin-driven product design framework. *International Journal of Production Research*, 3935--3953.

Boyes, H., & Watson, T. (2022). Digital twins: An analysis framework and open issues. *Computers in Industry*, 103763.

Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., & Calinescu, A. (2022). Digital twins: State of the art theory and practice, challenges, and open research questions. *Journal of Industrial Information Integration*, 100383.

Tao, F. a. (2018). Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 2405--2415.

Liu, M., Fang, S., & Dong, H. X. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, 346--361.

Wärmefjord, K., Söderberg, R., Schleich, B., & Wang, H. (2020). Digital twin for variation management: A general framework and identification of industrial challenges related to the implementation. *Applied Sciences*, 3342.

Wang, Z., Gupta, R., Han, K., Wang, H., Ganlath, A., Ammar, N., & Tiwari, P. (2022). Mobility digital twin: Concept, architecture, case study, and future challenges. *IEEE Internet of Things Journal*, 17452--17467.

Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE access*, 167653--167671.

Pantelidakis, M., Mykoniatis, K., Liu, J., & Harris, G. (2022). A digital twin ecosystem for additive manufacturing using a real-time development platform. *The International Journal of Advanced Manufacturing Technology*, 6547--6563.

Adamenko, D., Kunnen, S., & Nagarajah, A. (2020). Digital twin and product lifecycle management: What is the difference? *Product Lifecycle Management Enabling Smart X: 17th IFIP* (pp. 150--162). Switzerland: Springer .

Stark, R. (. (2020). WiGeP-Position Paper : “ Digital Twin ”.