ing.grid preprints

RESEARCH ARTICLE

A Digital Twin Model for Distributed Systems in the Field

Atefeh Gooran Orimi [1], Rayen Hamlaoui [1], Christian Backe [2], Veit Briken [2], Roland Lachmayer [1]

1. Institute of Product Development, Leibniz University of Hannover, Germany.
2. German Research Center for Artificial Intelligence, Robotics Innovation Center Bremen, Germany.

Date Submitted:
2024-03-30

Keywords:
Digital twin model, field data, distributed systems

Data availability:
None.

Software availability:
None.

Corresponding Author:
Atefeh Gooran Orimi
gooran-orimi@ipeg.uni-hannover.de

Abstract. Digital Twins (DTs) have received significant interest in recent years, particularly in industrial and engineering projects. Nonetheless, despite their considerable benefits, existing DTs are primarily domain-dependent and cannot be well generalized. This drawback hinders a straightforward implementation of DTs by researchers and is specifically evident when the physical twin corresponds to field-related entities within distributed systems. In this contribution, we propose a DT model for distributed systems and importantly focus on the communication between and within both physical and virtual spaces. Subsequently, we define the necessary components and elaborate on their specific characteristics and functionalities in the context of digital master and digital shadows. For our investigations here, we further explore a case study implemented in the archetype GOLO of the NFDI4Ing consortia. We provide a comprehensive description of the investigated distributed system and discuss how the proposed DT model can be efficiently applied.

1 Introduction

Digital Twin Models (DTMs) represent state-of-the-art technologies that have gained widespread application in recent years across various domains, including manufacturing, healthcare, transportation, aerospace, and more [1]–[10]. They serve as virtual replicas of physical entities or states, offering significant potential for addressing a multitude of challenges. Additionally, DTs can be used to develop new methodologies for technical products and assist engineers in optimizing the manufacturing processes, especially in the paradigm of technical inheritance [11], [12]. The evolution of DT concept has primarily based on the advancements in Industry 4.0 technologies such as the Internet of Things (IoT), cloud computing, sensor technology, additive manufacturing, artificial intelligence, machine learning, and simulation techniques (see [5], [10], [13]–[17] and references therein for a variety of studies in this direction).

Typically, a DTM comprises three key components: *Physical space* encompassing the physical entities to be modeled and analyzed; *virtual space* containing digital twins that replicate the physical entities and serve as the basis for analysis and investigations; and finally *digital thread* (a.k.a. *communication*) which bridges the physical and virtual spaces. The latter aspect plays a vital role in DT technologies, distinguishing them from simpler simulation techniques. Fig. 1 illustrates these three components of a DTM. One should note the difference between the terms DT and DTM. We refer to the virtual counterpart of the investigated physical entities as a DT,

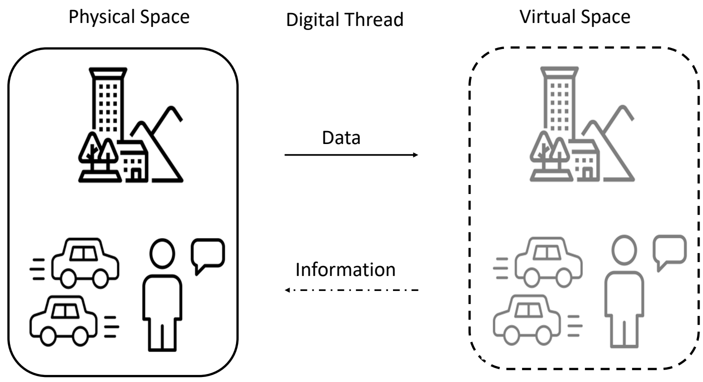


Figure 1: A DT model consisting of the investigated physical entities in the physical space, their DT counterparts in the virtual space, and digital thread as the communication between the two spaces. The image is inspired by the work [18].

and refer to the complete framework consisting the DT and the physical entities together with the dataflow that occurs between and within these two spaces as a DTM (or similarly a DT framework).

In recent years, many researchers have discussed DTs in different fields of study for various applications. Some examples include [2], [15], [17]–[19] and references therein. However, most previous works do not investigate distributed systems in the field and are primarily tailored to manufacturing processes or experiments conducted in controlled lab conditions. Here, we focus on investigating a DTM for field data and distributed systems. The goal is to establish a framework that digitally replicates physical distributed systems used in the field, emphasizing the enhancement of field data reusability. Our distributed systems typically consist of three components: technical systems, humans, and the environment (field). Such scenarios defined by distributed systems find application in various studies like autonomous driving, robotics in agricultural fields or diver monitoring with autonomous underwater vehicles, and traffic infrastructure development. Regardless of the scenario and research objectives, collecting data using technical systems in the field with human interaction should adhere to the FAIR principles (i.e., Findable, Accessible, Interoperable, and Reusable [20]) for maximum benefit.

Field conditions present unique challenges related to real-world environments that cannot be fully predicted and addressed. Consequently, an established DTM designed for field experiments should be capable of addressing these diverse challenges, including unpredicted weather conditions, synchronization of sensors and tools, data protection and security issues, and facilitating internet and electricity in challenging field environments.

One example of previous work on DTMs for field-related scenarios is presented by Wang et al. [21]. The work proposed a Mobility Digital Twin (MDT) and conducted the Personalized Adaptive Cruise Control (P-ACC) system to integrate the three digital building blocks of the MDT: the human DT, the vehicle DT, and the traffic DT. This integration aimed to fulfill MDT digital functionalities of storage, modeling, learning, simulation, and prediction. Another relevant work is by Ivanov et al. [22], which presents the DT concept of a city. Recognizing the urban economy as a complex multi-vector system, the authors proposed DTs of individuals linked in a single cooperative system that allows one DT to use data produced by other DTs to address this challenge. Further, the work by Pauwels et al. [7] considers a DT for the local repository of real-world buildings for robot navigation as autonomously as possible.

In this paper, we propose a DTM based on the concept of digital master and digital shadows for distributed systems employed in field conditions. Specifically, we investigate a use-case from the archetype GOLO of the NFDI4Ing consortia [23]. In this regard, we first discuss distributed systems that we aim to investigate in Sec. 2, and later, describe the specific characteristics of the proposed DT in connection to the investigated physical entities in Sec. 3. The functionality of such DT components will be further specified and detailed. Similar to previous studies, we specifically focus on communication as the exchange of data and information between and within the physical and virtual spaces. We will discuss such communications in more detail in Sec. 4 and show how the proposed DTM can be used in practice. Finally, we conclude the paper in Sec. 5.

2 Distributed Systems in the Field

In our investigations, we refer to a distribution of various physical entities, including humans (e.g., drivers and pedestrians), technical systems (e.g., robots, vehicles, etc.), and the field (the real-world environment) as a *distributed system*. We examine scenarios in which such distributed systems operate in the field, to make data FAIR (following the FAIR principles [20]) and optimize system performance for specific tasks. For instance, we investigate the locomotion systems of robots in challenging agricultural environments by collecting the relevant data [24], [25]. Additionally, we explore vehicles gathering data and information in real traffic conditions within urban areas.

In distributed systems, multiple physical entities are interconnected, making investigations or operations more challenging. Note that it is crucial to mirror the interaction between these physical entities in the virtual space as much as possible. Importantly, the intricate architecture of technical systems often poses a significant barrier to the straightforward implementation of a DT model. For instance, the use-case under investigation here involves a vehicle equipped with GPS, a camera, Light Detection and Ranging (LiDAR), and an Inertial Measurement Unit (IMU). Among other difficulties, synchronizing all these devices in field conditions and collecting the relevant data based on the FAIR principles poses a significant challenge.

Throughout the paper, we will utilize this use-case to provide insights into the details of the proposed model, focusing on the requirements pertinent to our equipped vehicle in urban environments. As a results, certain details and components discussed here may be tailored to the specifics of our investigated use-case. Nevertheless, it is important to note that the proposed DT model can be readily applied, with some modifications if necessary, to other similar technical systems deployed in various real-world environments.

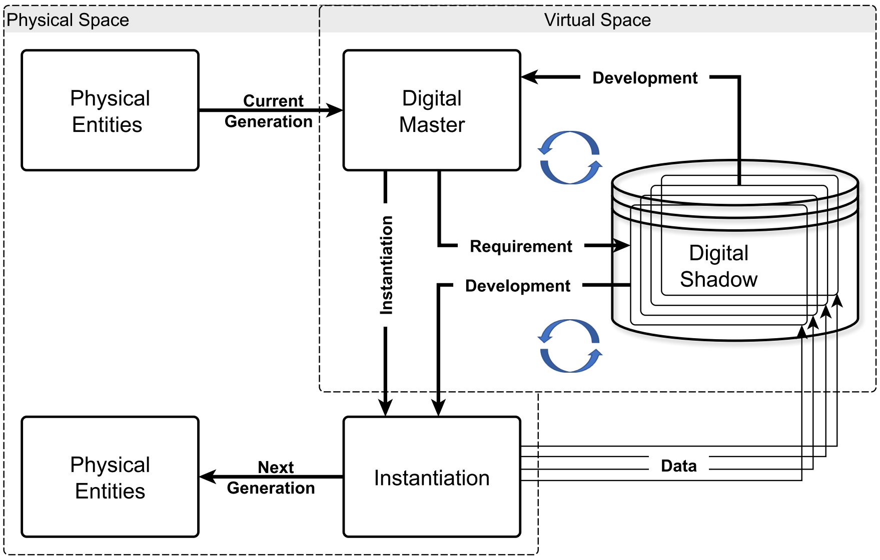


Figure 2: Proposed DT model with the communication between physical and virtual spaces.

3 Proposed Digital Twin Model

Given the discussion above on field-related experiments, we now present the details of the proposed DT model for distributed systems and explore the different characteristics that such a model possesses. 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. Previous work [19], [26] discuss the DT in a virtual space to encompass two main elements, a *Digital Master* and different *Digital Shadows*. We here follow such works and precisely define our DT with a digital master and different digital shadows in the virtual space. The former encapsulates requirements of the digital shadows, tools for instantiation, and other necessary infrastructure information, while the latter corresponds to storage, analysis, prediction, and monitoring the behavior and performance of the investigated physical counterparts by processing the collected data (e.g., collected from different sensors, devices, and other sources). The two elements will be discussed in detail further below.

Fig. 2 illustrates a visual representation of the proposed DT model and the connection between the different components of the physical and virtual spaces. In this model, we specifically explore dataflows between the digital master and digital shadows, as well as between the two spaces and entities. Dataflow serves as a way of communication that enables effective data and information exchange. It is particularly important since implementing an applicable model requires tailored approaches to successfully address all the complexities and challenges of the domain.

Given the architecture of the proposed DT model in Fig. 2, data and information of the corresponding physical entities (the current generation) will be first transferred to the digital master as an element of our DT in the virtual space. The ‘requirement’ for the digital shadows and the details for ‘instantiation’ are the initial information that we receive from the digital master. Next step will be producing the first instance of our use-case (in our investigations here, we establish an actual physical instance, i.e. a vehicle equipped with different sensors and cameras).

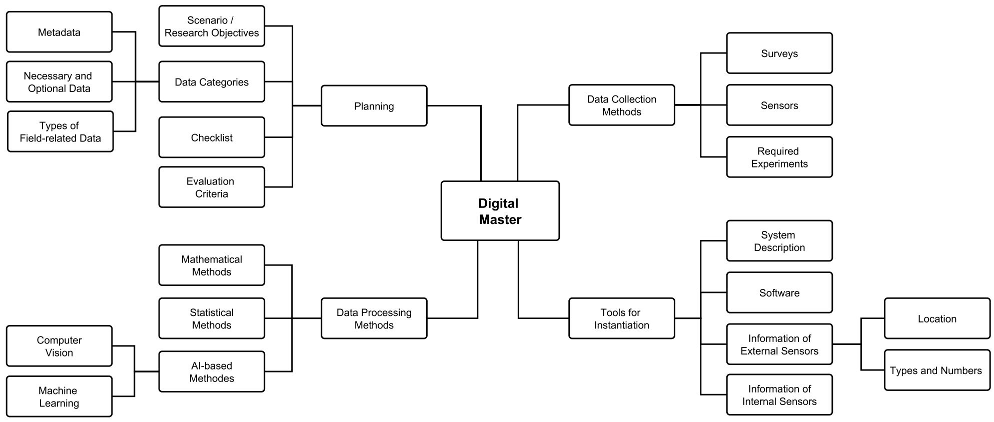


Figure 3: Different components of the digital master investigated here.

The established (physical) instance will be then used to collect data from the field, and such data will be transferred to our digital shadows for further processings. Later, the outcome of our digital shadows will be information for the 'development' of our digital master as well as our instance to improve the performance. In the initial loop, evaluation and outcomes refines the digital master. That is, if the chosen data analysis methods, 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 instance directly, leading to optimization, decision-making, or active monitoring. Finally, the information extracted from the data and details of the optimized physical instance will be transferred to the next generation of our physical entities, which contains the desired modifications for our physical entities.

3.1 Digital Master

The digital master encompasses 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 four key components within the master (as illustrated in Fig. 3), each containing distinct details for a successful implementation of the corresponding (physical) instance as well the shadows. As mentioned, these four components are tailored to the specifics of our investigations here, and one might consider other components to be more relevant for another study. Also, the components discussed in Fig. 3 might be semantically different from each other, but they provide the essential information required for our framework here.

Planning. The initial component is 'planning,' which involves addressing the fundamental requirements for creating our DT model. This may include the objective questions that need answering as well as the checklists outlining our research. Subsequently, we proceed to define specific categories of data, encompassing metadata and both optional and essential data. In fact, data should be categorized and collected according to its priority and relevance to specific contexts. Such priority of data varies from one project to another, depending on the different research objectives.

For instance, temperature may be considered optional data for investigating process timing, while time may be optional for examining temperature.

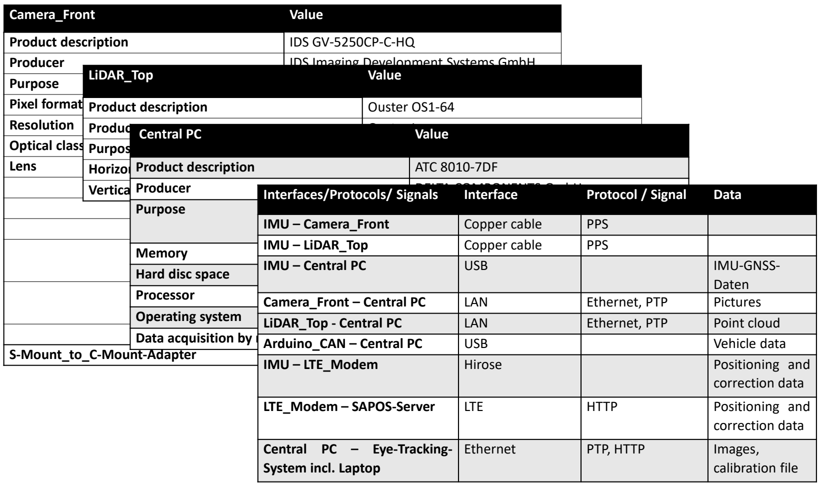
Among the components within the framework, metadata directly aligns with the FAIR principles, establishing a structured and coherent foundation for the entire process. Notably, metadata requires comprehensive exploration and refinement throughout most DT phases (see [27], [28] for relevant studies). Lastly, determining evaluation criteria is another crucial aspect of the master, which can be further utilized in the shadows.

Data Collection Methods. This is the systematic process of gathering important data and information from various sources, with the IoT playing a significant role. IoT is defined as a network of physical objects embedded with technology to communicate, sense, or interact with their environment. Sensors are fundamental to the IoT ecosystem which enable collecting data from humans, the field, and technical systems—the three main components of our distributed system. In addition to leveraging modern sensor technologies, we integrate alternative methods such as surveys and questionnaires to enrich our understanding of the field. These data sources complement sensor data and enhance our analysis. Experimental data is also incorporated when additional tests under real-world conditions are required. In our use-case, we utilize all of these methods for collecting data (i.e., sensors, surveys, and experimental data).

While our primary focus is on collecting necessary data to achieve research goals, it is also important to adopt a broader perspective and capture data that may be valuable for future investigations. This approach enhances data reusability. For example, acquiring environmental temperature data, even if not directly relevant and necessary to current objectives, can prove beneficial in the future. 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 unnecessary data already available is advantageous; and second, other researchers pursuing different goals in the same domain could utilize our temperature data. This illustrates the concept of data reusability, a principle that enhances data FAIR-ness and encourages collaboration and shared knowledge. This concept aligns with the necessary and optional data planning mentioned above.

Tools for Instantiation. In this step, our focus lies on investigating and identifying the essential tools and information in virtual space required for instantiation and production. This enables us to obtain the necessary data and information for the creation of our digital shadow(s), which further leads to performing the desired processing. In simpler terms, the master collects the actual information of our physical entities (distributed systems) and sets the initial requirements for creating our instance (either virtually or physically), which can be categorized into two important parts: the Bill of Material (BoM) and the Bill of Process (BoP). Once that is done, we can start collecting data using the established instance and proceed to the shadow.

Given our direct engagement with the distributed system in the field, examples of BoM in our use-case can be sensors such as LiDAR, cameras, IMU, and eye-tracking systems that equip our technical system and driver in the physical space for data collection, the instantiation process,



Interfaces/Protocols/ Signals	Interface	Protocol / Signal	Data
IMU – Camera_Front	Copper cable	PPS	
IMU – LiDAR_Top	Copper cable	PPS	
IMU – Central PC	USB		IMU-GNSS-Daten
Camera_Front – Central PC	LAN	Ethernet, PTP	Pictures
LiDAR_Top - Central PC	LAN	Ethernet, PTP	Point cloud
Arduino_CAN – Central PC	USB		Vehicle data
IMU – LTE_Modem	Hirose		Positioning and correction data
LTE_Modem – SAPOS-Server	LTE	HTTP	Positioning and correction data
Central PC – Eye-Tracking-System incl. Laptop	Ethernet	PTP, HTTP	Images, calibration file

Figure 4: Example of a BoM designed for the investigated use-case.

and, by extension, the establishment of our digital shadows. Fig. 4 shows an example of the details and list of materials that we need for creating our instance. To complete the BoM, we should also provide BoP (Fig. 5) to determine the details of the process and installation of the provided materials. These two categories cover every detail of the internal and external sensors that we considered for data collection, as well as different software and the process of installation that will be used for creating the instance.

Using simulation techniques and simulation software, real-world products can be completely or partially created, tested, and validated in virtual environments. In our use-case, we utilize CAD models to create a platform for fixing and installing the sensors on top of the vehicle optimally.

Data Processing 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 and the techniques we use for data collection.

Nowadays, a variety of different processing methods are available for researchers that can be utilized in different tasks within the context of DTs (see, e.g., [2], [16], [19], [29] for a discussion on different approaches for data processing in a DT model). This can help researchers get more accurate results, but it can also make it difficult to choose the right analysis method. In the master we sort different methods, with a special focus on basic mathematical and statistical techniques, as well as advanced artificial intelligence and machine learning approaches. These methods are used for tasks like pattern recognition, objects detection, optimization, classification, and in general extracting knowledge from the data collected in the field based on the objectives of the project as well as the properties of the collected data.

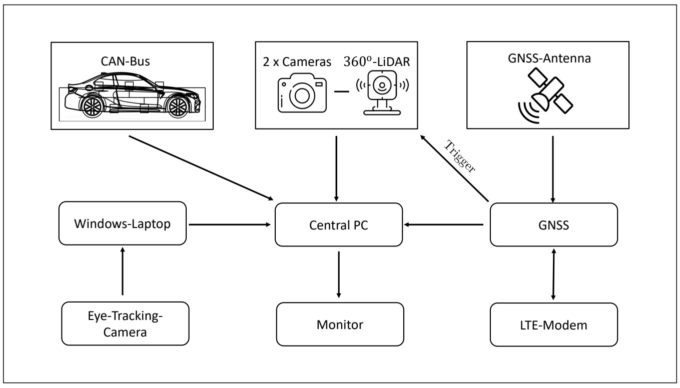


Figure 5: Example of a BoP designed for the investigated use-case.

3.2 The Considered Use-Case Instantiation

Having provided all the necessary details in the master, it is time to proceed to the next step and create our (physical) instance corresponding to our technical system. We do so by using the information of the master, specifically BoM and BoP, for examining each component of our distributed system. In fact, the planning conducted in the master, and specifically the prepared checklists and questionnaires can be used to ensure a well-documented experimental process. Through the use of a checklist, an ordered step-by-step process is designed. For some experiments, like performing roadway light projections, survey is used as a tool for gathering subjective feedback from participants and establishing a qualitative evaluation.

For our investigate use-case, we equip the vehicle with a GPS, camera, LiDAR, and IMU. Placement of these sensors is strategically determined beforehand in the master, with sensors in our use-case placed on the vehicle's roof. It should be noted that modern vehicles often encompass over 100 sensors and Electronic Control Units (ECUs), so our approach does not just focus on the vehicle alone but also on a composite system with advanced sensors. Hence, if necessary, data such as speed, acceleration and steering angle information can be collected from the CAN-Bus signals implemented in the vehicle.

In terms of the environment, we monitor traffic data (e.g., using Google Maps) and utilize GPS for geographical coordinates and route information. Additionally, we record weather data like temperature and humidity to ensure the reusability of the collected data. Finally for humans, in addition to other information like age, gender, and eye health, we use eye-tracking sensors in our use-case to collect further relevant data.

The diversity of sensors in our use-case ensures a wide range of collected data, some of which may be considered optional depending on our research objectives. Nonetheless, we collect all the relevant data here (without extra cost) to increase data reusability, providing options for engineers in other research projects. As an example, for any research aimed at enhancing the accuracy of autonomous parking, weather temperature can be considered optional data, while for research investigating sensor functionality in different weather conditions, temperature may be seemed necessary data.

Still, we again emphasize that data collection from the field faces a variety of different (and usually unpredicted) challenges that should be considered. In our case, using the established instance, we collect data while driving through the city environment, and the collected data will go through the next part, which is the shadow, for storage, processing, sharing, and access.

3.3 Digital Shadow

As mentioned above, the collected field data will be transferred to the digital shadow(s), where the primary focus is on the processing phase. That is, 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 instance, as well as enhancing the digital master based on the objectives of our research. This constructs an iterative procedure that continues until predefined objectives are achieved.

Digital shadows assume responsibilities encompassing data storage, data processing (including data pre-processing, quality assurance, analysis, etc.), and share and access. 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 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 or even repeating the data collection process to improve data quality.

In our investigations here, we consider four essential tasks for each shadow that will be explained in the following. Fig. 6 also illustrates the structure of a digital shadow and provides an 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 real-time accessibility, and local storage within a predefined memory space.

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. On the other hand, 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. Moreover, a hybrid approach that combines elements of both storage methods holds particular significance. Such a 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.

For our use-case, we utilize local data storage. Our sensors are synchronized, and the collected data is timestamped based on the recording time. This local data collection approach allows for more in-depth analysis due to the substantial data volume. Cloud storage and hybrid storage are planned to be implemented in the upcoming stages.

Data Processing. In the subsequent step, the digital shadow undertakes comprehensive data processing encompassing various stages These stages include data pre-processing methods such as cleaning, filtering, and data transformation to ensure the quality and compatibility of the data for the analysis techniques that will be applied later. It may also include tasks such as removing noise or outliers, handling missing data, normalization, and feature extraction.

Once the data is pre-processed, we exploit various analytical methods for data analysis to extract meaningful insights, patterns, or relationships. This stage often involves applying statistical analysis, mathematical modeling, machine learning algorithms, or other analytical techniques to uncover hidden information within the data. Note that the methods employed for data analysis (and in general data processing) depend on the research objectives and vary based on the applied analytical techniques (all these methods are specified in the master beforehand). We here refer to different statistical and mathematical approaches without any specifications of the models in order to convey the variety of the different methods that can be used.

After the analysis is conducted, the results are further processed or refined to meet specific objectives or requirements. This may involve summarizing the findings, visualizing the results through graphs or charts, or interpreting the outcomes in the context of the research or project goals. Post-processing aims to enhance the interpretability and reusability of the analyzed data.

For the use-case that we investigate here, given that our data is collected from different sources, ensuring coherence and consistency in the pre-processing phase is essential. For instance, images obtained in different weather conditions (e.g. those images acquired in night, or rainy days) are assumed to be noisy or blurry, and therefore image enhancement techniques based on state-of-the-art AI methods are required. Additionally, the synchronization of our sensors is crucial for optimal results. Furthermore, by employing graphical representations and dashboards, the results are visualized, making them more usable for other researchers working on field data and distributed systems.

Evaluation. The evaluation step involves assessing the outcome based on the specific performance standards and measures established in the master. This ensures that the system's outcome is reliable and can be later used to enhance both components, concluding the two loops displayed in Fig. 2.

In our use-case, evaluation is carried out in two parts: Data evaluation and system evaluation. For data evaluation, we ensure that our data adheres to quality metrics such as accuracy, completeness, reliability, and timeliness. In this regard, the established Data Quality Metrics website [30] can be used as an exemplary source which provides the necessary metrics for data quality assurance across engineering fields. System evaluation, which informs improvements to our digital master (and in general the investigated technical system), ensures that our system can adapt to increased requirements for specific analyses and experiments. This leads to optimized resource utilization and system sustainability. Based on these evaluations, we can achieve our objectives by creating optimization loops based on pre-determined evaluation criteria.

Share and Access. The data access phase focuses on communication between data providers and data consumers (stakeholders). During this phase, we determine and record which individuals or re-users are accessing specific data and methods. In our use-case, our stakeholders are mostly

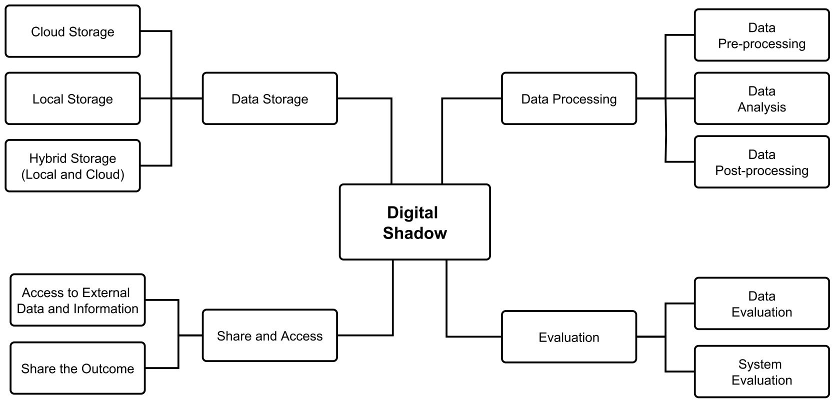


Figure 6: Different components of a digital shadow investigated here.

academia and scientific partners who are working with field data. Also, as consumers of other projects, access to external data to increase interoperability and data enrichment is another issue that should be considered in this part.

Moreover, sharing is defined by the distribution of data and information through web portals and database platforms. The results and the collected data (both raw and processed data) are (usually) publicly available. In this direction, a platform for providing the analysis and visualizing our collected field data will be made available which can be used alongside the established Data Quality Metrics website [30].

4 Dataflow in the Proposed DT Model

The core aspect of the proposed DT model, is the concept of 'digital thread' and the communication between and within physical and virtual spaces which occurs through the exchange of data and information. This component plays a significant role in the model, particularly in establishing it for distributed systems in the field. Following previous work (e.g., [21]), we consider four different types of communication that connect the various components of our model (these communications ensure an active dataflow throughout the proposed DT model):

Physical-to-physical communication. In distributed systems, multiple physical entities (e.g., technical systems or devices, field environments, humans, etc.) are interconnected, which poses challenges for investigations or operations. It is crucial to replicate the communication between these physical entities in the virtual space as accurately as possible. This entails collecting communication data among these entities using various sensors, because sensing and interacting with their internal states or the external environment plays a significant role in interpretation [31].

For instance, in our use-case, we examine the distraction potential with an eye-tracking system, which detects the direction of the subjects' gaze. Hence, data gathered by humans, specifically drivers, through visual observations, will be accurately represented within the virtual space [32].

Physical-to-virtual communication. This communication involves transferring data from the physical space to the virtual space. Sampled data from the physical space is transmitted to the virtual space using 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).

Virtual-to-virtual communication. The DT components are connected with each other to form a network for information and output exchange. We here proposed a digital master in combination with different digital shadows, each responsible for a certain task. The interaction between these different components of the model is defined using virtual-to-virtual communication. Overall, the communications and interactions between the master and shadow(s) not only facilitate data-driven improvements but also enable more effective monitoring, decision-making, and enhancements in both physical and virtual spaces.

Virtual-to-physical communication. Finally, the output is transferred from the virtual space to the physical space, proposing specific changes in the parameters of the system to optimize the model. This output involves monitoring or decision-making processes that can be implemented directly on the physical entities themselves, utilizing a combination of direct control mechanisms and feedback loops. This final step, referred to as virtual-to-physical communication, further completes the end-to-end framework.

5 Conclusion and Future Work

In this paper, we presented a DT model tailored to field and real-world environmental experiments. Our investigation specifically focused on field data and distributed systems within the field. The considered case study, which involved a vehicle equipped with various sensors and cameras in urban environments, further demonstrated the effectiveness of the proposed DT model in real-life applications. The use-case is defined under the archetype GOLO of the NFDI4Ing consortia [23]. It is specially designed to assess the performance of the proposed DT model alongside other aspects of the study in collecting data based on the FAIR principles. Consequently, the first obtained results showed the potential of such models in real-life applications by increasing the FAIR-ness of the collected data as well as improving the investigated technical system when functioning in the field.

Undoubtedly, data plays a fundamental role in the process of establishing the digital master and its shadow(s). Nonetheless, different applications of a DT model lead to distinct representations of data in the DT. For instance, in manufacturing, we commonly employ CAD simulations during a (digital) instantiation and the design phase of a DT. This enables virtual modeling and testing of physical products, significantly reducing time and costs. On the other hand, virtual modeling and instantiation, common in manufacturing applications, might be difficult for DT models in the field due to the complexity of real-world conditions. Consequently, the data representation and dataflow may differ substantially from manufacturing. Besides, field data is usually collected in unpredicted conditions from different sources which also differs from lab or other controlled environments.

Exchange of data in order to share and reuse it needs to save and archive the data with appropriate quality, and appropriate format, which will not be possible except with a well-executed data management strategy. This highlights the role of research data management and its various phases and processes, where previous studies attempt to investigate also under the concept of a DT model [33]–[35]. Intuitively, some of the components discussed within the master and shadows in Figs. 3 and 6 can be directly related to the different phases of a data management system. Components such as planning, data collection, data storage, data processing, and share and access are evidently shared across different data management systems. As a consequence, further investigations of the proposed DT model in conjunction with proper field data management systems is an important direction of study for future work. Other investigations may include the deployment of the current DT model in other field-related experiments and using other case studies such as the ones discussed in [24], [25] (the use of robots in agriculture and water scenes). Leveraging cloud-computing systems and using cloud-based architectures like the Amazon Web Services (AWS) is also planned as future work.

6 Acknowledgements

The authors would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713.

7 Roles and contributions

Atefeh Gooran Orimi: Conceptualization and design, methodology and idea development, literature review, manuscript preparation.

Rayen Hamlaoui: Conceptualization and design, methodology and idea development, critical review.

Christian Backe: Conceptualization and design, critical review.

Veit Briken: Conceptualization and design, critical review.

Roland Lachmayer: Conceptualization and design, project supervision.

References

[1] C. Demminger, I. Mozgova, M. Quirico, *et al.*, “The concept of technical inheritance in operation: Analysis of the information flow in the life cycle of smart products,” *Procedia Technology*, vol. 26, pp. 79–88, 2016.

[2] F. Tao and M. Zhang, “Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing,” *IEEE access*, vol. 5, pp. 20 418–20 427, 2017.

[3] Y. Liu, L. Zhang, Y. Yang, *et al.*, “A novel cloud-based framework for the elderly healthcare services using digital twin,” *IEEE access*, vol. 7, pp. 49 088–49 101, 2019.

[4] M. Bärring, B. Johansson, and G. Shao, “Digital twin for smart manufacturing: The practitioner’s perspective,” in *ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers, vol. 84492, 2020, V02BT02A015.

[5] M. Pantelidakis, K. Mykoniatis, J. Liu, and G. Harris, “A digital twin ecosystem for additive manufacturing using a real-time development platform,” *The International Journal of Advanced Manufacturing Technology*, vol. 120, no. 9, pp. 6547–6563, 2022.

[6] Ö. Aydın and E. Karaarslan, “Openai chatgpt generated literature review: Digital twin in healthcare,” *Aydın, Ö., Karaarslan, E.(2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In Ö. Aydın (Ed.), Emerging Computer Technologies*, vol. 2, 2022.

[7] P. Pauwels, R. de Koning, B. Hendrikx, and E. Torta, “Live semantic data from building digital twins for robot navigation: Overview of data transfer methods,” *Advanced Engineering Informatics*, vol. 56, p. 101 959, 2023.

[8] M. Jafari, A. Kavousi-Fard, T. Chen, and M. Karimi, “A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future,” *IEEE Access*, vol. 11, pp. 17 471–17 484, 2023.

[9] T. Sun, X. He, and Z. Li, “Digital twin in healthcare: Recent updates and challenges,” *Digital Health*, vol. 9, p. 20 552 076 221 149 651, 2023.

[10] S. Selvarajan, H. Manoharan, A. Shankar, A. O. Khadidos, A. O. Khadidos, *et al.*, “Pudt: Plummeting uncertainties in digital twins for aerospace applications using deep learning algorithms,” *Future Generation Computer Systems*, vol. 153, pp. 575–586, 2024.

[11] R. Lachmayer and I. Mozgova, “Technical inheritance as an approach to data-driven product development,” *Design Methodology for Future Products: Data Driven, Agile and Flexible*, pp. 47–64, 2022.

[12] R. Lachmayer, I. Mozgova, W. Reimche, F. Colditz, G. Mroz, and P. Gottwald, “Technical inheritance: A concept to adapt the evolution of nature to product engineering,” *Procedia Technology*, vol. 15, pp. 178–187, 2014.

[13] T. Tiedemann, C. Backe, T. Vögele, and P. Conradi, “An automotive distributed mobile sensor data collection with machine learning based data fusion and analysis on a central backend system,” *Procedia Technology*, vol. 26, pp. 570–579, 2016.

[14] R. Anderl, S. Haag, K. Schützer, and E. Zancul, “Digital twin technology–an approach for industrie 4.0 vertical and horizontal lifecycle integration,” *it-Information Technology*, vol. 60, no. 3, pp. 125–132, 2018.

[15] F. Tao, H. Zhang, A. Liu, and A. Y. Nee, “Digital twin in industry: State-of-the-art,” *IEEE Transactions on industrial informatics*, vol. 15, no. 4, pp. 2405–2415, 2018.

[16] M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, “The role of ai, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities,” *IEEE Access*, vol. 9, pp. 32 030–32 052, 2021.

[17] M. Liu, S. Fang, H. Dong, and C. Xu, “Review of digital twin about concepts, technologies, and industrial applications,” *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, 2021.

[18] M. Grieves and J. Vickers, “Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems,” *Transdisciplinary perspectives on complex systems: New findings and approaches*, pp. 85–113, 2017.

[19] R. Stark, C. Fresemann, and K. Lindow, “Development and operation of digital twins for technical systems and services,” *CIRP Annals*, vol. 68, no. 1, pp. 129–132, 2019.

[20] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[21] Z. Wang, R. Gupta, K. Han, *et al.*, “Mobility digital twin: Concept, architecture, case study, and future challenges,” *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 452–17 467, 2022.

[22] S. Ivanov, K. Nikolskaya, G. Radchenko, L. Sokolinsky, and M. Zymbler, “Digital twin of city: Concept overview,” in *2020 Global Smart Industry Conference (GloSIC)*, IEEE, 2020, pp. 178–186.

[23] *Nfdi4ing – the national research data infrastructure for engineering science*. [Online]. Available: https://nfdi4ing.de.

[24] DFKI GmbH, *DeeperSense - Deep-Learning for Multimodal Sensor Fusion*, en-US, website, 2021. [Online]. Available: https://robotik.dfki-bremen.de/en/research/projects/deepersense/ (visited on 12/10/2023).

[25] DFKI GmbH, *RoBivaL - Robot Soil Interaction Evaluation in Agriculture*, en-US, website, 2021. [Online]. Available: https://robotik.dfki-bremen.de/en/research/projects/robival/ (visited on 12/10/2023).

[26] S. Krückemeier and R. Anderl, “Concept for digital twin based virtual part inspection for additive manufacturing,” *Procedia CIRP*, vol. 107, pp. 458–462, 2022.

[27] C. Backe, A. Gooran Orimi, V. Briken, R. Hamlaoui, and H. Görner, *Creating rich metadata for collaborative research: Case studies and challenges*, NFDI4Ing Conference 2023. Zenodo., 2023. [Online]. Available: https://doi.org/10.5281/zenodo.8430753.

[28] C. Backe, B. Wehbe, M. Bande, N. Shah, D. Cesar, and M. Pribbernow, *Metadata of a large sonar and stereo camera dataset suitable for sonar-to-rgb image translation*, NFDI4Ing Conference 2023. Zenodo., 2023. [Online]. Available: https://doi.org/10.5281/zenodo.10373154.

[29] J.-F. Uhlenkamp, J. B. Hauge, E. Broda, M. Lütjen, M. Freitag, and K.-D. Thoben, “Digital twins: A maturity model for their classification and evaluation,” *IEEE Access*, vol. 10, pp. 69 605–69 635, 2022.

[30] *Data quality metrics*. [Online]. Available: https://quality.nfdi4ing.de/en/latest/.

[31] M. W. Grieves, “Virtually intelligent product systems: Digital and physical twins,” in 24, vol. 11, 2019, pp. 175–200. [Online]. Available: https://doi.org/10.2514/5.9781624105654.0175.0200.

[32] T. Glück, T. Biermann, A. Wolf, *et al.*, “Distraction potential of vehicle-based on-road projection,” *Applied Sciences*, vol. 11, no. 24, p. 12 030, 2021.

[33] J. Ríos, J. C. Hernandez, M. Oliva, and F. Mas, “Product avatar as digital counterpart of a physical individual product: Literature review and implications in an aircraft,” *Transdisciplinary Lifecycle Analysis of Systems*, pp. 657–666, 2015.

[34] G. Schroeder, C. Steinmetz, C. E. Pereira, *et al.*, "Visualising the digital twin using web services and augmented reality," in *2016 IEEE 14th international conference on industrial informatics (INDIN)*, IEEE, 2016, pp. 522–527.

[35] H. Dierend, O. Altun, I. Mozgova, and R. Lachmayer, "Management of research field data within the concept of digital twin," in *International Conference on System-Integrated Intelligence*, Springer, 2022, pp. 205–214.