IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 12 June 2024, accepted 30 June 2024, date of publication 4 July 2024, date of current version 15 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3423372

RESEARCH ARTICLE

A Data-Driven Framework for Digital Twin Creation in Industrial Environments

MARIETHERES DIETZ[1], THOMAS REICHVILSER[2], AND GÜNTHER PERNUL[1], (Member, IEEE)
[1]Department of Information Systems, University of Regensburg, 93053 Regensburg, Germany
[2]Deloitte, 81669 Munich, Germany

Corresponding author: Marietheres Dietz (marietheres.dietz@ur.de)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Krones AG.

ABSTRACT Smart manufacturing uses data generated within manufacturing systems to provide intelligent and flexible control of production processes. To realize the vision of smart manufacturing, digital twins play a key role. As a virtual representation of any real-world counterpart, the digital twin enhances various use cases by providing analytical, simulation and replication capabilities. Most research works focus on potential application scenarios for digital twins within manufacturing. Despite its great potential, few works address the generic creation of digital twins within an industrial environment. To fill this gap, we introduce a data-driven framework for digital twin creation in industrial environments. The core creation process of our framework consists of four vital parts, and the respective data required, to build a digital twin. Before data is even acquired, we argue that individual conditions must be set to determine the overall scope. We further address existing interactions and dependencies between components. To validate the framework, several semi-structured expert interviews are carried out. Furthermore, related works are identified using a systematic literature review - followed by a comparison to our proposed framework.

INDEX TERMS Creation, digital twin, simulation, smart manufacturing.

I. INTRODUCTION

The rapid development of modern information and communication technologies is having a disruptive effect on a wide range of areas. Particularly within manufacturing, this circumstance is driving the transformation from automation towards smart manufacturing systems [1]. To achieve that, Cyber-Physical System (CPS) play a crucial role. The main idea of CPS is to build up smart embedded and networked systems within manufacturing systems. By combining physical (e.g., sensors, actuators) and digital (e.g., software, network) components, this class of systems can communicate, perceive their environment, interpret information and act on the physical world. Thereby, the communication between the physical and virtual world bases on modern communication technologies [2], [3].

To build a CPS, the digital twin can act as a key technology [4]. As a digital representation of a physical system over its lifecycle, the digital twin holds great potential for realizing a CPS-based industrial environment that meets the emerging requirements of smart manufacturing [5]. Most publications focus on identifying potential application areas for industrial digital twins (e.g., [5], [6], [7], [8]). However, less attention is paid to the actual implementation or creation of a digital twin within an industrial environment. So far, only a few works deal with the data-oriented aspects for creating a digital twin in an industrial environment.

To fill this gap, this paper aims to provide a general data-driven framework for digital twin creation in industrial environments. Moreover, the proposed framework covers core technologies for the creation of a digital twin and can be used as a reference model for implementing digital twins in industrial environments. Our framework is further verified by interviews with industrial experts as well as

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

by a literature search to subsequently compare to related works. The innovative part of this research is our data- and -flow-centric focus for building digital twins. While other works may mention data regarding digital twin creation, they mostly do so either for very one specific example (e.g., smart building [9]) or do not holistically regard the whole flow from system (environment) to digital twin but mention one part of data involved (e.g., sensor data mentioned by [10]), while not even distinguishing between the different nature of data involved. Our strength further lies in combining two methods to provide research rigor on the one hand, while also investigating practical applicability.

The rest of the paper is structured as follows: Section II provides a brief explanation of the digital twin paradigm and smart manufacturing with CPS. Our research method follows a combined approach, which is detailed in Section III. Section IV introduces our proposed framework for the creation of a digital twin in an industrial environment and describes its functionalities. Section V compares our framework with related work identified through a systematic literature review. Finally, Section VI draws a conclusion and provides future research directions.

II. BACKGROUND

A. SMART MANUFACTURING AND CYBER-PHYSICAL SYSTEMS IN INDUSTRY

In the last decade, industries have focused on smart manufacturing to realize Industry 4.0. Thereto, sensors and other operational data is gathered to increase the knowledge of systems and optimize production [11]. Cyber-physical systems go hand-in-hand with this trend: They constitute the joining of computational (cyber) and physical parts [12]. To conclude, functionalities within industrial environments can be divided into a physical domain (e.g., machines, actuators) and a cyber domain (e.g., software, network traffic).

B. DIGITAL TWIN

Even though the digital twin paradigm encompasses the virtual twin, its physical counterpart and their data [8], it differs from CPS as the twin is a means of not only ensuring operability of a system but to improve, investigate and predict the future of its counterpart. A digital twin can be described as a virtual representation of any real-world counterpart (e.g., a CPS) over its lifecycle. It exchanges data bi-directionally [6] by collecting information from its counterpart (e.g., sensor values) and sending data (e.g., commands) towards it. The data the digital twin collects about its counterpart is commonly semantically enriched [13] and can then be used in the different operation modes of the digital twin [14]. Fig. 1 illustrates the described paradigm and shows the different operation modes as identified by Dietz and Pernul [14].

Analyses, the first operation mode, use real-world data to optimize the real-world object (e.g., predictive maintenance).

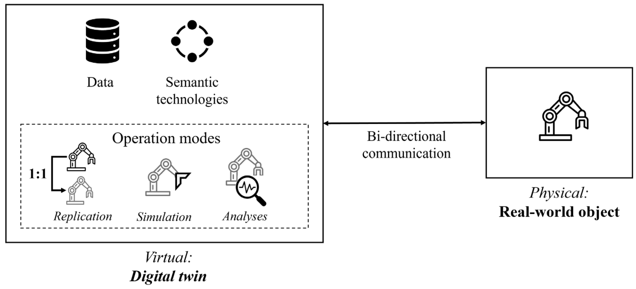

FIGURE 1. Digital twin paradigm including potential operation modes.

A second operation mode is simulation, where the real-world counterpart is virtually modelled and given different scenarios and parameters to change its state. In this mode, relevant problems can be tested without affecting the operation of the real-world counterpart [5], [15]. It is important to note that a simulation alone commonly does not equate a digital twin [5]. A third operation mode is replication. It aims not only at imitating the behavior, but also at providing the exact same state the real-world counterpart is currently in. This allows monitoring and controlling the counterpart with the help of the digital twin.

III. METHODOLOGICAL RESEARCH APPROACH

To address the issue on the data required for digital twin creation, our research approach included a combination of rigorous methods like semi-structured expert interviews and a structured literature search (SLR). The whole research procedure is depicted in Fig. 2. We opted to use a combined approach of those two methods not only as it sets us apart from related works but also because it efficiently allows us to assess our approach in theory as well as for practical use: The SLR strongly supports research rigor as it consists of a very structured and follow-through approach that can be reproduced. On the other hand, semi-structured expert interviews follow a less strict approach while still upholding a certain scope. Moreover, they provide input from practice, which is often neglected in research but nevertheless important, as it ensures applicability and avoids research to stay in the so-called ivory tower. Compared to related works, most research regarding digital twin creation do either regard the practical side by providing a prototypical implementation (e.g., [9], [16], [17]) or use case (e.g., [18]), while some provide a SLR (e.g., [19]) but neglect practical input.

We conducted a structured literature search to find and compare to related works. The quantitative results and information about the search can be found in Section III-A. The final qualitative results including the comparison are described in Section V. Furthermore, expert interviews were carried out to discuss and refine the established framework (see Section III-A).

A. EXPERT INTERVIEWS

To gain feedback on the practicability of our framework, we conducted semi-structured interviews [20] with

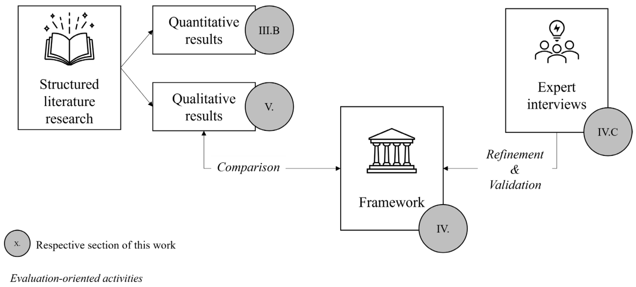

FIGURE 2. Combined research method including structured literature search and expert interviews.

TABLE 1. Focus of the interview according to the interviewee's knowledge (anonymized).

Interviewee	Department	Function	Focus
A	Intelligent Solutions	Head	Framework
B	Info Mgmt Digital Work	Employee	Processing
C	Analytics and Simulation	Head	Creation
D	Security	Head	Framework
E	Info Mgmt Processes and Governance	Head	Framework
F	Info Mgmt Digital Work	Head	Framework, Aquisition
G	Info Mgmt Digital Work	Employee	Application

professionals from the industrial sector. The interviews aim at conforming and refining our proposed approach for creating digital twins with the current requirements from practice.

The interviews are conducted in cooperation with a manufacturing firm (special engineering sector) from Germany. We chose our interview partners according to the main phases of our data-driven framework. We further interviewed executive personnel to gather intelligence on the coherence of these phases. Depending on the individual knowledge of the interviewee, we focused on either a phase of the framework (Aquisition, Processing, Creation or Application) or the entire framework (see Table 1).

Before conducting the semi-structured interviews, we developed a interview guide that was sent to all interviewees. This guide explains the overall aim of the questions as well as a short version of the proposed framework. It further informs about the interview procedure, its questions and gave contact details. The interviews with each expert were carried out via video telephony. The duration varied between 45-60 minutes. At the beginning, our proposed framework was explained to the experts and released for discussion afterwards. The discussion contained the predefined questions of the interview guide. In this way, the interview can also be seen as an expert survey.

Each interview was finally summarized in the form of key points. Afterwards, these were synthesized and structured thematically. The experts' feedback can be found in Section IV-C.

TABLE 2. Literature search protocol.

Databases	IEEE Xplore Digital Library, ACM Digital Library, ScienceDirect, AIS eLibrary
Search method	keyword search
Search criteria	keyword query in `title AND abstract`
Keyword query	`["digital twin*"] AND ["model*" OR "creat*" OR "build*" OR "framework*" OR "architecture*" OR "generat*"]`

B. LITERATURE SEARCH

We carried out a systematic literature review following Okoli and Schabram [21]. The purpose of our literature review is twofold: First, it aims at generating an overview of the current state of research on digital twin creation. The second purpose is the identification of suitable comparative work to evaluate our framework. The search protocol (see Table 2) summarizes the key points of our literature review.

To cover a very broad spectrum on research and gain results spanning from generic approaches to specific use cases, we of course specified "digital twin" in our search term and then queried for various terms covering a creation process to achieve a large amount of research papers to sort through instead of eliminating potential relevant works by choosing a restricted search. To determine the relevance and quality of the search results, an iterative screening process is followed. This screening process includes the following steps:

- 1. Screening: From the total results (1,213 results), existing duplicates are removed. Subsequently, the title of the publications are checked and sorted out according to their relevance (283 results).
- 2. Screening: Remaining publications are further narrowed down based on their respective abstract (84 results).
- 3. Screening: Full texts of each publication are examined in detail, followed by a final assessment of relevance. This led to 16 results: [9], [10], [16], [17], [18], [19], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]

Regarding the results, numerous publications can be identified within the literature for the general topic "digital twin creation". Thereby, it is worth having a closer look at the distribution of the years of the publications. Although no limitation for the time period of the considered publications was set, the search results only include papers starting from 2017. This indicates that the creation of a digital twin covers a modern field of research. Moreover, Table 3 shows that the main publications appear in more technology (IEEE) or practice-oriented (ScienceDirect) data bases.

Analyzing the individual publications, we eliminated results in our screening rounds that (a) proposed a very specialized or tailored digital twin for a certain system or the simulation thereof and (b) the creation of a digital twin multiverse and their networking characteristics, and hence, not a single twin. We further deemed not relevant all (c)

TABLE 3. Literature search results.

Database	Results	Relevant
IEEE Xplore DL	692	7
ACM DL	41	0
Science Direct	464	9
AIS eL	16	0
	1,213	16

machine-learning based approaches for predictive and other analytical tasks (e.g., fault diagnosis, anomaly detection) as well as (d) optimization and evaluation approaches that do not concern the overall creation of a twin.

During our systematic literature review it was shown that while research on "digital twin creation" has been addressed, commonly those results do not take a data-driven perspective and lack an overarching presentation. Our framework, which is detailed in the following (Section IV) meets this need. We further distinguish our research from existing works by following a bottom-up approach that explains the framework and its data (management) by starting from a general viewpoint, diving deeper into each of its parts subsequently.

IV. DATA-DRIVEN DIGITAL TWIN CREATION FRAMEWORK

Our framework is depicted in Fig. 3. It comprises two important parts: (1) The determination of individual conditions (see Section IV-A) is a prerequisite for further digital twin creation. (2) With these conditions determined, a digital twin can be built (see Section IV-B).

A. DETERMINATION OF INDIVIDUAL CONDITIONS

Determining individual conditions enables a targeted selection of required data for the creation of the digital twin. This positively affects the corresponding cost drivers within the acquisition and processing phase. The individual conditions can be broken down into the aspects of (a) the digital model - including simulation level and simulation technology, and (b) the use case. Fig. 4 illustrates these factors in greater detail.

1) DIGITAL MODEL

A digital model forms the basis for a digital twin. To find a suitable technology to build the digital model, it is vital to identify the required representation level. This level depends on the required functionalities to be mapped into the virtual world. As mentioned in Section II, the functionalities within an industrial environment can be divided into a physical domain (e.g., machines, sensors, actuators) and a cyber domain (e.g., software, network traffic). Both domains can again be segmented.

The physical domain comprises all physical parameters (e.g., physical processes, movements) of the regarded industrial environment. These, in turn, can be partitioned into individual physical sublevels: mechanical, electrical, chemical and thermal. Within industrial processes, there commonly exists a combination of physical sublevels. Such a combination is declared as a multi-physical level.

The cyber domain comprises all digital parameters (e.g., communication, programs) of an industrial environment. It can be further divided into information regarding the system's software and network traffic. In terms of representing system software, the circumstantiality needs to be considered: It can be differentiated between exact and logical representation of system software. The exact representation covers an identical copy of the software, including their parameters. The logical one reflects the behavior and the functions of the software. The logical representation can further be functional or identical. In the functional option, the systems software behavior is mimicked in any programming language. Whereas in the identical option, the software from the real system can be copied (same programming language). Once the required levels are selected, matching technologies have to be identified, to create the digital model.

Fig. 4 shows exemplary matching technologies for each representation level. For instance, OpenPLC[1] supports ladder logic and structured control language of real-world PLCs and thus matches the "logical representation - identical" category. In contrast, MiniCPS[2] does not support PLC programming languages. However, a Python-based code mimicking PLC logic can be written (e.g., stop the program when sensor value of machine is too hot). Furthermore, MiniCPS is able to simulate network traffic based on industrial communication standards like Modbus or Ethernet/IP. Therefore, it can be mapped to the "logical representation - functional" category as well as "network traffic". Please note that the tools and software provided in Fig. 4 are examples commonly used in research. Therefore, they have been proven to work for their intended use for building digital twins.

2) USE CASE

Another aspect that influences the creation of a digital twin is the use case concerning the deployment of the digital twin. To date, several papers have examined potential applications for digital twins in the industrial domain [7], [32]. Eckhart and Ekelhart [33] provide an excellent overview of potential digital twin applications. They further assign their identified use cases to the appropriate position in the industrial value chain. Taking into account the identified simulation levels and the specific use case, the appropriate simulation technology for the creation of the digital model can be identified. For instance, regarding the operation phase of the real-world counterpart, the digital twin could be used for system testing [33] or to optimize the manufactured product [8]. To conclude, the application and purpose of the digital twin influences digital model with its required simulation (sub)levels and technology.

In addition, considering the information regarding the individual conditions, a targeted selection of the required data

[1] https://openplcproject.com/
[2] https://github.com/scy-phy/minicps

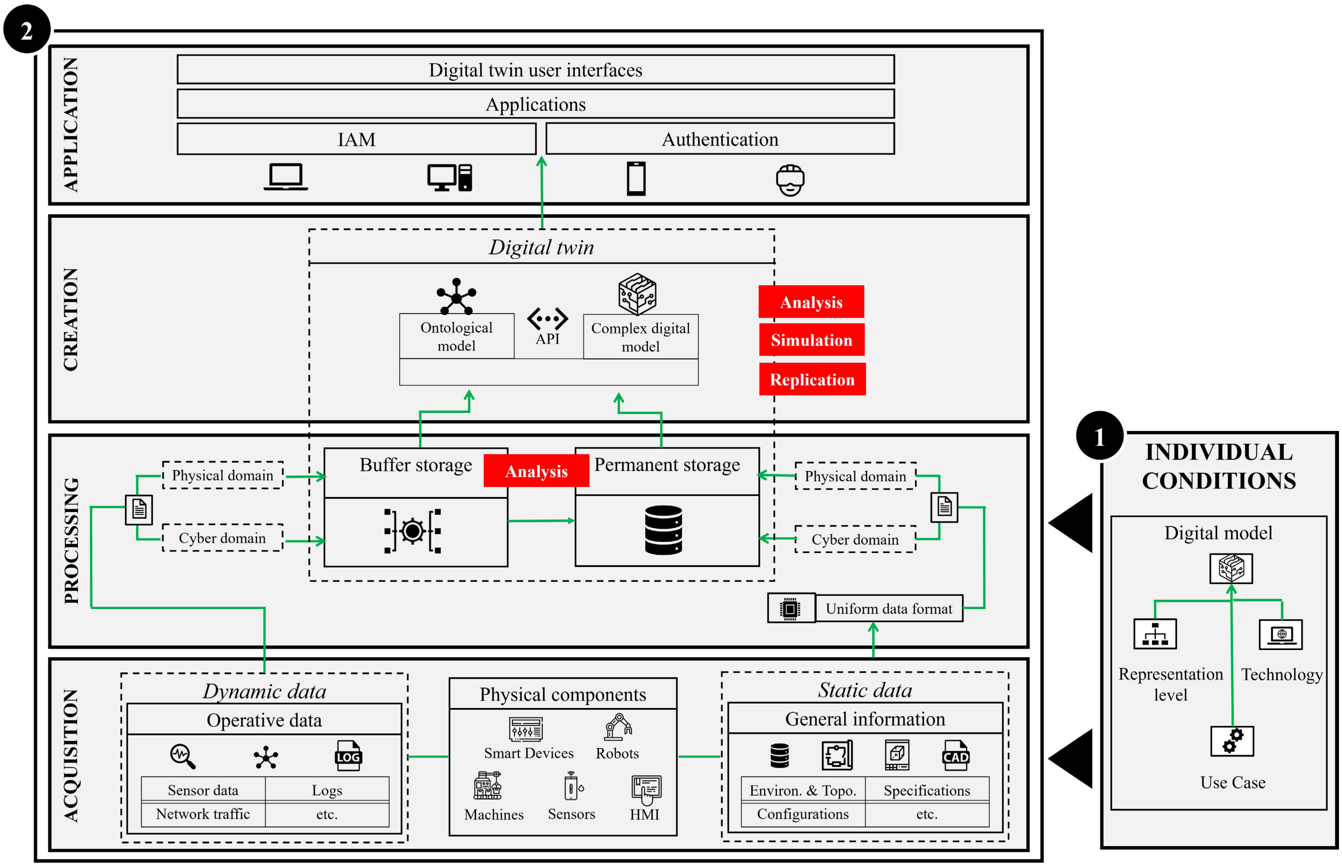

FIGURE 3. **Data-driven framework for digital twin creation in industrial environments.**

for the creation of the digital twin (Section IV-B) can be conducted.

B. CORE CREATION PROCESS

The core creation process (see (2) in Fig. 3) consists of four parts: Acquisition, Preparation, Creation, Application. These are detailed in the following.

1) ACQUISITION

To ensure that the physical components of the real-world counterpart (e.g., robots, machines, sensors) can be mapped correctly into the virtual world, different data - of static and dynamic nature - must be collected. The data required to be collected is dependent on the individual conditions set beforehand.

Static data commonly comprises general information about the respective physical components (e.g., topology, specifications, configurations). In contrast to dynamic data, static data changes far less over time, approximately every 1-2 years. Therefore, it only needs to be re-collected in the event of fundamental changes with regard to the real-world counterpart (e.g., replacement of a component) and its environment (e.g., adaptation of the topology). Existing database systems can be used as reference point for this data.

Dynamic data is continuously generated during the operation of the real-world counterpart (e.g., sensor data, network traffic). To incorporate valid data tuples, the tuples must be checked. This may cause short delays leading to a close-to real-time performance.

2) PROCESSING

The acquired data is subsequently processed and stored. Depending on the nature of data (static or dynamic), there are different procedures concerning data processing.

After their acquisition, static data must be brought into a uniform data format (e.g., AutomationML[3]). Thereafter the data can be broken down into either physical- (e.g., blueprint of the machine) or cyber-oriented (e.g., program code of a machine) data. Subsequently, static data are stored in a permanent storage.

Similarly, dynamic data are classified into a cyber (e.g., network traffic) or physical (e.g., sensor values) domain to facilitate subsequent processing. The dynamic data is then transferred to a buffer storage system. Once the data reaches the buffer storage, it is further analyzed and transformed. For instance, anomalies or patterns in the data can be identified. These are then transferred to the permanent storage as

[3]https://www.automationml.org/

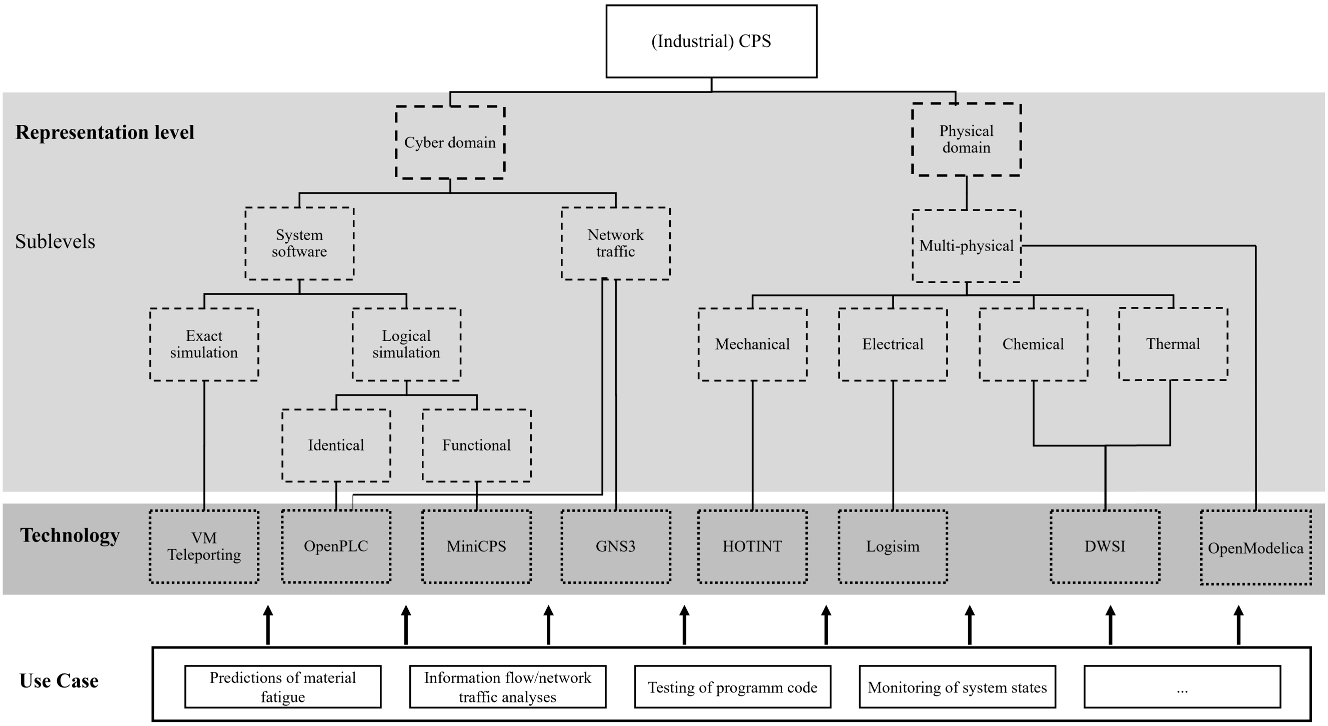

FIGURE 4. **Factors to determine individual conditions and build digital models.**

historical data records. In the permanent storage, data can be further analyzed to generate additional value (e.g., derivation of rules to identify an anomaly).

Based on information contained in respective storage systems, individual actions of the analysis process (see Section II) can already be implemented. The processing in general is also dependent on the determined individual conditions.

3) CREATION

Using previously processed data, the digital twin is created. The digital twin can be seen as a combination of several components - including an ontological model, digital models, API, buffer storage and permanent storage. The creation itself follows the subsequent logic:

1) Creation of digital models: A CPS is a combination of individual subsystems or components, which can be assigned to the cyber or physical domain. To obtain a virtual representation of a CPS, a digital model for each component must be created. Subsequently, these single models can be merged into a complex digital model. This can be realized by using Functional Mock-Up Interfaces (FMI) or Functional Mock-Up Units (FMU) [34] as illustrated in Fig. 5.
2) Mapping the behavior of the real-world counterpart: To map the behavior of the real-world counterpart, the assembled digital models should not act separately, but as a whole. An ontological model (c.f. Zheng and Sivabalan [31]) manages this feat. The model maps corresponding relations to static and dynamic data from the respective storage. This leads to drawing appropriate conclusions from the combination of static and dynamic data, which in turn enables the simulation of the current state and behavior of the real-world counterpart.
3) Integrating behavior into the complex digital model: In a final step, the information from the ontological model needs to be integrated into the complex digital model. To enable this integration, an API serves as a kind of command center between the ontological and the complex digital model. The API references individual standard libraries of the respective model. Thus, it can interact with both models, and control the data flow between them. The API further controls the retrieval and transfer of static and dynamic data from the two storage systems to the ontological and the complex digital model.

The created digital twin can be used in different operation modes such as analysis, simulation and replication (see Section II). The key points of the creation phase can be summarized as follows:

- The digital twin is a combination of a complex digital model, an ontological model, a buffer storage and a permanent storage
- By using appropriate standard interfaces (e.g., FMI, FMU), single digital models can be linked to form a complex digital model
- The permanent storage (static data) serves as main source for the creation of the complex digital model, while the buffer storage (dynamic data) commonly

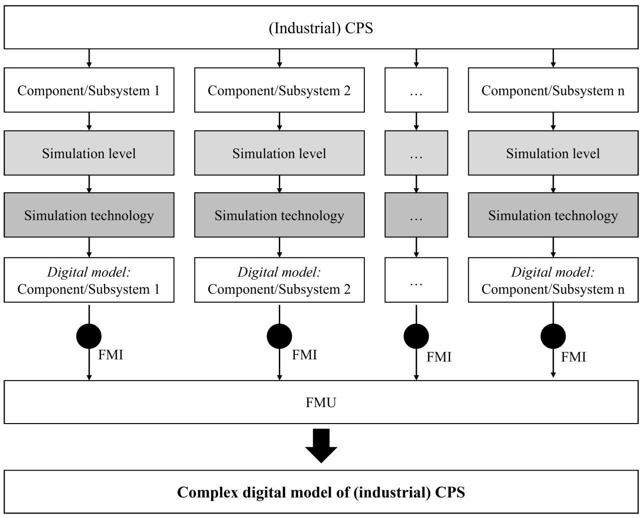

FIGURE 5. **Creation of a complex digital model of (industrial) CPS using FMI/FMU.**

enables the representation the current state of the real-world system. This is turn, manages the proper function of the operation modes relying on current data (analysis and replication)

- The ontological model maps the behavior of the real-world counterpart to the complex digital model
- The API controls and coordinates the complex digital model and ontological model. It further regulates the flow of information between the individual models and the storage systems

4) APPLICATION

Various applications and user interfaces can be provided on top of the previous parts to enable optimal use of the digital twin. The design of the applications and user interfaces depends on the specific use case. Often, individualized dashboards present valuable use case-dependent information in a compressed form. Furthermore, visual instructions for complex processes can be created by using appropriate technologies for visualization (e.g., augmented reality). An integrated Identity and Access Management (IAM) and an additionally required authentication mechanism ensure that only authorized users have access to the respective services of the digital twin. Please note that the application layer differentiates from the operation modes (analysis, simulation and replication) in the following manner: The operation modes are functional requirements of a digital twin, which form the basis for other applications. For instance, the analysis mode will run in a user dashboard that visually provides insight to the analyzed data. However, applications can be set to meet individual requirements, which can affect the operation modes settings. For instance, using AR to simulate a system requires user-defined parameters for the simulation, resulting in an information flow from the application layer towards the operation mode simulation.

C. FEEDBACK FROM EXPERTS

To refine our framework, expert interviews were carried out. The results of these interviews are summarized in thematically in the following.

1) GENERAL

All participants acknowledged the meaningfulness of the proposed layers. The actions within the phases were deemed similar to best-practices in other firms according to some of the interviewed professionals. The experts stated that the main costs will emerge during data acquisition.

2) FRAMEWORK STRUCTURE

All interviewees commented that individual conditions must be met before initiating the frameworks' main phases. They stated that these conditions affect the acquisition and processing phase, resulting in prerequisites for these layers.

3) ACQUISITION

Most interviewees perceived the realization of our acquisition as feasible as well as dependent on the underlying use case. Feedback was also given on static data, where an expert added examples of general information like circuit layouts, machine and process descriptions. To mimic the real-world counterparts, some interviewees mentioned the importance of dynamic log and sensor data. Few experts stated that from data acquisition to buffering and processing, latency may occur, which could hamper real-time processing.

4) PROCESSING

The proposed logic and potential technologies for implementation of the processing layer were considered reasonable and practicable. Some interviewees suggested that the buffer storage's intelligence is crucial: It has to select useful information and discard non-relevant data. Also, first analytical tasks can be performed in the storage systems.

5) CREATION

Few professionals stated that combining single digital models towards one complex digital model is already common practice in some firms and can be achieved by using FMI. Our proposition to integrate the ontological model in the form of a graph-based database was acknowledged by the experts. However, some interviewees mentioned that the technological implementation of this phase might be time-consuming and costly.

6) APPLICATION

Some professionals highlighted that the applications should be orientated towards the individual requirements of the end user. Therefore, the framework should not state individual applications but sum them up in a general term (Applications). To secure potentially sensitive data, authentication and IAM should be installed.

The results of the expert interviews led to the following modifications of our framework:

- Framework: Restructuring of individual conditions
- Acquisition: Additions to static data source examples, modification of term to general information
- Processing: Marking buffer and permanent storage with analytical capabilities
- Application: Integration of various exemplary applications into general term applications, addition of IAM as well as authentication

V. RELATED WORK

In this Section, the presented framework (Fig. 3) is compared to related work from scientific literature. The methodology used to select suitable publications for comparison is detailed in Section III-B.

A. COMPARISON CRITERIA

The related publications identified in Section III-B are compared to our proposed framework. To consistently compare the research works, we used the criteria shown in Table 4.

We divided between general criteria and conceptual factors. General criteria consider comprehensive research aspects. Conceptual criteria focus on the main artifact of individual papers. In concrete terms, they are based on the respective framework - with special focus on provided layers (see Section IV). We further included the criteria "creation" as it is the main focus of this work, and without it digital twins would not exist. Table 5 summarizes the result of the conducted comparison.

B. GENERAL FACTORS

Almost all of the related publications can be assigned to an industrial application area: Most works take an industrial focus on digital twins [10], [17], [24], [26], [28], where some further see the digital twin as a key technology for integrating the smart manufacturing approach [22], [27], [29], [30], [31]. Closely related to this area, are papers that either focus on the integration of a digital twin within an CPS [23], or regard digital twins within different IoT use cases [19]. Also, one work specifically regards the nuclear sector [16], while another shows how to use them for smart buildings [9]. Finally, two publications use digital twins in a general organizational context [18], [25]. Our work aligns with the majority by centering around industrial environments.

While some publications focus on the required functionalities of a digital twin (function-oriented) [18], [19], [22], [24], [25], [28], [30], others go into more detail and additionally provide generic proposals for the implementation (solution-oriented) of the mentioned functionalities [10], [27], [29]. The rest of the works rather present a technological implementation (technology-oriented) [9], [16], [17], [23], [26], [31]. In contrast, our work does not merely present some software-restricted solution but takes a bottom-up approach by first concentrating on a conceptual framework with a strong emphasis on the data involved as well as considering different use cases and scenarios a twin might be deployed. It therefore provides more flexibility while still providing enough detail (e.g., we provide an overview of potential technologies for implementation in Fig. 4).

TABLE 4. Comparison criteria.

General factors	Conceptual factors
Application area • Sector • Specific application area	Conditions
Classification (function-, solution- or technology-oriented)	Acquisition: • Information source • Breakdown (static/dynamic) Processing: • Storage technologies • Real-time requirements • Breakdown (cyber/physical) Creation: • Ontology • Combination • Operation modes Application: • Application examples • Security features

C. CONCEPTUAL FACTORS

Except for AboElhassan et al. [10], individual conditions are not considered in related works.

However, all authors [9], [10], [16], [18], [19], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31] (except one) consider and describe the acquisition of data to be the necessary first step for the creation of a digital twin. Nevertheless, the proposed realization of this phase varies greatly: While almost all refer to their data source at one point, only one other work regards the different nature of data velocity (static/dynamic) [9]. Others might mention certain dynamic data (e.g., sensor data) but fail to differentiate them from static data in the acquisition phase. We take into account that data from industrial environments are of different nature and can be mostly distinguished by its velocity. Generally, data generated during systems operations can be considered dynamic, while general information about the real-world counterpart is rather static. We think it is important to differentiate as it greatly impacts the subsequent processing of the data, especially considering storage options.

Processing acquired data is seen as a vital measure for digital twin creation by most other works. However, two works [25], [26] completely neglect the processing phase. Most authors provide information on storage technologies or regard the need for data storage. Few consider real-time

TABLE 5. Comparison of related woks concerning digital twin creation.

Comparison criteria	[10]	[16]	[9]	[22]	[17]	[23]	[19]	[18]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]
Application area	I	N	SB	SM	I	ICS	IoT	G	I	G	I	SM	I	SM	SM	SM
Classification	SO	TO	TO	FO	TO	TO	FO	FO	FO	FO	TO	SO	FO	SO	FO	TO
Conditions	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
AQUISITION	◔	◑	●	◔	○	◔	◑	◔	◑	◑	◑	◑	◑	◑	◔	◑
Information source	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓
BD (static/dynamic)	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
PROCESSING	◕	◑	◕	◔	◔	◑	◔	◔	◔	○	○	◔	◕	◔	◔	◑
Storage technologies	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	✓	✓	✗	✗	✓
Real-time requirements	✓	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
BD (cyber/physical)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
CREATION	◑	◔	◕	◔	◔	◑	◑	◔	◑	◔	◔	◑	◔	◑	◔	◑
Ontology	✓	✗	✓	✓	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓
Combination	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	✓	✓
Operation modes	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗
APPLICATION	◑	◑	◑	◑	○	◕	◑	●	○	○	◑	◕	◑	◑	○	◕
Application examples	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗	✓
Security features	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

BD: breakdown, G: general, I:industry, SB: smart building, SM: smart manufacturing, N: nuclear; FO: function-oriented, SO: solution-oriented, TO: technology-oriented
○ fully neglected, ◔ partly recognized (criteria neglected), ◑ partly recognized (criteria partly considered), ◕ partly recognized (all but few criteria considered), ● fully recognized (all criteria considered).

requirements [9], [10], [23], [28]. For instance, the MQTT publish/subscribe protocol [23], RabbitMQ [28] and data stream processing [10] present two technological solutions for implementing the real-time requirements of a digital twin. Like most related works, we suggest processing the data – including data storage, where we differentiate between buffer storage for data of dynamic nature and a permanent storage for static data and processed dynamic data. Thereby, we meet requirements for handling real-time data. In contrast to all other works, we further added a breakdown of data between cyber and physical: As most industrial systems are comprised of CPS, they have cyber parts and the respective data thereof (e.g., dynamic: network data, static: program codes). The same goes for its physical parts (e.g., dynamic: temperature of machine, static: specifications of systems). We think it of great importance to be aware of the domain the data belongs to while processing it.

Concerning creation, most (but not all) authors see the digital twin as a combination of several models or components [9], [10], [16], [17], [18], [19], [23], [24], [27], [29], [30], [31]. Those who do not describe the twin as such, mainly focus on the development of one specific (twin) model [22], [26], [28] that is not comprised of other models. Interestingly, about half of the related works [9], [10], [19], [22], [23], [25], [30], [31] recognize the need for adding context to data – most suggest using an ontology to meet this end [9], [10], [19], [23]. Two publications explicitly mention that digital twins can run analytical or optimization tasks and simulations [10], [24], [27]. Going one step further, we argue that the digital twin can even manage three operation modes (analysis, simulation, replication).

Part of the authors focus on specific applications resp. use cases within the industry [9], [10], [16], [22], [23], [26], [27], [28], [29], [31]. For instance, virtual commissioning of a semiconductor production use case is tackled by [10]. Others present more general or hypothetical industrial use cases [18], [19]. Four works do not provide information about the (potential) application of the created digital twin [17], [24], [25], [30]. In contrast to our work, none of the related publications integrates corresponding security features within this phase.[4]

In summary, related works confirm the comprehensive parts of our framework (see Section IV). Especially two publications ([28], [31]) follow a similar outline: Yang et al. divide their framework into five layers (perception, network, data, twin and application) [28], while Zheng and Sivabalan categorize their framework into four layers (physical, data extraction and consolidation, cyberspace and interaction) [31]. Thereby, perception and network layer [28] match with the physical layer [31], which in turn matches with our acquisition layer. To our processing layer align the data layer [28], and the data extraction and consolidation layer [31], respectively. The core part of our framework, the creation layer harmonizes with the twin [28] resp. cyberspace layer [31]. Finally, the application layer [28] and interaction layer [31] are in accordance with our own application layer.

However, all related works vary in the details. Specifically, the following points can be noted:

- Acquisition: Our explicit division of data into static and dynamic data is a feature commonly neglected by related works.
- Processing: The idea of combining different storage technologies is confirmed by related publications (e.g., [23], [31]). Our final breakdown of the processed data into a cyber and physical domain can be regarded as unique feature strengthening the CPS notion commonly found in industrial environments.
- Creation: The consideration of the digital twin as an amalgamation of several components is confirmed.

[4]Please note that works, where the implementation code is verified using security tools (e.g., [17]) do not fall into this category.

We are the first to consider three different operating modes of a digital twin (e.g., analyses, simulation, replication) in digital twin creation.

- Application: In general, all authors integrate corresponding application components into the presented models. We distinguish our framework from others by including security features.

VI. CONCLUSION

Although the digital twin technology promises great potential for realizing the smart manufacturing approach, few works deal with the creation of a digital twin. To bridge this gap, our framework provides a general approach for the digital twin creation in an industrial environment.

According to industrial experts (see Section IV-C), the creation of a digital twin has to be designed dependent on the individual conditions. Based on those, appropriate technologies can be identified. The digital twin itself consists of a meaningful combination of several models, and thus, technologies. A permanent storage and a buffer storage provide the basis for data processing. The complex digital model lies at the core of our framework and mirrors the real-world object into the virtual world. It is realized by merging digital models of individual CPS components. An ontological model prepares the database semantically and communicates with the complex digital model via an API. This collaboration enables mapping the context of the real-world object into the virtual world. Through a systematic literature review we identified related publications. By comparing with related work, the basic segmentation of our framework can be confirmed. However, we provide several useful additions that were neglected in previous works (see Section V-C).

Future research should be guided by current limitations. For instance, individual conditions could be worked through methodologically. Especially simulation levels and the subsequent assignment of suitable simulation technologies can be gathered by rigorously conducting separate systematic literature and technology reviews. Currently, our simulation technologies were identified from current publications and matched to respective levels. So, we cannot claim completeness in regard to these technologies. However, we found that those were the most academically used tools in regards to digital twins and simulations. Furthermore, the practical implementation of the proposed components can be considered in more detail. In our opinion, an emphasis should be put to the implementation of the complex digital model. For instance, a first approach is suggested by using FMI [34], which could be further explored in terms of limitations for usage and by comparison to potential other approaches for generating complex digital models.

ACKNOWLEDGMENT

The author would like to thank the Krones Group[5] and their industrial experts for the helpful statements during their interviews, and the subsequent valuable insights to the digital twin technology used in their firm as well as others.

[5]https://www.krones.com

REFERENCES

[1] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. D. Noh, "Smart manufacturing: Past research, present findings, and future directions," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 3, no. 1, pp. 111–128, Jan. 2016.

[2] Y. Qamsane, C.-Y. Chen, E. C. Balta, B.-C. Kao, S. Mohan, J. Moyne, D. Tilbury, and K. Barton, "A unified digital twin framework for real-time monitoring and evaluation of smart manufacturing systems," in *Proc. IEEE 15th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2019, pp. 1394–1401.

[3] S. Jeschke, C. Brecher, T. Meisen, D. Özdemir, and T. Eschert, *Industrial Internet of Things: Cybermanufacturing Systems*. Cham, Switzerland: Springer, 2017, pp. 3–19.

[4] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 3, pp. 18–23, Jan. 2015.

[5] E. Negri, L. Fumagalli, and M. Macchi, "A review of the roles of digital twin in CPS-based production systems," *Proc. Manuf.*, vol. 11, pp. 939–948, Jan. 2017.

[6] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018.

[7] C. Cimino, E. Negri, and L. Fumagalli, "Review of digital twin applications in manufacturing," *Comput. Ind.*, vol. 113, Dec. 2019, Art. no. 103130.

[8] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[9] D. D. Eneyew, M. A. M. Capretz, and G. T. Bitsuamlak, "Toward smart-building digital twins: BIM and IoT data integration," *IEEE Access*, vol. 10, pp. 130487–130506, 2022.

[10] A. AboElHassan, A. H. Sakr, and S. Yacout, "General purpose digital twin framework using digital shadow and distributed system concepts," *Comput. Ind. Eng.*, vol. 183, Sep. 2023, Art. no. 109534.

[11] H. Lasi, P. Fettke, H. G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Bus. Inf. Syst. Eng.*, vol. 6, no. 4, pp. 239–242, 2014.

[12] F.-Y. Wang, "The emergence of intelligent enterprises: From CPS to CPSS," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 85–88, Jul. 2010.

[13] M. Dietz and G. Pernul, "Digital twin: Empowering enterprises towards a system-of-systems approach," *Bus. Inf. Syst. Eng.*, vol. 62, no. 2, pp. 179–184, Apr. 2020.

[14] M. Dietz and G. Pernul, "Unleashing the digital twin's potential for ICS security," *IEEE Secur. Privacy*, vol. 18, no. 4, pp. 20–27, Jul. 2020.

[15] M. Grieves and J. Vickers, *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. Cham, Switzerland: Springer, 2017, pp. 85–113.

[16] D. Bowman, L. Dwyer, A. Levers, E. A. Patterson, S. Purdie, and K. Vikhorev, "A unified approach to digital twin architecture—Proof-of-concept activity in the nuclear sector," *IEEE Access*, vol. 10, pp. 44691–44709, 2022.

[17] H. R. Hasan, K. Salah, R. Jayaraman, M. Omar, I. Yaqoob, S. Pesic, T. Taylor, and D. Boscovic, "A blockchain-based approach for the creation of digital twins," *IEEE Access*, vol. 8, pp. 34113–34126, 2020.

[18] S. R. Newrzella, D. W. Franklin, and S. Haider, "Three-dimension digital twin reference architecture model for functionality, dependability, and life cycle development across industries," *IEEE Access*, vol. 10, pp. 95390–95410, 2022.

[19] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models," *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, Oct. 2020.

[20] J. Lazar, J. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, 2nd ed., San Mateo, CA, USA: Morgan Kaufmann, 2017.

[21] C. Okoli, "A guide to conducting a standalone systematic literature review," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 43, pp. 879–910, Nov. 2015, doi: 10.2139/ssrn.1954824. [Online]. Available: http://aisel.aisnet.org/cais/vol37/iss1/43

[22] J. Friederich, D. P. Francis, S. Lazarova-Molnar, and N. Mohamed, "A framework for data-driven digital twins of smart manufacturing systems," *Comput. Ind.*, vol. 136, Apr. 2022, Art. no. 103586.

[23] C. Liu, P. Jiang, and W. Jiang, "Web-based digital twin modeling and remote control of cyber-physical production systems," *Robot. Comput.-Integr. Manuf.*, vol. 64, Aug. 2020, Art. no. 101956.
[24] C. Nwogu, G. Lugaresi, A. Anagnostou, A. Matta, and S. J. E. Taylor, "Towards a requirement-driven digital twin architecture," *Proc. CIRP*, vol. 107, pp. 758–763, Jan. 2022.
[25] R. Parmar, A. Leiponen, and L. D. W. Thomas, "Building an organizational digital twin," *Bus. Horizons*, vol. 63, no. 6, pp. 725–736, Nov. 2020.
[26] E. Russo, G. Costa, G. Longo, A. Armando, and A. Merlo, "LiDiTE: A full-fledged and featherweight digital twin framework," *IEEE Trans. Depend. Sec. Comput.*, vol. 20, no. 6, pp. 4899–4912, Dec. 2023.
[27] P. Wang and M. Luo, "A digital twin-based big data virtual and real fusion learning reference framework supported by industrial internet towards smart manufacturing," *J. Manuf. Syst.*, vol. 58, pp. 16–32, Jan. 2021.
[28] M. Javaid, A. Haleem, and R. Suman, "Digital twin applications toward industry 4.0: A review," *Cogn. Robot.*, vol. 3, pp. 71–92, Jan. 2023.
[29] C. Zhang, G. Zhou, J. He, Z. Li, and W. Cheng, "A data- and knowledge-driven framework for digital twin manufacturing cell," *Proc. CIRP*, vol. 83, pp. 345–350, Jan. 2019.
[30] L. Zhang, L. Zhou, and B. K. P. Horn, "Building a right digital twin with model engineering," *J. Manuf. Syst.*, vol. 59, pp. 151–164, Apr. 2021.
[31] P. Zheng and A. S. Sivabalan, "A generic tri-model-based approach for product-level digital twin development in a smart manufacturing environment," *Robot. Comput.-Integr. Manuf.*, vol. 64, Aug. 2020, Art. no. 101958.
[32] M. R. Enders and N. Hoßbach, "Dimensions of digital twin applications—A literature review," in *Proc. 25th Americas Conf. Inf. Syst. (AMCIS)*, Cancun, Mexico: Organizational Transformation & Information Systems (SIGORSA), 2019, pp. 1–10. [Online]. Available: https://aisel.aisnet.org/amcis2019/org_transformation_is/org_transformation_is/20/
[33] M. Eckhart and A. Ekelhart, "Digital twins for cyber-physical systems security: State of the art and outlook," in *Security and Quality in Cyber-Physical Systems Engineering*, S. Biffl, M. Eckhart, A. Lüder, and E. Weippl, Eds., Cham, Switzerland: Springer, 2019, pp. 383–412. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-25312-7_14#citeas
[34] E. Negri, L. Fumagalli, C. Cimino, and M. Macchi, "FMU-supported simulation for CPS digital twin," *Proc. Manuf.*, vol. 28, pp. 201–206, Jan. 2019.

MARIETHERES DIETZ received the B.Sc., M.Sc., and Ph.D. degrees in business informatics from the University of Regensburg, Germany, in 2022. During a semester abroad, she studied management information systems at Linnaeus University, Sweden.

To date, she has authored almost a dozen articles in this regard. Her research interests include digital twins and information security.

THOMAS REICHVILSER received the B.Sc. degree in industrial engineering in Landshut, Germany, in 2018, and the M.Sc. degree in business informatics from the University of Regensburg, Germany, in 2021. He is currently the Senior Consultant of Cloud Analytics and Data Platforms, Deloitte, Munich, Germany.

GÜNTHER PERNUL (Member, IEEE) received the Diploma and Ph.D. degrees (Hons.) in business informatics from the University of Vienna, Austria. He is currently a Professor with the Department of Information Systems, University of Regensburg, Germany. Previously, he held positions with the University of Duisburg-Essen, Germany; the University of Vienna; the University of Florida, Gainesville, FL, USA; and the College of Computing, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include data and information security aspects, data protection and privacy, data analytics, and advanced data-centric applications.