

COMMUNICATION

Model-Driven Engineering for Digital Twins: Opportunities and Challenges

Judith Michael*¹ | Loek Cleophas^{2,3} | Steffen Zschaler⁴ | Tony Clark⁵ | Benoit Combemale⁶ | Thomas Godfrey⁴ | Djamel Eddine Khelladi⁷ | Vinay Kulkarni⁸ | Daniel Lehner⁹ | Bernhard Rumpe¹ | Manuel Wimmer⁹ | Andreas Wortmann¹⁰ | Shaukat Ali¹¹ | Balbir Barn¹² | Ion Barosan² | Nelly Bencomo¹³ | Francis Bordeleau¹⁴ | Georg Grossmann¹⁵ | Gabor Karsai¹⁶ | Oliver Kopp¹⁰ | Bernhard Mitschang¹⁰ | Paula Muñoz Ariza¹⁷ | Alfonso Pierantonio¹⁸ | Fiona A. C. Polack¹⁹ | † Matthias Riebisch²⁰ | Holger Schlingloff²¹ | Markus Stumptner¹⁵ | Antonio Vallecillo¹⁷ | Mark van den Brand² | Hans Vangheluwe^{22,23}

¹RWTH Aachen University, Aachen, DE

²TU Eindhoven, Eindhoven, NL

³Stellenbosch University, Stellenbosch, ZA

⁴King's College London, London, UK

⁵Aston University, Birmingham, UK

⁶Université de Rennes, Rennes, FR

⁷CNRS, Université de Rennes, Rennes, FR

⁸TATA Consulting, Pune, India

⁹JKU Linz, Linz, AT

¹⁰Universität Stuttgart, Stuttgart, DE

Simula Research Laboratory, Oslo, NO

¹²Middlesex University, London, UK

¹³Durham University, Durham, UK

¹⁴ETS, Montréal, CA

¹⁵University of South Australia, Adelaide, AU

¹⁶Vanderbilt University, Vanderbilt, USA

¹⁷University of Málaga, Málaga, ES

¹⁸University of L'Aquila, L'Aquila, IT

¹⁹University of Hull, Hull, UK

²⁰Universität Hamburg, Hamburg, DE

HU Berlin, Fraunhofer FOKUS, Berlin, DE

²²University of Antwerp, Antwerp, BE

²³Flanders Make, Lommel/Leuven/Kortrijk, BE

Correspondence

*Judith Michael, Software Engineering, RWTH Aachen University, Aachen, Germany.
Email: michael@se-rwth.de

Summary

Digital twins are increasingly used across a wide range of industries. Modelling is key to digital twin development – both when considering the models which a digital twin maintains of its real-world complement ('models *in* digital twin') and when considering models of the digital twin as a complex (software) system itself. Thus, systematic development and maintenance of these models is a key factor in effective and efficient digital twin development, maintenance, and use. We argue that model-driven engineering (MDE), a field with almost three decades of research, will be essential for improving the efficiency and reliability of future digital twin development. To do so, we present an overview of the digital twin life cycle, identifying the different types of models that should be used and reused at different life cycle stages (including systems engineering models of the actual system, domain-specific simulation models, models of data processing pipelines, etc.). We highlight some approaches in MDE that can help create and manage these models and present a roadmap for research towards MDE of digital twins.

KEYWORDS

Digital Twin, Model-Driven Engineering, Systems Engineering, Cyber-Physical Systems

1 | INTRODUCTION

Digital twins are among the key drivers of advanced manufacturing¹. Recent surveys^{2,3} have identified a wide range of application areas and technological approaches to building digital twins. Digital twins are software systems that provide services on top of virtual representations of actual systems. They typically enhance actual systems into cyber-physical systems by providing mechanisms for data collection, (partial) modelling, planning and decision making, and changing the state and behaviour of the actual system. Thus, digital twins are complex software systems, raising questions about how they can be engineered both efficiently and effectively.

In research and practice, there is no single definition of a digital twin although all leading definitions have common features, such as sending and receiving data from the actual system or providing a virtual representation⁴. To navigate the design options for twins it is important for digital twin engineers to have conceptual models⁵ and reference models that capture such twins' essential features and variation points and to collaborate with data scientists, as well as domain, system and software engineers. To date, there has been little attention paid in research and practice to digital twin development methods. We argue that models from software and systems engineering can be used to define methods that are used to guide development teams and could be used as the basis of digital twin development platforms⁶.

Digital twins manage models of the actual system and offer services including, but not limited to *analysis, diagnosis, prediction, fault detection and planning*. Digital twins could cover different capabilities⁷: (1) data fusion from sensors to detect properties; (2) consolidation of properties to populate models; (3) the use of model snapshots to provide dashboards and detect situations; (4) projection of situations to speculate about future states (simulation); (5) prediction of future situations from historical information (learning); (6) adaptation that provides decision-support (AI); up to (7) partial and total adaptive control. Each type of service will place requirements on the representation and processing of the system models.

The field of Model-Driven Engineering (MDE) has been argued to support the efficient and effective development, deployment and maintenance of software systems in general⁸. MDE could be used throughout the digital twin life cycle to *specify* and *design* a twin in order to provide a (possibly executable) prototype. These models might be used to automatically generate parts (or even all) of the twin, including configuring its connections to the actual system^{9,6}. An alternative might be that the twin uses models at run-time to execute the models directly in order to improve its adaptability or its ability to provide feedback.

This article reviews the opportunities for MDE to support the field of digital twins. It is organized as follows: Section 2 shows key aspects of digital twins and describes the state-of-the-art for their contexts and life cycle; Section 3 sketches challenges around models and modelling in digital twins; Section 4 provides an overview of the key aspects of

MDE relevant to the digital twin life cycle; finally, Section 5 performs a gap analysis leading to a research roadmap.

2 | DIGITAL TWINS

To understand where digital twin engineering can be supported by the field of MDE, we first describe what parts constitute digital twins, explain their life cycle, and describe relevant digital twin contexts.

2.1 | What is a Digital Twin

There are many different definitions of "digital twins" in the literature² and recent standards such as¹⁰, focusing on different aspects. For the purposes of this paper, we use the following conceptualization based on the 5D model from Tao et al.¹¹. This conceptualization differentiates three key components (see Figure 1):

1. The *actual system*, which is a system or object in the real world. Note that our understanding includes a wide range of systems and objects, including socio-technical or biological systems. The actual system offers interfaces for *output* (that is, sensing / extracting data about the state of the actual system) and *input* (that is, controlling the state of the actual system either directly or by supporting decision making processes of human controllers).
2. A *virtual representation* of the *digital twin*, which includes *models* and connected *data* digitally representing the configuration and state of the actual system.
3. A set of *services* provided on top of the information captured in the digital twin which includes, e.g., services for the synchronization of selected properties with the actual system at a defined synchronisation rate, for reasoning about information from the virtual representation, and visualization and reporting information to digital twin users⁴.

A digital twin is a software system, which includes the virtual representation of the actual system with models and data, and the services needed to fulfill the purpose a digital twin serves.

To realize the software system representing a digital twin and its services, industry and research suggest different reference architectures, e.g., by the Digital Twin Consortium¹³, the standard series ISO-23247 for manufacturing^{14,15}, Reference Architecture Model Industry 4.0 (RAMI 4.0)¹⁶, semantic Asset Administration Shells¹⁷, or more conceptual architectures such as the 3D model¹⁸, the 5D model^{11,12}, or the work by Newrzella et al.¹⁹, Kritzingner et al.²⁰, Josifovska et al.²¹, Boyes and Watson²² or Eramo et al.⁵. These architectures provide an overview of different functionalities a digital twin could cover. However, each concrete, domain-specific digital twin's software architecture has to be tailored for the purposes it aims to cover.

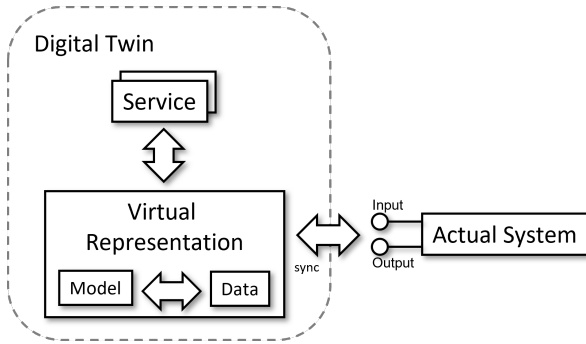


FIGURE 1 Digital twin schema derived from Tao et al.^{11,12}

2.2 | Digital Twin Contexts

Digital twins can be designed for different categories of actual systems. The actual system and its environment constitute the *context* of the digital twin. There are fundamental properties that characterize the context: whether the actual system is an engineered or a non-engineered system, whether it is a controlled or non-controlled system and whether it contains inanimate or animate components or human beings. Based on these properties, we can distinguish three main categories of actual systems which may form the context:

1. **Engineered systems** are systems whose design and realisation have been carefully planned. We differentiate three types of such systems, based on whether they have been developed with digital control in mind:
 - (a) **Cyber-physical (including embedded) systems** are technically engineered products or groups of products comprising physical components and computational control devices. An example could be a robot, a modern car, a wind turbine, an air plane, a production line, or a power plant. Crucially, often they already have digital components that provide an interface for a digital twin to interact with the actual system.
 - (b) **Technical systems without software.** This category contains artefacts such as tables, steam engines, classical bicycles, hydraulic presses, windmills, bridges, buildings, etc. These systems are engineered, but do not have a software component (yet). To allow the creation of digital twins, we need to enhance the system with digital interfaces for data collection and for detecting changing states and behaviours.
 - (c) **Software** systems sit somewhere halfway between the previous two. Software is naturally digital, making the creation of a digital twin of the software easier in principle. However, the information captured digitally is not typically the information one wants to reflect in the digital twin, necessitating the creation of additional interfaces (e.g., for the monitoring of build processes).
2. **Biological / natural systems** which are not designed by human engineers and can exist without human intervention, such as biological organisms and ecosystems, e.g., trees, forests, plantations, animal

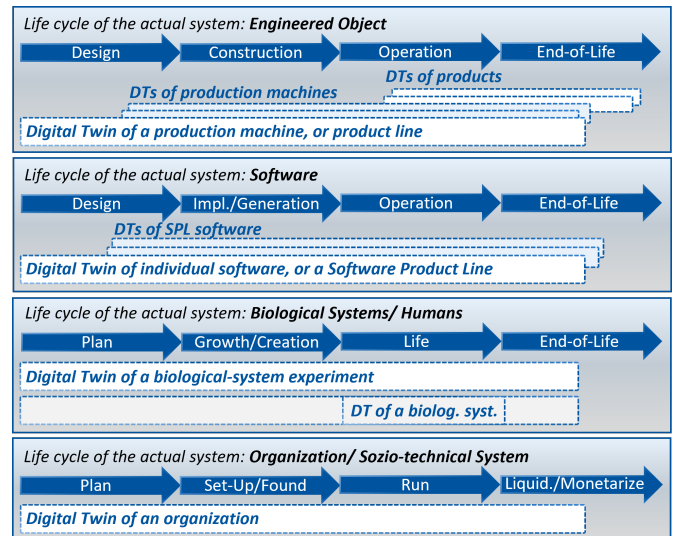


FIGURE 2 Digital twins for different types of actual systems and their life cycle.

populations, or weather phenomena. Here, creating digital interfaces is much more challenging and often means a transformation from pure biological to bio-technical systems.

3. **Socio-technical systems and processes, including organisations**, i.e., systems that involve human behaviour and interaction between humans and technical systems. An example could be a hospital, the flow of patients through a healthcare system, or the organizational structure of a company. They might already provide digital interfaces for some aspects, e.g., technical devices, but also require to add new ones, e.g., to detect human behavior.

These different categories each come with their own requirements as well as broader aims, such as supporting sustainability, privacy, or maintainability.

2.3 | Digital Twin Life Cycle

The life cycle of a digital twin is deeply connected to the life cycle of the actual system it is twinning. The life cycle of the digital twin and its actual system does not have to run in lockstep. For example, consider digital twins used for simulation-based prediction during system design: the life cycle of such a digital twin will be fully completed alongside the design phase of the actual system and the digital twin may no longer be used once the actual system has been constructed and is in operation. Figure 2 shows the life cycle of some examples for such actual systems and some examples for digital twins.

For an *engineered object*, the life cycle includes the design, construction, operation, and maintenance until the end of life. A digital twin of a product line can accompany the whole life cycle starting with, e.g., engineering models created during model-based systems engineering²³, simulation and forecasting services, simulation data in the

design, continuous updating of data and models with the reality during construction, monitoring, and forecasting during operation until their demolition or refurbishment during the end-of-life phase. The same phases occur for a concrete production machine or one concrete product produced in such a production machine. However, in relation to the life cycle of the product line, their life cycles start later.

The life cycle of *biological systems* such as humans, plants or animals includes the (optional) plan, growth, life, and end-of-life phase. If we take a biological-system experiment as an example—for example, to create a lab-grown ear—a digital twin already accompanies the planning of the experiment and uses services to check the plans with regulators, observe the ear during its growing phase, and regulate the environment of the ear to provide better growth. After transfer to a human, the ear might be further observed and stimulated to determine whether this experiment is successful or not. The same life cycle occurs for other biological processes such as 3D bio-printing or in vitro fertilization. Furthermore, a digital twin could accompany humans or organs of humans such as a digital twin for the brain²⁴, a human heart²⁵ or lungs. For sensing and actuating a biological system synchronized with its digital twin, digital twin developers might consider the whole bio-technical system including relevant technological solutions as the actual system.

The life cycle of *software systems* is similar to the one of engineered objects. In waterfall-like development processes, developers start with the analysis and design of a system, implement, generate, operate, reengineer, stop, or replace the system in the end-of-life phase. In agile processes, the first three phases are continuously repeated for each software product increment. During the whole life cycle, a digital twin could accompany such a software system. If we create a software product line (SPL)—used for developing families of similar software products—we could have a digital twin for the SPL reference design and implementations as well as each concrete realization of the SPL in an individual software product.

The life cycle of *socio-technical systems* such as an *organization* can be summarized as planning, set-up or founding an organization, running it, and liquidating or monetizing it. If a structured planning phase exists, e.g., in start-up incubators, if founders start by elaborating a business plan, or if a crowdfunding campaign should support the funding of a company, a corresponding digital twin could already exist from the beginning. If a company already exists, the digital twin might only exist during the operation of a company, e.g., to monitor business processes or key performance indicators.

Changes in the actual system and its environment may lead to re-adjustments and changes in the digital twin, such as adding new services, handling new types of models, or changing data structures and visualization needs. These changes lead to the need for continuous DevOps²⁶ cycles during the lifetime of the actual system and its digital twin.

3 | CHALLENGES AROUND MODELS AND MODELLING IN DIGITAL TWINS

The previous section presented typical elements of digital twins and their life cycle, highlighting some of the challenges in developing and maintaining digital twins. Importantly, digital twins include *models* (in the form of digital representations that can be used to derive insights) of some actual system. This section examines these models in detail: what are the different types of models involved in a digital twin's life cycle and how can making these models explicit help us engineer and maintain digital twins more efficiently and effectively.

Models are more than data: they capture structure and provide context for the interpretation of data. Digital twins include data gathered from the actual system and models that support the interpretation and analysis of these data. Some models are derived from data—either *a posteriori* from historical data or (updating) dynamically. Other models are provided as *a priori* descriptions/prescriptions²⁷ independent of data, though they may be parameterized with data from the actual system. Where information cannot be easily measured or computed, one can use approximation models, also known as surrogate models. These aim to learn functional relationships between data inputs and outputs statistically, but can sometimes lead to a lack of explainability. Models are on the knowledge and information planes of the well-known data-information-knowledge hierarchy^{28,29}.

In our discussion, we make a fundamental distinction between two types of models:

1. *Models in digital twins* are models of the actual system that are maintained inside the digital twin. They capture information about the actual system and enable analysis, simulation, inspection etc. These models can be descriptive, predictive, or prescriptive (see⁵).
2. *Models of digital twins*, on the other hand, consider the digital twin system as a complex software-intensive system and capture information about the structure, behaviour, and state of the digital twin system.

The following subsections discuss each of these two types in turn.

3.1 | Models in Digital Twins

These are the models we often think of first when considering a digital twin – after all, models of an actual system are an integral part of every digital twin. Considering them as explicit artefacts, rather than as an implicit model that perhaps *emerges* from data in a database together with a particular analysis algorithm (say, a deep-learning-based classifier), allows us to think about their structure, purpose, and life cycle more deeply.

3.1.1 | Kinds of models

Digital twins include different kinds of models. In particular, modelling experts differentiate structural and behavioural models:

1. *Structural models*, which capture the logical structure of the actual system—i.e., its sub-systems, components, and actors. Examples are, e.g., structural models of railwork networks, citizen energy communities, or production plants. Structural models can be expressed in general purpose modelling languages—for example Systems Modelling Language (SysML) block definition diagrams—as well as domain-specific modelling languages, e.g., domain-specific architectural component-and-connector diagrams.
2. *Behavioral models*, which capture key processes and interactions in the actual system and between the actual system and its environment. A typical modelling language for such processes in the business domain is the Business Process Model and Notation (BPMN) standard. In the systems engineering domain, state machines are a familiar modelling language, but more domain-specific models also exist.

Orthogonal to these two kinds, modelling experts also structure models based on properties of the system engineers aim to develop models for: engineered or natural systems. These models could have structural and behavioural aspects.

1. *Models for engineered systems* capture detailed mechanisms, the tangible structure and behavior of these systems. Examples include CAD models, building information models (BIM) in the construction domain, geographic information system (GIS) models (for example of the geographical layout of a railway network), or SysML models describing the structure and behavior of a production machine.
2. *Models for natural systems* aim to describe the structure and behavior of these existing systems. Examples include mathematical models (e.g., ODE-style descriptions of chemical reaction networks), models describing the topology and geographical structure of the world, or weather models as simulations of the future state of the atmosphere through time.

Models in digital twins can be captured in general-purpose modelling languages but also in highly domain-specific modelling languages^{30,31}. This broad variety of languages raises challenges around integrating models captured in different modelling languages, created by experts in different organizational units, or focused on different aspects of the actual system, ensuring their consistency and coherent evolution over time.

Systems of interest for representation through a digital twin are often systems of systems³². Different sub-systems in such a system of systems are controlled by different organisations or actors, making the idea of a single, centrally managed digital twin of the complete system unrealistic. Instead, for a practical realization of digital twins, there is a need

for federating digital twins, reflecting the compositional construction of the system being twinned and making the digital twin itself a system of systems³³. In such a situation, each of the “sub-twins” will have its own ways of referring to key elements of the system being represented. Centrally aligning these model concepts to a commonly agreed ontology is usually difficult because of the nature of distributed organisational control in the actual system. Instead, digital twin engineers need to be able to map representations onto each other. Capturing the decisions we make in such a mapping explicitly is important for quality assurance purposes.

3.1.2 | Purpose

Every model is constructed for a purpose³⁴. The purpose determines what aspects of reality we include in the model and what we abstract away. This abstraction, in turn, determines what the model can be used for and where its limitations lie. For example, a simulation model created for analysing the flow of streams of passengers through an airport may not have enough information to analyse individual passenger decision making.

This focus on a purpose and different abstractions is interesting in the context of digital twins, where the lifetime can be long. The questions digital twin users ask of a digital twin—that is, the digital twin use cases—are likely to change, leading to a slow change in the purpose of the digital twin over time. The services in the digital twin need to evolve with this change. As a result, the models underlying these services may no longer be sufficiently accurate or include all the information required and may need to change, too. This need for change creates challenges around maintainability and consistency: large, monolithic models will be more difficult to evolve consistently, efficiently, and systematically³⁵. Even the modelling tools and formalisms may change over time, given the potentially very long lifetimes of digital twins and their actual systems. Equally, evolving models will likely change the alignment between models and the underlying data. One solution attempt might be to collect as much data as possible and speculatively integrate as much of it as possible into models even if this information is not required for the current purpose. This solution attempt may provide flexibility when the purpose changes (assuming changes in purpose can be predicted reasonably well), but will make the use of the models more difficult for the initial purpose.

For digital twins to be used as decision-support systems (for example, in socio-technical digital twins³⁶) or even more autonomously (as in the definition of digital twins proposed by Kritzinger et al.²⁰), digital twin users must be able to trust them³⁷. This means digital twin users must be able to understand how the digital twin has arrived at its decisions and recommendations³⁸ and that the models inside the digital twin are sufficiently accurate representations of the actual system for the recommendations and decisions to be appropriate. Simulation is a core service provided by digital twins and fundamental to its planning and decision-making capabilities. In the context of engineered systems, trust in models and simulations has been discussed under the labels

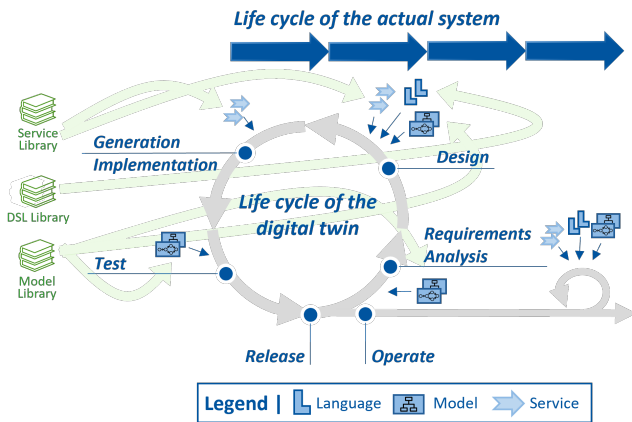


FIGURE 3 Life cycle of a digital twin in detail.

of “experimental” or “validity” frame^{39,40}. These take the perspective that the essential validation required is whether a simulation can reproduce the input–output behaviour of an actual system to a desired level of accuracy. In the context of complex systems (including biological and socio-technical systems), research suggests⁴¹ that a richer set of information is required to ensure trust in the model, linking modelling decisions to the scientific literature, real-world experimentation, expert decision making etc. In particular, this approach suggests that we need to be able to trust the structure of the models and not just their input–output behaviour. Establishing trust at this level requires making a structured *argument*^{42,43}, similar to how one might construct a safety-assurance case. Where digital twins integrate artificial intelligence, in particular, machine learning, similar arguments are required to ensure trust⁴⁴.

3.2 | Models of Digital Twins

Here, the focus is on some of the modelling concerns relevant as part of the development of a digital twin as a complex software system. This development can be connected to the model-based systems engineering process⁴⁵, however, it is a process on its own requiring agile development methods.

Figure 3 takes a closer look at the continuum formed by the whole life cycle of a digital twin, as supported by DevOps principles, i.e., the requirements analysis, design, generation and implementation, test, release and operate phase. This continuum enables the seamless transition from one step to another in the life cycle. From a software engineering point of view, a high degree of reuse is to be aimed for; libraries of models, domain-specific modelling languages, and digital twin services foster such reuse (the green arrows in Figure 3 show the library reuse).

Digital twin engineers collect relevant requirements for the digital twin and create requirements models within the requirements analysis phase of a digital twin. This collection can include functional requirements of the digital twin to be built up to specific user needs and wishes.

The requirements are then used to design the digital twin. In the design phase, we plan the first version of the digital twin and create design models such as for the system architecture or data structure, select which languages should be available in the digital twin during its runtime, and select which services from the service library to include in the digital twin. The production code and test code that uses the selected digital twin services and functionalities are derived from the design models by hand or using generative techniques as in Model-Driven Engineering (see Section 4). After successful tests, the digital twin can be released and start its operation. The operation of the digital twin is characterized by the continuous monitoring of the actual system and its context.

3.3 | Models Across the Digital Twin Life Cycle

As digital twins exist for long-living actual systems, e.g., up to 20 years for injection molding machines or wind-turbine parks, and up to 50 years for bridges or buildings, it is clear that requirements for the actual system and the digital twin of it, and its context may change massively. This continuous need to react to changes requires an engineering approach with continuous DevOps cycles for the digital twin which influences both models in digital twins and models of digital twins.

Since these continuous DevOps cycles of the digital twin are triggered by changes in the system or its context, we are focusing on these changes in the ongoing relationship between the actual system and the digital twin. Figure 4 takes a closer look into the four main phases of the life cycle of an engineered actual system: Design where the actual system is planned; Construction where it is built; Operation where it is used; and End-of-Life where it is refurbished, demolished, and recycled. In an ideal world, all of these phases are accompanied by a digital twin which transforms continuously from a design digital twin to a construction digital twin, or monitoring digital twin. This transformation requires continuous DevOps cycles for the digital twin (the DevOps cycle from Figure 3 is shown as continuous grey loops below the actual system in Figure 4).

Already at the beginning of the design phase of the actual system ① in Figure 4, we have to engineer the digital twin by reusing models of the actual system and digital twin services as discussed in Figure 3, to have a digital twin already available when the design of the actual system starts.

Depending on the type of actual system represented by the digital twin, models of the system may already play a role independently of the existence of a digital twin. In particular, where the actual system is an engineered system (including, for example, production line systems), models are likely to exist that were used in *developing* the actual system. To ensure high fidelity of the digital twin as well as efficient development of the digital twin, it is helpful for digital twin engineers to be able to reuse these models directly inside the digital twin. However, reuse can be challenging, as it requires the models to be kept consistent with the actual system throughout its design and implementation. It also requires the definition of mechanisms for how elements in the model connect to

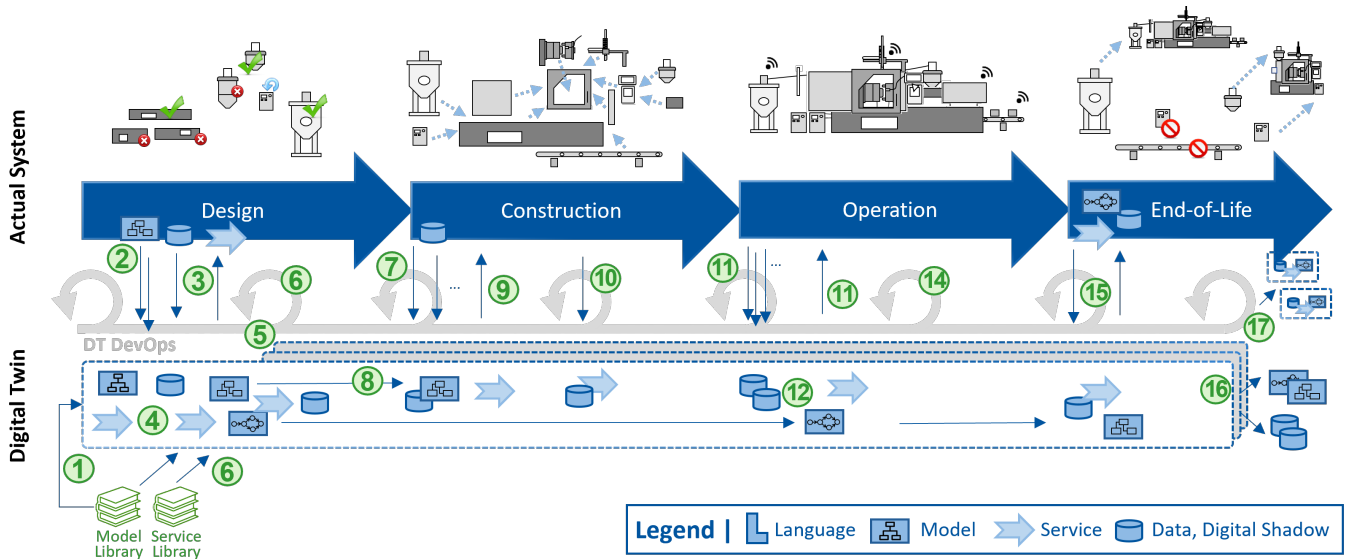


FIGURE 4 Life cycle of the actual system accompanied by the life cycle of its digital twin.

elements in the actual system. This connection is a general requirement and may become easier where design models are reused, as the relevant traceability information may already have been captured.

In the design phase (2 in Figure 4), models about the actual system are created, e.g., systems engineering models. These models are continuously updated in the digital twin as they evolve in reality over time. Moreover, (3) data about the engineered system is created in third-party applications—for example, sets of parameters in simulation and optimization software—and shared with the digital twin and communicated back in other stages of the design phase. The digital twin can handle data and models as it (4) includes services—for example, for data and parameter visualization, simulation, or model validation.

The engineering process of the digital twin has to take into account that the design process of the actual system often starts with the design of objects which are then realized not in lots of size 1 but where several actual systems are constructed and operated. Thus, multiple digital twins have to be realized, based on the same initial set of engineering models, data and services at the beginning, which are then adapted to the actual context during construction. When the planning of a product line moves on to a concrete engineered product, (5) we can transfer the planning data and models of a product line digital twin to the digital twins of concrete production machines. Those are then further detailed during their design phase and receive updated models, and data.

Changes in this design process or changing requirements might make it necessary to (6) start a new dev-cycle of the digital twin and, e.g., update existing models of digital twins, or include additional models or services from the libraries.

Complex actual systems take some time to be actually constructed—from several months up to several years for production lines and factories. In this construction phase, we receive (7) data on the construction process and have to (8) update models if the planned ones deviate

from reality. Based on such changes, engineers might have to (9) update parameters in the cyber-part of an actual cyber-physical object.

As in the design phase, also in this phase changing management, changing laws, new key performance indicators, new services, and new data might require changing the digital twin and (10) starting new dev-cycles.

In the operation phase, the digital twin (11) monitors the actual system and receives data continuously. One can (12) compare planned runtime models with real runtime models and if deviations occur, the digital twin (13) sends execution commands to adapt the actual system. In additional development cycles (14), one can add new languages—for example, to add runtime models, or new services (e.g., for process discovery from data or process conformance checking).

In the end-of-life phase, the actual system and its parts can take different paths: the system can be deconstructed, parts can be recycled, reused, or thrown away. This change in the actual system might require (15) to reconfigure our digital twin, e.g., based on new data or the absence of expected data. Due to legal reasons (16), we might need to keep data and models (or the whole digital twin) longer than the actual system. Moreover, we might (17) reuse parts of the actual system and provide related data, models, and services for the digital twin of the other actual system.

To summarize some main challenges when developing and evolving digital twins: Besides the current practice of ad-hoc development of digital twins, several problems are related to the underlying complexities of the actual system^{46,47}: The system complexity due to a large number of heterogeneous subsystems, time complexity related to the long lifespan of the actual system and changing realities, integration complexity due to heterogeneous artefacts, different stakeholders and heterogeneous views, as well as information complexity due to heterogeneous

data and models. In addition, budget challenges might influence the development and maintenance of digital twins. Moreover, current digital twin practices face a gap in information and feedback flows between the different life cycle phases.

4 | MODEL-DRIVEN ENGINEERING

As we have seen, models play a key role in many aspects of digital twin development. This prominent role of models raises important questions about how these models should be created, maintained, manipulated, and used (see also Tao et al.^{48, Sect. V.B}). Some of these challenges have already surfaced in the discussion above. For example, we have seen the importance of being able to reuse models across life cycle stages.

MDE⁸ is a sub-field of software engineering that focuses on using models as first-class artefacts in the software engineering process. As a result, the field has developed many techniques for the efficient development of modelling languages, effective manipulation and validation of models, and techniques for maintenance and optimisation of models. These techniques will also be useful in the creation and use of digital twins. The next section discusses some of the ways in which MDE can help address digital twin challenges and outlines some open challenges that should be addressed between the MDE and digital twin communities.

To set the context for this discussion, we briefly set out the key concepts in MDE. For space reasons, this section cannot be an exhaustive introduction to the field. For more detail, Brambilla et al.⁸ provides a good introduction.

- *Models* are the key ingredients of MDE. A model is an abstraction of parts of the real world for a purpose.
- In MDE, models need to be expressed in a *modelling language* that has been formalised (i.e., exists in a way that is unambiguously processable by a computer). Modelling languages can be general-purpose, like the Unified Modelling Language or SysML. For many projects, domain-specific modelling languages (DSMLs), created specifically for the development of applications in a particular domain, can be more useful.
- Models can be transformed into other models or into text (e.g., program code, documentation, reports, etc.) by automated *model transformations*, often expressed in dedicated transformation languages.
- A large range of tools and mechanisms exist to *manage models*. These support model versioning, model comparison, model analysis, model validation, model composition and reuse, linking models, and many more.

5 | MDE OF DIGITAL TWINS: A RESEARCH ROADMAP

The previous sections described the different aspects and processes involved in MDE of digital twins. They covered the topics of development, execution and analysis of a digital twin, and throughout, highlighted the concerns and considerations developers should have in developing a digital twin. In this section, we perform a gap analysis, discussing the remaining challenges facing digital twin development and potential avenues for future research topics. Figure 5 shows an overview of current challenges, existing solutions from the MDE Body of Knowledge and needed action items for each topic further discussed in the following subsections.

5.1 | Development Guidelines and Standardisation for Digital Twins and their Engineering

Digital twin development is still largely 'artisanal' and would benefit from the possibility of reuse and from standardisation of development knowledge⁴⁹, such as ongoing in ISO/IEC JTC 1/SC 41 for IoT and DTs[†] or for cities and municipalities in DIN SPEC 91607 - 2024-11⁵⁰. Even though implementation differences exist between digital twins in different contexts and environments, the potential to standardise reference architectures or requirements capture for digital twins remains. In an MDE-based approach to digital twin development, such knowledge would be captured in DSMLs, models, and model transformations. Research challenges remain around specifying the concrete DSMLs and transformations: For example, what are the concepts required in DSMLs that can capture the typical requirements, assumptions, and scope of digital twin systems? What are the most important non-functional properties of digital twin systems, including specific properties such as synchronization frequencies and fidelity? How can they be captured in models of digital twins and can such models be automatically transformed for analysis at digital twin design time or even to generate digital twin implementations that achieve the required non-functional properties by construction?

5.2 | Digital Twin Interoperability

Integration and data exchange between different components is a key technical challenge in digital twin development. Correct and timely data conversion when synchronizing the information in the digital twin with the status of the actual system is crucial for digital twin efficacy, particularly where virtual models and services within a digital twin are largely heterogeneous and use different data formats. Time-based qualities, security, and modularity are core digital twin qualities and are closely tied to design and data transfer concerns. An MDE approach

[†] <https://www.iso.org/committee/6483279.html>

Topic	Challenges	MDE BoK	Action Items
Guidelines and Standards	Ad-hoc development	Modeling languages, model transformations, reference models	Contextualization for digital twin domains with domain experts
Interoperability	Integration complexity due to heterogeneous artefacts, different stakeholders, heterogeneous views	Interchange formats, model transformations, code generation, interoperability models	Formalizing structures currently used in specific domains for realizing bidirectional mappings between different digital twin implementation approaches
Multi-paradigm/ Multi-view	System complexity and heterogeneous domain knowledge and needs	Multi-paradigm modeling/multi-viewpoint modeling frameworks including concepts, methods, and tools	Efficient methods to define multiple views on digital twins with consistency assurance between different views, abstraction levels, etc.
Evolution	Time complexity due to changing realities and long lifetimes	Model evolution and co-evolution approaches, change types, change impact, and repairs	Continuous evolution support for running digital twins/actual systems, change categorizations for digital twins and consistency requirements
Continuous Engineering	Costs, system complexity and heterogeneous domain expert backgrounds, gap between design and operation	Model-based DevOps based on models@runtime techniques and tools	Abstraction levels for digital twin data and scalable models@runtime to deal with long-living systems, feedback methods from operation to design
Composition of Digital Twins	Information complexity due to heterogeneous data and models	Model/modeling language composition operators and theories	Development of language support for composing digital twins (both during design-time and runtime)

FIGURE 5 Overview of the research roadmap: challenges, existing solutions from the MDE Body of Knowledge and needed action items

could help address these challenges in two ways: (i) using DSMLs, data formats could be standardised across components for more efficient information interchange; and (ii) where such alignment is not possible, the translations required could be implemented using model transformation technology, making them explicit and amenable to analysis and verification. Where digital twin components are provided externally as 'black boxes', MDE offers opportunities for automating the generation of wrappers and glue-code to increase reuse efficiency. In the area of simulation—keeping in mind that many digital twins include a simulation capability—distributed co-simulation⁵¹ and the high-level architecture⁵² are established approaches for black-box composition. These approaches can be combined with the generation capabilities of MDE to achieve even greater flexibility across domains⁵³. Challenges remain, of course. For example, only a limited understanding currently exists of what the structure of the information to be interchanged should be (though work such as the Asset Administration Shell¹⁷ has made an MDE-based start on these questions). Equally, many data interchange transformations need to be bidirectional. The theory and tooling for bidirectional transformation is an active research area within MDE.

5.3 | Multi-Paradigm and Multi-View Modelling for Digital Twins

Models in a digital twin often span multiple levels of abstraction. For example, a digital twin for agriculture may include a model concerning the qualities and properties of fields such as turnover rate and yield,

but may also include a model on individual crops and their optimal growth conditions, etc. Integrating these different abstraction levels into a single digital twin is often far from trivial. Communication protocols between models at different abstraction levels, and aggregation of data collection from those models, need to be considered carefully. Often, a digital twin involves a diverse range of stakeholders during development, maintenance, and use. Different stakeholders may require different viewpoints on the digital twin (both in relation to models *in*, and models *of* the digital twin), including the type of data exposed to the user and the format/syntax of data presentation. In MDE, the areas of multi-paradigm and multi-view modelling address challenges like the two above, making the theory and tools developed highly relevant for digital twin development. Significant open research challenges remain. For example, research is needed on how to minimise development overheads such as facilitating stakeholder engagement, implementation time for different data views, and view integration. Bidirectional synchronization and transformations play a key role here⁵⁴.

5.4 | Model, Object, and Digital Twin Evolution

Section 2.3 discussed the relationship between the life cycle of the actual system and the life cycle of the digital twin. We highlighted that this relationship requires constant co-evolution between the digital twin and the actual system – beyond just bi-directional data updates, this also needs to support updates to the structure of any models kept in the digital twin⁵⁵. For example, when new services and components are

added to the actual system, these need to also be represented in the digital twin, requiring appropriate amendments to the structure of any models in the digital twin. MDE research has a significant body of work exploring aspects of co-evolution between models and the reality they represent or reactive updates from one model to connected models⁵⁶. These ideas are also applicable in the digital twin context. However, new challenges arise related to the need for the digital twin to be updated automatically when changes in the actual system occur. For example, how and when a digital twin should be 'disconnected from/reconnected to' the actual system and how the virtual representation and services of the digital twin can be updated appropriately and automatically. Digital twin engineers also have to consider how these changes may impact the consistency of models/meta-models and data during transition periods.

5.5 | Model-Based DevOps for DTs

Digital twins require models of the actual system. Creating these models initially can be challenging. For engineered actual systems in particular, models often already exist from the system design phase. However, these can currently be difficult to reuse when creating digital twins. In MDE, the idea of models@runtime^{57,58} provides tools and techniques for managing system (design) models at runtime of a software system. This idea has led to the extension of DevOps ideas to the space of run-time model-based DevOps^{59,60}, including the smooth transition from design models to runtime models⁶¹. Digital twins need to manage very large models and data sets efficiently and effectively, and this management continues to be a research challenge for models@runtime approaches, too. Similarly, digital twin engineering still raises challenges around identifying appropriate abstraction and refinement mechanisms to synthesise the rich detail of actual data into processable model information and *vice versa*⁶². Moreover, feedback from the operation phase of the actual system and its digital twin back to the system design and its design models could improve engineered systems.

5.6 | Composition of Digital Twins

Actual systems are complex, and as discussed before, represent actually systems of systems. As a consequence, digital twins cannot be developed monolithically, but rather need to be considered as federations of sub-twins³³. This federation raises challenges about the safe and robust composition of different digital twins, their models and simulations. MDE has a long tradition of exploring issues of modularity and composition/integration of models, transformations, and DSMLs, both in foundational terms and practical tools. All of these are applicable to digital twin development and can provide significant benefits. MDE is also able to capture typical digital twin composition operators in appropriate DSMLs, which would enable more efficient expression of new composition scenarios, including their dynamic manipulation and optimisation. However, significant challenges remain⁶¹, not least around managing information protection between different organisations in

charge of different sub-digital twins, and in relation to supporting dynamic federation (and disbandment) of digital twins. This challenge requires dynamically adjusting models and simulations, possibly part-way through simulation runs, while making sure that real-time and historic data (which may be based on different twin federations) are managed robustly and reliably.

6 | CONCLUSIONS

To create, maintain and evolve digital twins along with their actual system over a long lifespan is a significant engineering challenge. Modelling is at the core of digital twins⁴⁸. In this paper, we have discussed different forms of modelling in and of digital twins and how MDE provides the basic techniques to support the digital twin life cycle in an agile manner. We have identified different ways of applying MDE to digital twin development and operation and have identified future research challenges for MDE-based digital twin development.

ACKNOWLEDGEMENTS

This paper is the result of discussions during Dagstuhl Seminar 22362 "Model-Driven Engineering of Digital Twins". We thank the team at Dagstuhl for supporting our discussions.

Partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and the Agence Nationale De La Recherche (ANR) – France – Model-Based DevOps – 505496753. Website: <https://mbdo.github.io>.

References

1. Xian W, Yu K, Han F, Fang L, He D, Han QL. Advanced Manufacturing in Industry 5.0: A Survey of Key Enabling Technologies and Future Trends. *IEEE Transactions on Industrial Informatics* 2023; 1-15. doi: 10.1109/TII.2023.3274224
2. Dalibor M, Jansen N, Rumpe B, et al. A cross-domain systematic mapping study on software engineering for Digital Twins. *Journal of Systems and Software* 2022.
3. Muctadir HM, Manrique Negrin DA, Gunasekaran R, Cleophas L, van den Brand M, Haverkort BR. Current trends in digital twin development, maintenance, and operation: an interview study. *Software and Systems Modeling* 2024; 23(5): 1275–1305. doi: 10.1007/s10270-024-01167-z
4. Zhang J, Ellwein C, Heithoff M, Michael J, Wortmann A. Digital Twin and the Asset Administration Shell: An Analysis of 3 AASs Types and their Feasibility for Digital Twin Engineering. *Journal Software and Systems Modeling (SoSyM)* 2025. doi: 10.1007/s10270-024-01255-0
5. Eramo R, Bordeleau F, Combemale B, Van Den Brand M, Wimmer M, Wortmann A. Conceptualizing digital twins. *IEEE Software* 2021; 39(2): 39–46.

6. Dalibor M, Heithoff M, Michael J, et al. Generating Customized Low-Code Development Platforms for Digital Twins. *Journal of Computer Languages (COLA)* 2022; 70. doi: 10.1016/j.cola.2022.101117
7. Wagg DJ, Worden K, Barthorpe RJ, Gardner P. Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* 2020; 6(3): 030901. doi: 10.1115/1.4046739
8. Brambilla M, Cabot J, Wimmer M, Baresi L. *Model-Driven Software Engineering in Practice*. Morgan & Claypool. second edition ed. 2017.
9. Kirchhof JC, Michael J, Rumpe B, Varga S, Wortmann A. Model-driven Digital Twin Construction: Synthesizing the Integration of Cyber-Physical Systems with Their Information Systems. In: *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems* ACM; 2020: 90-101.
10. ISO/IEC 30173:2023 Digital twin – Concepts and terminology.; 2023.
11. Tao F, Liu W, Zhang M, et al. Five-dimension digital twin model and its ten applications. *Computer integrated manufacturing systems* 2019; 25(1): 1–18.
12. Tao F, Zhang M, Nee A. Chapter 3 - Five-Dimension Digital Twin Modeling and Its Key Technologies. In: Tao F, Zhang M, Nee A., eds. *Digital Twin Driven Smart Manufacturing* Academic Press. 2019 (pp. 63-81)
13. McKee D. Platform Stack Architectural Framework: An Introductory Guide.; 2023.
14. ISO/DIS 23247-1. Automation systems and integration—digital twin framework for manufacturing—Part 1: overview and general principles.; 2020.
15. Ferko E, Bucaioni A, Pelliccione P, Behnam M. Standardisation in Digital Twin Architectures in Manufacturing. In: *2023 IEEE 20th International Conference on Software Architecture (ICSA)*; 2023: 70-81
16. Schweichhart K. Reference Architectural Model Industrie 4.0 (RAMI 4.0)-An Introduction. *Publikationen der Plattform Industrie* 2016; 4(0).
17. Bader SR, Maleshkova M. The Semantic Asset Administration Shell. In: *Semantic Systems. The Power of AI and Knowledge Graphs: 15th International Conference, SEMANTiCS 2019*. Springer; 2019: 159–174.
18. Grieves M, Vickers J. *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*: 85-113; 2017
19. Newrzella SR, Franklin DW, Haider S. Three-Dimension Digital Twin Reference Architecture Model for Functionality, Dependability, and Life Cycle Development Across Industries. *IEEE Access* 2022; 10: 95390-95410. doi: 10.1109/ACCESS.2022.3202941
20. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in manufacturing: A categorical literature review and classification. In: *IFAC* 51; 2018: 1016–1022.
21. Josifovska K, Yigitbas E, Engels G. Reference framework for digital twins within cyber-physical systems. In: *2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS)* IEEE. ; 2019: 25–31.
22. Boyes H, Watson T. Digital twins: An analysis framework and open issues. *Computers in Industry* 2022; 143: 103763. doi: 10.1016/j.compind.2022.103763
23. Shoshany-Tavory S, Peleg E, Zonnenshain A, Yudilevitch G. Model-based-systems-engineering for conceptual design: An integrative approach. *Systems Engineering* 2023; 26(6): 783-799. doi: 10.1002/sys.21688
24. Amunts K, al. e. The coming decade of digital brain research - A vision for neuroscience at the intersection of technology and computing. *Imaging Neuroscience* 2024; 2. doi: 10.1162/imag_a_00137
25. Coorey G, Figtree GA, Fletcher DF, et al. The health digital twin to tackle cardiovascular disease: a review of an emerging interdisciplinary field. *npj Digital Medicine* 2022; 5(1). doi: 10.1038/s41746-022-00640-7
26. Kim G, Humble J, Debois P, Willis J, Forsgren N. . IT Revolution Press. 2nd ed. 2022.
27. Combemale B, Kienzle J, Mussbacher G, et al. A Hitchhiker's Guide to Model-Driven Engineering for Data-Centric Systems. *IEEE Software* 2020; 38(4): 71–84.
28. Ackoff R. From data to wisdom: Presidential address to ISGSR, June 1988. *Journal of Applied Systems Analysis* 1989; 16: 3–9.
29. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science* 2007; 33(2): 163–180. doi: 10.1177/0165551506070706
30. Cheng BHC, Combemale B, France RB, Jézéquel JM, Rumpe B. On the Globalization of Domain Specific Languages. In: *Globalizing Domain-Specific Languages* Springer; 2015.
31. Combemale B, France R, Jézéquel JM, Rumpe B, Steel J, Vojtisek D. *Engineering Modeling Languages: Turning Domain Knowledge into Tools*. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series . 2016.
32. Michael J, Pfeiffer J, Rumpe B, Wortmann A. Integration Challenges for Digital Twin Systems-of-Systems. In: *10th IEEE/ACM Int. Workshop on Software Engineering for Systems-of-Systems and Software Ecosystems* IEEE; 2022
33. Olsson T, Axelsson J. Systems-of-Systems and Digital Twins: A Survey and Analysis of the Current Knowledge. In: *18th Annual System of Systems Engineering Conference (SoSe'23)*; 2023: 1–6
34. Stachowiak H. *Allgemeine Modelltheorie*. Wien: Springer . 1973.
35. Torres W, Brand v. dMGJ, Serebrenik A. A systematic literature review of cross-domain model consistency checking by model management tools. *Softw. Syst. Model.* 2021; 20(3): 897–916. doi: 10.1007/S10270-020-00834-1
36. Barn BS. The Socio-technical Digital Twin: On the Gap Between Social and Technical Feasibility. *IEEE Conference on Business Informatics* 2022.
37. Laplante P. Trusting Digital Twins. *IEEE Computer* 2022; 55(7): 73–77. doi: 10.1109/MC.2022.3149448

38. Michael J, Schwammberger M, Wortmann A. Explaining Cyberphysical System Behavior with Digital Twins. *IEEE Software* 2024; 41(1): 55–63. doi: 10.1109/MS.2023.3319580
39. Zeigler BP, Muzy A, Kofman E. *Theory of Modeling and Simulation*. Elsevier. 3rd edition ed. 2018.
40. Denil J, Klikovits S, Mosterman PJ, Vallecillo A, Vangheluwe H. The experiment model and validity frame in M&S. In: *Proceedings of the Symposium on Theory of Modeling & Simulation*; 2017: 1–12.
41. Stepney S, Polack FAC. *Engineering Simulations as Scientific Instruments: A Pattern Language*. Cham: Springer . 2018.
42. Zschaler S, Polack FAC. Trustworthy agent-based simulation: the case for domain-specific modelling languages. *Software and Systems Modeling* 2023; 22(2): 455–470. doi: 10.1007/s10270-023-01082-9
43. Wilsdorf P, Zschaler S, Haack F, Uhrmacher AM. Potential and Challenges of Assurance Cases for Simulation Validation. In: *2024 Winter Simulation Conference*; 2024.
44. Burton S, Herd B. Addressing uncertainty in the safety assurance of machine-learning. *Front. Comput. Sci.* 2023; 5: 1132580. doi: 10.3389/fcomp.2023.1132580
45. Bickford J, Van Bossuyt DL, Beery P, Pollman A. Operationalizing digital twins through model-based systems engineering methods. *Systems Engineering* 2020; 23(6): 724–750. doi: 10.1002/sys.21559
46. Feichtinger K, Meixner K, Rinker F, et al. Industry Voices on Software Engineering Challenges in Cyber-Physical Production Systems Engineering. In: *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*IEEE; 2022.
47. Michael J. *Model-Driven Engineering of Digital Twins with Informative and Assistive Services*. Aachener Informatik-Berichte, Software Engineering. Shaker Verlag . 2025.
48. Tao F, Zhang H, Liu A, Nee AYC. Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics* 2019; 15(4): 2405–2415. doi: 10.1109/TII.2018.2873186
49. Niederer SA, Sacks MS, Girolami M, Willcox K. Scaling digital twins from the artisanal to the industrial. *Nature Computational Science* 2021; 1(5): 313–320.
50. DIN . *DIN SPEC 91607:2024-11: Digital twins for cities and municipalities* . 2024. Available online: <https://dx.doi.org/10.31030/3575521>.
51. Gil S, Kamburjan E, Talasila P, Larsen PG. An architecture for coupled digital twins with semantic lifting. *Software and Systems Modeling* 2024; 1–26. doi: 10.1007/s10270-024-01221-w
52. IEEE . *IEEE 1516-2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) – Framework and Rules* . 2010. Available online: <https://standards.ieee.org/ieee/1516/3744/>.
53. Zschaler S, Mustafee N, Harper A, et al. On Simulation Reuse in Healthcare Applications. *Health Systems* 2025. Under review.
54. Bordeleau F, Combemale B, Eramo R, Van den Brand M, Wimmer M. Towards Model-Driven Digital Twin Engineering: Current Opportunities and Future Challenges. In: *Systems Modelling and Management - First International Conference, ICSMM 2020*. 1262 of *Communications in Computer and Information Science*. Springer; 2020: 43–54
55. David I, Bork D. Towards a Taxonomy of Digital Twin Evolution for Technical Sustainability. In: *Int. Conf. on Model Driven Engineering Languages and Systems Companion (MODELSC)ACM/IEEE*; 2023.
56. Rațiu CC, Assunção WKG, Herac E, Haas R, Lauwerys C, Egyed A. Using reactive links to propagate changes across engineering models. *Software and Systems Modeling* 2024. doi: 10.1007/s10270-024-01186-w
57. Giese H, Bencomo N, Pasquale L, et al. Living with Uncertainty in the Age of Runtime Models. In: *Models@run.time - Foundations, Applications, and Roadmaps [Dagstuhl Seminar 11481, November 27 - December 2, 2011]*. 8378 of LNCS. Springer; 2011: 47–100
58. Bencomo N, Götz S, Song H. Models@run.time: a guided tour of the state of the art and research challenges. *Journal Software and Systems Modeling (SoSyM)* 2019; 18(5): 3049–3082. doi: 10.1007/s10270-018-00712-x
59. Hugues J, Hristosov A, Hudak JJ, Yankel J. TwinOps - DevOps Meets Model-Based Engineering and Digital Twins for the Engineering of CPS. In: *23rd ACM/IEEE Int. Conf. on Model Driven Engineering Languages and Systems: Comp.ACM*; 2020.
60. Combemale B, Wimmer M. Towards a Model-Based DevOps for Cyber-Physical Systems. In: *Software Engineering Aspects of Continuous Development and New Paradigms of Software Production and Deployment*Springer International Publishing; 2020: 84–94.
61. Combemale B, Jézéquel JM, Perez Q, et al. Model-Based DevOps: Foundations and Challenges. In: *Int. Conf. on Model Driven Engineering Languages and Systems Companion (MODELSC)ACM/IEEE*; 2023
62. Bencomo N, Cabot J, Chechik M, et al. Abstraction Engineering. arXiv; 2024. Available online: <https://arxiv.org/abs/2408.14074>.