#8

Digital Twins

An enabler for digital transformation

Scan the QR for the animation of this chapter

Andrés Tello (RUG) and Viktoriya Degeler (RUG)

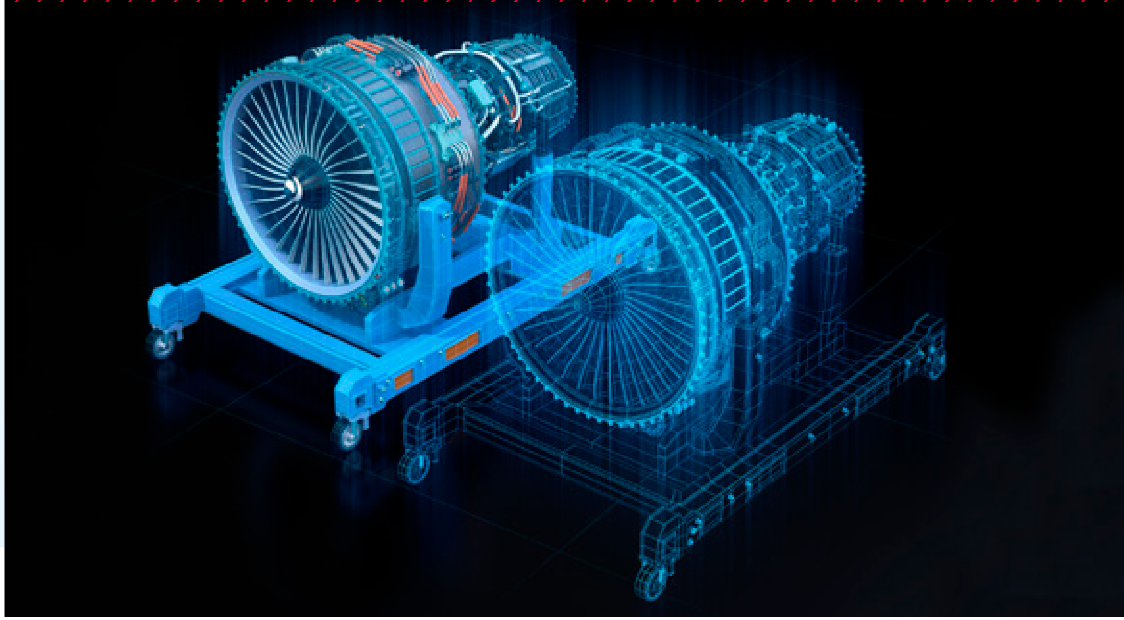
Digital Twins are a virtual representation of anything of value for an organization that create a link between the real and virtual worlds by a continuous bidirectional data/ information exchange. In this chapter we present the origins of the concept and how it evolved with the advent of new technological trends. In addition, we describe the main characteristics of a Digital Twin, the benefits of its use, and real-world examples of the usage of digital twins'. Finally, the challenges for its adoption, and the elements to be considered for managing the quality of the Digital Twin are presented to give a complete overview of this new technology.

Andrés Tello is a PhD Student at the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence at the University of Groningen. His research is focused on Knowledge Graphs for Context modeling and representation in smart environments.

Dr. Viktoriya Degeler is an Assistant Professor at the University of Amsterdam. Prior to that, she was affiliated with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence of the University of Groningen. Her research is focused on intelligent pervasive systems and context modeling and representation. She is active in promoting AI approaches in the industry, training in AI for SMEs program, and acted as an AI technical mentor for startups.

The Idea in brief

The Issue	The Response	The Bottom Line
Digital transformation fostered the evolution towards the digital representation of physical entities.	Physical entities can be represented in digital. Subsequently, optimization and algorithms applied in the digital world improve the physical entity. But also, the other way around.	Digital twins allow organizations to execute digital transformation and link the physcial and digital world.



Demystifying the Digital Twins

Digital Twins is a concept that has been around for quite a few years. Recently, it has gained a lot of attention and it was included in Gardner's and Deloitte's reports of top technological trends since 2018, for three consecutive years (Hartmann & der Auweraer, 2021). But where did the concept of Digital Twins come from and just how disruptive is it really?

The term Digital Twins came into sight in the early 2000s with the digitization of machinery and production systems in the manufacturing industry. At its origin, Grieves and Vickers defined a Digital Twin (DT) as a virtual representation of a product, comprising the physical product, the virtual representation of it, and the data and information flow between the real and virtual spaces (Grieves, 2014). Later, the concept was extended adding what they coined a Digital Twin Prototype (DTP), a Digital Twin Instance (DTI) and a Digital Twin Environment (DTE) (Grieves & Vickers, 2017). The DTP is a prototypical version of the physical product, including all the information required to *describe and produce* the physical version of it. The DTI is the virtual representation of a specific, existing, physical product. The DTE is a multi-domain physics application space for operating on Digital Twins, either for performance prediction or information querying about the state of the physical twin. In addition, in this new conceptual view of the Digital Twin there is a linking between the real and virtual spaces throughout the entire life cycle of the physical product. Hence, the DT enables the design, prototyping, testing, production, and use of its physical counterpart (Minerva, Lee, & Crespi, 2020).

The advent of new technological trends, e.g., Internet of Things (IoT), Big Data, Artificial Intelligence (AI), Machine Learning (ML), fostered the evolution of the original conception of Digital Twins. The digitization era reshaped the concept, from a virtual representation of a physical product, to a digital representation of a physical entity, a process, a service, a system, or other intangible asset (Madni, Madni, & Lucero, 2019; Shao & Helu, 2020; Stark, Kind, & Neumeyer, 2017; Tao & Qi, 2019). This broader vision of the concept, and the vast amount of data that can be acquired from the physical world today, enable the creation of Digital Twins of both, living and non-living things, even humans (El Saddik, 2018; Lu et al., 2020; Saracco, 2019). In this evolution of the concept, Academia and Industry have proposed a myriad of definitions for Digital Twins. Nonetheless, all of them are based on its original principle: a digital representation of a physical asset —understanding by asset anything of value for an organization (Malakuti et al., 2019)— and the linking of the physical and virtual spaces through data and information exchange. The virtual model in the Digital Twin includes the design documents, engineering models, simulations, data analysis and other data that describe the structure, performance, health status and maintenance history of its physical pair. Then, during the operation of the Digital Twin, the virtual model is updated in real-time with the latest data provided by the onboard and surrounding IoT sensors in the physical twin, as well as external data sources from the surrounding environment. All those data readings are used to train AI algorithms in the Digital Twin to provide the information that allows to optimize the performance of the physical twin, perform preventive or corrective adjustments, or support *managers on data/knowledge driven decision making. This continuous data-information exchange is what distinguishes Digital Twins from other similar concepts such as digital model and digital shadow. The digital model is just a digital replica of their physical pair with no data exchange happening between them. In the digital shadow, the data flow occurs in one-way from the physical to the digital object. On the contrary, in the Digital Twin the data-information flow is in both directions, and a change of state in either side can produce a change in their counterpart. This continuous bidirectional flow of data and information, along its entire life cycle, is what makes a Digital Twin a dynamic and evolving entity and not merely a high-fidelity copy. Figure 1 shows the conceptual overview of a Digital Twin.*

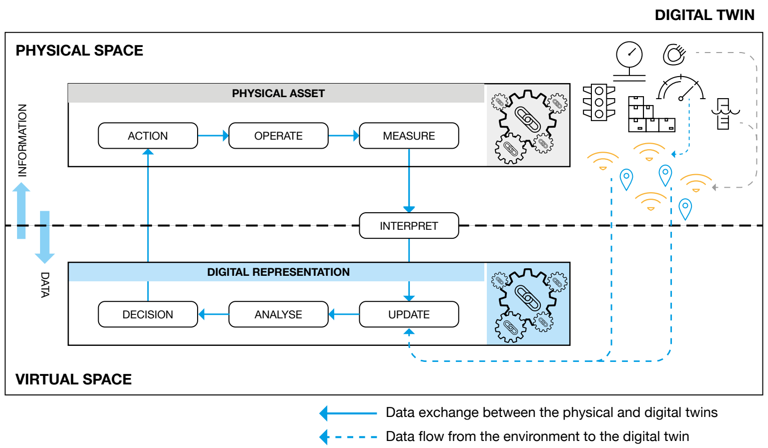


Figure 1. Digital Twin conceptual overview. (Adapted from (VanDerHorn & Mahadevan, 2021))

The different concepts and implementation details proposed during its evolution allows to identify the most important characteristics of Digital Twins. According to (Barricelli, Casiraghi, & Fogli, 2019), a Digital Twin implementation should have the following characteristics. *A seamless connection and continuous data exchange* between the physical and digital world. This is the key to capture and provide the most accurate representation of the physical counterpart. A seamless connection is what allows to keep the twins updated in real-time when a change of status occurs in either pair. A dynamic, eventually sensed, physical twin data. This is the data that describes the status of the physical asset being replicated. It is dynamic because it changes over time, and the frequency of collection depends on the use case. *A dynamic environment data* describing the surrounding environment status. The collected data should describe not only the physical twin, but also its environment. This can include data from other IoT sensors, weather, traffic, noise, and so on. Such data is used to feed prediction algorithms, or it is delivered to domain experts to support decision-making. *Communication processes* that allows the bidirectional data exchange previously mentioned, either between the physical and virtual spaces, between Digital Twins or between the Digital Twin and domain experts. *A data storage system* for descriptive, historical, and real-time generated during the entire life cycle of the Digital Twin. *Ontologies* to equip the Digital Twins with semantics which enable reasoning and inference capabilities. *Data fusion algorithms* to integrate the vast amount of data generated from heterogeneous data sources. *Continuously improving* AI which provide to the Digital Twins the predictive capabilities and insightful data analytics.

Benefits of Digital Twins

Key Points

The use of Digital Twins helps to:

- Integrating data silos
- Anticipate and prevent unwanted situations
- What-if simulations to evaluate different scenarios
- More informed decision-making

Any industry or application area can benefit from the added value associated with the adoption of digital twins . The following are the most important benefits of their use.

Integrating data/information silos (VanDerHorn & Mahadevan, 2021). Continuous bidirectional communication between the physical and real world is a characteristic that must be present in a Digital Twin implementation. In addition, this data-information exchange takes place during the whole life cycle of the implemented system. This implies that all the components that belong to the environment and are replicated by the Digital Twin should share their data within a common platform. Usually, within organizations, all the information systems, operational data, or any other datasets generated by daily processes are kept isolated, each of them serving particular use cases (e.g., finance, marketing, planning) and specific types of users (e.g., managers, operators). On the contrary, in a Digital Twin all this data is collected and shared, either to other Digital Twins, other systems, subsystems, domain experts or decision- makers that make use of it at different levels of granularity. This way, all the actions taken automatically by the Digital Twin itself, or by human users, are based on a complete picture of the environment being replicated.

Digital Twins allow to **foresee unwanted situations**. Usually, in an industrial manufacturing setting, corrective maintenance is performed after a failure occurred, causing a business downtime because of unscheduled maintenance operations. Other companies improved their practices and try to estimate the lifetime of the physical assets to apply preventive maintenance before the actual failure occur. Although with this approach the maintenance operations are programmed in advance, there is still a downtime in business operations. Hence, in the ideal scenario, the interruptions of business operations should be reduced as much as possible. Digital Twins facilitate the achievement of this goal with their predictive capabilities. AI and ML algorithms use the data gathered from the physical environment combined with the operational historical data to predict the behavior of the emulated asset. Hence, **a predictive maintenance model** in the Digital Twin anticipates a machinery failure right before it happens. This provides a balance between corrective and preventive maintenance approaches, avoiding long operation interruptions and thereby saving time and costs.

Digital Twins allow to **evaluate the effects of decisions before applying them in practice**. Simulation capabilities of Digital Twins enable testing what-if scenarios to evaluate the effects of decisions. Digital Twins simulations allow to identify unexpected scenarios, analyze the system response to them, and test actions to mitigate the unwanted behavior. All of this without affecting the real assets (Rasheed, San, & Kvamsdal, 2020). For example, the Digital Twin of a service could be used to simulate the demand and evaluate different resource utilization scenarios to identify the most optimal one. In the healthcare industry, a Digital Twin of a patient could be used to test drugs and assess their effects before to apply them in the real patient. In a smart city Digital Twin, traffic flow simulations allow to evaluate strategies to reduce traffic congestion.

Efficiency and safety are improved with the use of Digital Twins. For example, in automating repetitive tasks which are better performed by machines. In addition, risky situations can be avoided by humans by using a Digital Twin to operate dangerous environments remotely. Hence, humans focus on creative and innovative work, just monitoring and controlling the actions performed by the Digital Twin in threatening environments (Rasheed et al., 2020).

More informed decision-making is another benefit of Digital Twins utilization (Rasheed et al., 2020). IoT sensors monitoring the physical assets in real-time, historical data about the performance and operation of the replicated entity, and the power of AI and ML algorithms to perform advanced analytics, provides decision-makers with a complete view of the dynamics of the twinned environment. The Digital Twin provides the insights for data-driven decision- making. Hence, the decisions are based on precise and accurate data which reflects the whole picture of the evaluated situation.

As a result, the capabilities of Digital Twins enable **process optimization, higher efficiency, and savings**.

Digital Twins applications

Key Points

- This section describes real examples of how Digital Twins are being used in Manufacturing, Smart Cities, Health Care, Supply Chain, Construction and Education.
- These examples include Digital Twin implementations in the Netherlands and Projects being executed in the Northern Region.

While Digital Twins initially appeared in manufacturing, their value can now be seen in other areas way beyond it. The following are a few examples of how Digital Twins are being used today.

Manufacturing

Digital Twins originated and grew in the manufacturing industry. Today Digital Twins are one of the enabling technologies for smart manufacturing and Industry 4.0 (Tao, Zhang, Liu, & Nee, 2018).

In the area of Product Life Cycle Management, Digital Twins are presented as an effective approach to overcome data isolation and data duplication among the different stages of a product's life cycle, i.e., design, manufacturing and service (Tao, Cheng, et al., 2018). According Tao et al. (2018), during the design phase Digital Twins allow testing and validating the models before manufacturing the real product. In addition, inefficiencies in manufacturing arise from the lack of an optimization mechanism for resource management, the difference between the production plan and its actual execution, and inaccurate manufacturing process control. A Digital Twin Shop Floor is proposed as a solution to integrate the physical space and data generated during the product manufacturing stage. Moreover, the product service is related to the product use and maintenance phases. At this stage Digital Twins can be used to detect, understand and foresee anomalous events by means of data collected from the physical product. At SIEMENS, they realized the importance of Digital Twins in the manufacturing industry as an enabler of seamless integration of the data generated at each stage of a product's life cycle (Rosen, Von Wichert, Lo, & Bettenhausen, 2015).

Another example of Digital Twins in practice is at General Electrics, where they use it for predictive maintenance in the Power Generation, Oil and Gas industries. They have created Digital Twins of jet engines, wind-farms, off-shore oil rigs, power generation equipment, pumps, compressors, chillers, among other things, for what they have over 300 of DT blueprints. The company reports 1.5 billion US dollars on savings to their customers with the use of Digital Twins (Parris, n.d.). The automotive industry is also leveraging the potential of Digital Twins. TESLA creates a Digital Twin for each car they manufacture. Then, the factory receives real-time data about the status of the car and uses it later for automatic adjustments or reconfiguration (Barricelli et al., 2019; Schleich, Anwer, Mathieu, & Wartzack, 2017).

Smart Cities

The rapid development of sensing and communications technologies has enabled the conception and realization of a Smart City. A city where real-time data coming from IoT sensors and other data sources allows to record and monitor the state of buildings, parks, roads, bridges or any built infrastructure. This data in combination with operational data from urban- related processes allows to assess and optimize the performance of the services offered to the citizens. In this context Digital Twins of Smart Cities are used to monitor and control physical urban activities (Austin, Delgoshaei, Coelho, & Heidarinejad, 2020; Dembski, Wössner, Letzgus, Ruddat, & Yamu, 2020). The Digital Twins support decision makers in the planning and operation of complex urban-related processes, e.g., urban mobility, energy efficiency, water management, urban planning, control and risk management.

An interesting example is the implementation of a Digital Twin for the city of Herrenberg, Germany (Dembski et al., 2020). In this use case, the Digital Twin of Herrenberg comprises a 3D model of the built environment, a mathematical street network model, urban mobility and wind flow simulations, people's movements patterns, IoT sensor's data, and quantitative and qualitative data, about people's perception gathered in a collaborative system. The implementation of the Digital Twin allowed testing and evaluating the impacts of different scenarios and potential solutions for the challenges of the city (e.g., high traffic load and air pollution). In addition, having a virtual space where citizens can interact and be part in the solution of their problems, makes them more willing to get involved in public participation processes.

The city of Porto in Portugal, implemented H2Porto, a "Technology Platform for the Integrated Water Management of the Urban Water Cycle" (Bentley Systems, 2019). H2Porto is a Digital Twin that integrates over dozens of software systems, including data from water supply, wastewater drainage and treatment, stormwater drainage, surface waters, and coastal bathing water quality. The Digital Twin is used in predictive operational analytics to forecast flooding and water quality issues, for early failure identification and prescriptive actions to ensure resilience of water infrastructure, and to improve city services and responsiveness. The outcomes of the H2Porto Digital Twin are operation gains of 25%, reduction of failures in water supply of about 30%, reduction of duration of pipe bursts repairs of about 8%, improved decision-making due to real-time data access, and data stability and reliability close to 99%.

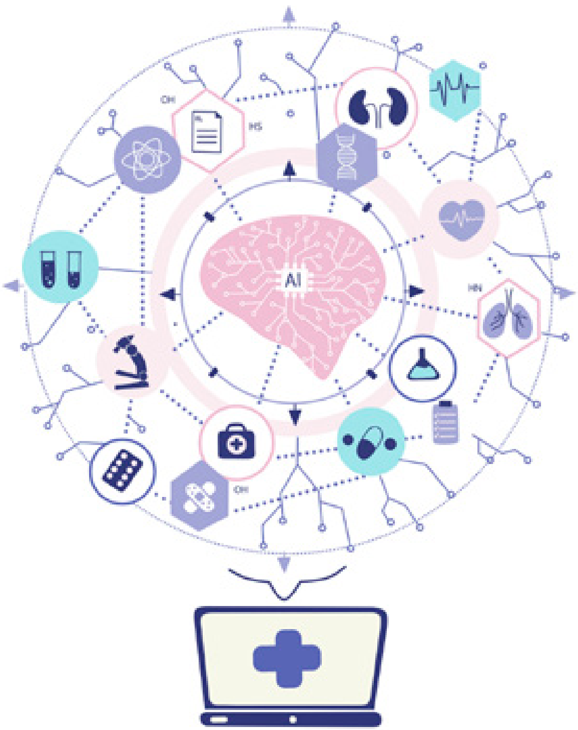
Virtual Singapore is another example of a Digital Twin that relies on 3D semantic modelling to create an exact virtual copy of the city-state Singapore (National Research Foundation of Singapore, 2019). Virtual Singapore integrates data about terrain and water, infrastructure, transportation, and even vegetation information. In addition, data about demographics, traffic, temperature, humidity, noise and light intensity is also collected in real-time by IoT sensors worn by students. Thereby, the Digital Twin of Singapore is used to support test-bedding concepts and services, planning and decision-making. For instance, Virtual Singapore is used to test the coverage areas of 3G/4G Networks. It allows to analyze transport and pedestrian movement patterns and simulate crowd dispersion in order to test evacuation protocols during an emergency event. Moreover, to prevent heat waves affect the comfort of its citizens, urban planners use the Digital Twin to assess how the placement of new buildings will affect wind flow across the city, and to evaluate measures to mitigate that. Furthermore, Virtual Singapore is leveraged by the research community through partnerships between government agencies, universities and the private sector to use the Digital Twin as an experimentation environment for innovation, research and development projects.

Amsterdam is another use case for Digital Twins that aims to become the city more livable and 'smart'. The Digital Twin is focused on monitoring traffic flows in the surroundings of the ArenA stadium. The application guides the visitors to right parking place when it is very busy. KPN will play an important role in this implementation by means of its 5G Network (KPN, n.d.).

Healthcare

The first applications of Digital Twins in healthcare were predictive maintenance and performance optimization of medical devices (Barricelli et al., 2019). Today, big companies like General Electrics and Siemens extended its use to complete coverage in the healthcare industry.

GE Healthcare, together with GE's Global Research Center, created what they called 'Hospital of the Future Simulation Suite', a Digital Twin to mimic hospital operations (GE Healthcare, n.d.). The Digital Twin allows to model behavior of patient and staff, variation in demand and supply, and patient pathways with the aim of optimizing patient flow, capital investment, and operational efficiency. Care givers and hospital administrators can test what-if scenarios that help them to determine which actions to take in a scenario-based, data-driven approach. The GE's Digital Twin implementation has helped to optimize bed use, surgical block schedule, and staffing planning to meet the dynamics of the predicted demand.



Siemens Healthineers in a partnership with the Mater Private Hospital from Dublin, created a Digital Twin of the hospital's radiology department (Siemens Healthineers, 2019). The idea of a Digital Twin to optimize the operations at radiology department came from the inability to meet the increased demand for medical scans. Growing the department's size was unfeasible due to the lack of space and the overall costs that would be incurred. In this use case, the Digital Twin leverages simulations, and it is enhanced with data coming from real-time location systems to provide a detailed view of how the hospital is working. Once again, what-if scenarios allowed to evaluate the impact of increasing patient demand within the radiology department, and how it affects on resource utilization, staff, and other areas of the hospital. The outcomes were shorter waiting time for patients, faster patient turn around, increased equipment utilization, and lower staffing costs by reducing the overtime work per day.

In addition, Siemens is revolutionizing healthcare industry with the creation of digital models of human organs, being the ultimate goal the Digital Twin of the entire human body (Siemens Healthineers, 2018). They started with a Digital Twin of the heart, enriched with deep learning algorithms to simulate the multi-scale physiological processes of this organ. Cardiologists at the University of Heidelberg are testing those algorithms, evaluating responses to cardiac re-synchronization therapy on the Digital Twin before the actual intervention in the real patient.

Moreover, the digital revolution have enabled healthcare practitioners to use Digital Twins for tailored treatments evaluation, facilitating the expansion of precision medicine. The use of Digital Twins for personalized medicine, i.e., precision medicine, goes from Digital Twins for testing and identifying the best performing drugs on specific diseases (Björnsson et al., 2020), to Digital Twins of patients for tailored nutrition guide (Gkouskou et al., 2020), and to a Digital Twin reference model for the design, development, and operation of personalized treatment management systems (Rivera et al., 2019). The ability to execute trial-error tests on computer models instead of people, and to virtualize hospitals' processes and workflows, have made Digital Twins to become a disruptive technology in the healthcare industry.

Supply Chain

One of the challenges faced by Supply Chains is the management of disruptions and its recovery mechanisms. Prior identification of disruption scenarios and nodes in the supply chain prone to fail is essential for determining proper actions and recovery mechanisms in the presence of disruptions. Model-based and data-driven approaches implemented in a Supply Chain Digital Twin (SCDT) are proposed for disruption risk management (Ivanov & Dolgui, 2020). The SCDT allows to monitor the supply chain network state. Real-time data from IoT sensors, track and trace systems, RFID systems, third-party data about natural, financial, or political risks, and historical disruption data feed disturbance prediction AI algorithms to support decision-makers to anticipate interruptions. If detected, simulations of disruptions' dynamics and the evaluation of alternative supply network topologies, enable recovery policies optimization.

Information asymmetry, in remanufacturing supply chain, refers to the fact that some actors in the chain have more or better information and the lack of data sharing upstream and downstream enterprises in the chain. The inability to obtain updated, real-time data at every stage in the supply chain hinders decision-making and process optimization. According to the study carried out in (Chen & Huang, 2020), the current solutions to this problem are not systematic and too theoretical. Thus, Digital Twins are proposed as a solution for solving information asymmetry in the supply chain of the remanufacturing industry. By concept, Digital Twins implementations involve a continuous bidirectional flow of real-time data between the physical and digital world. Therefore, having a Digital Twin of the remanufacturing supply chain imply all the data and information, generated at any enterprise in the chain, will be shared and available in real-time. It is important to point out that its applicability depends on enterprise demand and application willingness, information security, and data integration technology (Chen & Huang, 2020). However, the solution for the latter is straightforward due to the technological advancements today, e.g., IoT, Big Data, Cloud Computing.

Construction

In the era of digital transformation, construction industry is considered among the least digitized (Greif, Stein, & Flath, 2020; Leviäkangas, Paik, & Moon, 2017). To cope with the current demand within the industry 4.0, i.e., real-time data-driven decision-making, Digital Twins are a viable option to revolutionize construction industry as well.

A Digital Twin-based decision support system for non-high-tech industries (e.g., construction) is proposed in (Greif et al., 2020), to monitor and control the bulk silo dispatch and replacement processes. Currently, when a construction company requests a bulk silo, it is filled and weighted at the production plant. Then, the silo is transported to the construction site and returned to the plant after its use. The bill is calculated based on the current weight compared to the weight at delivery. On the other hand, if the company requires more material, it requests a replenishment. Then, a truck from the production plant is sent with a weighted amount of material to the construction site to refill the original silo. The bill is calculated in the same way, after the silo is returned to the production plant. There are clear limitations in the current dispatch and replenishment processes. First, the silos need to be returned to the plant for billing. They can not be used by other customers directly, even if the silo has enough material to meet those requirements. Second, the construction company is not aware of the fill level of the silo. Hence, a replenishment is requested after they realize the silo is empty. This leads to work interruption until the new material arrives to construction site. After a silo is delivered, the production plant is not aware of the usage of material at the construction site. Thus, remaining material could be stored at construction site for a long period, until the contractor decides to return the silo, possibly resulting in that material losing its good condition for further use.

The aforementioned real-world scenario from the industry raises the bar for improvements. A Digital Twin of these processes should include data about orders, customers, silos, and trucks. Of course, the implementation of such a system would require large investments, but the benefits of its implementation could be seen at short, medium, and long term (Greif et al., 2020). Simulations performed on the Digital Twin using three years of historical data showed that the costs of transportation could be reduced by 25%. Reducing transport by means of intelligent data-driven decision-making not only represents monetary benefits but also environmental gains because of the reduction in CO_2 emissions (Greif et al., 2020).

Structural Health Monitoring (SHM) is another use case for Digital Twins in construction industry. An example is a framework for creating Digital Twins of bridges, implemented using two monitored railway bridges in Staffordshire, UK (Ye et al., 2019). According to the study, there are several limitations of current practices in SHM of bridges. For instance, the difficulties for storing, processing and interpreting the available data because of the size and heterogeneity of the bridges' datasets. The heterogeneity of bridges' data management systems hinders querying relevant data. As a consequence, there is a limited interoperability across such systems because of different data formats. Thus, based on the conceptual principle of Digital Twins, they are seen as the technological tool that could enable integration, sharing, analysis and interpretation of bridges' health status data at every stage of their life-cycle.

In the Netherlands, the "Rijkswaterstaat" (Directorate-General for Public Works and Water Management) are the forerunners in the field of Digital Twins. They created a Digital Twin of the tunnel under construction that connects the A16 and A13 in Rotterdam Noord. The Digital Twin is being used to simulate an identical replica of the tunnel and anticipate possible incidents. It also allows to enhance the coordination between the stakeholders of the project (KPN, n.d.).

Education

The Digital Twins' capability of being a high-fidelity representation of its physical counterpart allows them to be used as an educational tool as well. In this context, Digital Twins are presented as a pedagogic tool to assist the learning process in certain courses at university level (David, Lobov, & Lanz, 2018a, 2018b). Leveraging the simulation capabilities of Digital Twins and their ability to provide real-time information, they are used to apply the knowledge acquired on theoretical lectures in the classroom, by carrying out practical laboratory exercises. All the exercises are performed on a virtual representation of the physical process. This allows the students to understand every detail by playing and manipulating the virtual asset, as it would be the physical one, without affecting the real operation. The practical sessions using the Digital Twin are supervised by the professor, and only after successful performance of the students, they are taken to the physical site to apply the learned skills.

In addition, the notion of "cognitive Digital Twins" allows them to be an assistive tool to enhance the way that we, humans, learn. The wealth of data available today, makes it difficult to keep the pace with the explosion and obsolescence of knowledge (Saracco, 2019). While the overwhelming amount of information makes it difficult to know what we do not know, the demand for skilled professionals pushes us to continue learning. Hence, having a Digital Twin of our learning process, featured with cognition capabilities, would be the right tool to help us to organize and command the way we acquire knowledge. As stated in (Saracco, 2019), it would depend on us to feed the Digital Twin with the most accurate information about our learning path, e.g., our curricular education, conferences attended, papers read, papers submitted and published, training courses, etc. This way, our cognitive Digital Twin could infer the knowledge we have, what is missing and how to fill those gaps. Moreover, with all that information at hand, the Digital Twin could be used for implementing customized education programs.

Digital Twins in the Northern Netherlands

Six Dutch Universities (RUG, TU/e, TU Delft, Twente, Leiden and Tilburg), under the leadership of the University of Groningen, in a joint project with twelve industrial partners develop Digital Twins if High-Tech Systems. The Digital Twin are used to simulate and monitor those systems function and anticipate unwanted behavior using AI-based predictions. The Digital Twins can be used by different industries with the aim to accelerate time-to-market or enhance production systems (Innovatiecluster high tech systems Drachten, n.d.).

The challenges for its adoption

Key Points

- This section describes the challenges that should be taken into account before starting a Digital Twin implementation.
- In addition, it depicts the efforts of local and European organizations to support Small to Medium Enterprises to embrace Digital Twin technologies.

Although the benefits of Digital Twins are clear, their implementation requires a deep understanding of the burden associated with its adoption. Recent developments on Digital Twins are the result of the advancements of other technologies such as AI and IoT. Therefore, the challenges for a successful implementation and widespread adoption of Digital Twins go hand in hand with those of its enabler technologies.

The potential of Digital Twins goes beyond storing all available data about the past and present of its physical pair, but the capacity to leverage such data to anticipate the future, i.e., to get insights for supporting data-driven decision-making. Data Analytics plays an important role to achieve this goal. Insightful data analytics relies on complex and computationally demanding AI deep learning algorithms. Hence, the data collected from the real-world by the Digital Twin is fed to AI and ML algorithms to find patterns and extract knowledge from it.

Unfortunately, all that processing power comes at a price, a sometimes highly-expensive IT infrastructure. State-of-the-art AI deep learning algorithms run on GPUs, which can costs from 1,000 to 10,000 USD (Fuller, Fan, Day, & Barlow, 2020). Besides computational power, it is important to think of strategies to leverage the infrastructure already in place combined with the new IoT technology available in the market. Although today there are several options for cheap IoT sensors, changing all the IT infrastructure may imply high additional costs. Then, the challenge is how to integrate old IT infrastructure and legacy systems with current IoT devices and applications. This shows that the initial investments for running a Digital Twin, and exploit it at its fullest capacity, could be a decisive factor for a go-no-go decision in a small or even medium-sized enterprise.

Today, companies can decrease its IT infrastructure costs by contracting cloud service providers. Current cloud solutions offer storage, computation capacity, relational and NoSQL database services, developer tools, machine learning pipelines, virtual servers, or even entire data centers according to the needs of their clients. However, a common concern when opting for such solutions is the mistrust of giving the control of their data to third party companies and that privacy and security over its information assets could be compromised.

Then, another challenge of Digital Twins' implementation is to guarantee **privacy and security** of the data collected. Regarding data analytics and AI, the challenge falls on Law and Regulations' compliance. The power of AI has disrupted almost every industry today, but together with the success comes uncertainties. To what degree the results obtained through AI is ethical and fully compliant with all local, regional, national and international regulations. For instance, the General Data Protection Regulation (GDPR), the Europe's new data privacy and security law, is "the toughest privacy and security law in the world" (GDPR.eu, n.d.). Therefore, designing and implementing a Digital Twin compliant with such regulations may impose significant challenges. Federated learning is proposed as a possible solution to overcome privacy and security issues (Fuller et al., 2020). This is a concept proposed by Google, where machine learning models are trained in a distributed fashion. Data owners share the model, but its data is kept locally during training. At the end, the optimized ML model is shared back to the users. Hence, data analytics in the Digital Twin could be implemented using federated learning to avoid data sharing beyond the place where it is being created.

Trust is another concern that comes hand in hand with privacy and security. AI is a relatively new technology and there is still a lot of debate around the possible negative effects of its use. For example, the case in the UK where a machine learning algorithm used to profile visa applicants was criticized for being "racist", created skepticism about AI credibility. After a legal challenge against such system, the UK's Home Office stopped using the algorithm in August 2020 (BBC news, n.d.). The "Mirai" botnet, considered the largest DDoS attack recorded until now, infected about 15 million of IoT devices by attempting to login, by brute force, the devices configured with default usernames and passwords (Vishwakarma & Jain, 2020). The intent of these examples is to show that trust on Digital Twins depends on how the organization and end-users are aware of the challenges of its implementation and use. From the organization's point of view, it has to ensure that their algorithms comply with privacy and security regulations. The European GDPR requires that automated decision-making should provide meaningful explanation of the logic involved in the process. Thus, model validation and Explainable AI could raise trust on the insights provided by the Digital Twin. Understanding where the outputs of ML algorithms come from guarantees compliant systems and can make decision-makers appealing to exploit the Digital Twins potential.

Connectivity and **useful data** are other major challenges to be addressed when designing and implementing Digital Twins (Fuller et al., 2020; Rasheed et al., 2020). The key for a successful implementation of a Digital Twin is the readiness and efficiency of the bidirectional communication mechanism between the physical and virtual spaces. The monitoring and actuating capabilities of the Digital Twin rely on real-time data obtained from the physical world. The accuracy of the predictions, simulations and the overall performance of the Digital Twin depends on the timeliness and quality of the data used for analysis. The large number of IoT devices that could be deployed on a Digital Twin environment generates Big Data. It is "Big" not only because of the amount of data generated, but its large variety and high generation rate. Sensors' data is faulty which requires pre-processing to ensure its quality. Latency in communication can hinder the timeliness needed for Digital Twins data provisioning. This, added to unavoidable power outages, and unintentional software and deployments errors, poses a challenge for Digital Twins' implementation. The type of data collected, the data collection rate, and the number and placing of sensors depend on the application domain. Too few data points can produce inaccurate predictions while too many can originate redundant and unnecessary details causing transmission bottlenecks. Therefore, implementation of Digital Twins requires a thoroughly planned connectivity and data collection strategies, guaranteeing that the virtual model evolves together with the physical one and vice versa.

Currently, Digital Twins are being used for solving real-world problems in different industries. However, each project, vendor or implementer of Digital Twins propose their own models and architectures. There is **a lack of standards** for designing, modeling and implementing Digital Twins (Fuller et al., 2020; VanDerHorn & Mahadevan, 2021). The ISO/DIS 23247-1 is an advancement on standardization of Digital Twins. However, this standard is focused specifically on Digital Twins for manufacturing. Another effort towards Digital Twin standardization is the Digital Twin Consortium, which brings together academia, government and industry to provide a common vocabulary, architectures, and security and interoperability directives for Digital Twins implementation (Object Management Group, n.d.). Their members include companies, leading in their respective industries e.g., GE Digital, Bentley Systems, Microsoft, Dell, working on technical guidelines and taxonomies, reference frameworks, and requirements for new standards on Digital Twins.

The aforementioned challenges might seem unmanageable for Small to Medium Enterprises (SME). However, there are regional and European initiatives to enable SME to embrace Digital Twins technologies. TNO combines data, models, and business processes to create a digital replica of a physical life cycle of an asset. TNO provides guidance to SME offering them an assessment tool to identify their digitization needs. Then, TNO support goes from creating the digital models SME requires, assisting SME in the data collection process, or any other tasks related with the implementation of Digital Twins (TNO innovation for life, n.d.). DIGITbrain is a European initiative to give SME access to Digital Twin technology. The aim is to offer AI-based Manufacturing-as-a-Service to SME to simulate any stage of a manufacturing process. This will allow SME to make predictions and to run what-if scenarios to make data- driven decisions (EGI.eu, n.d.).

The challenges for Digital Twins adoption presented in this section provide a general overview of the details that need to be taken into account before embracing on a Digital Twin implementation project. Users and implementers need to be conscious that Digital Twins should not be used to merely follow a trend, but a deep understanding of the pros and cons of this technology is required to ensure a successful Digital Twin application.

Managing the quality of Digital Twins

Key Points

This section describes the main factors to ensure the quality of the Digital Twin:

- Sensors composition and coverage
- Model creation and configuration
- Machine Learning pipelines
- Hyperparameter tuning
- Model quality optimization

Assessing the quality of the Digital Twin is one of the essential questions faced by any system designer and developer. There are many parameters that can compose the quality of the Digital Twin. One can think of expressiveness – how simplified is the digital version of the physical environment? Does it capture well all the intricacies, complexities, chaoticism of the physical world with minute precision, or does it provide only a rough model for a couple of variables? Another potential quality parameter is the computational efficiency – how many computational resources do we need to keep the real-time simulation link up and running?

But arguably the most important quality parameter of a Digital Twin for simulations and scenario predictions is the degree of trust we can put into the system and its risk of failing to model the real physical conditions correctly. In other words, the Digital Twin's accuracy.

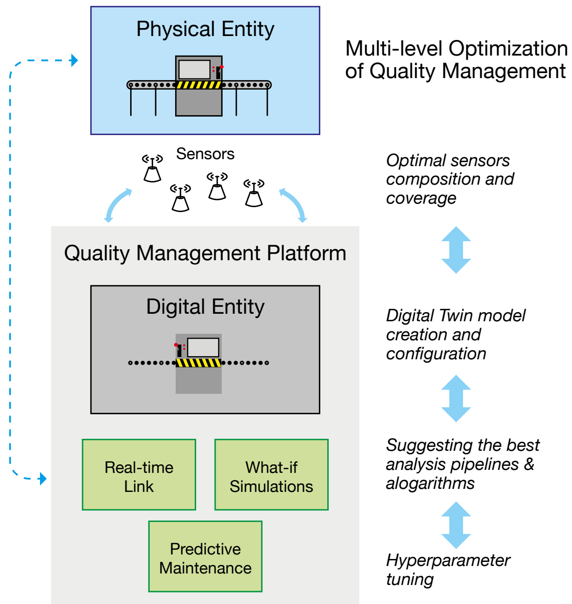


Figure 2. Quality Management platform and multi-level optimization

Optimal sensors composition and coverage

This is arguably the most important level of Digital Twin construction. The physical world is perceived through the collection of sensors, and the final quality and accuracy of the twin is hard-constrained by the quality of coverage that these sensors provide. Sensors can be of diverse nature: we are talking not just about the standard IoT sensors, such as passive-infrared for movement detection, sound sensors, videocameras, pressure, air quality sensors and similar, but also about other ways to collect information, such as digital room schedules, device specifications. In the absence of a proper device to measure certain features, the system may even be constructed in such a way that allows to ask humans, operating the system, to provide missing information. While this may be undesirable, due to the strive for full automation, for some cases this may be the most practical way to gather the required knowledge.

The original physical entity is called fully observable, when existing sensor coverage can detect all potential changes in the state of the entity and its surroundings. While, at the first glance, it may look like it is necessary to have a fully observable system to create a perfect Digital Twin, in fact, in many cases full observability is not required. For example, think of a system that wants to keep the temperature in a room within a comfortable range, but only when there are people inside the room. The Digital Twin of such a system does not need to know the exact number of people in it for the perfectly optimal thermostat control. It only has be able to distinguish between zero (no one is present) and non-zero (at least one person is present) states. On the other hand, if we are to make the system a bit more complex, and require it to control the air quality as well, the number of people in the room becomes a meaningful variable that will impact potential control actions of the system. Therefore, in this case additional sensors to measure the number of people will be required.

In the work by Degeler and Curry (2014), a system with partial observability of the physical environment is considered. It is shown that in many situations it can be computed in real-time whether the available incomplete information is enough to decide on the actions that have to be performed, and often optimal decisions can still be made even with gaps in understanding of the environment. On the other hand, there are situations when missing information becomes critical to the correct decision making. The system can then create a decision tree that is used to guide humans to provide missing data or to perform certain actions.

Model creation and configuration

Once the sensor coverage is established, the next level of managing quality of the Digital Twin is *creating and configuring* a well-suited model of the physical entity. In general, any model should provide a certain representation of the states in which the physical object can be present and transitions between those states. The most common representation of the state is with a *feature vector* and transitions is with a change vector, that can be added to the initial state feature vector to result in the new feature vector, that describes the state after the transition (the event).

Representation with *ontologies* is also very common for Digital Twin environments. Unlike standard structured data representation in a table-like manner (with a collection of data points each having a set of the same features), knowledge representation with an ontology captures a collection of entities and relations between those entities, and represents them in a graph-like manner. The dataset in such a representation is often called a *knowledge graph*. A standard ontological statement consists of three parts: subject–relation–object. For example, Figure 3 shows, how basic ontological relations can look like for an office activity recognition scenario (Nguyen et al., 2013). Web Ontology Language (OWL) provides a common standard notation for ontological representation (McGuinness, Van Harmelen, 2004).

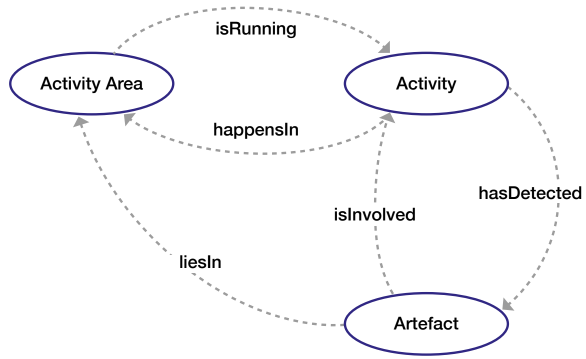


Figure 3. Office activity basic ontology

Machine learning pipelines

Whatever model representation is chosen, this data should be usable by the Digital Twin system to simulate the current physical world and, often, to predict possible future states and model what-if scenarios. For that, most commonly, a variety of different machine learning algorithms is employed, such as logistic regression, support vector machines, (deep) neural networks, clustering and pattern detection, etc. For Digital Twins that control the physical entity, machine learning steps are often followed by additional reasoning steps to decide on the best course of actions, commonly modeled as a constraint satisfaction problem, or an automated AI planning problem.

The whole reasoning process can be split into a number of reasoning steps, which together are normally called a *pipeline*. Think of the most common pipeline scenario: (1) Raw data is collected from sensors. Each sensor has its own data processing step, which may include preliminary data cleaning, error detection, data transfer to the central hub, data transformations, etc. (2) Data from different sources is collected together and further processed to form a coherent view of the physical world. This may include consistency analysis, feature extraction, further cleaning and logical inference. The data is transformed into the chosen model representation. (3) Chosen machine learning algorithms are applied to the data. This pipeline step can normally be further split into separate training and detection or prediction steps. It is also common to employ several different machine learning algorithms, in which case each of them can be represented as their own pipeline step. (4) Once the trained machine learning algorithm is applied, the system can make decisions on the possible execution of actions. This pipeline step can involve something as simple as showing the results of the prediction in a dashboard, but can also contain further complex reasoning steps, that finish with passing the execution of the decided steps to the IoT actuators.

Hyperparameter tuning

The final level of improving the quality of the Digital Twin is the traditional machine learning *hyperparameter* tuning approach. This allows us to assess, under the current conditions, what the performance metrics (accuracy, f1-score, mean square error, and so on) of the chosen predictive approach are, as well as find the best configuration of parameters for the chosen predictive approach. Among the most commonly chosen hyperparameter tuning methods are grid search, randomized parameter estimator, and Bayesian optimization approach applied to the parameters feature vector.

However, traditional machine learning hyperparameter tuning operates within a very constrained world of possibilities. In particular, it assumes the permanency in the data collection, and dataset quality and features.

Model quality optimization: bringing all levels together

When dealing with the construction of a Digital Twin, a typical system has many degrees of freedom, normally coming from the ability to change the configuration of sensors, add new physical sensors, change the representation, change the simulation parameters of the Digital Twin, and so on. Think of a case, where, after extensive data collection and predictive algorithm construction, a hyperparameter tuning is performed that allows to improve accuracy by 2%. Meanwhile, a new vibration sensor should have been added to a part of the mechanism that is not being adequately monitored, and adding data readings from that sensor to the predictive feature set would have resulted in a considerably larger increase in the accuracy of failure detection. Moreover, the variable data availability may affect the choice of the Digital Twin model, and by updating the model, we may need to re-evaluate the best suited data science algorithm used for a given purpose.

Therefore, it is important to design a system that optimizes the quality of the Digital Twin model on several levels, as shown in Figure 2. Starting at the physical level, the system optimizes the placement of sensors to minimize their initial installation and maintenance costs, while maximizing the observability of the physical units and surrounding environment. At the next level, the Digital Twin model itself is constructed based on the incoming data sources, and the real-time link is created to its physical counterpart. At the third level, the constructed Digital Twin model and the collected dataset are used to find the best combination of data science algorithms for a given question (e.g. for the prediction of the future state of the unit or its part), while at the last level, the hyperparameter tuning of the chosen algorithm is performed.

Important to note, that it is not a strictly unidirectional top-to-bottom optimization system, and decisions done at every level are back propagated, because they can affect the optimality of the solution on previous levels. As such, it is a close-loop system that continuously re- evaluates decisions done at every level and provides suggestions for real-time re-configuration when necessary.

Digital Twins: the future is now

Digital Twins have allowed to materialize what was envisioned and seen as science fiction just a few years ago. Their original conception as a model of a physical device evolved to a “living” digital representation of anything of value for an organization, being either a device, a process, a service or the entire organization itself. Today, digital twins are considered the key for digital transformation and organizations that want to embrace the wave of industry 4.0 are betting on them. Digital twins have disrupted the way organizations operate and function today. Their benefits have popularized their use in almost every industry, from manufacturing, to aerospace, to healthcare, education and so forth. Assisting end-users from design to monitoring, analysis and prediction, they provide insights and support decision-making based on real-time data from the real world.

Technologies like IoT and AI are shaping the way digital twins are implemented. Thus, the benefits and challenges of a digital twin implementation go hand in hand with the benefits and challenges of those other technological trends. Organizations and end-users should be aware and acquire deep understanding of the pros and cons of these technologies in order to be able to engage as soon as they can with a digital twin implementation project. With the ubiquitous nature of these technologies nowadays, the future is now.

References

- Austin, M., Delgoshaei, P., Coelho, M., & Heidarinejad, M. 2020. Architecting smart city digital twins: combined semantic model and machine learning approach. **Journal of Management in Engineering**, 36(4): 4020026.
- Barricelli, B. R., Casiraghi, E., & Fogli, D. 2019. A survey on digital twin: Definitions, characteristics, applications, and design implications. **IEEE Access**, 7: 167653–167671.
- BBC news. n.d. **Home Office drops `racist' algorithm from visa decisions**. https://www.bbc.com/news/technology-53650758.
- Bentley Systems. 2019. **Cities Leverage Digital Twin Initiatives to Advance BIM and GIS Improving the Quality of Life for Their Constituents**. https://www.bentley.com/en/about- us/news/2019/october/21/1026-ai-cities-advance-digitally-to-support-smart-city- initiatives.
- Björnsson, B., Borrebaeck, C., Elander, N., Gasslander, T., Gawel, D. R., et al. 2020. Digital twins to personalize medicine. **Genome Medicine**, 12(1): 1–4.
- Chen, Z., & Huang, L. 2020. Digital twins for information-sharing in remanufacturing supply chain: a review. **Energy**, 119712.
- David, J., Lobov, A., & Lanz, M. 2018a. Learning experiences involving digital twins. **IECON 2018- 44th Annual Conference of the IEEE Industrial Electronics Society**, 3681–3686.
- David, J., Lobov, A., & Lanz, M. 2018b. Leveraging digital twins for assisted learning of flexible manufacturing systems. **2018 IEEE 16th International Conference on Industrial Informatics (INDIN)**, 529–535.
- Degeler, V., & Curry, E. 2014. Human-assisted rule satisfaction in partially observable environments. 2014 **IEEE 11th International Conference on Ubiquitous Intelligence and Computing**, 171–178.
- Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., & Yamu, C. 2020. Urban digital twins for smart cities and citizens: the case study of Herrenberg, Germany. **Sustainability**, 12(6): 2307.
- EGI.eu. n.d. **DIGITbrain**: giving SMEs access to digital twins. https://www.egi.eu/news/digitbrain-giving-smes-access-to-digital-twins/.
- El Saddik, A. 2018. Digital twins: The convergence of multimedia technologies. **IEEE Multimedi**a, 25(2): 87–92.
- Fuller, A., Fan, Z., Day, C., & Barlow, C. 2020. Digital twin: Enabling technologies, challenges and open research. **IEEE Access**, 8: 108952–108971.
- GDPR.eu. n.d. **What is GDPR, the EU's new data protection law?** https://gdpr.eu/what-is-gdpr. GE Healthcare. n.d. Digital Twin. https://www.gehccommandcenter.com/digital-twin.
- Gkouskou, K., Vlastos, I., Karkalousos, P., Chaniotis, D., Sanoudou, D., et al. 2020. The "Virtual Digital Twins" concept in precision nutrition. **Advances in Nutrition**, 11(6): 1405–1413.
- Greif, T., Stein, N., & Flath, C. M. 2020. Peeking into the void: Digital twins for construction site logistics. **Computers in Industry**, 121: 103264.
- Grieves, M. 2014. Digital twin: manufacturing excellence through virtual factory replication. **White Paper**, 1: 1–7.
- Grieves, M., & Vickers, J. 2017. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. **Transdisciplinary perspectives on complex system**s: 85–113. Springer.
- Hartmann, D., & der Auweraer, H. 2021. Digital twins. **Progress in Industrial Mathematics: Success Stories: The Industry and the Academia Points of View**, 5: 3.
- Innovatiecluster high tech systems Drachten. n.d. **Northern Netherlands leader in breakthrough technology Digital Twins**. https://www.icdrachten.nl/en/nieuws/noord- nederland-koploper-in-doorbraaktechnologie-digital-twins/.
- Ivanov, D., & Dolgui, A. 2020. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. **Production Planning \& Control**, 1–14.
- KPN. n.d. **Digital twins: de volgende stap in digitalisering**. https://www.kpn.com/zakelijk/blog/digital-twins-de-volgende-stap-in-digitalisering.htm.
- Leviäkangas, P., Paik, S. M., & Moon, S. 2017. Keeping up with the pace of digitization: The case of the Australian construction industry. **Technology in Society**, 50: 33–43.
- Lu, Y., Liu, C., Kevin, I., Wang, K., Huang, H., et al. 2020. Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. **Robotics and Computer-Integrated Manufacturing**, 61: 101837.
- Madni, A. M., Madni, C. C., & Lucero, S. D. 2019. Leveraging digital twin technology in model- based systems engineering. **Systems**, 7(1): 7.
- Malakuti, S., Schmitt, J., Platenius-Mohr, M., Grüner, S., Gitzel, R., et al. 2019. A four-layer architecture pattern for constructing and managing digital twins. **European Conference on Software Architectur**e, 231–246.
- McGuinness, D. L., Van Harmelen, F., 2004. OWL web ontology language overview. **W3C Recommendation**, 10(10): 2004.
- Minerva, R., Lee, G. M., & Crespi, N. 2020. Digital twin in the IoT context: a survey on technical features, scenarios, and architectural models. **Proceedings of the IEEE**, 108(10): 1785– 1824.

- National Research Foundation of Singapore. 2019. Virtual Singapore. https://www.nrf.gov.sg/programmes/virtual-singapore.
- Nguyen, T. A., Degeler, V., Contarino, R., Lazovik, A., Bucur, D., et al. 2013. Towards context consistency in a rule-based activity recognition architecture. **2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computin**g, 625–630.
- Object Management Group. n.d. **Digital Twin Consortium - About us**. https://www.digitaltwinconsortium.org/about-us/.
- Parris, C. n.d. What is a Digital Twin? **GE Digita**l. https://www.ge.com/digital/blog/what-digital- twin.
- Rasheed, A., San, O., & Kvamsdal, T. 2020. Digital twin: Values, challenges and enablers from a modeling perspective. **IEEE Access**, 8: 21980–22012.
- Rivera, L. F., Jiménez, M., Angara, P., Villegas, N. M., Tamura, G., et al. 2019. Towards continuous monitoring in personalized healthcare through digital twins. **Proceedings of the 29th Annual International Conference on Computer Science and Software Engineerin**g, 329– 335.
- Rosen, R., Von Wichert, G., Lo, G., & Bettenhausen, K. D. 2015. About the importance of autonomy and digital twins for the future of manufacturing. **IFAC-PapersOnLine**, 48(3): 567–572.
- Saracco, R. 2019. Digital twins: Bridging physical space and cyberspace. **Computer**, 52(12): 58– 64.
- Schleich, B., Anwer, N., Mathieu, L., & Wartzack, S. 2017. Shaping the digital twin for design and production engineering. **CIRP Annal**s, 66(1): 141–144.
- Shao, G., & Helu, M. 2020. Framework for a digital twin in manufacturing: Scope and requirements. **Manufacturing Letters**, 24: 105–107.
- Siemens Healthineers. 2018. Exploring the possibilities offered by digital twins in medical technology. **Siemens Healthcare GmbH**. https://cdn0.scrvt.com/39b415fb07de4d9656c7b516d8e2d907/1800000005899262/d3fa3e835e2d/Exploring-the-possibilities-offered-by-digital-twins-in-medical- technology_1800000005899262.pdf.
- Siemens Healthineers. 2019. The value of digital twin technology. **Siemens Healthcare GmbH**. https://cdn0.scrvt.com/39b415fb07de4d9656c7b516d8e2d907/1800000007115634/143ec96042c1/Siemens-Healthineers_Whitepaper_Digital-Twin- Technology3_1800000007115634.pdf.
- Stark, R., Kind, S., & Neumeyer, S. 2017. Innovations in digital modelling for next generation manufacturing system design. **CIRP Annals**, 66(1): 169–172.
- Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., et al. 2018. Digital twin-driven product design, manufacturing and service with big data. **The International Journal of Advanced Manufacturing Technolog**y, 94(9): 3563–3576.
- Tao, F., & Qi, Q. 2019. **Make more digital twin**s. Nature Publishing Group.
- Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. 2018. **Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics**, 15(4): 2405–2415.
- TNO innovation for life. n.d. DIGITAL TWINNING. https://www.tno.nl/en/focus-areas/industry/roadmaps/smart-industry/digital-twinning/.
- VanDerHorn, E., & Mahadevan, S. 2021. Digital Twin: Generalization, characterization and implementation. **Decision Support Systems**, 113524.
- Vishwakarma, R., & Jain, A. K. 2020. A survey of DDoS attacking techniques and defence mechanisms in the IoT network. **Telecommunication Systems**, 73(1): 3–25.
- Ye, C., Butler, L., Bartek, C., Iangurazov, M., Lu, Q., et al. 2019. A digital twin of bridges for structural health monitoring. **12th International Workshop on Structural Health Monitoring 2019.**