Oil and Gas digital twins after twenty years. How can they be made sustainable, maintainable and useful?

David B. Cameron1 Arild Waaler1 Tiina M. Komulainen2

1SIRIUS Centre, Department of Informatics, University of Oslo, Norway, {davidbc,arild}@uio.no
2Faculty of Technology, Art and Design, Oslo Metropolitan University, Norway, timako@oslomet.no

Abstract

The digital twin offers a potentially powerful way of using simulation to support business and change the way industrial operations are done. The idea of the digital twin is not new but recent changes in information technology make implementation of digital twins a natural next step in the application of simulation technologies. Simulation practitioners will find that their models are increasingly embedded in complex systems that combine simulations with operational data to solve a business problem. However, the successful adoption of this approach is challenging. This paper asks the question: “How can digital twins be made sustainable, maintainable and useful?”. We focus primarily on the development of twins in the oil and gas industry. Most academic work in this area has been done in the manufacturing industries. We review this literature and propose a simple model of digital twins. This allows us to identify challenges with current implementations and propose a research agenda that will allow future twins to be sustainable, maintainable and usable.

Keywords: digital twin, semantics, integration, on-line simulation, Industrie 4.0, hybrid analytics.

1 Introduction

The digital twin is one of the buzzwords of the last year. For example, Gartner group identified digital twins as one of the top ten strategic technology trends for 2017 (Panetta, 2016) and 2018 (Panetta, 2017). As Gartner themselves note, the idea of a digital twin is not new, as the discipline of engineering simulation is at least thirty years old. Klostermeier, Haag & Benlian (2018) attribute the twin concept to researchers in NASA, with roots back into the Apollo program. They also note that the digital twin may be seen as a fourth milestone in the application of simulation technology, as shown in Figure 1. This model is useful, but it obscures work done in the last decades where simulations were aligned with operational data to solve localized control and monitoring tasks.

Figure 1. Milestones in simulator application (from Klostermeier et al. (2018))

For example, in the oil and gas sector, it is now twenty years since the first implementation of an on-line flow assurance system (Hyllseth & Cameron, 2003). This provided a digital twin of a multiphase pipeline from the Oseberg field in the North Sea.

As simulation and automation engineers, we should welcome this new, high level focus on using simulation to support operational decisions. However, we see that there are important issues related to sustainability, maintainability and usability that must be addressed if digital twins are to realize their potential in industrial operations. This paper presents these challenges and proposes a program of research, standardization and best practice that addresses them.

We propose the idea of a semantic backbone that supports the integration of several different types of digital twin application around a shared understanding of a facilities design. To be successful, this backbone will need to find a pragmatic balance between comprehensiveness and maintainability. It will also need to build on standards in a way that prevents reinvention of the wheel and allows modular construction of semantic models.

A successful digital twin also requires the correlation of structured data from designs, measurements and simulations with unstructured data from logs and documents. In addition, there remain many issues to be resolved around hybrid modelling: the linkage between data science - applied statistics and empirical modelling - and physical and engineering simulations. Our conclusion is that useful digital twins require the building of an interdisciplinary collaboration between computer scientists, data scientists and their engineering and operational colleagues.
2 Materials and Methods

This paper builds on the authors’ own and diverse experiences in the field of integration of complex technical systems and online simulation. Each author brings a different perspective to the problem. One of us is a computer scientist, while the other two are chemical engineers with experience from developing and implementing on-line simulation systems. The analysis is supported by a literature search performed around Industrie 4.0 and digital twins (in English and German) using ScienceDirect, OnePetro and Web of Science, with a focus on recent papers.

Digital twins have been identified as a core research topic by the SIRIUS Centre for Research-Based Innovation. This centre held a workshop on digital twins, with participants from oil companies, system vendors and consultancies on 20th March 2018. This paper is a partial summary of discussions at this meeting.

3 What is a Digital Twin?

3.1 Defining a Digital Twin

As noted above, the digital twin term had its origin in the American aerospace industry. Thus, the Defense Acquisition University defines a digital twin to be “An integrated multi-physics, multi-scale, probabilistic simulation of an as-built system, enabled by Digital Thread that uses the best available models, sensor information, and input data to mirror and predict activities/performance over the life of its corresponding physical twin.” (DAU, 2018a). This definition references another concept, namely the digital thread. This is defined as “An extensible, configurable and component enterprise-level analytical framework that seamlessly expedites the controlled interplay of authoritative technical data, software, information, and knowledge in the enterprise data-information-knowledge systems, based on the Digital System Model template, to inform decision makers throughout a system’s life cycle by providing the capability to access, integrate and transform disparate data into actionable information.” (DAU, 2018b). This definition then points to a specific data model, namely the Digital System Model, which is “A digital representation of a … system, generated by all stakeholders that integrates the authoritative technical data and associated artefacts which define all aspects of the system for the specific activities throughout the system lifecycle” (DAU, 2018c).

We can work this definition into the framework shown in Figure 2. In this figure we use concepts from the oil and gas industry to identify the parts of a twin.

Figure 2. A conceptual framework for a digital twin, with concepts from the process industries.

Three types of data must be coordinated in a digital twin: (1) measurements of a system and its surroundings, (2) a description of the system’s configuration and construction (the asset configuration) and (3) simulations and analyses of the system. We combine these three data sources to produce information about the system state that is useful for one or more users of the system.

For an oil platform or process plant, the asset configuration is contained in a Life-cycle Information (LCI) system. This allows a digital twin to exploit the virtual reality possibilities of a detailed 3D geometric model. Measurements come from the facility’s control and automation system, specialized monitoring systems and the laboratory information management system. Many types of simulation and analysis are possible. We need to ask, what are we simulating? Are we interested in structural integrity? Then we need a structural simulation. Are we interested in process behaviour? Then we need a dynamic process simulator. Are we interested in the state of rotating equipment? Then we need an analysis tool that finds events in vibrational, temperature and tribology data.

3.2 Digital Twins in Oil and Gas

Explicit references to digital twins in the oil and gas industry are few and are often high-level, non-technical presentations. This may be due to the term only having come into use in the last few years, as a commercial term used to transfer technology from the “digital sector”. Poddar (2018) defines a twin as a “virtual and simulated model or a true replica of a physical asset”. What a true replica is, is left undefined. A list of design, operations and optimisation applications is given without examples and commercial integration platforms are listed. Sharma et al. (2017) give a longer but similar presentation. They describe a twin of a supply boat that can be used to optimise operations and maintenance. They also note

https://doi.org/10.3384/ecp181539

Proceedings of The 59th Conference on Simulation and Modelling (SIMS 59), 26-28 September 2018, Oslo Metropolitan University, Norway
that a twin can support real-time barrier analysis for risk assessment.

A recent textbook on digital oil and gas technologies (Carvajal et al., 2018) does not mention digital twins. However, the concept is implicit in the on-line use and updating of an integrated asset model (p236ff) and in introducing the not-yet-realized concept of a closed loop asset model (p336). They point out challenges in implementing such a framework. We will return to this in section 4.

An early description of on-line integrated asset surveillance is given by Deaton et al. (2007). Their list of challenges is still valid: disparate data sources, non-integrated and unfiltered data. They note that data problems can be overcome, but a successful system requires embedding in the operators work processes.

Ten years later, Sankaran et al. (2017) describe Anardarko Petroleum’s implementation of an integrated performance and optimization system that integrates control data and telemetry with simulation and optimization to support operational decisions. They noted that care and system design was needed to ensure that the application remained scalable and maintainable.

Digital twins for specific disciplines, such as structural and marine performance of floating platforms (Renzi et al., 2017), drilling fluids in a well (Mayani et al., 2017) and flow assurance in a gas pipeline (Lunde et al., 2013) emphasise further that the idea of the digital twin is not new. What is new is the idea of the comprehensive, interdisciplinary twin.

### 3.3 Digital twins in manufacturing

As noted above, the digital twin had its origin in manufacturing, in particular aerospace. It is therefore worth looking at the literature from this domain in considering how digital twins can work in the oil and gas domain. Digital twins in manufacturing focus on a product and its lifecycle, from concept generation and to recycling or disposal (Tao et al. 2017). A digital twin can make a product smart, personalized, sustainable, green, and service-oriented in an interconnected and interactive way (Tao and Zhang 2017). In oil and gas, we rather focus on the asset and its lifecycle, from concept through engineering, procurement and construction to operations and, eventually decommissioning.

The data for a product twin comes from the design and from the production process. Production data includes equipment states, materials and work-in-process. It is collected in real-time from RFID tags, sensors, actuators, controllers, lasers, vision systems, scanning systems and coordinate measuring machines. (Tao and Zhang 2017, Grieves 2014).

This product focus is supported by product life cycle management (PLM) systems. This approach was first proposed by Grieves in 2003 (Mayani et al., 2018; Grieves 2014; Grieves and Vickers 2017).

*Industrie 4.0* is a German initiative in developing and applying cyber-physical systems in manufacturing (Lu 2017, Trappey et al. 2017). The digital twin is a core component in these systems (Uhlmann, Hohlweiler, and Geisert 2017, Parrott and Warshaw 2017). These authors note that two types of twins can be built: a product twin, that simulates the product’s form and function, and a process twin that simulates how a product is made. These twins converge for the digital twin or an oil and gas platform.

According to Tao et al. (2017), a digital twin is a real-time reflection of a physical system in virtual space, as per the DAU definition above. It can interact and converge with historical data, interacts seamlessly with real-time data and updates its models continuously (self-evolution). In the conceptual design phase, a digital twin can give quick overview of the different product-related data and customer’s feedback, enable communication between customers and designers, and thereby, help the designers to determine the product functionalities. At the detailed design phase, the digital twin enables repeated simulation tests and ensure that the desired performance is achieved for the product design scheme and the production line. At the verification stage the digital twin allows virtual verification of the different aspects of the product, and thereby cut costs for extensive physical verification and testing.

A wide variety of architectures have been proposed for digital twins (Tao and Zhang 2017, Alam and Saddik 2017, Uhlmann, Lehmann, and Steinhilper 2017, Yun, Park, and Kim 2017, Rosen et al. 2015, Parrott and Warshaw 2017). It is difficult to find a consensus architecture, but we can see that digital twin is placed in the automation hierarchy between level 3 and 4, as shown in Figure 3.

![Figure 3: Digital twins in the automation hierarchy.](https://doi.org/10.3384/ecp181539)

The digital twin acquires physical plant data through the SCADA system or the process database: what is now the Industrial Internet of Things (IIoT). This data is
compared with design data and is used to drive simulations of product and process behaviour. Results from these analyses are sent to the information and data analysis system of the manufacturing process to provide recommendations to operators and management.

Peer-reviewed papers on digital twins seem to focus on the mechanical behaviour of single products. Thus, Schroeder et al. (2016) report a very simple digital twin for a valve. The different components of the valve including sensors and actuator, were modelled using an AutomationML tool. The data was made available for the user applications via FIWARE middleware. Cai et al. (2017) presented a digital twin for a CNC machine. The graphical simulation of the machine was CAD modelled in SolidWorks software, transferred into STL format and the graphical simulation was created with OpenGL library. The manufacturing data and sensory data were stored in a PostgreSQL database. Analytical tools were used to monitor the operation conditions, extract machining features and to predict product characteristics.

The opposite extreme of a manufacturing digital twin is the F35 digital twin. West and Blackburn (2017) describe the ambitions of Lockheed and the US government to build an on-line structural and functional model of each F35 fighter and its components. West and Blackburn’s analysis is deeply sceptical to the feasibility of this approach. It is simply too complex and too computationally intensive. We will return to these challenges in the next section.

So, what can we learn from the manufacturing industry? Manufacturers of the components in and oil and gas platform – instruments, valves, pumps, compressors, transformers – will offer twins of these components that will support capture of operation data for improving designs and optimising maintenance. These individual product twins will need to be incorporated into the oil company’s facility or process twin. The facility twin, on the other hand is more than the sum of its components. It will need to simulate the processing of oil and gas, the electrical, mechanical and structural behaviour of the facility. This integration of diverse and complex components into a complex system is difficult.

4 Challenges in digital twins

We noted above West and Blackburn’s justified but slightly polemic scepticism to the F35 digital twin. We share some of their fears. A digital twin of an oil and gas facility is a complex system. It requires the integration of a patchwork of different solutions from multiple vendors.

In this section we will describe eight challenges that need to be addressed to allow digital twins to fulfil their promised potential.

1. Business Models, Security and Confidentiality. The first challenge is to find a business model that is sustainable. Who benefits from a digital twin? We believe that all participants in the supply chain can benefit from digital twins. However, this requires clarity in the sharing of data and new models for procurement of facilities and services. An operator has a strong incentive to have a comprehensive twin for monitoring and optimization. It is linked to their license to operate. They are responsible for the asset and its data. Vendors can benefit from using operational data in their product twins. This sharing and collaboration for simulation and analytics is a good thing. However, mechanisms must be found for fair rewards, allocating responsibility, ensuring secure access and protecting IPR.

2. Work practices. The digital twin can change work practices, but only if it offers the users tangible and measurable benefits. The user must be convinced that the system gives benefits, is safe and is usable. The user must help, not hinder. Carvajal et al. (2018, p236) also pointed to this challenge: the digital twin must support the engineering and management organization.

3. Scope. A digital twin that aims to do everything is likely to do nothing well. We need to use fit for purpose modelling, and the principle of parsimony (i.e. just enough for functionality, but no more). However, we cannot do this at the cost of capturing necessary interactions and interdisciplinary effects. We need to support models with different granularity, that allow a user zoom in on the digital twin as needed. We also need different, but consistent models for processes with widely varying time constants, e.g. compressors (milliseconds) compared with a multiphase pipeline (days) or a reservoir (weeks).

4. Usability. How do I, as a user, get the information I need without being distracted by the information I do not need? How can I do this quickly and easily? A comprehensive digital twin has multiple concerns and can produce massive amounts of data. A specialized user must be able to find the information she needs quickly, easily and without distraction.

5. Integration. A digital twin will consist of many data sources and simulation models. How can we connect different models together without creating a “point-to-point nightmare”? Many vendors are offering platforms for data integration (Poddar, 2018). Commercial platforms are offered by vendors as diverse as GE, SAP, Siemens, Kongsberg, DNV GL, Schlumberger and Cognite (just to focus on the Norwegian market). An operating company will have to cope with several vendor platforms, both proprietary and open access, in addition to their own, internal platform. Carvajal et al. (ibid.) note that data is stored in “disparate locations, are in structured and unstructured formats, are not linked to a common data base, span long production and acquisition times, and have not been adequately and consistently quality checked”.

https://doi.org/10.3384.ecp181539
6. Maintenance. A digital twin is a complicated software and hardware system. How can the maintenance of the necessary data, software and hardware be cost-effective? As noted above, a twin combines life-cycle information, measurements of the assert state and simulations. All these must be maintained so that they reflect the as-built state of the facility. They must be included in the facility’s management of change.

7. Computational overload, edge and cloud. A comprehensive digital twin will require extensive computational resources. These resources will likely be distributed over a hybrid cloud that combines private clouds with vendor platforms and HPC resources. This system needs good design and operation to work. This design will need to carefully define what should be done at the edge of the system and what is done in the cloud.

8. Uncertainty, Validation and Data Science. Finally, a digital twin is only as good as the data and models used in the system. Data must be cleaned and reconciled. Models must be validated and tuned to ensure that they follow the state of the facility. Even “first principles” models must be tuned. Setting up effective methods for tackling uncertainty and model alignment should be a fruitful area for collaboration between control and process engineers and the data science community.

5 A research program for digital twins

How can digital twins be made sustainable, maintainable and useful despite these challenges? We believe that a solution will collaboration between computer scientists, control and simulation engineers, data scientists and end-user technical specialists. Here we will present the computer science and data science parts of solution. This research program combines the sub-disciplines of knowledge representation, natural language technologies, formal methods, scalable computing and data science. This knowledge of technologies must be informed by the deep domain knowledge that is embedded in the digital twins’ simulation models and is owned by the facility’s engineers – chemical, petroleum, mechanical, electrical and control – and managers.

Challenges related to scope, usability, integration and maintenance can be addressed if we adopt a semantic backbone that supports the integration of several different types of digital twin application around a shared understanding of a facility’s design. To be successful, this backbone will need to find a pragmatic balance between comprehensiveness and maintainability. It will also need to build on standards in a way that prevents reinvention of the wheel and allows modular construction of semantic models.

Recent advances in the construction of ontologies using templates (Forsell, 2017) promise to allow this.

Simulation providers can use these ontologies to allow exchange of model configurations with engineering databases and the exchange of calculated results with the monitoring and optimization layers of the digital twins. Point-to-point connections through tag cross-reference lists can be replaced with declarative mappings – where data in the simulation results or configuration is mapped to items in the semantic backbone.

Pragmatic standards are needed. Unfortunately, semantic standards such as ISO15926 have not lived up to their promise as tools for integration and data exchange. This is partly due to a lack of tool support and to a choice of technology, Express, that became dated and hampered progress. We also believe that the standard development attempted to be comprehensive at the expense of solving smaller, realistic problems along the way. Semantic models used by our group aim to take the best from ISO15926 where it is possible. We also will take account of the industry initiatives in DEXPI (www.dexpi.org) and CFIHOS (http://uspi-global.org/).

Good semantic models can also address the usability problem. Mapping data to concepts that are used by the end-user allows automatic generation of graphical interfaces that meet a specific user’s needs. The OptiqueVQS (Soylu et al., 2018) framework is just such a tool. Our research group is working further with this approach to implement faceted search (Klungre & Giese, 2017).

A successful digital twin also requires the correlation of structured data from designs, measurements and simulations with unstructured data from logs and documents. This requires the structured data to be interpreted and transformed into structured data in the process model. There are now many commercial and academic tools for parsing and processing text. The IBM Watson framework is one such commercial offering (Chen et al. 2016) and can be used to parse and process texts. However, language algorithms that are trained on general data sets do not perform well when confronted with oil and gas terminology. A process of domain adaption is needed to improve performance. This process is made difficult by the smaller corpuses of data that are available for training. The SIRIUS centre is working on this domain adaption challenge, with promising results and good performance in solving standard challenge problems (Nooralahzadeh, Øvrelid and Lønning, 2018).

The challenges of maintenance and computational overload can also be addressed through using formal methods to design and monitor the deployment of a digital twin. As we noted above, a digital twin is a collection of interacting computational components, deployed across one or more cloud platforms and including edge devices. The behaviour of this system is
difficult to predict, especially at design time. However, simulation tools for the computer systems themselves can be used to test different deployment plans and resolve challenges. The same model can also be used as a monitoring tool for the deployed system—a digital twin of the digital twin (Johnsen, Lin & Yu, 2016).

Finally, there remains the challenge of uncertainty, validation and data science. The digital twin is built on models. To quote George Box (Box, 1979), all these models are wrong, but some are useful. A digital twin will contain many models. Some will be based on physical principles: structural, geometrical and process simulations. Others will be purely empirical, based on machine learning. These models must be validated against observed facility behaviour and aligned so that they mirror observed normal behaviour. Aligning models to observed data is difficult and remains an art. Finding out whether a discrepancy is due to an error in data, a wrong parameter, poor model structure or an actual malfunction in the facility requires a good understanding of the facility and well-developed judgment. This is true whether the models a rigorous physical model or a machine-learnt empirical representation.

A maintainable digital twin will contain structured tools that allow validation and tuning of all the models in the system. We believe that hybrid analytics—the combination of data science with physical and engineering simulations—is a valuable and fruitful area of research. Machine learning can benefit from being constrained by the laws of physics, while the laws of physics contain parameters that are uncertain or expensive to measure. Good statistical practice is needed in the engineering communities and engineering knowledge is needed among data scientists.

We are working on two problems related to data science for digital twins. The first of these is related to data access. Data science projects in industry are currently not scalable. Each new implementation needs to start from scratch, finding data, checking it and making it available. Our proposed semantic backbone will allow data science solutions to be rapidly transferred to similar sites in an organization (Kharmalov et al., 2017).

The second area of research is related to the use of sensor data in data science. When a data scientist talks about streaming data, they usually mean a sequence of discrete event records—like tweets or sales transactions. The stream of data from analogue sensors is subtly different. The underlying signal from a sensor is continuous. The process of digitization itself introduces uncertainty and error in the calculation. Filtering and data compression provide further sources for error. Common data science frameworks expect data in vectors at common times. Production of this from a time series data base requires interpolation. All these details increase the cost and decrease the usefulness of data science work. A well-defined semantics and query tool for time-series data from sensors could solve many of these challenges.

Companies in the oil and gas sector are installing digital twins now, using commercially available platforms and siloed applications. This provides academics with an opportunity to engage with the observed problems of our colleagues in operations and maintenance. This means that we need a research program that engages with operations and today’s digital twins through pilots.

Each pilot has a narrow enough focus to be tractable. The companies we are collaborating with have linked these installations to a well-defined business case. Current pilots are ambitious: if successful they will bring previously unachieved levels of interdisciplinarity, effectiveness and access to data in design, operation and maintenance. At the same time, the pilots are focused on one specific business problem. By working with existing pilots and proposing new pilots we plan to establish a virtuous cycle, where shortcomings in today’s technology and methods can be filled with research-driven innovations.

6 Conclusion

This paper has attempted to give a synthetic overview of an important, if hyped, element of today’s digitalization landscape. The digital twin is, in many ways a rebranding of several generations of on-line systems for simulation and decision support.

Marketing and vendor communication is raising expectations about digital twins. We welcome this interest and believe that it gives an opportunity to integrate operations technology and information technology. However, we also believe that implementing digital twins is inherently difficult. Solving the challenges that we have identified will require the best efforts of many computer scientists, data scientists, engineers and managers. The role and importance of simulation can only be expected to grow in the next decade. Our research agenda can help ensure that we do not disappoint our managers.
Acknowledgements
The SIRIUS centre is funded by Norwegian Research Council, project number 237898. It is co-funded by partner companies: Statoil, Schlumberger, DNV GL, IBM, SAP, Computas, OSIsoft, Numascale, Dolphin Interconnect Solutions, Kadme, Evry and Fluid Operations AG.

We thank the participants from SIRIUS partner companies and others at the workshop on digital twins held in Oslo on 20th March 2018.

References
https://doi.org/10.1109/ACCESS.2017.2657006


Panetta, K., 2016, Gartner Top 10 Strategic Technology Trends for 2017,  


https://doi.org/10.4043/28480-MS

https://doi.org/10.4043/28012-MS


https://doi.org/10.2118/187222-MS

https://doi.org/10.1016/j.ifacol.2016.11.115

https://doi.org/10.4043/27638-MS

https://doi.org/10.3233/SW-180293

https://doi.org/10.1007/s00170-017-2333-1

https://doi.org/10.1109/ACCESS.2017.2756069 .


