Journal of Innovation

16th Edition

INNOVATIONS IN DIGITAL TWINS

March 2021

industrial internet®
CONSORTIUM

LETTER FROM THE THOUGHT LEADERSHIP CO-CHAIRS

Dear Reader:

Digital twin technology enables you to see what happens as a result of your decisions *before* you invest in them - *without* the risk.

Digital twin spans hundreds of potentially blockbuster applications that bridge the gap between the physical world and the virtual world in industries such as aerospace, manufacturing, and energy. It is already fulfilling its potential to improve process efficiency and optimize operational sustainability while ensuring a better digital experience for customers.

Given these remarkable outcomes, it's no wonder that the digital twin market is expected to climb at a 38% CAGR to reach $26B by 2025[1].

A digital twin is defined[2] as a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity:

- Digital twin systems transform business by accelerating holistic understanding, optimal decision-making, and effective action.
- Digital twins use real-time and historical data to represent the past and present and simulate predicted futures.
- Digital twins are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and implemented in IT/OT systems.

The Industrial Internet Consortium (IIC) has taken a market leadership position by identifying the defining characteristics of a digital twin, relations among digital twins to form composite systems, and the role of digital twin in the lifecycle of entities.[3] Additionally, the IIC has set the pace in enabling the cross-industry collaboration necessary to drive the success of digital twins.

Toward that end, we have compiled the following informative, enlightening, and thought-provoking articles that look at digital twin from a variety of technical perspectives. We hope you will find this compendium beneficial to your own digital twin initiatives:

- *Web-Based Digital Twin*, by several authors from the Technical University of Darmstadt, presents current solutions for implementing digital twins over web technologies, which many see as a prerequisite for establishing digital twins on a broad scale.

[1] https://www.grandviewresearch.com/press-release/global-digital-twin-market
[2] https://www.digitaltwinconsortium.org/hot-topics/the-definition-of-a-digital-twin.htm
[3] https://www.iiconsortium.org/stay-informed/digital-twins-for-industrial-applications.htm

- *Open Source Drives Digital Twin Adoption*, by a cross-section of industry leaders from ABB, Bosch, Digital Twin Consortium, Microsoft, SAP, XMPro Inc., and Yo-i Information Technology, Ltd., elaborates on the challenges in adopting the digital twin concept in practice and the motivations for having open standards and Open Source activities to address these challenges.

- *Design and Implementation of a Digital Twin for Live Petroleum Production Optimization: Data Processing and Simulation*, by a trio of authors from OspreyData Inc., delves into how digital twins can be built to represent systems of assets that can be valuable in determining optimal operational set-points.

- *Digital Twin and IIoT in Optimizing Manufacturing Process and Quality Management*, by a host of authors representing Yo-i Information Technology, Ltd., presents the concepts and practices on the design, implementation and some preliminary outcomes based on a real-world use case in a production process control and quality management application in the steel industry.

We hope you enjoy this edition. Please reach out to us to publish your research and business outcomes in future editions.

Best regards,

Mark Crawford
Director, Standards Strategy, SAP
Thought Leadership Task Group Co-Chair

Edy Liongosari
Chief Research Scientist, Accenture Labs
Thought Leadership Task Group Co-Chair

Cover image courtesy of Dreamstime.

LETTER FROM THE IIC

Dear Reader:

The Industrial Internet Consortium is pleased to publish this 16th edition of *The IIC Journal of Innovation*. This collaborative effort is the sum of very many parts, and we would like to take this opportunity to recognize and thank the team of editors and peer reviewers who lent their time and expertise to the editorial process of enhancing each of the articles contained in this edition with their unique perspectives and wisdom.

Sincere thanks goes to this edition's editors and peer reviewers:

Mr. Mark Crawford, Director, Standards Strategy, SAP
Mr. Edy Liongosari, Chief Research Scientist, Accenture Labs
Mr. Abhik Chaudhuri, Chevening Fellow (UK), FCSA (US), Tata Consultancy Services
Mr. Christian Beier, Azure IoT Engineering – Strategic Engagement Lead, Microsoft
Mr. Jens Lachenmaier, Senior Researcher, Ferdinand-Steinbeis-Institut
Ms. Kathy Walsh, Content Marketing Manager, Real-Time Innovations
Mr. Kevin Kalish, Lead IoT Technology Strategist, Europe, SAS
Mr. Kym Watson, Principal Scientist, Fraunhofer IOSB
Mr. Rahul Sinha, Scientist, Tata Consultancy Services
Mr. Russell Waddell, Managing Director, MTConnect Institute
Dr. Vincent Bemmel, Industry Technology Lead, Corlina
Mr. Brian Jones, Senior Marketing Specialist, Industrial Internet Consortium
Mr. Evan Birkhead, Strategic Marketing, Industrial Internet Consortium

As a global, not-for-profit, public-private partnership of nearly 200-member organizations, we encourage every organization across all industries to get involved and be proactive in shaping the Industrial Internet. We welcome your feedback and participation.

Many thanks,

Cheryl Rocheleau
Director of Marketing
Industrial Internet Consortium

TABLE OF CONTENTS

Web-Based Digital Twin

Authors:

Vladimir Kutscher
Technical University of Darmstadt
kutscher@dik.tu-darmstadt.de

Christian Plesker
Technical University of Darmstadt
christian.plesker@stud.tu-darmstadt.de

Johannes Olbort
Technical University of Darmstadt,
olbort@dik.tu-darmstadt.de

Reiner Anderl
Technical University of Darmstadt
sekretariat@dik.tu-darmstadt.de

Benjamin Röhm
Technical University of Darmstadt
roehm@dik.tu-darmstadt.de

INTRODUCTION

The digital twin is gaining acceptance in a variety of industries. It is a cornerstone that enables the next step of extensive digitization. Satellites, airplanes, automobiles [1], production machines, energy systems, chemical processes [2], sports shoes [3] and humans themselves [4] receive a digital twin, opening new business models and use cases.

In many cases, the digital twin or parts of it are implemented using proprietary and specialized technologies. As a result, it is manufacturer-dependent and requires specialist knowledge to use. This hinders the interaction of different disciplines and makes it difficult to use the digital twin throughout the entire lifecycle. At the same time, open, standardized and widely used technologies also offer many opportunities for implementing a digital twin. Web technologies have been ubiquitous for a long time, connecting both people and technical systems. It is largely platform-independent, has many users, offers strong security and a high degree of maturity. Computers that have a web interface can use basic functions on the Internet immediately, usually without supplementary software. In addition, many technologies such as cloud computing, big data and Internet of Things are already available within the web technologies and are relevant for digital twins. We see this as the prerequisite for the digital twin to become established on a broad scale in practice.

In the following, we would like to present the solutions to some problems in current practice with the approach of a web-based digital twin. Thereby, we focus on industrial systems by choosing a computer-controlled laser plotter as demonstrator for a prototypical implementation. From this context, we set requirements for the concept and implementation and use them for validation. Fig. 1 shows the physical laser plotter, which is connected to its web-based digital twin bidirectionally.

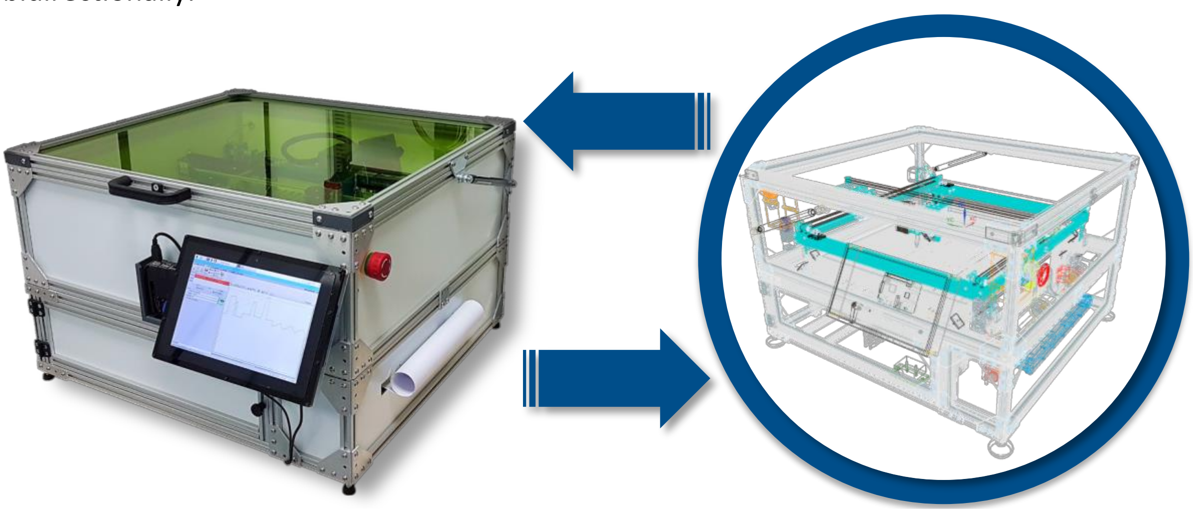


Fig. 1: Computer-controlled laser plotter as physical asset and the symbol of its web-based digital twin.

STATE OF THE ART

Digital Twin

The digital twin permeates diverse sectors and found its way into almost all areas of industry [5]. As a result, the counterpart of a digital twin is no longer just the physical twin, but rather an asset in general, including a system or a process. By means of a bidirectional connection, both are in contact and the digital twin can represent the current, future and historical state of its counterpart. Via an interface, the user can interact with the digital twin and thus influence the asset.

The *Industrial Internet Consortium* (*IIC*) is working on an overarching and concise definition of the digital twin [2] and is cooperating with the German *Plattform Industrie 4.0 [6]*, since the digital twin is a key concept for *Industrie 4.0* (*eng.: Industry 4.0*) [7]. Progressive digitization and networking can create intelligent and individual products that meet current trends and customer requirements.

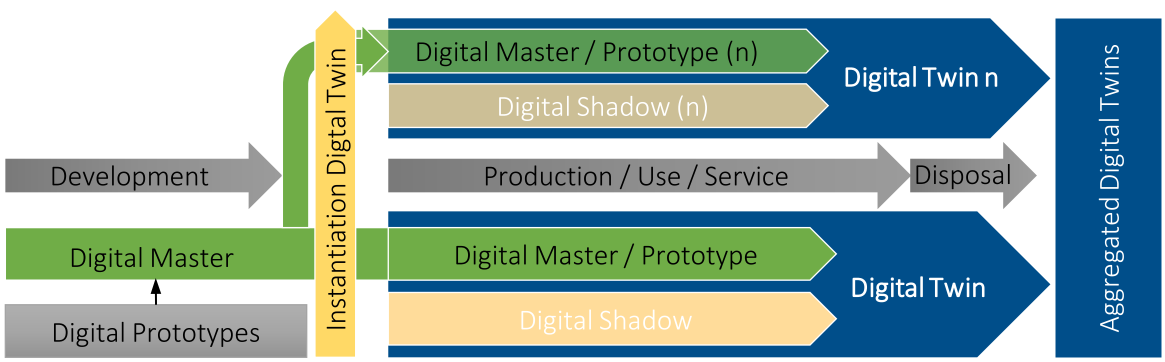


Fig. 2: Instantiation of digital twins following [9].

The German *Scientific Society for Product Development,* called *WiGeP,* provides a possible description of the life cycle of a digital twin of a product [8]. Fig. 2 shows an adapted illustration. In the presented scenario, the digital twins are instantiated from so-called digital masters, which result from product development and serve as a blueprint for the digital twin [8]. The digital master includes the digital prototypes from product development and continues to exist in the digital twin in the form of individualized models.

The entire data from the life cycle of a physical product form a so-called digital shadow. The instantiated digital twins exist throughout the product lifecycle [1] and provide aggregated data that can be reused as a knowledge pool for example to optimize the next product generation. For existing assets (brownfield), there is also the possibility of retrofitting a digital twin. The existing environment provides constraints that need to be analyzed and incorporated into the necessary retrofit measures [9]. In any case, the product-specific digital twin is linked to the asset [10].

Digital twins can be discrete or composite during their existence [2]. The discrete digital twin is an entity that generates added value on its own. In a composite digital twin, multiple elements serve a common purpose. Multiple digital twins can further act in a system of digital twins and serve an overarching purpose. In the digital twin, the internal networking of individual components and the external interaction with other twins is thus an important aspect that addressed by suitable communication technologies and interfaces.

In addition to communication, the digital twin consists of further key aspects like data, models and computing. As Fig. 3 already indicates, the aspects are not clearly separable and overlap with each other, all of which influence the digital twin. The data of a digital twin is of different nature and may cover all phases of the product lifecycle. The types of data considered in the following are operational data and simulation data, since these are relevant in the presented context.

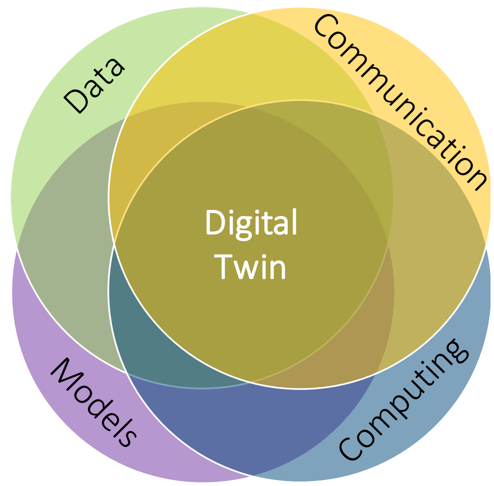


Fig. 3: Key aspects of the digital twin.

Data Modelling and OPC UA

For the operational data of an asset, the *Open Platform Communications Unified Architecture* (OPC UA) [11] attains a special position. It is a platform-independent, service-oriented architecture that is being standardized for a wide range of industries. The standard builds on two basic pillars: the transport mechanism and the data modeling. Various communication technologies are available as transports, which differ depending on the approach (client-server/publish-subscribe).

But the distinctive feature is the data modeling, which focuses on layer-wise information models. The information models have sub-models for devices in general, as well as extensibility for specific industries and vendor-specific information [12]. The industry-specific models are the result of collaborations between various companies and organizations and manifest in so-called Companion Specifications. The OPC UA information models represent the state data of the asset. All information and functions are thus available via standardized interfaces, to which the digital twin can also connect [10].

Simulation and FMI

Another key aspect in the digital twin is simulation [13] and integration various simulations into a multi-domain simulation chain. The simulation techniques and tools that exist today are mostly designed to be used during product development. On the one hand, these proprietary tools are not suitable for use in the digital twin due to frequent incompatibility with simulations from other domains and, on the other hand, due to the required licenses and expertise. The problem of compatibility is addressed in corresponding standardization organizations [14] and scientific publications [15, 16]. Vendor and domain independence can be achieved through open cross-domain modeling languages such as *Modelica* [17] and exchange formats like the *Functional Mock-Up Interface* (FMI). FMI is an open standard that defines a container and interface for exchanging dynamic models [18] and is particularly relevant in the presented context, since simulations can be supplied in the digital twin and integrated into a web-based environment. The goal is to provide a uniform interface to simplify the creation, storage, exchange, and reuse of simulations from different simulation environments. FMI can be divided into two process models: *FMI for model exchange* and *FMI for co-simulation.*

The dynamic models are packaged in a *Functional Mock-Up Unit* (*FMU*). An *Extensible Markup Language* (*XML*) file configures the FMU and C-code files describe the function. An FMU for model exchange only contains the model of the simulation. Accordingly, the tool for using an FMU must have its own solver for equation-solving. In an FMU for co-simulation, the model is compiled and exported with its own solver. This allows the tool to run the FMU independently with the solver provided and operate it without knowledge and ownership of solvers [18].

With the application of sensor data-based simulation models, the need for a dedicated *Simulation Data Management for Digital Twins* (SDM-DT) also arises. Derived from classical simulation data management systems from product development, the SDM-DT provides, manages and archives sensor-based simulation data [19]. Consequently, sensor-based simulation data can be managed in the digital twin to describe predictive and prescriptive scenarios. The performed simulations are virtually backed up and decision traceability is ensured [20].

Web Technologies

Both FMI and OPC UA already utilize web technologies such as various transfer protocols or offer corresponding interfaces. The multitude of protocols involved in the network connection between two end points can be divided into groups, which are represented, for example, in the OSI (*Open Systems Interconnection*) layer model [21]. Within individual groups, the protocols are interchangeable, resulting in many possible combinations. Widely used protocols include the *Hypertext Transfer Protocol* (*HTTP*), the *Transmission Control Protocol* (*TCP*) and the *Internet Protocol* (*IP*). Internet technologies are being further developed especially for industrial applications. Under the collective term *Industrial Internet of Things* (*IIoT*), the IIC offers concepts and architectures that make Internet technologies ready for industry and corresponds with activities within the *Plattform Industrie 4.0* [22].

A variety of technologies exists not only for communication via the Internet, but also for the development of web resources. These can be classified into the three basic areas of frontend, backend and database. While the frontend is close to the user and is processed on his computer, the backend and database is executed on the server. For the development of web applications, web frameworks specialize either on development of the backend, frontend or both (full-stack). These libraries provide pre-programmed functions and simplify the development of new applications. Some of the most popular web frameworks include *jQuery*, *React*, and *Angular* [23]. These support the developer in the creation of a web application, which for the frontend mostly contain the programming languages *Hypertext Markup Language* (*HTML*), Cascading Style Sheets (*CSS*) and *JavaScript*. HTML determines the structure, CSS the presentation and JavaScript the logic of web applications. The programming language chosen to implement the backend depends on the web application. Languages used here include C, C++, Java, Python, PHP and others.

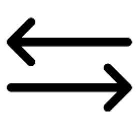
A considerable number of scientific papers already use web technologies for the conceptualization and implementation of the digital twin. Schroeder et. al. [24] employ web services in combination with augmented reality for visualizing the data of the digital twin. Souza et. al. [25] present an architecture for a digital twin based on Industrial Internet of Things technologies. Liu et. al. [26] utilize the digital twin for modeling and web-based remote control of cyber-physical production systems. In contrary to approaches that use web technologies only partially, we would like to present a completely web-based digital twin in the following.

Expected Features

What features do we expect from a web-based digital twin? In the following, we would like to list some essential characteristics we determined for our approach condensed from scientific literature [24, 27–29].

Representation of data: The digital twin must represent consistently all relevant data of the specific asset. This includes the characteristics and information about the asset, data collected by sensors and any asset-related data from the entire lifecycle from development to operation and maintenance. The data can either be located on the digital twin itself or accessed by reference to external sources.

Representation of behavior: The digital twin must be able to represent the behavior of the asset. The behavior of an asset can usually be represented by mathematical equations that describe fundamental physical, chemical, and biological phenomena. It can be analytical models that run on a computer. Or it may be software in general, which describes the logic of a system and dictates it by electronics. All these models and computation determine the changes of an asset over time and thus its dynamic behavior.



Controlled bidirectional connection: The digital twin must be able to establish a bidirectional connection with the asset. This connection does not necessarily have to exist all the time, as tests may be performed on the digital twin without directly influencing the asset, for example. For this reason, we do not consider the bidirectional connection to be present all the time but controlled, to also enable independence of the systems. This is also beneficial during failure of a system.

Standardized external interfaces: Digital twins must be able to provide their data and functions to each other and to other entities. To enable the exchange, a network connection must be provided and communication must be based on vendor-independent interface standards.

Standardized internal interfaces: Internal communication within a composite digital twin must also be based on standardized interfaces. Since the individual elements come from different domains, their need a unifying interface.

Quality of services: The digital twin does not have any benefit if it cannot provide sufficient quality of service. Especially in the industrial environment, the reliability of the services, for example in terms of real-time requirements, stability under load, or accuracy, are crucial factors that the digital twin must guarantee.

WEB-BASED DIGITAL TWIN

In the following, we want to discuss the concept of a web-based digital twin. For the generation of the web-based digital twin we used a methodology with four steps: requirements analysis, system design, implementation, and verification. The requirements analysis is used to determine the requirements on a digital twin in general and a web-based digital twin in particular. We then used these requirements in the system design to determine suitable technologies for implementing a web-based digital twin. In the third step, the prototypical implementation serves to test the viability of the final model. The verification of the requirements in the last step verifies the suitability of the generated web-based digital twin.

Concept of a Web-Based Digital Twin

Using the described requirements and the methodological approach, we have developed a distributed web-based digital twin. Fig. 4 represents the components relevant during our design. The web-based digital twin represents an asset in its current and past state throughout its lifecycle. The state is recorded in the form of data and stored in databases. To obtain an estimate of a possible state in the future, the digital twin also represents behavior using models and simulations. Since on the one hand the simulations should run mainly autonomously in the background and on the other hand several simulation types interact with each other to represent the entire system, a

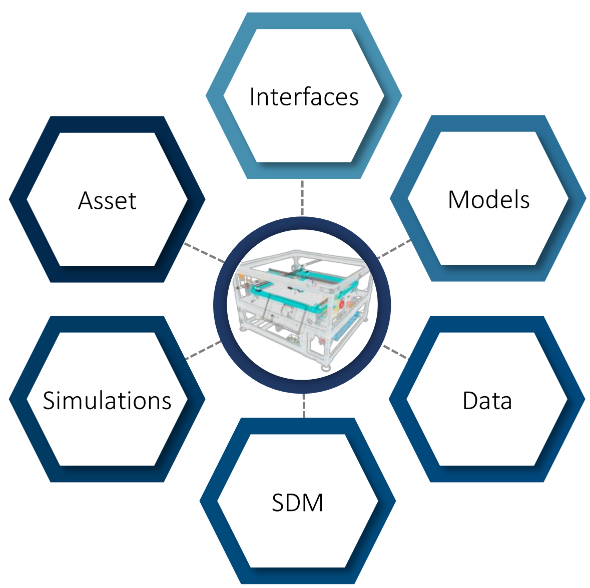


Fig. 4: Components of a web-based digital twin.

simulation data management (SDM) is needed. The digital twin makes its data and functions available to external entities via interfaces and accesses their resources. These entities include, for example, other digital twins, assets, and humans.

Fig. 5 shows the structure of the digital twin. In the following, we would like to discuss the individual aspects. As a web-based application, the digital twin must ensure access to resources via HTTP-interface. All information and functions provided must be able to be called, changed, saved or controlled in this way. The web server makes it possible to provide resources, for example, by means of web pages via HTTP.

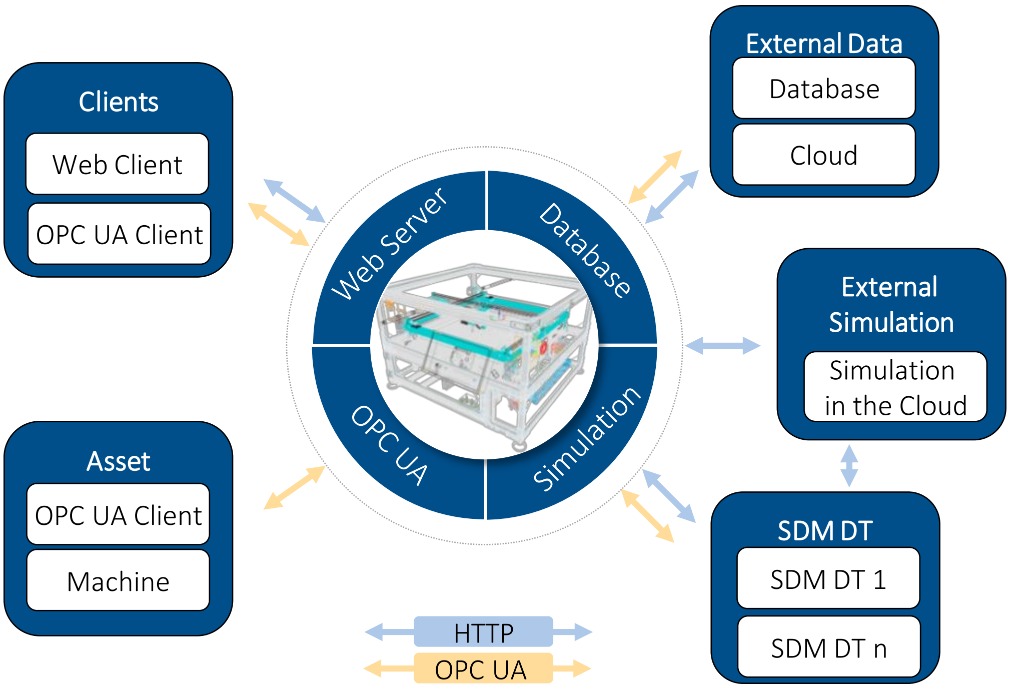


Fig. 5: Web-based digital twin.

The web server stores the data of the web server itself, as well as of the digital twin, in databases. This can be, for example, user login data, data from product development or the operating data of the asset. If the size of the data load exceeds the available memory, additional external databases are used. The web server has access to the asset's data and presents the information by means of web pages. Simple texts, images, graphics or videos can be integrated directly.

The visualization of special file formats such as CAD (*Computer Aided Design*) models is enabled by libraries of the JavaScript scripting language, which although does not support all file formats like most native ones. Neutral or standardized file formats, such as *STEP* (*STandard for the Exchange of Product model data*), *IGES* (*Initial Graphics Exchange Specification*), *STL* (*Standard Triangulation Language*) or *VRML* (*Virtual Reality Modeling Language*) differ in their information content. While STEP transfer most of the information from a native format, STL as a 3D format describes only the surface. With the possibility to visualize the data of the linked asset, the web-based digital twin can now digitally represent the state.

The bidirectional connection between the digital twin and the asset is established on the basis of OPC UA information models, which are integrated in an OPC UA server [10]. Due to the vertical integration of OPC UA and the renunciation of the classic automation pyramid, the OPC UA server can be located anywhere. The OPC UA server contains the status information and provides them to any OPC UA clients. In addition to the simple display of the information units, the state of the asset can be visualized with the help of geometric models.

Another characteristic of a digital twin is the representation of behavior. Representing behavior digitally means replicating how it works through simulations. Performing simulations within the web-based digital twin can be divided into three categories:

- Simulations on the client side
- Simulations on the server side
- Simulations on external servers

The three categories differ in the use of computing resources to perform the simulations. Simulations on the client side use the resources of the requesting client for this purpose. To run the simulations, the desired solution algorithms must be developed in JavaScript. The shift to the client side has advantages if the client has large computing resources and drawbacks if the client does not, e.g. a smartphone. The opposite approach is simulation on the server side. The simulation accesses the same resources as the web server of the web-based digital twin for computation. This also reverses the pros and cons compared to frontend simulation. A third option is to move the computation to external hardware. In this case, the simulation runs on an external server and the results return to the web server. The external server can be self-managed or provided by cloud computing solutions.

The integration of simulations into the web-based digital twin requires the selection of a suitable interface. The multitude of possible interfaces from different simulation environments is a challenge that is addressed by the FMI standard. Currently, more than 100 tools are available for exporting and importing FMUs. With the increasing number of FMUs in the digital twin, the management of the simulation models becomes more important and should be addressed by SDM-DT.

From the corresponding metadata of the models, the respective open simulation parameters with their functional values as well as the output parameters can be read out. The different available simulation models can put together into simulation networks, knowing the input- and output parameters. The use of simulation networks enables time-efficient high-resolution simulations. Crucial for the calculation of the simulations, is the linking of the open simulation parameters with the corresponding sensor values of the asset. Furthermore, discrete simulation data must be generated from the continuous sensor values, which are calculated together with the models.

Depending on the sampling rate of the sensors, different measured values from different sensors and different time steps are combined to a coherent data set. The fusion with data that may not be captured by the asset, such as environmental values, must also be ensured. The use of a simulation data management system in the digital twin meets these requirements. Only through the central interface between models, sensor data and solver, an efficient sensor data-based simulation can be realized.

The current FMI technology is suitable for simulation on the server side and on external servers. Client-side simulation cannot be performed with the current state of the art, since no tools for integrating FMI technology are available for JavaScript.

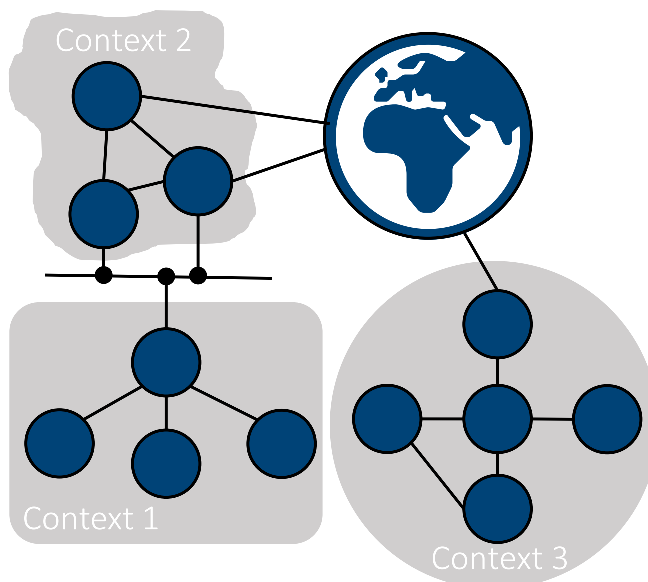


Fig. 6: Possibilities of networking digital twins.

For internal and external communication, the developed web-based digital twin basically offers two interfaces. As shown in Fig. 5, the digital twin uses both HTTP and OPC UA. For the bidirectional connection between the digital twin and the asset, the web-based digital twin utilizes an OPC UA server, which enables a uniform, cross-industry interface. All accessible information and functions are available via this server and can be requested by a client.

Networking between digital twins, which can have different relationships to each other, also plays a key role. Fig. 6 shows some possibilities for the networking. First, the digital twins are arranged in their environments or contexts. In the hierarchical structure of context 1, a digital twin connects across the boundaries of the context by means of an internal network to several digital twins of a second context. The fully networked digital twins of the second context are partially connected to the Web and can interact with a third context with its own structure. This flexible networking is fundamental for a digital twin and can be implemented in the web-based digital twin by means of the web technologies.

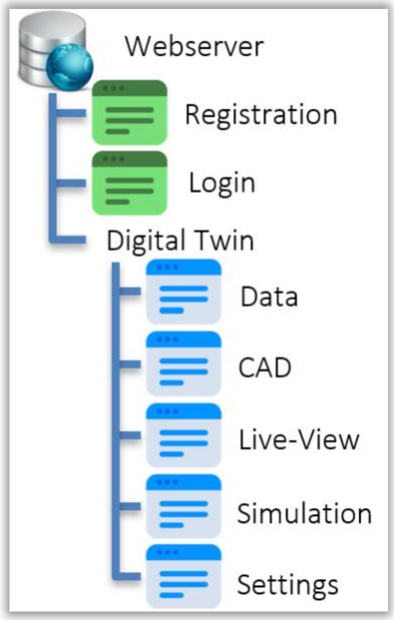


Fig. 7: Structure of the web-based digital twin.

PROTOTYPE OF A WEB-BASED DIGITAL TWIN

We implemented a prototype to validate the concept. For the technical implementation, we used the Python programming language and numerous libraries.

The web server as the core of a web-based digital twin was developed using the open-source full-stack framework *Django*. The individual functional areas of the web-based digital twin are implemented as apps in Django and thus enable the modularization of the digital twin. Fig. 7 shows the structure of the web-based digital twin. In addition to the public pages Login and Registration, six additional pages are added for the digital twin. The Data page enables access to asset data coming from multiple origins like databases and cloud services. The CAD view contains 3D models of the asset. Live View displays the current state of the asset. The Simulation tab provides access to simulations, their execution and their results. The Settings view provides the preferences of the web-based digital twin.

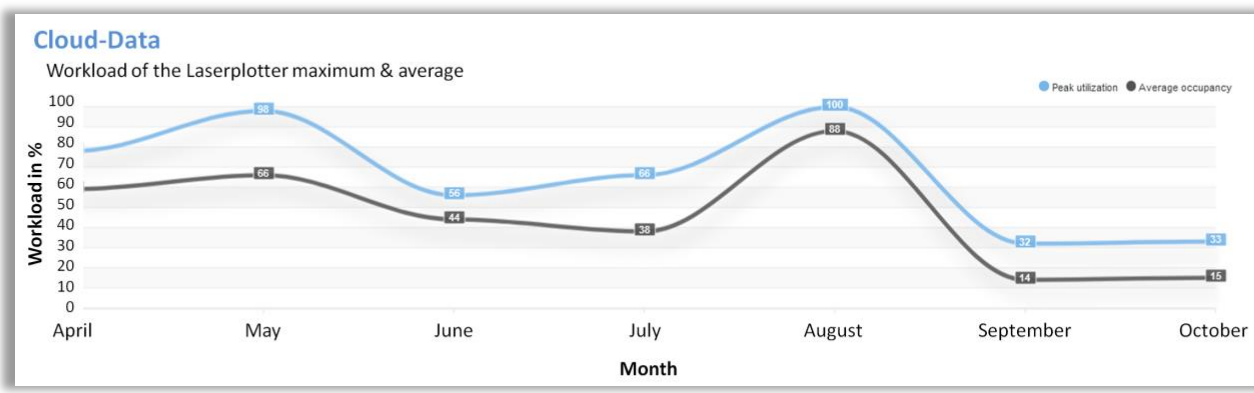


Fig. 8: Exemplary operation history of the laser plotter using ApexCharts library (https://apexcharts.com/).

The data for the Data View originates from different sources. On the one hand the digital twin has its own database system, which can be one of the databases the utilized framework Django supports, for example: PostgreSQL, MariaDB, MySQL, Oracle or SQLite. On the other hand, external databases, also from the cloud, can be integrated via HTTP. Fig. 8 shows exemplary data on the operation history of the laser plotter.

Furthermore, the Data View links the OPC UA information model of the laser plotter to the web service. Fig. 9 shows a section of the data node structure created using the *FreeOpcUa*[1] library. The JavaScript library *jsTree*[2] was used to display the data [10] of the OPC UA Server. In addition to the system information structured according to use cases, the control variables of the laser plotter are also represented.

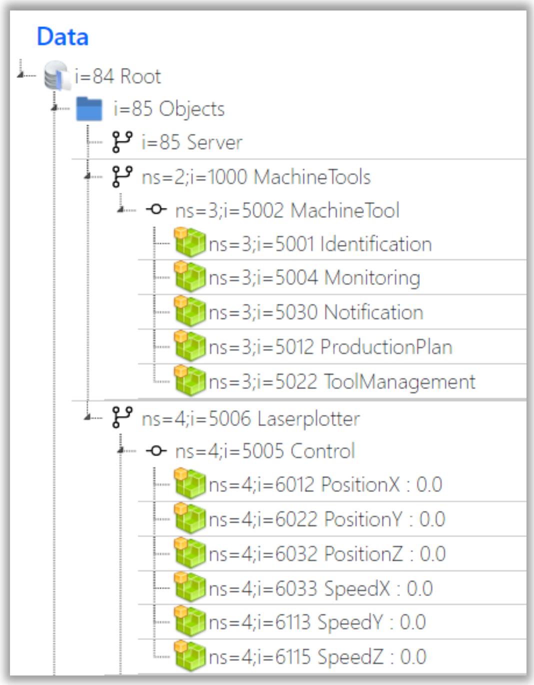


Fig. 9: OPC UA data structure of the laser plotter integrated into the web-based digital twin (ns – namespace, i – Identifier).

For the basic representation of geometric models there are some JavaScript libraries. However, the import of standardized CAD exchange formats such as STEP are not supported as the preferred option. Babylon.js was identified as an alternative which supports the file formats .glTF, .glb, .obj and .stl. The CAD models are represented by vector graphics on the web pages and can be viewed in the Internet browser from different perspectives. The calculations are performed on the computing resources of the client and thus determine the performance. Fig. 10 shows an example of the implemented CAD user interface using the plot and laser unit.

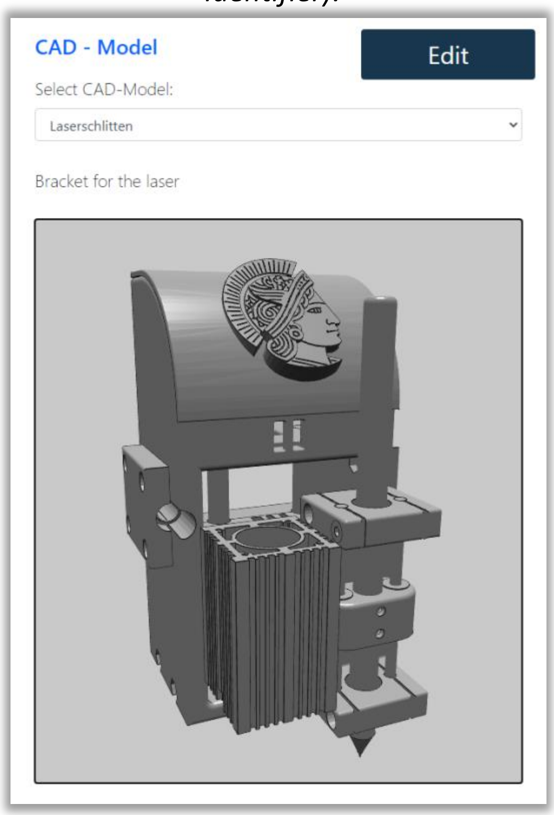


Fig. 10: Visualization of 3D geometry of a laser plotter component.

The live view combines the up-to-date data and the 3D geometry models. The representation of the

[1] https://github.com/FreeOpcUa/freeopcua

[2] https://www.jstree.com/

models is based on the continuously updated variables from the OPC UA server within the bidirectional connection to the physical asset and thus reflects the current, spatial state of the system. Fig. 11 shows the geometry model of the laser plotter mechanics, whose position of the cartesian axes is animated according to the position variables from the OPC UA server.

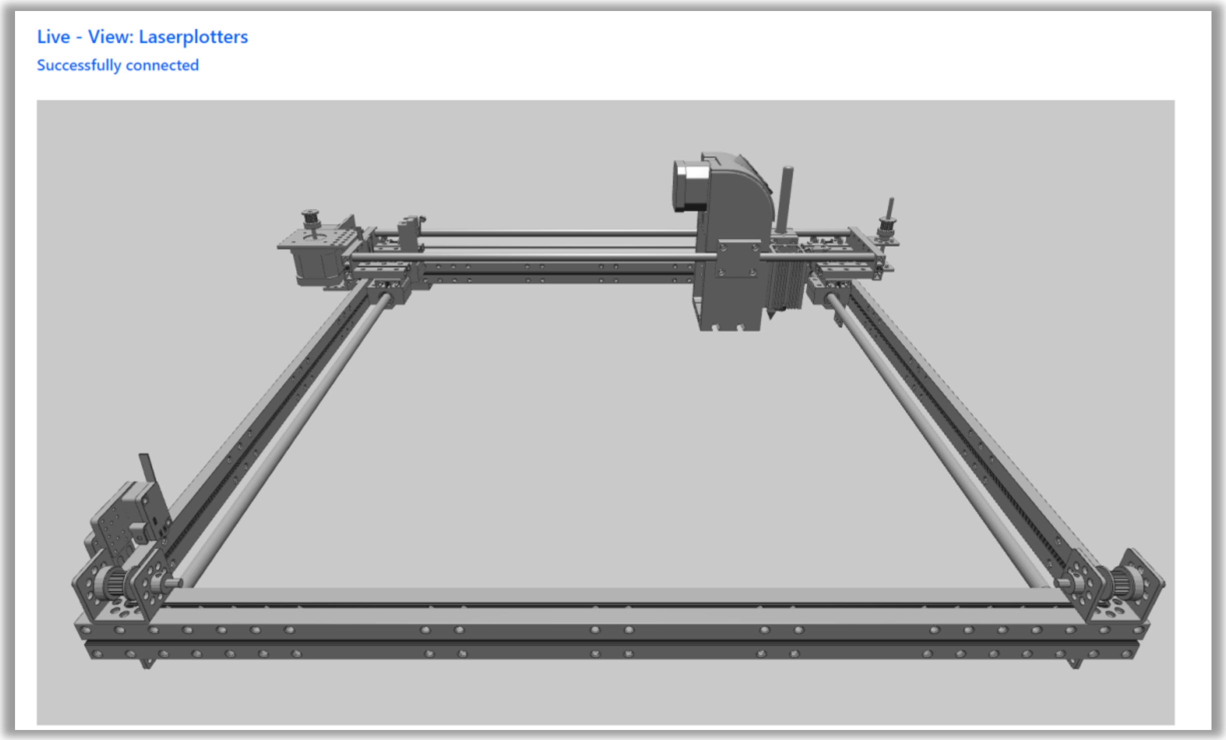


Fig. 11: Live view of the up-to-date state of the asset. The animation is moving accordingly to the current position variables in the OPC UA server. The model is stripped down to the essentials to speed up the animation.

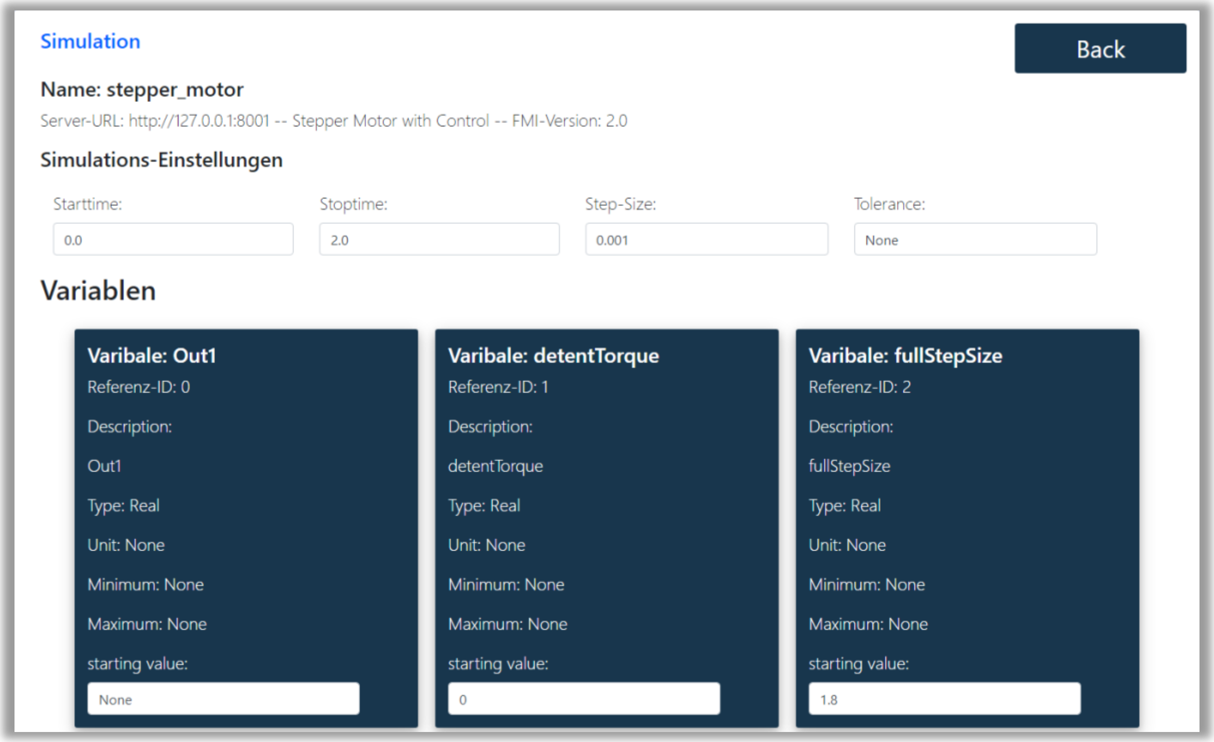


Fig. 12: Web-based user interface for the FMU simulation of a stepper motor.

The web-based digital twin can also include different types of simulations using FMI. For the laser plotter, for example, the behavior of the stepper motors used is relevant. The modelling and simulation tool MATLAB/Simulink was used to create a suitable simulation. Subsequently, the simulation was exported as a stand-alone FMU with solver. Variables were defined, which serve the user as input parameters for the simulation. The simulation runs on an external server and is integrated into the web-based digital twin using HTTP. After successfully adding a simulation, the web-based digital twin creates a simulation interface automatically, as shown in Fig. 12. An input field displays the default start value, which can be adjusted by the user. By pressing the start button the digital twin sends the corresponding data to a simulation server via an HTTP request to call the corresponding simulation function. Fig. 13 shows the simulated steps of the stepper motor over time. The simulations can thus be used, for example, to analyze the behavior of individual components under specific conditions.

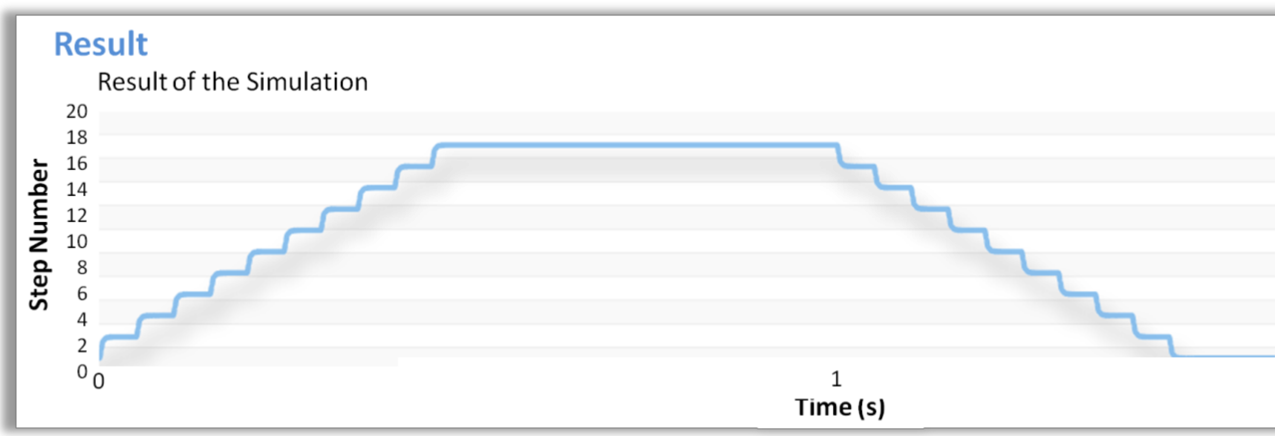


Fig. 13: Simulation results of an FMU stepper motor simulation on an external simulation server.

DISCUSSION OF THE RESULTS

In the following, we would like to discuss the presented concept and the implementation regarding the requirements for a web-based digital twin and the challenges and research gaps. Table 1 lists our assessment of the requirements.

Table 1: Assessment of the elaborated requirements on a web-based digital twin.

Requirements	
Representation of data	◐
Representation of behavior	◐
Controlled bidirectional connection	●
Standardized external interfaces	●
Standardized internal interfaces	●
Quality of Services	◐

As we have shown in the concept and implementation, different types of data can be represented on the web to describe the asset. However, the integration of CAD models is limited at the current time. Exchange formats such as STL can transfer the geometry - but information about the material or information about product manufacturing is lost. Extensive formats such as STEP cannot be imported with the libraries examined.

After the system is represented in the web-based digital twin, the current status and behavior can be mapped by coupling the status information. The FMI standard can be used to bridge the incompatibilities between different simulation types when interconnecting them to build a simulation chain. The support of the FMI import and export is different in the proprietary simulation tools and not consistently possible. In addition, the interfaces between simulations of different domains are not standardized and the communicated data must be adapted manually. There is still a great need for research in the use of simulation data management in connection with digital twins. The conceptually elaborated idea could not be tested in the prototype, since no supporting development tools could be identified in this area.

The web-based digital twin, which is similar in state and behavior, can be coupled bidirectionally with the asset. However, due to the conceptual consideration, the directional connection was not implemented rigidly, but can also be unidirectional or completely disconnected depending on the use case.

By using standardized web interfaces when providing data and functions, the web-based digital twin can interact with other entities. The web server provides all functions via HTTP using all HTTP standards.

The digital twin must offer sufficient service quality depending on the use case in order to be used reliably. The quality of service depends on many factors, which must be analyzed depending on the implementation. Despite the advanced technologies, it will not be possible to ensure the guaranteed and very short response time required in some industrial applications. The communication between the individual components of the digital twin, as well as the processing of individual programs, and the execution of simulations reaches a composition of the digital twin that cannot be used for real-time applications at the current state. Further research needs to be done in this area and the technologies used need to be made more capable for this type of application.

Overall, we believe that the web-based digital twin offers an interesting opportunity to implement new use cases and business models. In this context, some suitable web technologies are already available, while others need research and development first.

References

[1] M. Grieves and J. Vickers, “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems” in F.-J. Kahlen et al., Eds.: Springer International Publishing, 2017.

[2] Industrial Internet Consortium, “Digital Twins for Industrial Applications: Definitions, Business Values, Design Aspects, Standards and Use Cases”, IIC, Feb. 2020.

[3] U. Weber and H. Grosser, “Digital Twins - Leading the Way to Tomorrow’s Ecosystems: Detecon Study”, Detecon, 2019.

[4] B. R. Barricelli, E. Casiraghi, J. Gliozzo, A. Petrini, and S. Valtolina, “Human Digital Twin for Fitness Management”, *IEEE Access*, 2020.

[5] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital Twin in manufacturing: A categorical literature review and classification”, 2018.

[6] P. I. 4.0. IIC, “Digital Twin and Asset Administration Shell Concepts and Application in the Industrial Internet and Industrie 4.0: An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper”, 2020.

[7] R. Anderl, S. Haag, K. Schützer, and E. Zancul, “Digital twin technology – An approach for Industrie 4.0 vertical and horizontal lifecycle integration”, *Information Technology*, 2018.

[8] R. Stark, R. Anderl, K.-D. Thoben, and S. Wartzack, “WiGeP-Positionspapier: Digitaler Zwilling”, *ZWF*, 2020.

[9] V. Kutscher, J. Olbort, O. Anokhin, L. Bambach, and R. Anderl, “Upgrading of legacy systems to cyber-physical systems”, *Proceedings of TMCE 2020*, 2020.

[10] V. Kutscher, J. Olbort, C. Steinhauer, and R. Anderl, “Model-Based Interconnection of Digital and Physical Twins Using OPC UA” in *Advances in Intelligent Systems and Computing*, Springer International Publishing, 2020.

[11] DIN, “DIN 62541-1 OPC UA: Teil 1: Übersicht und Konzepte” 2011.

[12] W. Mahnke, S.-H. Leitner, and M. Damm, *OPC Unified Architecture,* Springer Vieweg, 2009.

[13] S. Boschert and R. Rosen, "Digital Twin—The Simulation Aspect", in P. Hehenberger and D. Bradley, Eds.: Springer International Publishing, 2016.

[14] Simulation Interoperability Standards Organization. *Official Website* [Online]. Available: https://www.sisostds.org/ (accessed: Feb. 27 2020).

[15] C. Gomes, C. Thule, D. Broman, P. G. Larsen, and H. Vangheluwe, "Co-simulation: State of the Art" in *Systems and Control*, 2017.

[16] D. Broman, L. Greenberg, E. A. Lee, M. Masin, S. Tripakis, and M. Wetter, "Requirements for hybrid cosimulation standards" in *Proceedings of the 18th International Conference on Hybrid Systems Computation and Control*, 2015.

[17] Modelica Association, "Modelica - A Unified Object-Oriented Language for Systems Modeling: Language Specification", 2017.

[18] Modelica Association Project FMI, "Functional Mock-up Interface for Model Exchange and Co-Simulation: Version 2.0.1" 2019.

[19] Arbeitskreis PLM, "Simulation Data Management: Integration of Simulation and Computation in a PDMEnvironment (SimPDM)" 2008.

[20] F. Rieg and R. Steinhilper, *Handbuch Konstruktion*: Hanser, 2012.

[21] *Information technology - Open Systems Interconnection: Basic Reference*, 7498-1, EN ISO/IEC, 1995.

[22] P. I. 4.0. IIC, "Architecture Alignment and Interoperability: An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper" 2017.

[23] Stack Overflow, *Developer Survey 2020.* [Online]. Available: https://insights.stackoverflow.com/survey/2020 (accessed: Jan. 25 2021).

[24] G. Schroeder *et al.,* "Visualising the digital twin using web services and augmented reality" in *Proceedings of the International Conference on Industrial Informatics (INDIN)*, 2016.

[25] V. Souza, R. Cruz, W. Silva, S. Lins, and V. Lucena, "A Digital Twin Architecture Based on the Industrial Internet of Things Technologies" in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, 2019.

[26] C. Liu, P. Jiang, and W. Jiang, "Web-based digital twin modeling and remote control of cyber-physical production systems" *Robotics and Computer-Integrated Manufacturing*, 2020.

[27] L. F. C. S. Durão, S. Haag, R. Anderl, K. Schützer, and E. Zancul, "Digital Twin Requirements in the Context of Industry 4.0" *Proceedings of the 15th IFIP International Conference on Product Lifecycle Management*, 2018.

[28] B. Ashtari Talkhestani *et al.,* “An architecture of an Intelligent Digital Twin in a Cyber-Physical Production System” *at - Automatisierungstechnik*, 2019.

[29] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “Characterising the Digital Twin: A systematic literature review” *CIRP Journal of Manufacturing Science and Technology*, 2020.

ACKNOWLEDGEMENTS

This work was funded by the Hessian LOEWE initiative within the Software-Factory 4.0 project.

- Return to the Beginning of this Article
- Return to the Table of Contents

Open Source Drives Digital Twin Adoption

Authors:

Erich Barnstedt
Chief Architect, Standards & Consortia, Azure IoT
Microsoft Corporation
erichb@microsoft.com

Birgit Boss
Standardization, Open Source and IP-Management,
Bosch Connected Industry
Robert Bosch GmbH
Birgit.Boss@de.bosch.com

Erich Clauer
VP, Industry Standards & Open Source
SAP SE
erich.clauer@sap.com

Dan Isaacs
CTO
Digital Twin Consortium
isaacs@digitaltwinconsortium.org

Shi-Wan Lin
CTO
Yo-i Information Technology, Ltd.
shiwanlin@thingswise.com

Somayeh Malakuti
Senior Scientist
ABB AG Corporate Research Center, Germany
somayeh.malakuti@de.abb.com

Pieter van Schalkwyk
CEO
XMPro Inc
pieter.vanschalkwyk@xmpro.com

Thiago Weber Martins
Industry 4.0 Standards & Open Source Architect
SAP SE
thiago.weber.martins@sap.com

INTRODUCTION

"Digital Twin" has been listed as an emerging technology during past years and more and more companies offer Digital Twin as part of a foundation for their digital offerings. However, there has not been consensus on what a Digital Twin is and what characteristics it should have. Some refer the origin of Digital Twin to an academic paper, where Digital Twin was defined as a combination of physical product, virtual product and their connections [1]. Some also claim that Digital Twin existed in industrial applications under different names[2]. One could observe that lack of a common definition for Digital Twin has led to many flavors of this concept, sometimes making readers wonder whether Digital Twin is more than just a buzzword at all. Several literatures discussed the definitions[3] [4] [5] and requirements[6] [7] that Digital Twins could have.

As the first step to clarify the role of Digital Twin in various sectors, various consortiums and committees such as the Industrial Internet Consortium (IIC)[8] and Plattform Industrie 4.0[9] each

---

[1] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," White paper, 2014. [Online]. Available: *http://www.apriso.com/library/Whitepaper_Dr_Grieves_DigitalTwin_ManufacturingExcellence.php*

[2] Somayeh Malakuti, Jan Schlake, Christopher Ganz, Eric Harper, Heiko Petersen, "Digital Twin: An Enabler for New Business Models", Automation 2019

[3] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and U.S. Air Force vehicles," in Proc. 53rd AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., 2012. [Online]. Available: https://arc.aiaa.org/doi/pdf/10.2514/6.2012-1818

[4] Stark R, Anderl R, Thoben K-D, Wartzack S. WiGeP-Positionspapier: „Digitaler Zwilling". ZWF 2020.

[5] Elisa Negri, Luca Fumagalli, Marco Macchi, editors. A Review of the Roles of Digital Twin in CPS-based Production Systems. Elsevier 2017.

[6] Moyne, James & Qamsane, Yassine & Balta, Efe & Kovalenko, Ilya & Faris, John & Barton, Kira & Tilbury, Dawn. (2020). A Requirements Driven Digital Twin Framework: Specification and Opportunities. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3000437.

[7] Digital Twin Requirements in the Context of Industry 4.0, Product Lifecycle Management to Support Industry 4.0, 2018, Volume 540, ISBN: 978-3-030-01613-5, Luiz Fernando C. S. Durão, Sebastian Haag, Reiner Anderl, Klaus Schützer, Eduardo Zancul

[8] IIC Industrial Internet of Things Vocabulary, https://www.iiconsortium.org/vocab/

[9] Digital Twin and Asset Administration Shell Concepts and Application in the Industrial Internet and Industrie 4.0 - An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper - https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/Digital-Twin-and-Asset-Administration-Shell-Concepts.html

provided a definition of this concept; likewise, a definition is provided by the Digital Twin Consortium (DTC)[10]. Followed by the definition, various architectural aspects that must be considered in the design of Digital Twin have been proposed[11]. These include, for example, information models, APIs, connectivity to physical twins, data ingestion mechanisms, security and interoperability of Digital Twins. Other organizations, such as the Open Industry 4.0 Alliance[12], focused on creating a network of manufacturers to drive industries into adopting the Digital Twin technologies of Industry 4.0 based on existing definitions and standards.

Each vendor may offer its proprietary Digital Twin solution. However, systems of systems, or, in general, complex systems, consist of components offered by different vendors. Therefore, individual Digital Twins need to eventually interact towards forming system-level Digital Twins and/or to enable cross-vendor interactions. Consequently, interoperability of Digital Twins (e.g., their meta model and APIs) in complex systems becomes inevitable.

To accelerate the development of Digital Twins within or across companies, there is an inevitable need for open standards for Digital Twins, as well as Open Source implementations to promote the agile development and adoption of those standards [13] [14] [15]. Alongside the IIC and Plattform Industrie 4.0, as the importance of this topic has grown, multiple new organizations such as DTC, Industrial Digital Twin Association (IDTA), and Open Manufacturing Platform (OMP) have recently been founded to develop these topics further.

This paper elaborates on the challenges in adopting the Digital Twin concept in practice and the motivations for having open standards and Open Source activities to address these challenges. In addition, the paper provides an overview of the activities of some organizations around these topics in more details. We discuss the challenges that will arise if there will be multiple open standards and Open Source implementations of Digital Twins.

[10] https://www.digitaltwinconsortium.org

[11] Industrial Internet Consortium, "Digital Twins for Industrial Applications White Paper", https://hub.iiconsortium.org/portal/Whitepapers/5e95c68a34c8fe0012e7d91b

[12] https://openindustry4.com

[13] https://www.din.de/resource/blob/65354/619baa1958b89b8a7b6cd9be2b79f223/roadmap-i4-0-e-data.pdf

[14] https://www.ietfjournal.org/open-standards-open-source-open-loop/

[15] https://www.ietfjournal.org/three-years-on-open-standards-open-source-open-loop/

CHALLENGES OF DIGITAL TWINS

The challenges to adopt Digital Twin in the industry can be summarized as follows:

- **Lack of a common definition and understanding of the Digital Twin concept:** The term "Digital Twin" has been coined in academia in 2003, and ever since, many different definitions and flavors of it have emerged, which usually focus on particular use cases in which a Digital Twin is adopted. However, there is also the view that the concept of Digital Twin has existed in industry for several decades, under different names. The lack of a common definition and understanding of the Digital Twin concept risks it being degraded to the level of a buzzword. Various consortiums such as IIC, Plattform Industrie 4.0 and DTC have tried to address these challenges.
- **Backend data integration:** Digital twin is a means for providing holistic access to the otherwise dispersed and siloed data. This requires architectures and solutions for integrating data within and across boundaries of organizations, dealing with heterogeneity of data sources, data formats, APIs and security rules.
- **Standardized information model:** To increase interoperability across data sources within and across the boundaries of organizations, standardized information, standardized meta-models, and APIs are required to describe Digital Twins. As for other industrial practices, companies may face the "standardization deadlock" problem, where on one hand there might be insufficient practical experience to drive standardization needs, and on the other hand the practical experiences might lead to proprietary solutions. This requires that architectures and solutions adopt the "on-demand interoperability", where proprietary Digital Twins can be translated to standard ones on-demand, upon the availability of standards.

Finally, as the fast-changing market is accelerating the innovation cycle, it is crucial to ensure the adoption of a new technology, such as the Digital Twin, as fast as possible. Otherwise, Digital Twin might lose its momentum in making an impact on the current cycle of digital transformation. Although current standardization efforts driven by multiple organizations around the globe aim at establishing the Digital Twin and its adoption in the market, the standardization processes are usually lengthy and slow. However, an aligned collaboration between these organizations could potentially accelerate this process and would lead to higher quality. To address this challenge, an open collaborative approach to leverage the synergies of these organizations and make the standardization more agile is important and beneficial to a faster adoption of Digital Twins. We introduce and discuss such an approach in the following sections.

DRIVING DIGITAL TWINS THE OPEN SOURCE WAY

In 2019, the Gartner Hyper Cycle listed Digital Twins as being on the "Peak of Inflated Expectations." Some still feel it is just a buzzword but the world is moving on. Digital twins now can be reflected on the "Slope of Enlightenment". In 2020, several new organizations were founded or kicked-off to establish Digital Twin technology in industrial settings. Hence, the aim is clear: Smooth the way to reach the "Plateau of Productivity."

A remarkable realization common to almost all of the Digital Twin related organizations is the consideration of Open Source development as an important pillar for promoting the faster adoption of Digital Twins.

These organizations, in addition to IIC, Plattform Industrie 4.0, and Open Industry 4.0 Alliance, are involved in the development and promotion of Digital Twins and are briefly described below. They include: the OMP, the DTC, GAIA-X, Clean Energy and The Smart Manufacturing Innovation Institute (CESMII) [16] and the IDTA. Although CESMII has been founded earlier, CESMII is additionally listed in this section because it recently started Open Source activities.

Furthermore, a remarkable trend in the general Open Source community is that, increasingly, they do not aim just for Open Source code development, but for Open Source code and open specification development together.

For the Linux Foundation and Eclipse, it was a learning curve to not just focus on Open Source code development. The Joint Development Foundation (JDF) that focuses on open specification development joined the Linux Foundation in 2018[17]. Eclipse defined its own specification process in 2019[18]. The Internet Engineering Task Force (IETF) also identified Open Source as an essential development model to overcome the challenges of the tedious traditional standards processes and keep them on pace with fast innovation cycles [14] [15].

OASIS Open Foundation follows a similar hybrid approach to support open collaboration for developing standards[19]. Hence, the understanding now seems to be that standardization and Open Source are coming closer together to be more effective in both developing and promoting open specifications, open standards and Open Source.

---

[16] https://www.cesmii.org/production-of-zero-defect-zd-slabs/

[17] https://www.linuxfoundation.org/en/press-release/jdf-joins-lf-family/

[18] https://www.eclipse.org/community/eclipse_newsletter/2019/january/EFSP_vs_JCP.php

[19] https://www.oasis-open.org

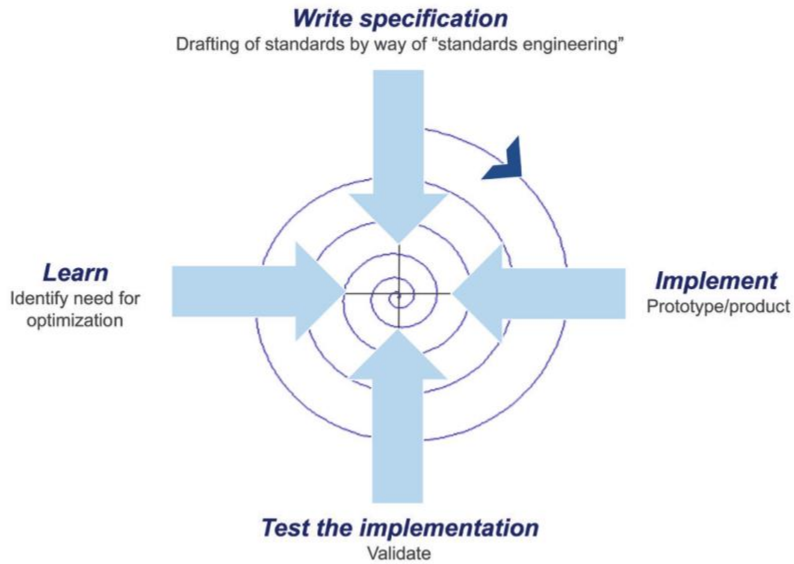


Fig. 1: Agile standardization using Open Source as a development model according to SCI 4.0 [13] *(with permission from DIN Deutsches Institut für Normung e.V.)*

In 2020, the Standardization Council Industrie 4.0 (SCI 4.0) that is creating the German standardization roadmap[13] claimed "Agile Standardization" to be the way to go. Agile Standardization means a joint development of the consensus-based standard development used in Standardization Development Organizations (SDO) together with Open Source communities. Thus, from both sides, faster innovations and reactions are possible: The standard is adapted from experiences gained during Open Source development and vice versa; the standard can be tested very quickly to see whether it really fulfills all the requirements. This is illustrated in Figure 1.

In joining the work of code and specification development, however, it is neither "code first" nor "specification first", but how to create synergy between the two parallel efforts by working very closely together. This is the approach of OMP[20], DTC, IDTA, GAIA-X, and CESMII.

Open Manufacturing Platform (OMP)

The OMP was founded in 2019 by Microsoft and BMW. Bosch, ZF and AB InBev joined at the beginning of 2020. OMP has a different strategy than IDTA, DTC, CESMII, and GAIA-X. OMP is a project under the umbrella of the Joint Development Foundation. It is the only initiative in this area set up as an Open Source project from the very beginning. The goal is to "accelerate

[20] https://open-manufacturing.org/

innovation in Industrial IoT to shorten time to value."[21] OMP aims for both published best practices and encourage open code development. It is not dedicated to Digital Twin technology only, but the Semantic Data Structuring working group, contributes to it. This working group addresses the "*needs to share, join and reuse heterogeneous data of the manufacturing domain by applying common semantics for various stakeholders through comprehensive semantic data homogenization and by conveying manufacturing data along with contextual information*."

The OMP GitHub repository provides open publications, specifications, and source code: https://github.com/OpenManufacturingPlatform.

Digital Twin Consortium (DTC)

DTC was formed under the umbrella of the Object Management Group (OMG), a standards organization that supports the development of open standards. The DTC is not a standards organization, but provides guidance to address market confusion, interoperability challenges, and managing the risk of implementing Digital Twins in industrial and commercial settings.

The DTC industry working groups have a vertical market focus and currently consist of Manufacturing, Infrastructure, Healthcare and Life Sciences, Aviation and Defense and Natural Resources groups.

The horizontal working groups include the Marketing and the Technology, Terminology and Taxonomy (3T) focus areas. The 3T Working Group has further subgroups that focus on Security and Trustworthiness, as well as Open Source and Platform Stacks.

A primary focus for the DTC is the overall Digital Twin lifecycle and the different requirements of a Digital Twin as it evolves through its lifecycle. The Digital Twin lifecycle includes the digital thread, as well as that of the physical entity or asset that it supports.

The DTC provides an Open Source repository where vendors and end users can contribute various elements to accelerate the adoption of Digital Twins. Contributions may include Open Source code implementations, collaborative documents for guidance and training, Open Source models, or other assets that are of value to the Digital Twin community.

The cross-pollination by the different industry-focused working groups provides the opportunity for different industries to learn from each other and improve collaboration around architectures, tools and open-source contributions.

Industrial Digital Twin Association (IDTA)

[21] https://azure.microsoft.com/en-us/blog/the-future-of-manufacturing-is-open/

The IDTA[22] was founded in September 2020[23] by Verband Deutscher Maschinen- und Anlagenbau[24] (VDMA)[25] – the largest network organization in the European mechanical engineering industry – and Zentralverband Elektrotechnik- und Elektronikindustrie[26] (ZVEI)[27] – the German Electrical and Electronic Manufacturers' Association - together with Bitkom[28] – Germany's digital association - and 20 companies: ABB, Asentics, Bitkom, Bosch, Bosch Rexroth, Danfoss, Endress Hauser, Festo, Homag, KUKA, Lenze, Pepperl Fuchs, Phoenix Contact, SAP, Schneider Electric, Schunk, Siemens, Trumpf, Turck, Volkswagen and Wittenstein.

The goal of the IDTA is to bring the Digital Twin to the next level, make is enterprise ready by utilizing Open Source development model and to provide a one-stop-shop for the industrial Digital Twin, known as the Asset Administration Shell in Industrie 4.0. Five areas of focus were identified: Open Technology, Integration, Quality Management, Training and Marketing.

Besides specifications of Digital Twins[29], with focus on cross-company interoperability ("Open Technology"), the IDTA will also push the further standardization of so-called submodel templates ("Integration"), i.e., standardized (domain) knowledge needed for different use cases and data-driven business models that can be shared via the standardized API of the Asset Administration Shell. In addition to genuine submodel templates for Asset Administration Shells, existing and future OPC UA Information Models[30] will also be considered as a basis.

Existing Open Source activities in the context of the Asset Administration Shell will be supported under the umbrella of IDTA. There are already several Open Source projects in place that implement the Asset Administration Shell (AAS):

- **admin-shell-io** (https://github.com/admin-shell-io): This project hosts an open editor with a graphical user interface for Asset Administration Shells (AASX Package Explorer),

[22] www.industrialdigitaltwin.org

[23] https://www.zvei.org/en/press-media/pressarea/user-organization-industrial-digital-twin-association-founded

[24] German: Mechanical Engineering Industry Association

[25] http://www.vdma.org/

[26] German: electrical and electronic manufacturers' association

[27] https://www.zvei.org/

[28] https://www.bitkom.org/EN/EN

[29] https://www.plattform-i40.de/PI40/Redaktion/EN/Standardartikel/specification-administrationshell.html

[30] https://opcfoundation.org/developer-tools/specifications-opc-ua-information-models

an AAS Server, AAS specifications as well as schemata (aas-specs), training screencasts, and FAQs.

- **BaSyx** (https://projects.eclipse.org/projects/technology.basyx): This project hosts - besides other modules - SDKs or the Asset Administration Shell.
- **PyI40AAS** (https://git.rwth-aachen.de/acplt/pyi40aas): This project hosts a Python module for manipulating and validating AAS.
- **SAP AAS Service** (https://github.com/SAP/i40-aas): This project hosts a system based on Docker images implementing the RAMI 4.0 reference architecture (including AAS).
- **NOVAAS** (https://gitlab.com/gidouninova/novaas): This project provides an implementation of the AAS concept by using JavaScript and Low-code development platform (LCDP) Node-Red.

Building an attractive and active Open Source community based on these existing and new activities is one of the important goals of the IDTA.

GAIA-X

GAIA-X[31] was officially founded in 2020[32]. GAIA-X is a European project aiming at developing common requirements for a European data infrastructure. Therefore, interoperability, transparency, and openness are key factors for success. Digital twins will play an enabler role for data exchange across company borders. This is illustrated in the use case “Collaborative Condition Monitoring”[33]. Open standards and its Open Source derivatives such as the one developed within the Industrial Data Spaces (IDS)[34] are playing a major role in GAIA-X.

CESMII

CESMII, the Clean Energy and Smart Manufacturing Innovation Institute, was founded in 2016. In January 2021 CESMII and the IIC agreed to a liaison to “accelerate the development, adoption, and monetization of Industrial IoT (IIoT) technologies, infrastructure and solutions to deliver

[31] https://www.data-infrastructure.eu/GAIAX/

[32] https://www.isst.fraunhofer.de/en/news/press_releases/2020/PI_GAIA-X_foundation.html

[33] https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/collaborative-condition-monitoring.html

[34] GAIA-X and IDS Position Paper: https://internationaldataspaces.org/download/19016/

transformative business value for manufacturers through digital transformation"[35]. One of its goals is to realize interoperability via collaboration and standardization. Digital twins play an important role in this effort. The focus in this area is on so-called smart manufacturing (SM) profiles. An OPC UA information model serialization was chosen for these profile specifications.

Together with the OPC Foundation, CESMII is working on the UA Cloud library[36]. The goal is to provide open data models via a globally hosted cloud database. These data models shall be useable in Digital Twin implementations. SM Profiles, a code sample for the CESMII API, and more are available in CESMII GitHub repository (https://github.com/cesmii).

Comparison of Activities and Organizations

As the organizations introduced above share a similar goal – provide the foundations for an interoperable Digital Twin – they differentiate themselves in the focus of their activities as well as how "open" they are. Table 1 summarizes the activities of the organizations mentioned in this paper, highlighting their main contribution and focus[37].

[35] https://www.automation.com/en-us/articles/january-2021/industrial-internet-consortium-smart-manufacturing

[36] https://www.automation.com/en-us/articles/october-2020/new-international-initiative-opc-ua-cloud-library

[37] Note: The authors acknowledge that the metrics introduced in Table 1 are purely based on heuristic evaluations with the object to estimate the overlap of the activities within a criterion. The authors stress that no statistical analysis have been carried out.

Table 1 Overview of Organizations Driving the Digital Twin Technology

Organization	General Guidance	Focus on Digital Twins	Focus on Manufacturing Application	Open Code	Open Specification	Open Development Model
OMP	○	◔	●	●	◑	●
IDTA	○	●	●	●	●	●
DTC	●	●	◑	●	○	●
GAIA-X	○	◔	◑	●	●	●
CESMII	○	◔	●	●	●	●
IIC	●	◔	◑	○	○	○
Plattform Industrie 4.0	●	◔	●	○	●	○
Open Industry 4.0 Alliance	○	◔	●	○	○	○

Legend:
○ None / no activity
◔ Few activities
◑ Many activities
◕ Most activities
● All activities

From all organizations evaluated in this paper, IDTA and DTC are the ones that exclusively focus their activities on Digital Twins. However, while IDTA focuses on manufacturing applications of Digital Twin, DTC also considers its wider usage across different vertical domains spanning from natural resources to infrastructure, healthcare, and life science. Both are covering full deployment with security and trustworthiness from edge to cloud. Other organizations such as OMP, GAIA-X, CESMII, Open Industry 4.0 Alliance, IIC, and Plattform Industrie 4.0 do not limit their activities around Digital Twins, but also consider topics such as Cloud, Edge, data sovereignty, etc. Of these organizations, only the IIC and GAIA-X has a wider focus beyond manufacturing.

Regarding the development model driving the activities within these organizations, we recognize an increasing trend for open, collaborative models using Open Source as an approach to drive interoperability and standardization. As mentioned above, the newly funded organizations OMP, IDTA, DTC, GAIA-X, CESMII are publishing the outcomes of their activities in repositories that are open for the community. In particular, their codes are available in repositories such as GitHub with the corresponding Open Source Licenses (e.g. Apache v.2.0, Eclipse v2.0, Mozilla Public License (MPL), MIT). Additionally, organizations such as OMP, IDTA, GAIA-X, CESMII are also “open sourcing” their specifications. Both lead to an open development where the community can both access the codes or the specifications and provide feedbacks on them as well. Furthermore, if the Open Source license allows, they can use, modify and share their contributions.

Organizations such as OMP, IDTA and DTC have recognized that the Open Source development model provides additional capabilities to simplify and to accelerate the specification and standardization process of new technologies. The Open Source model adds agility, quicker cycles of evolution and open collaboration, and vendor-neutral governance to the development of standards. In particular, as more information and communication technologies are embedded into products (e.g., industrial machines and parts), the demand for software and communication protocols are increasing. The concept of Digital Twin represents a suitable example of this paradigm shift and, accordingly, the concurrent development of standards and code via the Open Source model is a perfect match. Other technologies, such as Cloud Computing, already have proven that Open Source played an essential role in their successful adoption (e. g. Cloud Foundry[38], Kubernetes[39], Gardener[40]) in the market.

The Open Source activities of DTC and IDTA will provide valuable contribution via reference implementations for existing use cases and standards defining Digital Twins. As shown in Figure 1, the Open Source implementations will provide a validation platform to test and to optimize respective standards and specifications. Especially, the Open Source community can provide direct feedback to the projects and new requirements as well as improvement suggestions by creating issues or pull requests in the corresponding repositories. For example, Plattform Industrie 4.0 has published several documents specifying the details of Asset Administration Shell and submodel templates that can be used as a reference for other organizations in their activities in the last few years and is continuing to do so. At the same time, these organizations could contribute to the newer versions of these documents based on their knowledge gained through the activities around Open Source projects or transferring the Asset Administration Shell specifications and Open Source codes to other applications beyond manufacturing. While IDTA is focusing on manufacturing applications of Digital Twins and has already decided which standard to use as basis for further development and improvement, DTC is looking across

[38] https://www.cloudfoundry.org

[39] https://kubernetes.io

[40] https://gardener.cloud

different domains and does not recommend single standards but aims to give advice to enable clearer understanding of the current concepts of Digital Twins and available standards and technologies. DTC contributions seek to include Open Source code implementations, collaborative documents for guidance and training, Open Source models, or other assets of value to the Digital Twin community.

Additionally, the Open Source projects of the organizations OMP, CESMII, and GAIA-X will also be providing valuable contributions for the Digital Twins. OMP will be contributing with comprehensive semantic data homogenization for Digital Twins in the manufacturing context. Also, in the manufacturing context, CESMII efforts on the OPC UA Cloud Library, by providing open data models, is a crucial contribution to enable interoperable machine-to-machine (M2M) communication via Digital Twins. In a wider context, GAIA-X will contribute with use cases and implementations of components for Digital Twins to fulfill security and data sovereignty requirements.

Analogous to the modern development approach for software applications, interoperable Digital Twins will be assembled using the components developed within Open Source communities. Enterprises and vendors will be able to combine these Open Source components with their proprietary technologies to build applications and systems that make a differentiation in the market and that provide added value to their customers.

CONCLUSION

The development of standards for Digital Twin is an imperative for setting the necessary foundation to ensure its successful adoption in the market. However, it is clear that an effort based solely on establishing open standards is not enough. In particular, as in the case of Digital Twins and Industry 4.0 / IIoT, where machines are connecting to and converging into the IT world, the demand for enabling software is increasing. Adequate, widely supported and widely adapted Digital Twin Open Source software can establish de facto standards for the underlying architecture of Digital Twins. To keep the pace with faster innovation cycles, it is crucial to leverage the opportunity made available by Open Source as a development approach that has been raised in several works [41] [42] [43]. In this context, driving Digital Twin development with the converged Open Source and open standard approach brings significant benefits, including:

- higher co-innovation and interoperability due to better collaboration between organizations, vendors, end-users and communities,
- faster validation of the specifications due to agile development of open standards by providing production-ready code,
- faster adoption of the technology due to the establishment of de facto standards,
- scalability of reference implementations into interoperable products for Digital Twins, and
- better cost-effectiveness and faster responsiveness to market demands and changes.

More than ever, the convergence of open standards and Open Source is crucial to enable interoperable Digital Twins since this convergence provides an open platform capable of dealing with technological complexity and market dynamics. Following the approach shown in Figure 1, early versions of the specifications, standards, and code of reference implementations are available to the Open Source community. Based on the feedback gained during the agile development, specifications, standards, and code of reference, implementations can achieve their necessary maturity level faster. As a result, the development cycles necessary to develop open standards become shorter. These open standards and specifications can be potentially submitted to ISO, IEC, DIN, or other SDOs. It is expected that the processes within these SDOs are also accelerated.

However, one must acknowledge the challenges that arises when multiple organizations provide different open standards and Open Source implementations for Digital Twins. Whether in an open collaborative model with the community or not, several organizations are developing specifications for Digital Twins and the information they provide is shaping their foundations.

These different organizations seem overlapping and competing at a first glance, and it may become a challenge for them to position themselves with each other and provide clear direction to their stakeholders. Therefore, the setup of collaborations and liaisons to align and complement activities between these different organizations are highly desirable for the good of the general community. Otherwise, there is the risk that the goal of faster implementation of

---

[41] https://iot.eclipse.org/community/resources/white-papers/pdf/Eclipse%20IoT%20White%20Paper%20-%20Open%20Source%20Software%20for%20Industry%204.0.pdf

[42] https://plc4x.apache.org/users/industry40.html

[43] https://openforumeurope.org/publications/standards-and-open-source-bringing-them-together/

interoperable Digital Twin remains elusive. As a result, fragmentation of the market due to lack of interoperability and confusion regarding "which standards should I use?" can be expected. Additionally, the complexity of Digital Twins might increase if all competing standards are needed, although they provide the same value. Similar to multiple organizations investing their efforts in defining and specifying similar standards, different Open Source projects aiming at the same or overlapping solutions might lead to confusion in the market and fragmentation of the technology. Moreover, different Open Source projects would miss the opportunity of having a sufficient number of active contributors leveraging the quality of the code instead of creating similar solutions competing with each other.

Going beyond setting up collaboration and liaison to align activities across organizations, the following recommendations could be explored in future discussions:

- A more detailed assessment of individual Open Source projects driven by the organizations evaluated in this paper could be addressed in future works (similar to the joint paper of IIC and Plattform 4.0 on the Digital Twin and AAS topic[9]). Based on this detailed assessment, a mapping of Open Source projects based on the identification of overlapping and unique features following predefined criteria (e.g. technology being used, open standard being implemented, manufacturing focus, etc.) of these projects could be performed. Using this mapping, activities and efforts of Open Source projects could be closely aligned in terms of their overlapping elements and refocused on the unique challenges they are solving. At the same time, such mapping can serve as a guide for the contributors as well as for users of the Open Source projects as to which Open Source project to contribute, or which Open Source Project should to be integrated in the products.
- Open source can be tested in testbeds to validate their applicability and interoperability by building, for example, Proof of Concepts (PoCs). Such testbeds provide a neutral environment where members, developers, and community exchange their experiences and align their activities.

Finally, in developing and promoting interoperable Digital Twin, it is neither "code first" nor "specification first", but the joint effort and synergy between the two parallel developments that will bring the expected accomplishment in the context of IIoT / Industry 4.0.

Disclosure

Authors of this paper are members of the organizations classified in this whitepaper: IIC, DTC, GAIA-X, IDTA, CESMII, Open Industry 4.0 Alliance, or/and Plattform Industrie 4.0.

ACKNOWLEDGEMENTS

The views expressed in the *IIC Journal of Innovation* are the contributing authors' views and do not necessarily represent the views of their respective employers nor those of the Industrial Internet Consortium.

- Return to the Beginning of this Article
- Return to the Table of Contents

industrial internet CONSORTIUM

Design and Implementation of a Digital Twin for Live Petroleum Production Optimization: Data Processing and Simulation

Authors:

Venkat Putcha
Director of Data Science
OspreyData Inc.
venkat.putcha@ospreydata.com

Nhan Le
Data Scientist
OspreyData Inc.
nhan.le@ospreydata.com

Jeffrey Hsiung
Principal Data Scientist
OspreyData Inc.
jeff.hsiung@ospreydata.com

INTRODUCTION

High-frequency live connection to data feed from sensors on equipment offers several benefits such as facilitating human surveillance for better asset management, identification and diagnosis of abnormalities and suboptimal operation. However, normal operation does not necessarily imply optimal operation. To obtain optimal production, the asset must be operated at appropriate set-points. For dynamic assets or systems, the optimum set-point changes with time. In such cases, the live data feed associated with the equipment's IIOT network can be harnessed to develop a dynamic set-point optimization mechanism.

Existing literature on Petroleum Production optimization on Artificial lift wells set-points is heavily focused on manual simulation, design and recommendation by experts, or through semi-automated batch implementation of Evolutionary[1], statistical or machine learning models. Current literature on Digital Twin implementations[2] in Oil and Gas present a broader picture on overall production process optimization[3], but not on dynamic individual asset level set-point optimization. Fully automated set-point recommendation requires a data processing engine integrated with a simulation engine that can manage, process and generate large volumes of data. Current literature does not provide design details of critical individual components to implement a fully-automated data processing and simulation engine. This paper attempts to address this deficiency.

Digital Twins representing systems of assets can be valuable in determining optimal operating set-points. The integration of live IIOT data-feed with a Digital twin system offers several challenges and requires a detailed design for effective implementation. The following aspects of the implementation are detailed in this paper: Data processing: profiling, clean-up, transformation and cloud-database maintenance for multiple assets, high frequency data Simulation: Automated cloud-database triggered field data relevant massive scale simulation (60,000 + per day).

---

[1] Garcia, Artur Posenato, and Vinícius Ramos Rosa. "A Genetic Algorithm for Gas Lift Optimization with Compression Capacity Limitation." Paper presented at the SPE Latin America and Caribbean Petroleum Engineering Conference, Mexico City, Mexico, April 2012.doi: https://doi.org/10.2118/153175-MS

[2] LaGrange, Elgonda "Developing a Digital Twin: The Roadmap for Oil and Gas Optimization." Paper presented at the SPE Offshore Europe Conference and Exhibition, Aberdeen, UK, September 2019.doi: https://doi.org/10.2118/195790-MS

[3] Okhuijsen, Bob, and Kevin Wade. "Real-Time Production Optimization - Applying a Digital Twin Model to Optimize the Entire Upstream Value Chain." Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019.doi: https://doi.org/10.2118/197693-MS

The specific use case chosen in the paper for showcasing the methodology is related to set-point changes on the artificial lift[4] equipment of the well for optimizing the production system. Examples of such artificial lift equipment include: Electrical Submersible Pumps (ESP), Rod Pumps, Gas Lift, and Plunger Lift.

DIGITAL TWIN SCHEMATIC

The design and implementation of a digital twin for dynamic set-point optimization on a petroleum production system using live IIOT data consists of several steps. These are:

1. Field Data processing: Collection, profiling, clean-up, transformation and cloud-database maintenance
2. Simulation: Automated cloud-database triggered field data relevant simulation
3. Inverse modeling:
    a. Connecting real-world IIOT data with simulations to learn system unknowns
    b. Evaluation: Estimate how closely the digital twin mimics the real-world asset from history
    c. Calibration: Implement initial steps using insights from digital twin to account for uncertainty
4. AI Model Recommendation: Deploy automated recommendations for set-point adjustments with updates based on dynamic trending of the asset

This paper focuses on the automation of the first two items of this process: Data Processing and Simulation. These steps are described in the context of feeding an AI engine that further consists of inverse modeling and model generated recommendation system. The details of the Inverse modeling and AI model recommendation components are beyond the scope of this paper.

Figure 1 represents a schematic of the overall process. It is important to note that this is a closed-loop ongoing process and not a feedforward sequence of steps that ends in a recommendation. This distinction is important for two main reasons:

a. After a model-generated recommendation has been implemented, an effective digital twin that is a live virtual representation of a physical system needs to identify changes in operating state, record and evaluate the response and trigger an ongoing cycle involving data processing, simulation, inverse modeling to adjust the system in case if the previously provided recommendation needs to be followed up with a new recommendation.
b. The digital twin can evaluate the impact of all historic set point changes and fine-tune the recommendation system.

[4] https://www.rigzone.com/training/insight.asp?insight_id=315&c_id=

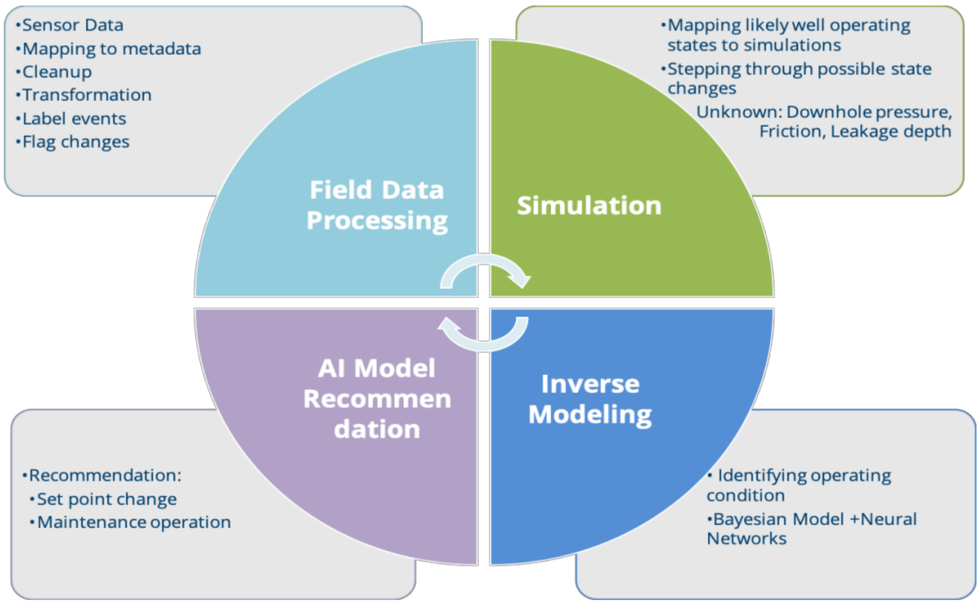


Fig. 1: Schematic of the closed-loop process providing an overview of the Digital Twin.

THE CURRENT STATE

Before going further into the details of the Digital Twin implementation, it is necessary to understand the scope and potential impact of this paper. In this section, it is attempted to set a baseline by describing the typical current state of operation in the Oil and Gas Industry.

The above described process in *Figure 1* is already implemented widely in the oil and gas industry, albeit, every step in the process is performed manually by subject matter experts, and on a well by well basis. The manual process is time-intensive, it takes several hours to implement it on one well at a given point of time for data collection, processing, simulation, inverse modeling and generating a set-point recommendation using a history matched model. Petroleum wells are dynamic entities that change their underlying operating conditions over time.

Further, wells are subjected to discontinuities in behavior due to design changes, workovers, re-stimulation and impact from nearby operations such as hydraulic fracture hits. Set-point reviews or changes are required as the well behavior changes. If diligently performed, the time investment required to optimize set-points is approximately 7-10 days per well per year. Due to the significant time investment, typically, the simulation based set-point optimization is employed semi-annually or annually per well, or when there has been a redesign or a workover. The details of the typical procedure are represented in the Manual Optimization section of *Figure 2*.

The automated version of the manual optimization results in dynamic optimization. Such a system can go through the entire process from data collection through to set-point recommendation for all wells on a continuous basis. Changes in well operating states are programmed to be detected, and the underlying models that are used to generate the set-point recommendations get updated with changing well conditions.

The term "automation" in this paper refers to the implementation of a system designed to minimize human intervention by identifying, templatizing, storing, scheduling and executing repeatable processes. As described above, it is feasible to generate manual set-point recommendations, however, to manually update and generate simulations every time there is a change in the system for a field consisting a few hundred wells becomes intractable. As a result, thumb-rules or intuition based set-point optimization takes the forefront position when compared to physics-based-modeling.

The scope and intent of this paper is to describe a system capable of automating massive scale simulation and the associated data processing using a cloud based distributed system. Such a system stores, transfers and processes data, and subsequently queues, executes and stores the results from thousands of simulations relevant to several hundreds of wells on an ongoing basis.

The expected impact of the paper is to provide motivation to set up pipelines for automating components or entirety of the data processing, simulation, inverse modeling and set-point recommendation workflow in the Oil and Gas industry and other analogous spaces capable of utilizing an IIOT network for developing a digital twin as defined in the introduction section of this paper.

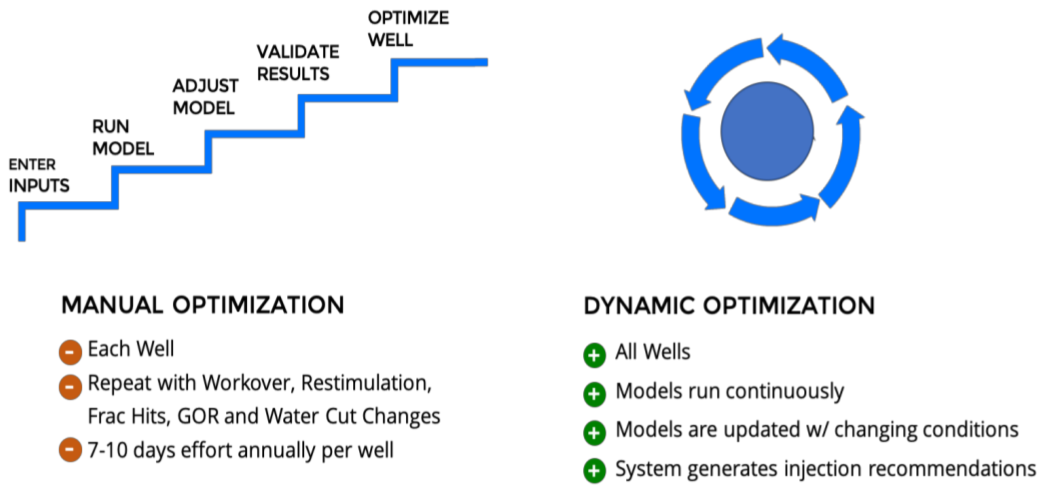


Fig. 2: Comparison between manual optimization (current state) versus dynamic optimization that can be achieved through the digital twin.

Data Collection

The first component of our digital twin as described in *Figure 1* is: Field Data Processing. Field data contains variety, *Figure 3* highlights this variety through some examples.

In regards to field data, there are two primary types of data: **sensor data** and **metadata**.

1. **Sensor Data:** The live signal from various sensors on assets was collected by a SCADA (Supervisory Control and Data Acquisition) system. Some general examples of sensor data include:
    - pressures, flow rates, temperatures for the well site, surface facilities and well downhole.
    - Other examples specific to artificial lift equipment may include details such as: Pump frequency, voltage, compressor discharge and intake pressures.

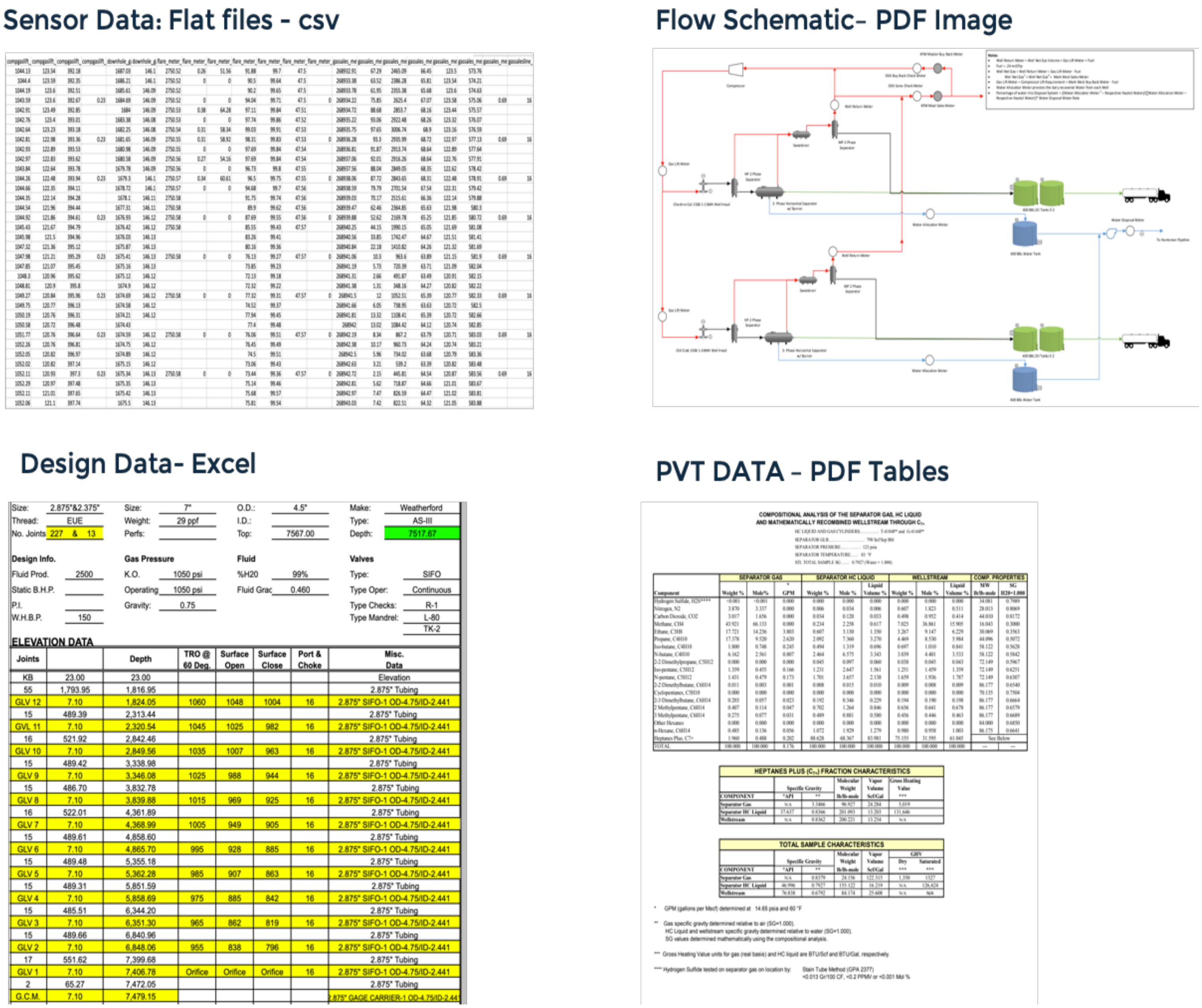


Fig. 3: Data variety – Sensor Data, Metadata.

These sensor signals were collected through wellsite telemetry and dropped onto a cloud based storage by the well operator's SCADA provider. This data was typically accessed by the operator for monitoring and surveillance. The sensor data is universally accessible once access credentials are shared for data transfer. Such data is time-series data and is typically transferred in files titled after a time-stamp, and containing the sensor signal for all the assets during a chunk of time. It might be worth noting that the sensor streams can be intermittent due network or telecommunication issues, delivered out of order or on a time delayed basis. These time-specific files can be aggregated and restructured to tables as suited by the user. In this particular case, this data was converted to tables indexed by asset ID and timestamp, with corresponding sensor signals stored in columns. An example of such a file is displayed in Figure 3, in the image titled "Sensor Data: Flat files – csv".

2. **Metadata:** To further represent the system which the sensor data is associated with, the metadata of the system is necessary. Examples of metadata can include but not restricted to:
    - Name, location, deviation and completion data of the well
    - Design data for the artificial lift equipment
    - Piping and Instrumentation schematic of the well and facilities sites
    - Fluid properties: PVT (Pressure, Volume, Temperature) data

   Figure 3 highlights how the metadata may be available in various formats such as images, excel sheets, PDF documents etc. A key component of automation is to digitize the various forms of data into a uniform data store. This involves creating a template that records the quantitative details of the metadata. An example of such a digitized template of the metadata is displayed in *Figure 4*. Once digitized, metadata is extractable through an API as a hierarchical data format such as JSON.

   It is important to note that metadata is not usually available at a single location or from a single source. There is a significant amount of manual effort that goes into contacting the field operators, the SCADA company, the equipment manufacturers to gather this data and to digitize it. A uniform reference data store space was created on the cloud to maintain and access the raw metadata as a source of truth.

   Often metadata is manually entered into the system because it is associated with the well when it is installed and does not change afterwards. When there is a maintenance event that changes a physical equipment of the well, the metadata may need to be updated depending on the changes being performed.

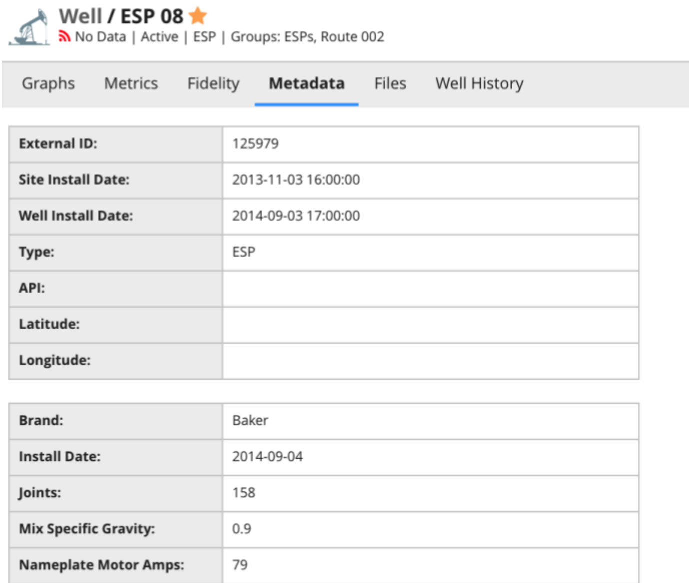


Fig. 4: Template for Digitized well metadata.

DATA PROCESSING

The sensor data and metadata were recorded and digitized from various sources. Further the data processing steps subsequent to this includes:

- Mapping data
- Profiling
- Cleanup
- Transformation
- Labeling events
- Flagging changes

Mapping: As described in the previous section, Sensor data and metadata come in different formats and are stored in different formats. Similar devices from different providers or SCADA systems often have different naming schemes, these sensor tags need to be mapped to their appropriate devices and assets. After the name mapping the time series sensor data is mapped to its asset metadata by merging the two data sources by asset ID.

Data profiling: After mapping sensor data and metadata, sensor data will be going through an EDA (Exploratory Data Analysis) process which we call as data profiling step. In this step, we look into some specific properties of well's sensor data to determine if it meets our model's requirements and in result we can create a cohort of wells which have sufficient data properties to be used in our models. Those properties are including but not limited to Monitored Time, Lapse Time/Monitor Time ratio, Zeros ratio, and Frequency. Monitor Time quantifies how much historical data that a well has. Lapse Time is defined as a period of time we don't receive any signal from the well, therefore Lapse Time/Monitor Time ratio is to determine the proportion of data availability over the life of a well. Similar to Lapse Time/Monitor ratio, Zeros ratio is used to see the actual variance in data. Frequency of data is also a key factor to the success of our models, the more granular data we have the better our models' performances are. A set of thresholds will be applied on each of the properties mentioned above. Sensors that meet these thresholds will be examined to see if they are sufficient for our models. Wells that include these qualified sensors are included in a cohort.

An example of Monitored Time threshold is shown below where Monitored Time threshold is set at 90 days which consequently reduces the number of wells in the cohort from 217 to 56 wells.

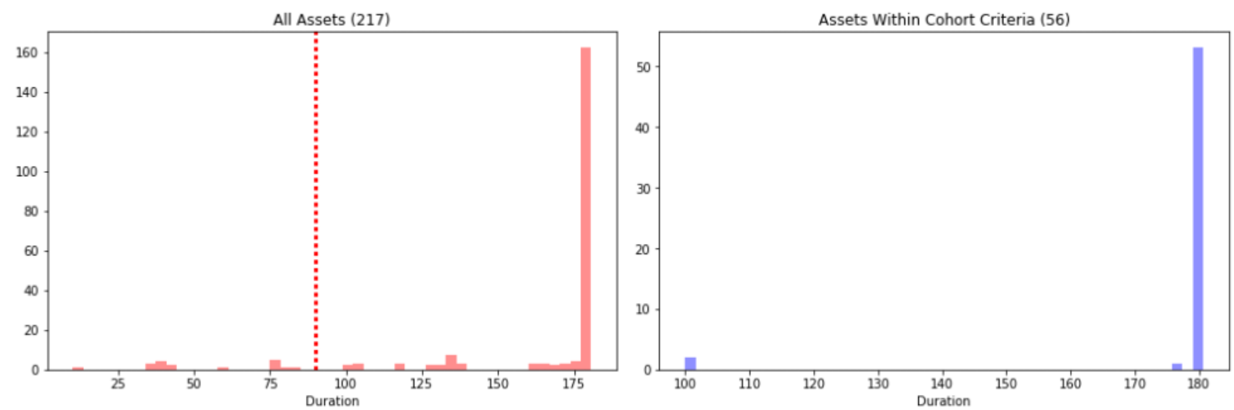


Fig. 5: Cohort selection based on monitored time.

Cleanup: Data coming from several wells over a period of time may include various challenges such as abnormal states of operation of the well (for example: shut-in, maintenance job, inconsistent performance), faulty signal, signal lapses, signal names that are inconsistent from convention. It is necessary to identify these inconsistencies. Some of these can be identified and eliminated systematically, such as identification of dummy values of a signal, removal of outliers, or identifying that a well is shut-in. There are other abnormalities that require human review, for instance, a signal that was named inconsistently, or a metadata element that does not fit the template.

Transformation: It is not beneficial to evaluate meaning out of data, or generate simulations at the frequency of live data. Signals may be recorded at different timestamps and variable frequencies. In the work presented in this paper, data was resampled to a daily frequency to match the frequency of the production rates of the wells. Further, the variation related features such as the oscillation frequency/wavelength, level of stability when compared to the normal signal was captured through a rolling normalized coefficient of variance. Unstable states were treated separately from stable states of operation. *Figures 6 and 7* respectively represent the use of box plots, and normalization of data, for outlier detection and identification of stable states.

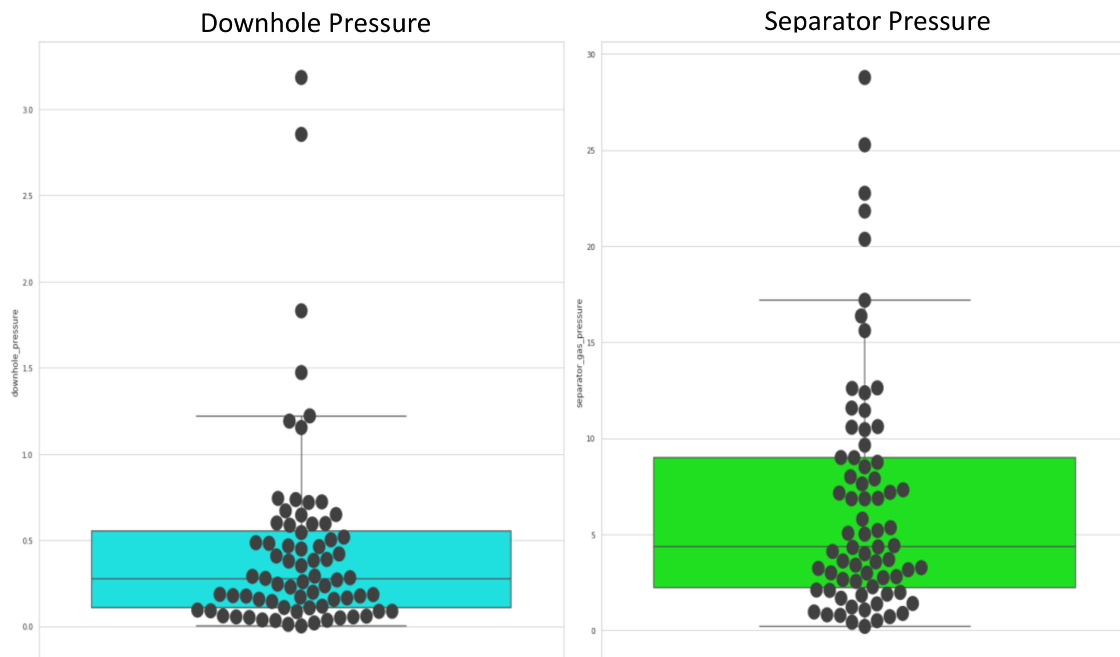


Fig. 6: Box plots indicating distribution of change in measured variables as a result of set-point (gas injection rate) changes.

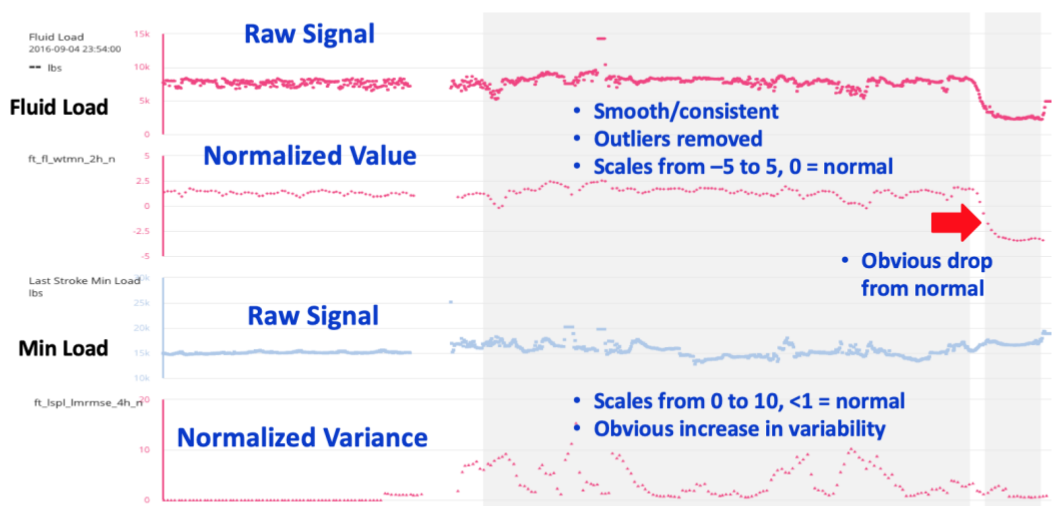


Fig. 7: Identification of stable operating states through normalization.

Labeling Events: When assets undergo gradual or prolonged states of abnormal behavior, the normalization filters may not be sufficient to identify and isolate such time zones. A monitoring platform that facilitates labeling can be a great tool. In the use case presented in this paper such a platform was used for Identification of events of interest by subject matter expert review. This method can be fast, effective and multipurpose. Data associated with the timestamps of the label can be easily extracted and utilized for further analysis, and supervised machine learning models can be trained for detecting such complex anomalies.[5] *Figure 8* shows an image of such an expert reviewed label to identify abnormal time.

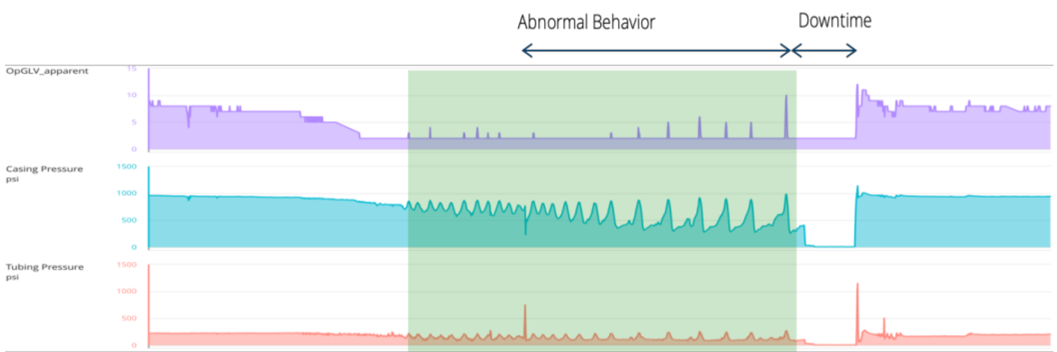


Fig. 8: Labeling of abnormal events.

Flagging changes: To have a closed loop system for generating set-point change recommendations, it is imperative record historical and life set-point changes in the system to identify and evaluate the response to the stimulus. An illustration of recording flagging changes and their responses is presented in *Figure 9*. The Gas Injection Rate displays the set point value, Oil and Water sensors represent the responses measures to a set-point change. In *Figure 9*, the periods across set-point change indicated by greyed zones are marked such that tubing and casing pressures are operating in a stable state. Measuring the response of a set point change while the well is operating abnormally leads to incorrect evaluation.

Flagging changes has its similarities and differences when to identification of abnormalities. The difference is that abnormalities are usually not human controlled and are usually unintentional. The similarity is that both set-point change and abnormalities indicate a change in operating state. These can be identified through supervised methods such as labeling followed by machine learning models, and unsupervised methods such as measuring deviation beyond a threshold. The approach in this paper is a supervised "human-in-the-loop" approach, where set-point recommendations generated by the system are monitored and reviewed by the operator prior

[5] Pennel, Mike, Hsiung, Jeffrey, and V. B. Putcha. "Detecting Failures and Optimizing Performance in Artificial Lift Using Machine Learning Models." Paper presented at the SPE Western Regional Meeting, Garden Grove, California, USA, April 2018. doi: https://doi.org/10.2118/190090-MS

to implementing the change. This is also a step towards expert augmented machine learning[6], where the feedback provided by the expert on the model generated recommendations is utilized to adapt the model to provide improved recommendations in subsequent rounds.

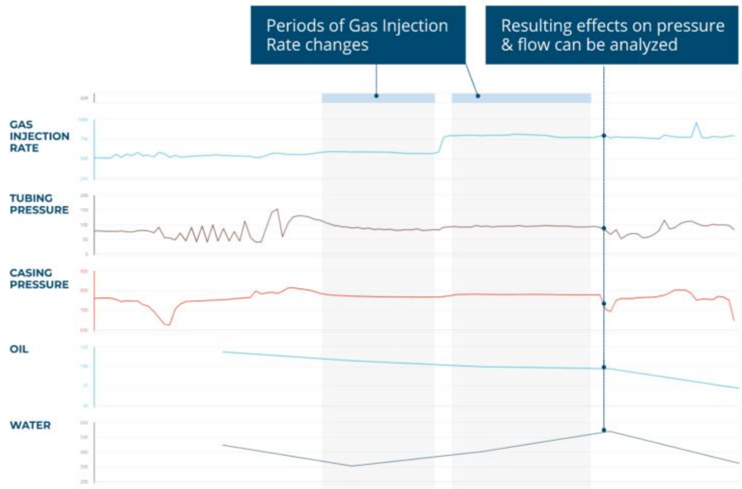


Fig. 9: Flagging changes and recording the response to the stimulus

SIMULATION

Having covered the collection and processing of physical data, in this section we proceed to describe the methodology for automated generation of virtual data. Simulation represents the virtual data generating component of the digital twin. The oil and gas optimization literature is rich in description of simulation based artificial lift and gas lift set-point optimization approaches. The paper by Rashif et al.[7] summarizes a survey of different gas lift optimization techniques using simulation as a basis. Borden et al.[8] presented a surveillance and workflows-based approach for gas lift optimization. Surendra et al.[9] presented further progress in the field by automating combining the analytics and physics-based modeling/simulation approaches through a case study

[6] arXiv:1903.09731

[7] https://doi.org/10.1155/2012/516807

[8] https://doi.org/10.2118/181094-MS

[9] https://doi.org/10.2118/201298-MS

in a Middle east oilfield, this study included the approach to simulating 50 wells in less than two weeks' time.

The work presented in the current paper takes a step further any creates a live-connection between field IIOT sensor data and a cloud-based simulation system that can generate 60,000+ simulations per day taking input from field data. This coupling between the physical and virtual data represents a key component of the digital twin. An effective digital twin is expected to represent and mimic a physical system. Hence, the simulation scheme described in this paper is closely connected to the physical data to maintain relevance.

The simulation section in this paper covers the following topics:

- Simulation Schematic
- Simulation Input
- Simulation Output

Simulation Schematic:

A commercial transient physics-based simulation software was employed to represent the fluid flow from the oil and gas reservoir through wellbore and the surface facilities that include the separation system and pipelines in order to transport fluids.

Figure 10 represents the process flow diagram of the physical setup and its corresponding simulation setup. Well and surface facilities schematics may vary among operators, fields, even between well pads. To generate a new exact simulation schematic for every variation in physical process flow is a very time-intensive process. Hence, for templatization purposes, the simulation scheme has been simplified to include only those components that are relevant to the objective. For example: in the process flow diagram in *Figure 10*, there are multiple tanks per fluid, this is reduced to a single tank per fluid in the simulation schematic since the back pressure created by the downstream tanks is negligible.

It is to be noted that the templatization process can in some cases also add complexity to the system. For instance, the process flow diagram in *Figure 10* represents a single well system with no connections to other wells. However, the corresponding simulation schematic includes a gas source coming from other wells downstream of the separator. This complexity has been included to make the simulation template a general one that can be utilized on systems with gas lines commingling from multiple wells. In the case of a single well system, the value of this parameter is set to zero.

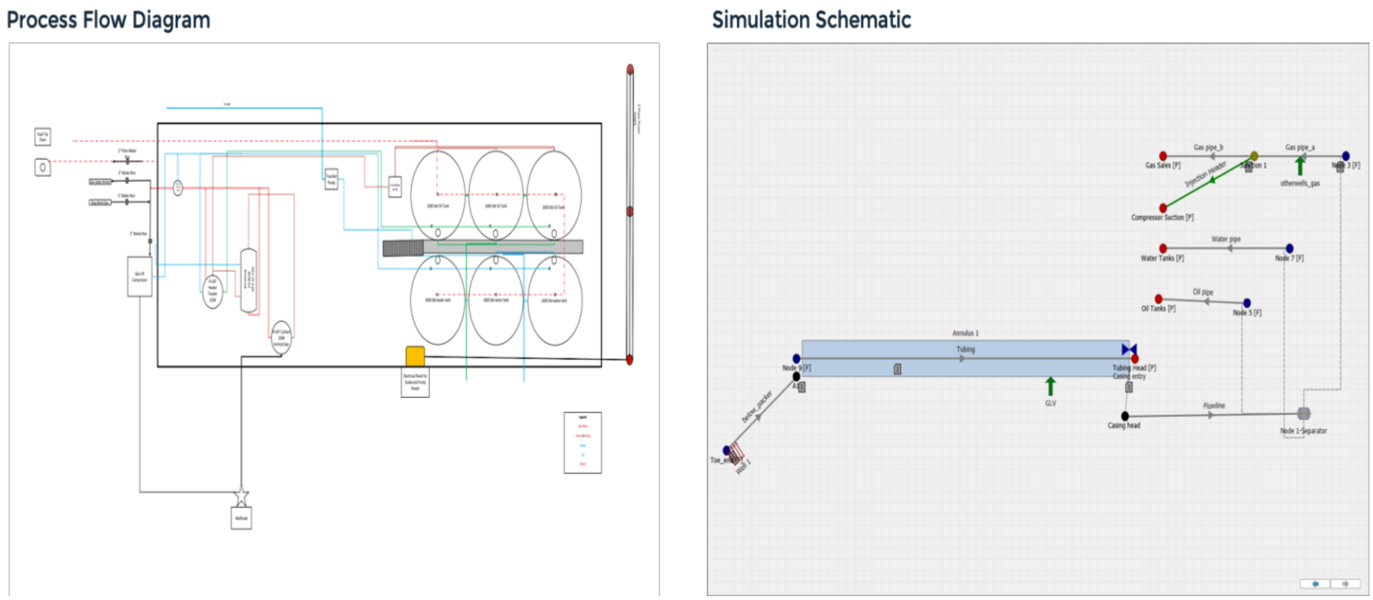


Fig. 10: Field data process flow diagram and its corresponding simulation schematic.

Simulation parameter sensitivity was assessed using box plots similar to the examples shown in *Figure 6*. It was observed that the spread of change (0-30 psi) in separator pressure was higher than that of bottom-hole pressure (0-3 psi) during set-point changes. This exploratory data analysis was helpful in the design of the simulation setup that led to some crucial decisions. For example, reduction of bottom hole pressure is considered to be one of the primary objectives[10] of artificial lift. In the nodal analysis performed through simulation software it is common to find literature with separator or the wellhead specified to be the end node[11].

Such systems assume the end node pressure to be constant during a set-point change. In a previous version of this work, several challenges were observed in mimicking the out pressures of physical system due to the assumption that the simulated wellhead/separator pressures were to be held constant during a gas injection change. Based on the data from the box plots, it was demonstrated to be an incorrect assumption. T

o mimic the physical system that accommodates separator pressures to change with changes in gas injection rates, the gas sales and compressor nodes downstream of the separator were set to be the end nodes. This resulted in a better correlation between simulation output and physical sensor output during gas injection changes.

---

[10] https://oilfieldteam.com/en/a/learning/gas-lift-28072018

[11] Camargo, Edgar & Aguilar, Jose & Rios, Addison & Rivas, Francklin & Aguilar-Martin, Joseph. (2008). Nodal analysis-based design for improving gas lift wells production.

Simulation Input:

The simulation schematic shown in *Figure 10* is Graphic User Interface (GUI) representation of a simulation file of the commercial physics-based simulator, such as Ledaflow or Olga. Each GUI based case has an associated input file that can be broadly divided into:

1. Static parameters: Inputs that are fixed for the entire run of the simulations, such as the well completion and design data, reservoir fluid properties, pipeline and separator properties.
2. Dynamic parameters: Inputs that can vary as a function of time such as the gas injection rate, reservoir pressure, produced gas to liquid ratio and water cut, sales gas back pressure, other wells gas.
3. A system has been set up to write simulation input files based on the parameters obtained from a queue of simulations stored on a No-Sql Database such as MongoDB or PostgreSQL. The architecture of this system is further elaborated in a subsequent section.

Simulation inputs are queued on the No-Sql Database based on the type of parameters as described in **Table 1** below. The parameters described in the "knowns" section in **Table 1** are directly recorded from field data. These parameters are updated in the simulation queue based on timely trends in field data. The value of these parameters is based on the exact operating range observed in the processed field data. The "unknowns" correspond to parameters whose values are difficult to measure yet have a significant sensitivity.

These may include static parameters such as the tubing friction factor, or, dynamic parameters which vary at a high rate such as reservoir pressure. Since the input values for these parameters are unknown, a wide range of possibilities within the bounds of physical guardrails are input for these parameters. The "approximations" column in **Table 1** refers to parameters that have some sample data from the field, but not precise, live, or well-specific data. These parameters can be approximated within a smaller range because of their static nature and relative insensitivity.

The simulation queue consists of combinations of the known, unknown and approximate parameters. New simulations are added to the queue based on the rate of change of the field data. After a point of time, further simulation may not be necessary on a well, as historical simulation may have covered the operating range. The ranges of the unknown parameters also narrow down with time as the inverse model provides estimates based on history matching. The details of the inverse model are beyond the scope of this paper, and shall be elaborated in subsequent publications.

Table 1: Details of simulation parameters and relationship to field data.

	Types of Simulation Parameters		
	Knowns	**Unknowns**	**Approximations**
Field data type	Sensor data and Metadata	Field data unavailable	Sample field data available. Live data unavailable
Simulation input type	Static data and dynamic data	Static data and dynamic data	Static data
Example sensors	• Injection pressure • Gas injection rate • Design operating pressures • Well deviation • Sales gas pressure • Produced gas to oil ratio • Produced water cut	• Reservoir pressure • Productivity index • Tubing friction factor • Gas injection depth • Solution gas to oil ratio	• Location of pipe sensors • Fluid specific gravities • Valve pressure loss coefficients • Less sensitive parameters (such as ambient temperature)
Range of input	Exact operating range as observed in field data	Wide range of input based on physical possibility	Narrow range of input

The commercial physics-based simulator employed in the case presented is a transient simulator. This implies that some of the input parameters can be entered as a time series, and the response to changes in these parameters can be obtained as a time series output. This detail is a key component of the simulation setup because the startup of a simulation and the post-processing associated with writing outputs to files has a non-trivial overhead on the simulation schedule. To reduce the input-output overhead and transition associated delays between simulations, dynamic parameters such as Reservoir pressure, Productivity index etc. are stepped through time series.

In the current case, 240 simulation cases were connected together as a time series, and input into a simulation queue as a single item. Each item read from the queue generates a commercial simulator input file with 240 cases in series. The number of cases per simulation file has been heuristically arrived at as an optimum scenario. If a very large number of cases are input per file, that may overload the system memory and output large files, this may lead to a system crash. A smaller number of cases input per simulation file results in underutilization of the transient capabilities of the simulator, and loss of time in the simulation input-output process.

Simulation output:

Each simulation file generates a time series output for the 240 cases as described in the simulation input section. The simulation output is updated to the No-Sql database item associated with its input. Python scripts are set up to extract the data from a No-Sql database, and post-process the time series data and extract individual case outputs corresponding to each input. These simulation outputs are matched with field observations to identify simulations representing likely operating states of the well. The "unknown" parameter values in simulation cases associated with output values irrelevant to field data are disincentivized in further rounds of simulation as a part of the inverse modeling process. A rough visual representation of the response comparison between field data and simulation is presented in *Figure 11*.

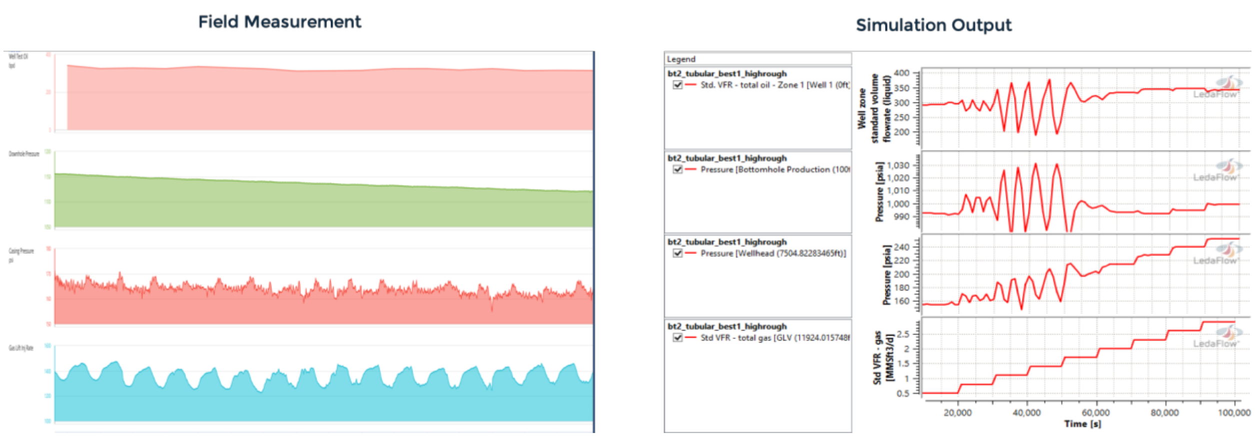


Fig. 11: Response comparison between field measurement and simulation output.

In the use case presented in this paper the response variables being matched with field data include:

- Oil production rate
- Wellhead pressure
- Downhole pressure

In the inverse modeling process subsequent to the simulation, further constraints are implemented on the relationship between simulation output and field responses including the historical trends that identify the likelihood of a combination of unknown parameters. Thus, the operating state of a well at a given time is estimated through this process, and this knowledge is used to generate set-point recommendations.

COMPUTATIONAL WORKFLOW: LIVE SENSOR DATA PROCESSING & SIMULATION

Live sensor data processing:

Live sensor data processing starts with sensor data being added to an on-cloud source location where it can be read by a monitoring workflow system. The monitoring workflow system copies the newly changed files and starts the ETL (extract, transform, load) process. The system typically starts with a scheduling system such as Apache Airflow or Spotify's Luigi which allow for workflows to be written as DAGs (directed acyclic graphs) of tasks. The scheduler executes these tasks on multiple workers following the specified dependencies between tasks and can be elastically scaled depending on load.

The ETL processes act as producers in a common messaging system workflow. The ETL process adds encoded sensor values as messages to a queue to be later read by consumers which store the data into a time series database. Processing queues such as Apache Kafka or Apache Flink create distributed durable queues for processing of queued data. Individual queue consumers can have purpose developed functionality for persisting sensor streams, creating new values and derived or calculated sensors. These durable queues provide a significant buffer of messages to be added if there is a spike in demand and consumers are not able to keep up with producers.

Eventually, the time series based sensor stream needs to be persisted in a time series aware storage system such as OpenTSDB or TimescaleDB. These time series storage solutions are purpose built data stores that store and query temporal data. Some of these stores can scale to millions of operations per second. Having a time series or temporal query engine becomes critical to effectively process sensor streams.

The design of the system allows for reprocessing of data if needed. The repeatable transformation process allows for better recovery from errors and bugs. The system is also performant. It is not uncommon to process 600k sensors per minute and the system can also scale to higher throughput by adding workers, queue partitions, or database nodes.

Simulation Workflow

After data processing, we are ready to use the data to generate simulation cases. The data is analyzed for parameter ranges to explore in simulation. Simulation cases are partitioned by field data and queued in a document database collection. Cloud instances configured with the commercial simulator software and a Python process consume the queue and process cases. Our commercial simulator can be called using a command line interface using a JSON file for input. The output from the simulator is saved to the same document with the case and the status is marked as completed.

A single simulation generates 63 MB of uncompressed data and 467 data points (approximately 4TB of data per day). However, our inverse modeling process requires only 40 of those data points and we use compression to store the portion of the results needed. A single instance of the simulator can process about 15K cases a day (approximately 10 cases a minute). We use 4 instances to process 60K cases a day and this process can be scaled up to more instances if needed.

CONCLUSION

The typical processes of the oil and gas industry with respect to data processing, simulation for well modeling and artificial lift set point optimization are time-intensive due to their manual nature. With the limitations of number of engineers per well[12], and with high decline rates[13] contributing to highly transient behavior, continuous optimization is a challenge. Inability to update set-points along with changes in well behavior may result in sub-optimal production rates. There may be significant economic benefit by optimizing set-points through increase in production, and/or reduction in operational costs[14].

By harnessing live data feed, on-cloud processing power in combination with simulation and data science tools, it is possible to develop a digital twin for scalable set-point optimization based on physics-based models on fields with hundreds of wells. In the digital twin, there is an interactive system between the field data from the physical world and the virtual data from the simulations. An overall framework for developing such a digital twin has been presented in this paper.

The design and implementation details along with the architecture of the system required to automate continuous field data processing and a massive scale simulation engine that can generate 60,000+ simulations per day has been described. The benefits and challenges in minimizing human involvement for automating key components of the system have been addressed, while also highlighting the components that benefit from a human-in-the-loop. The on-cloud solution also provides an opportunity to scale-up the capacity on an as needed basis by multiplying the computational units.

---

[12] https://jpt.spe.org/so-many-wells-so-few-engineersscaling-production-engineering-all-those-shale-wells

[13] https://www.hartenergy.com/exclusives/why-us-shale-production-declines-are-higher-you-might-think-188251

[14] Redden, J. David, Sherman, T.A. Glen, and Jack R. Blann. "Optimizing Gas-Lift Systems." Paper presented at the Fall Meeting of the Society of Petroleum Engineers of AIME, Houston, Texas, October 1974. doi: https://doi.org/10.2118/5150-MS

Several key decisions were made in terms of choosing the architecture for the system, integration with a platform that supports event labeling, factors to consider for templating the simulation setup, and facilitating the simulation output to be compared with field data. The importance and examples of these decisions were presented in the paper to help in adopting and replicating this work.

The methodology and experience shared in this paper is expected to help the community get one step closer to scalable and generalized automated set-point optimization, and also help engineers free-up their time for important decision making, rather than spending it on templatizable and repetitive tasks such as setting up simulations manually.

This infrastructure live massive scale simulation coupled with field IIOT sensor data paves the way towards the next steps required for closed loop dynamic set-point optimization. These include: Inverse modeling using a combination of machine learning and probabilistic models, set-point recommendation and evaluation system. These topics need separate individual papers and will be presented in future publications.

ACKNOWLEDGEMENTS

The views expressed in the *IIC Journal of Innovation* are the contributing authors' views and do not necessarily represent the views of their respective employers nor those of the Industrial Internet Consortium.

- Return to the Beginning of this Article
- Return to the Table of Contents

Digital Twin and IIoT in Optimizing Manufacturing Process and Quality Management

Authors:

Shi-Wan Lin
CTO
Yo-i Information Technology, Ltd.
shiwan.lin@yo-i.net

Maxine Fu
Chairman of Board
Yo-i Information Technology, Ltd.
maxine.fu@yo-i.net

Jin Zhou
Director of Smart Manufacturing Solutions
Yo-i Information Technology, Ltd.
zhou.jin@yo-i.net

Liu Zhenqi
Director of Industrial Applications
Yo-i Information Technology, Ltd.
zhenqi.liu@yo-i.net

Li Kun
Director of Industrial Solution Implementation
Yo-i Information Technology, Ltd.
kun.li@yo-i.net

Jia Yangkai
Industrial Application Business Analyst
Yo-i Information Technology, Ltd.
yangkai.jia@yo-i.net

Isabella Li
General Manager
Industrial Digitalization Solutions
Yo-i Information Technology, Ltd.
Isabella.li@yo-i-net

CONTENTS

1. IIOT AND DIGITAL TWIN IN PRODUCTION ENVIRONMENTS

The drive-chain in the Industrial Internet of Things (IIoT) involves three key elements of data, analytics and applications [1]: data are collected from production environments including equipment, products and processes enabled by IoT, data analytics is facilitated by advanced analytics including machine learning, and smart industrial applications are driven by data and analytics to provide a closed feedback loop to support optimal business decisions, production processes and equipment states, as illustrated in Fig. 1-1. Digital twin technologies [2] that systematically organize and synchronize data from the equipment, enable streaming data analytics, hence provide a dynamic representation of the production environment in the digital space. They further enhance the capability of smart industrial applications[3].

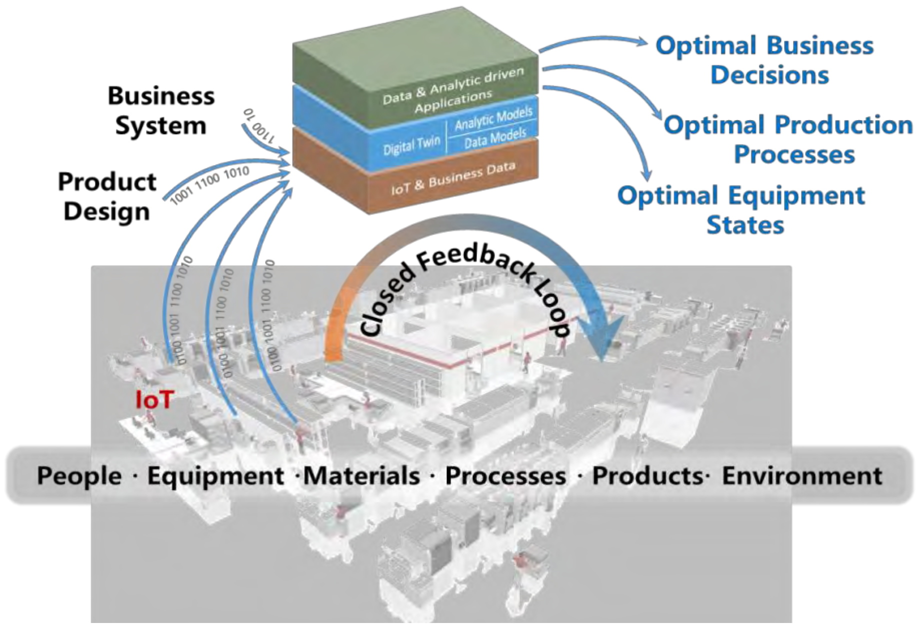


Fig. 1-1: IIoT and digital twin in production environment.

This article presents the concepts and practices on the design, implementation and some preliminary outcomes based on a real-world use case in a production process control and quality

---

[1] Industrial Internet Reference Architecture. V1.9. Available: https://www.iiconsortium.org/IIRA.htm

[2] Digital Twins for Industrial Applications. Definition, Business Values, Design Aspects, Standards and Use Cases. An Industrial Internet Consortium, White Paper. Version 1.0 2020-02-18. Available: https://www.iiconsortium.org/pdf/IIC_Digital_Twins_Industrial_Apps_White_Paper_2020-02-18.pdf

[3] M. Bärring, B. Johansson, G. Shao, Digital Twin for Smart Manufacturing: The Practitioner's Perspective, SME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE), Nov, 2020.

management application in the steel industry, guided by Lean management and Six Sigma concepts and best practice, built on an Industrial Internet platform embedded with a digital twin framework, extending the application of IIoT and digital twin beyond the commonly known domain of asset management and predictive maintenance.

The production process control and quality management application brings Industrial Internet, digital twin (including equipment and product digital twins), streaming and batch data analytics, machine learning, fused with Lean management and Six Sigma concepts and best practice into manufacturing process and quality management. By integrating and analyzing data from the equipment and other manufacturing IT systems, the system enables the correlation and monitoring of process design specification and actual process data, and quality data in near real time

- to promptly identify and resolve production problems,
- identify potential improvements in current and new manufacturing processes through the quality-process feedback loop, and
- to provide track and trace capability on product quality on individual products through the implementation of product digital twin.

To the end, the system helps the manufacture operator to ensure product quality and reduce production cost.

The concepts and practice of the Industrial Internet have been widely applied in various industries. The implementations of digital twins are starting to pick up in industries as well. At this moment, however, these implementations are disproportionally in the area of equipment maintenance in the large domain of asset management with novel features such as predictive maintenance supported by advanced analytics. There are few and far between real world use cases penetrating in the production processes covering broader application domains such as production process, quality, energy, equipment and operation safety management, each of which presents a deep and rich field where IIoT and digital twin technologies can offer great potential benefits.

In a previous volume of Journal of Innovation published by the Industrial Internet Consortium, part of the authors of this article co-authored an article reporting a real-world use case of the Industrial Internet and digital twin in energy management in the steel industry[4]. In the same vein, we report in this article the application of IIoT and digital twin technologies in production process and quality management in steel production processes that has the following characteristics:

[4] Shi-Wan Lin, Maxine Fu and Kebin Li, Digital Twin + Industrial Internet for Smart Manufacturing: A Case Study in the Steel Industry, Journal of Innovation, Industrial Internet Consortium, Nov. 2019.

- Integrate process design data, quality specification data, equipment operational real time data, quality measurement data into a holistic end-to-end closed-loop system, enabling comprehensive online monitoring and analytics of production process and supporting product quality traceability.
- Combine digital twin and Industrial Internet technology seamlessly into a holistic platform to support such an application.
- Enable digital twin for both equipment and product alike, dynamically bind product digital twins with equipment digital twins to enabling product process and quality online tracking, monitoring and traceability.
- Combine online data and analytic technologies with Lean management and Six Sigma concepts and best practice for production process and quality management, creating a digital Lean capability.

2. Problem Domain

The usage scenario involves production processes in steelmaking, an important continuous process heavy industry. More specifically, it involves the later processes such as forging and heat treatments in special steel production after raw molten steel has been casted into slabs, blooms or billets. However, the basic approach and technologies used in this scenario can be extended and applied to other similar processes in steelmaking or other similar industries as well.

2.1 Complexity in Steelmaking Processes

Steelmaking involves with many production processes that are long and complex. It takes in iron ore, coal, limestone and recycle steel as raw materials and produces the final products that include common steel (e.g., used in construction) and a phalanx of special steel used for making parts for various of machines. In between, it goes through many major processes, such as coke oven, blast furnace or direct reduction furnace to produce iron, basic oxygen furnace or electric arc furnace to produce molten steel, continuous casting to produce slabs, blooms and billets as intermediate products in the primary production processes. Then in the secondary production processes, the intermediate products are reheated, forged or pressed into various forms, such as pipes, sheets, bars, rods, and other structural shapes, which may be further heat treated, surfacecoated and finished before being released as products. Please refer to Fig. 2-1 and Fig. 2-2 for graphical illustrations [5] of primary and secondary production processes, respectively and additional steelmaking information thither in the footnote reference.

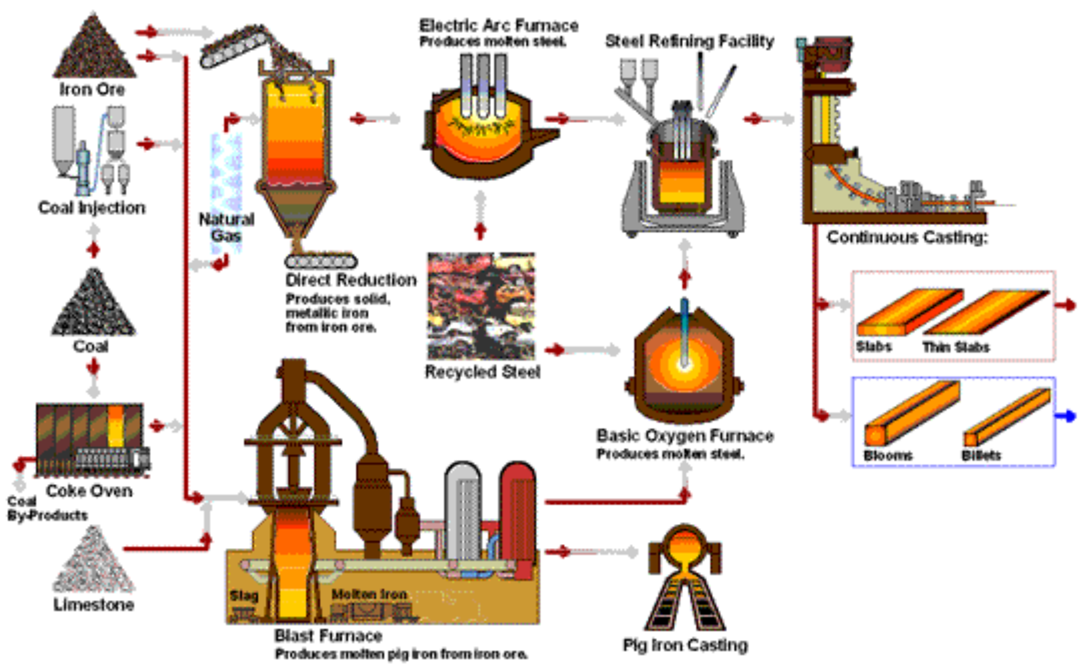


Fig. 2-1: Steelmaking primary production processes.

Further complexity in Steelmaking arises from that each of the major production processes may consist many subprocesses, each of which may be supported by various production lines that employ many large and sophisticated equipment with complex upstream and downstream supply chain relationship, as in part illustrated in Fig. 2-1 and Fig. 2-2. In the primary processes, steelmaking is largely continuous process. However, it becomes increasingly discrete process in nature in its secondary production processes.

Moreover, production process parameters in the upstream processes have strong influence over the downstream processes. The production process of high-end special steel is even more complicated involving more processes and sub-processes. Another rather unique aspect in steelmaking is that its products in process subject not only to mechanical but also chemical changes, making its quality monitoring and tacking more challenging. The making of special steel tends to consist of small batches of large varieties of products. At the same time, because special steels are used in important equipment of various industries that has higher reliability and safety requirements, and in turn the requirements for quality are more demanding and refined.

[5] https://www.steel.org/steel-technology/steel-production/

Consequently, the need for traceability of product quality in its full production lifecycle is also stronger.

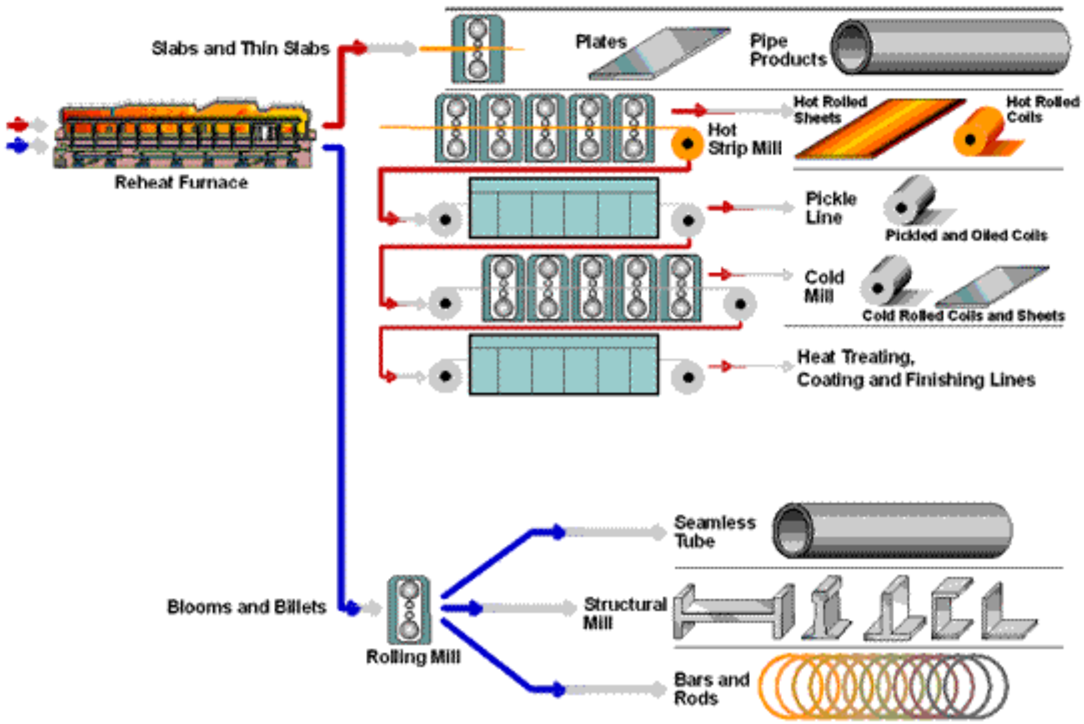


Fig. 2-2: Steelmaking secondary production processes.

2.2 Achievement and Challenges in Digital Systems in the Steel Industry

For the past few decades, steelmaking, like many other continuous process heavy industries, has been highly automated at the equipment level (ISA-95 level 1 and 2). PLCs and SCADA systems are widely used to control and monitor the production processes. In the recent decades, the steel industry has also invested in establishing various IT application systems managing one or other aspects of productions such as production planning and execution (manufacturing execution system – MES), production process control management system, quality management system (QMS), equipment management system, all belonging to at the ISA-95 level 3. Today, the steel industry at large is operating based on these systems that have significantly contributed to the increase of productivity, improvement in product quality and reduction of production cost.

However, a number of major technical challenges remain:

- At ISA-95 Level 1 and 2: Many PLCs often remain unconnected to higher level systems, so it is impossible to collect data from them. For PLCs that have been connected to SCADA systems, it is often that the SCADA systems remain isolated so the data that have been collected or alerts that have been generated can only be viewed locally using associated HMIs. Data and alerts in one SCADA system cannot be readily used to correlate or otherwise analyze with data from other systems. Analytics, if any, often requires manual work that are technically involved, time-consuming and unreliable, and remain to be ad hoc, incomplete, inconsistent, usually carried out too late to be useful.

- At ISA-95 Level 3: the existing application systems tend to be built over different periods of time, often sponsored by different departments with a strong focus on solving specific domain problems such as equipment maintenance without enough thoughts given to forming holistic view and considering the interaction of different aspects of the overall operations. Worse still, these systems were contracted out to third-party software integrators who built these systems independently, often with completely different designs and data models supported by different technology stacks with few functions exposed as APIs. Consequently, these systems operate on their own and has little integration with other systems, forming stacks of independent application silos and unreachable and reusable data islands. Furthermore, there are domain areas that are not yet covered by modern software application systems; thus, its operations still require manual data entry and tracking.
- Between ISA-95 Levels: few Level 3 application systems have comprehensive connectivity to the Level 1 and Level 2 systems making it difficult to obtain near real time data for advanced analytics in and across the application systems.

At present, the production operation management of steelmaking have been established on the basis of automation systems and IT application systems as briefly described above. However, there exist connectivity and integration barriers in the current systems preventing further improvement in operation efficiency from realization. These barriers in a major part are resulted from the technical challenges outlined above, between the production operation management application systems and the equipment automation control systems (PLCs and SCADAs), and among the application systems themselves.

The first type of barriers exists between the IT application systems and the automation systems because the PLCs and SCADAs are not nearly sufficiently connected to the application systems, causing many application systems separated from, or only weakly linked to, the physical reality of the production environment where equipment is in operation and products are being processed. The result is that the data in the underlying automation systems has not been fully collected and utilized, making it almost impossible to gain transparency over the status of production processes leaving alone the possibility for data and analytics-driven management of the operations.

The second type of barriers exists among the applications systems in various applications domains (e.g., process, quality, equipment, energy, production planning and execution), and in various production processes and lines, (e.g., those described in the steelmaking processes). This is largely the consequence of history in that the technologies of the past were not conducive to enable interworking of various applications in the very complex environments such as those in steelmaking. This leads to the situation in which application systems in different domains, or in the same domain but in different upstream and downstream processes, have not achieved the level of information integration and exchange that are needed (e.g., product quality management requires information about equipment operation status, and quality management systems in adjacent upstream and downstream are required to exchange information), so it is very difficult to support cross-process and cross-domain holistic and dynamic data- and analytics-driven operation management.

2.3 Specific Challenges in Steelmaking Process Control and Quality Management

The challenges and potential opportunity for improvements in the smaller scope of the two closely linked application domains: production process and product quality management, are similar.

- It is a common practice that the existing process management applications are mainly responsible for information management of product specification, production process and quality design. Too often these systems are not connected to the production management systems, and therefore the transmission of detailed process control specification for specific products still relies on conventional electronic documents exchange to reach the on-site production management systems and thus requires manual configuration of the processes.
- Additionally, most of the on-site operation process data and quality data cannot be automatically collected but rather to rely on manual reports to track, trace, and archive. As result, it is difficult to discover quality issues soon enough so that they can be addressed in time for corrective actions to advert or minimize the impact. Therefore, problem root cause analysis and solving if any are often after the fact.
- Quality analysis is still a manual exercise on static historical data using conventional SPC-like toolsets that are not connected to the production environment.
- Production process control specification and on-site operation data and quality outcome are not easily correlated to provide guidance on product and process control design.
- Product full lifecycle quality traceability is hard to achieve.
- Moreover, the process management put too much manual operational and recording workload on workers that are not efficient and prone to error.

The main quality attributes of steel products include structure performance determined by micro-structure of the material, surface quality and geometric dimensions, which mainly depend on the composition of the material and the processing technologies involving chemical reactions, thermal dynamic and mechanical processing during the production processes. Therefore, it is necessary to obtain data and information related to the three main categories of quality attributes and process control design with respect to the customer's requirements.

The quality pass rate of steel products, especially high-end special steel products, is often low. The main cause for this is the relatively weakness in process control that leads to large fluctuations in the processes that inadvertently affect the quality of the final products. There are many challenges contributing to the weak process control, such as the difficulty in maintaining stability of process control parameters (e.g., continuous casting speed and temperature of molten steel), in detecting interference of uncontrollable factors (e.g., the floating of inclusions in molten steel), and that in practice, due to the lack of means, there are many influential process parameters are not only uncontrolled but often not yet visible (e.g., in the rolling process, the temperature of the entire process in between the start and the final rolling point).

Additionally, many product quality issues are related to equipment status and operation methods. Therefore, in order to obtain quality products, standards need to be formulated and quantified for ingredients, processes, equipment, and operation methods, and equally crucial, there need to be real-time measurement and monitoring the actual process to ensure these standards are met, and any exceptions are detected promptly so they can be corrected in time.

Furthermore, the quality of steel product is affected by many coupling factors. Constrained by practical conditions, it is impossible to control every factor that may affect quality. To improve product quality, the key is then to identify the key factors and place strong emphasis on controlling them. However, the cause-and-effect relationship of various factors are often complex and coupled with each other, thus brings great challenge in identifying what the key factors are without in-depth analytics. On the other hand, the analytics for this and other similar purposes depends on large amount of quality data that contain the correlations between what seem to be randomly occurring defects to process control specifications and actual process data, equipment status and operational methods that may cause the defects.

These obviously cannot be accomplished with the conventional approach of relying heavily on spotted manual inspection and recording, and by counting on the independent and isolated automation systems and application systems that cannot share and align data. A totally new approach supported by a new set of technologies is required to address these challenges to achieve total quality control in the production environment. At this time, the IIoT and digital twin technologies are the right choice for addressing these challenges, as demonstrated by our practice.

3. THE SOLUTION

This article reports a successful case, using an industrial Internet platform embedded with a digital twin framework to break through the barriers at different architecture levels and among different domain application systems as described above. The final solution system

- carries out collection of process control and quality data, enables near real-time data analytic, establishes a new data- and analytics-driven production process and product quality management application.
- performs online dynamic comparative analysis on actual process data in reference to process design specification, carries out dynamic process quality monitoring and quality

traceability of products, discovers process exceptions and quality anomalies and enables quality issues resolution in time.
- On this basis, moreover, through statistical analysis of data, enables process verifications and improvements.

It ultimately helps production operators to achieve the goal of avoiding defective products, improving product quality, and reducing production costs.

Although the technical approach, technologies and solution are applicable in the wider use scenarios in the steelmaking processes, the project reported here initially focus one of the steelmaking processes, namely, the forging process to get started, verify and improve the solution before expanding into other processes[6].

3.1 The Design Process - Lean Six Sigma Management Driven

After a development contract was completed with the customer during which a preliminary use scenario research was carried on site, a comprehensive on-site discovery-research was carried out by a team with expertise in various domains, which mainly include:

- Industrial domain experts with deep insights about steelmaking processes who help to understand the production operation processes and are capable of evaluating the business needs of the customers from which to derive business requirements.
- Industrial software application analysists who work with industrial domain experts to translate business requirements into software requirements and assist the design of the software solution.
- Industrial data analytic engineers who are experienced with data modeling techniques to understand the data analytic requirements and map them into input data and modeling requirements.
- Industrial control/automation engineers who are responsible for gathering information about the deployment structure and configuration of the industrial equipment and other systems involved in the project, evaluating the connectivity and data collection requirements and design the IoT portion of the software solution.

Among the multi-dimensional on-site research, the most critical work is the evaluation of customer requirements in establishing this new production process and product quality management application which is to become a major tool to be used by production operators and managers alike in their respective daily work. The Lean management and Six Sigma concepts and best practice, having long been proven to be effective in in production operation management, are used as guiding principles to evaluate the production process and product

[6] As of the writing of this article, the first phase of the project has been completed and customer acceptance has obtained. A related project in a different steelmaking process has been launched.

quality management practice in order to come up with the business requirements, which partly requiring changes in the operation processes and organization, and partly to be implemented by the software solution, in a way to solidify and augment the operational management methodology and best practice in software.

The Lean management approach relies on three key ideas[7]:

- identify the value and value stream in an operation flow or process to drive value delivery
- eliminate waste (things that do not bring value to the end product)
- continuous improvement

with key operatives in eliminate waste and continuous improvement.

The Six Sigma[8] approach aims to management processes that can be defined, measured, analyzed, improved, and controlled (DMAIC) using a set of advanced statistical and project management toolset with a goal of keeping tolerance of fluctuation under ±6σ (standard deviation of the requirement).

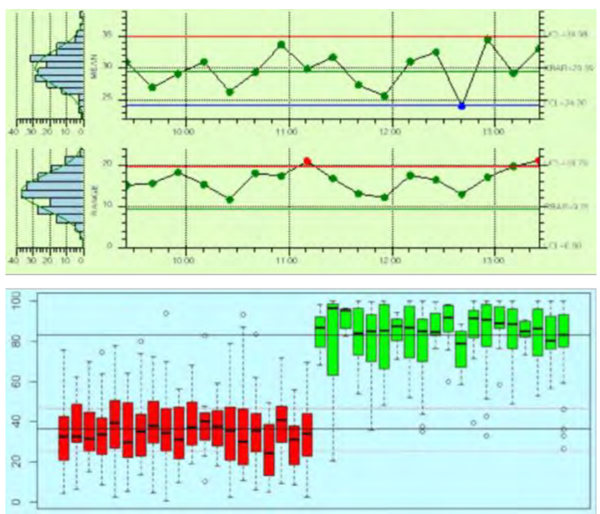


Fig. 3-1: Examples of process control parameter real time monitoring with built-in statistical analysis and alerts.

Lean management is keen on driving out waste, promotes standardization in workflows and procedures, and continuing improvement, whereas Six Sigma focuses on reducing process variation and enhancing process control. These two methodologies in management overlap and otherwise complement each other with diminishing distinction.

---

[7] https://kanbanize.com/lean-management/what-is-lean-management

[8] https://asq.org/quality-resources/six-sigma#:~:text=Six%20Sigma%20is%20a%20method,quality%20of%20products%20or%20services.

Based on the Lean management and Six Sigma principles, the on-site discovery-research and evaluation of customer requirements focus on：

- *standardization of operation*: standardize operation workflows and procedures that deliver the values – quality products, against a set of constraints and uncertainties, and seek to maintain the operation standard with the software as much as possible.
- *waste reduction*: reduce the occurrence of defects in the products and the amount of re-work needed to address quality issues. The key approach include:
    - to exercise the best data-driven control over the production system to maintain stability of the processes avoiding unwanted or unnecessary fluctuations;
    - to discover process exceptions and quality anomalies in near real time and identify their root causes so they can be addressed in time to avoid downstream wastes;
- *continuous Improvements*: to provide online visibility to the quality performance including using conventional Statistical Process Control (SPC) methods, however, based on real-time data so quality performance issues can be identified and addressed by enabling:
    - per-operator/workgroup longitudinal quality performance comparative analysis;
    - cross-operator/workgroup lateral quality performance comparative analysis;
    - per-product type longitudinal quality performance comparative analysis;
    - cross-product type lateral quality performance comparative analysis;
    - product defect multivariant-correlation analysis, considering factors in product design specification, process specification, actual process data, work procedure details, and individual operator skill levels, etc., to identify principal factors that contribute to the occurrence of defects so measures can be taken across different areas to improve design, processes and procedures and operator skills.

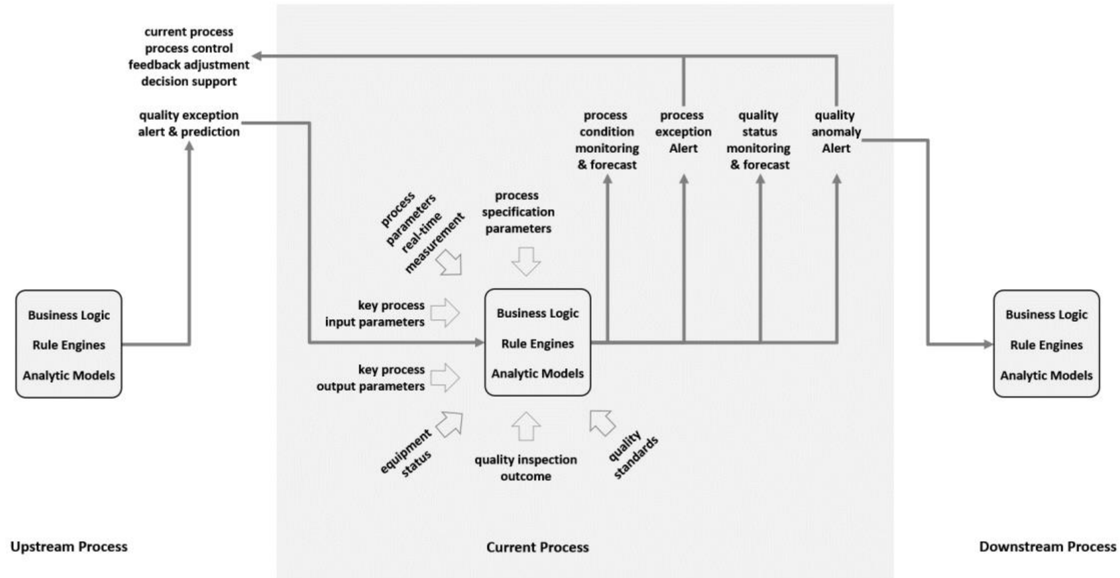


Fig. 3-2: Information flow of data and analytic-driven total quality management supported by real time process control.

Fig. 3-1, Fig. 3-2 and Fig. 3-3 illustrate some of the design ideas resulting from the Lean management and Six Sigma principles with regard to product process control and product quality management.

Most of the capabilities described above can and must be designed into and implemented in the software application to cement the methodology and best practice so they are exercised in the production processes largely automatically. The foundation of realizing most of these capabilities is data collection across all equipment and other operational systems involved in the production processes. On this basis, the project use software to comprehensively digitize the operation processes and procedures, enforcing to a large degree standardized operation workflow. It then leverages advanced analytics to help to optimize the operation management.

In a strong sense, the project leverages the power of data and analytics, and the capabilities of software to implement and enforce Lean management and Six Sigma principles, realizing a data- and analytic-driven digital Lean management and Six Sigma system for production process and product quality management.

3.2 Production Environment Configuration

The phase I of the project reported here limits its scope to a special steel plant, part of a larger steelmaking process, that consists of multiple heat treatment and forging processes and production lines that include electric furnace production line, electroslag furnace production line, double vacuum production line and forging production line. The furnace equipment involved includes electric arc furnace, induction furnace, refining furnace, conventional electroslag furnace, gas-protected electroslag furnace, vacuum induction furnace, vacuum consumable

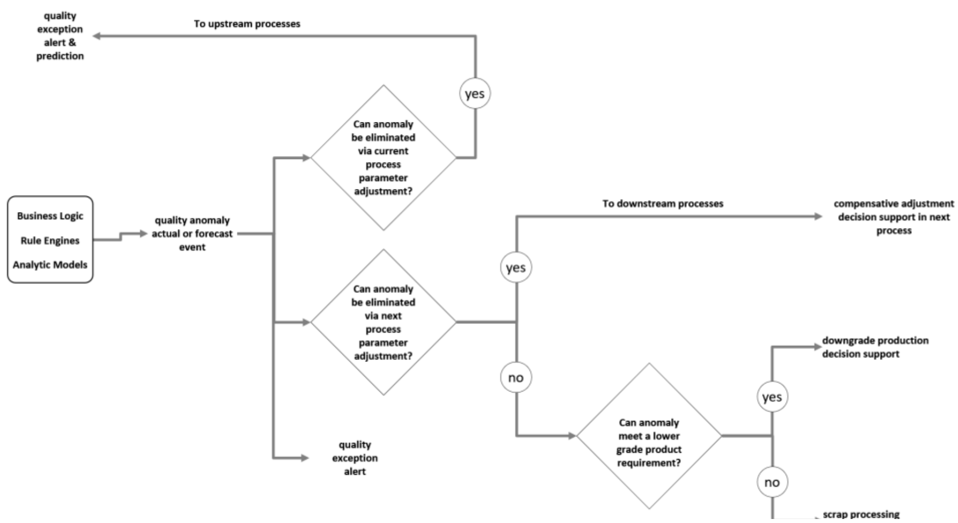


Fig. 3-3: Logical and information flows for quality exception handling to reduce wastes.

furnace, etc. that come with different forms, such as trolley furnace, indoor furnace, ring furnace

etc. The configuration of the equipment is schematically illustrated in Fig. 3-4. Each furnace can perform heat treatment of multiple products in production (PiP) at the same time. However, each of the PiP may be treated according to different process control requirements (e.g., under the same temperature but for different durations with strict temperature ramping up and down constraints).

Each type of equipment deployed in a given process may be of different models with varying capability and capacity. All the equipment as a whole can be roughly considered as a node in matrix or a graph[9] in which each PiP moves through during the production process. Because special steel products have many varieties and are being made in small batches, with each product type needing to meet specific product specification and to be produced with specific process control parameters. Furthermore, multiple products are often being produced at the same time. Consequently, the production scheduling for each PiP is dynamic to maximizing equipment Overall Equipment Effectiveness (OEE) and production capacity. Therefore, each of

[9] as a concept in computer science

PiP, even of the product type, may take a specific and unique path in the equipment matrix during its lifecycle of production.

In order to achieve the design goal of the application, it is required keeping track not just the path of the PiP in its lifecycle of production, but also its own set of actual process data when it is being processed by each piece of equipment and its characteristics (such as its temperature) in the staging area between equipment.

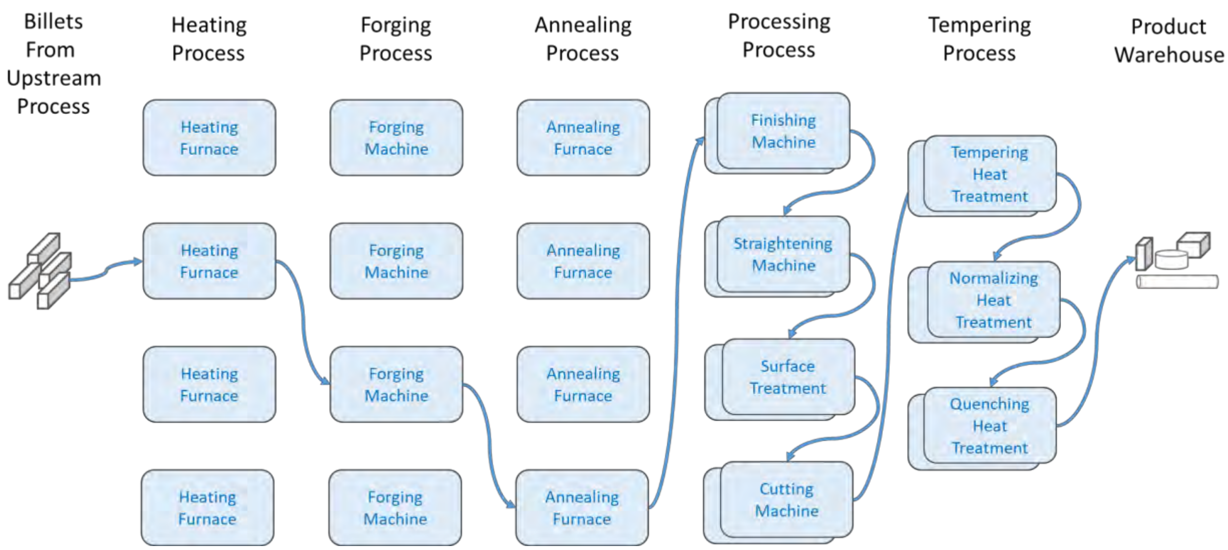


Fig. 3-4: A equipment and workflow process in special alloy steelmaking plant.

3.3 The System Design Requirements

The Lean management and Six Sigma-oriented customer requirement evaluation presents a few key requirements on the system design:

- The need to realize full lifecycle tracking of product in production (PiP), including its movement in the production processes, its dynamic association with production equipment during the time when it is being processed and the corresponding actual process data, and for some situation, its conditions (e.g., temperature) during transitional staging periods.
- The need to align each piece of PiP's actual process data with its corresponding process specification parameters to ensure the actual process data are within the required design range;
- The need to automatically detect and provide alert on process exceptions and product anomalies, and in some cases to provide automatic or semi-automatic root cause analysis for the detected exceptions and anomalies.
- The need to connect to work order system to obtain work dispatch information; the need to connect to production process specification management system to obtain process

specification for each PiP; the need to connect to product quality test system to gather and record test outcome associated with individual PiP.

- The need to provide online analytics based on the data collected to support online and batch quality performance and other analytics.

Evidently, this is a use case that can clearly leverage the capabilities an IIoT platform with a digital twin framework can offer:

- IoT capability to connect to equipment and sensors to collect data;
- comprehensive digital representation, namely, digital twins, of the production environment, most importantly, the equipment twins and the product twins to track their status and process data, and the dynamic relationship between PiPs with the equipment so actually process data can be collected and correlated with the PiPs;
- analytic capability that can be digital twin-embedded to analyze data associated with equipment and PiPs
- application DevOp environment for develop and running the application that implement the business logic and user interfaces.

3.4 IIoT Platform and Digital Twin Framework-based Design

The system design is based on the Yo-i Thingswise Industrial Data OS (iDOS), an Industrial Internet Platform embedded with a digital twin framework, purposely designed, in reference to

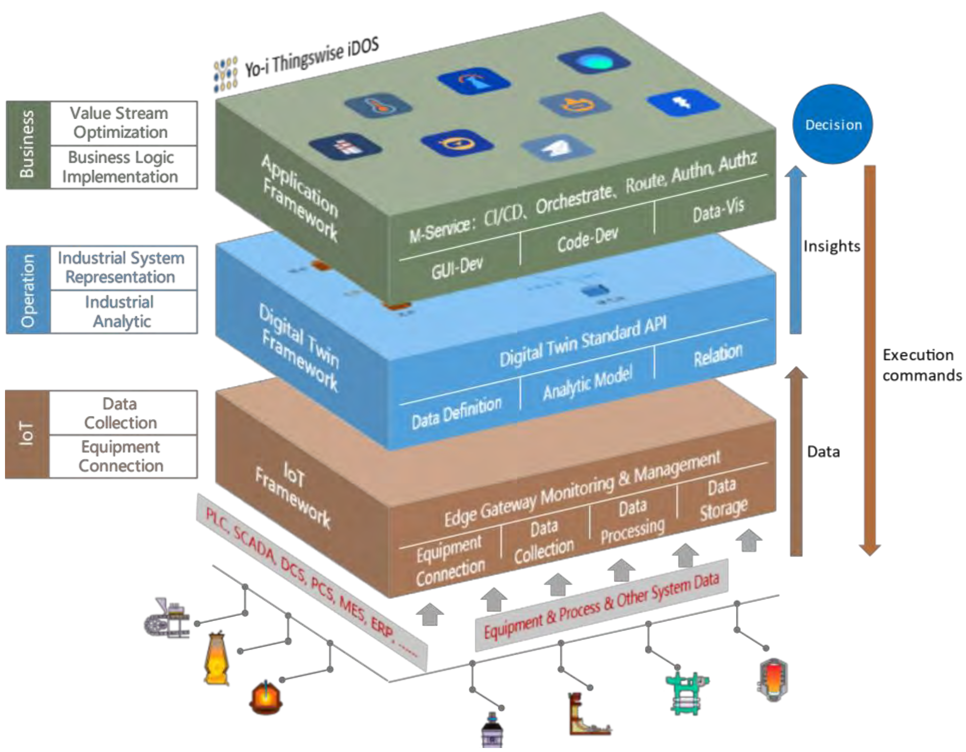


Fig. 3-5: Yo-i Thingswise iDOS platform functional architecture.

Industrial Internet Reference Architecture[10], for supporting digital transformation of industrial operational management. It is built on a core architecture that integrates the latest cloud-native technology, Big Data, machine learning and micro-service technologies as its foundation on which a set of innovative key functional components are developed to provide a flexible to deploy, easy to use, self-contained end-to-end industrial PaaS stack. As an industrial PaaS platform, it provides:

- *an IoT Framework* to support connectivity to equipment and devices for data collection and system control. It offers the following features:
  - Equipment connectivity
  - Data collection
  - Data preprocessing
  - Data storage
  - IoT gateway remote management service
- *a digital twin Framework* to digitally represent production environment, manage equipment data, and integrate analytic models associated with the equipment.
  - Digital twin design GUI
  - Digital twin to IoT data mapping
  - Digital twin analytic modeling
  - Digital twin API
- *an Application (DevOp) Framework* to enable GUI-based codeless and normal code-based application development, deployment and runtime management.
  - Data visualization tool
  - GUI-based application rapid development tool
  - Multi-language code-based application development SDK
  - Micro-service framework

The full technical stack of the platform, as shown in Fig. 3-5, supports end-to-end streaming data processing and analysis, enabling near-realtime responses to events occurring in the industrial operational environment. It is designed to enable industrial enterprises to develop and run data and analytic-driven industrial smart application.

[10] Industrial Internet Reference Architecture. V1.9. Available: https://www.iiconsortium.org/IIRA.htm

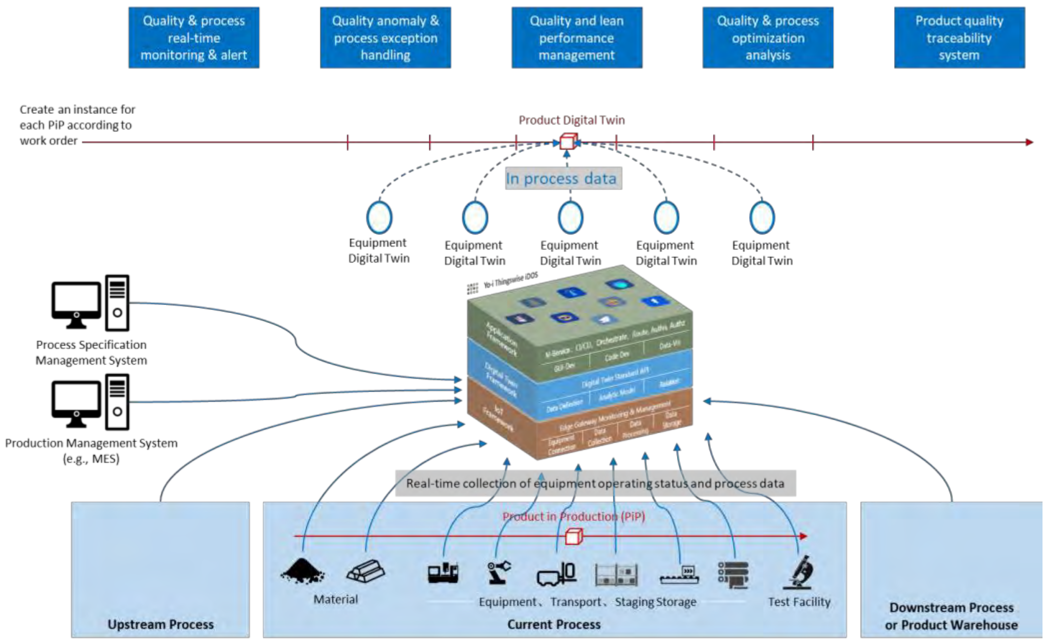


Fig. 3-6: System architecture.

3.5 Implementation key points

The overall system architecture based on the platform described above is illustrated in Fig. 3-6.

- Equipment connectivity and data collection, integration to other systems, such as Process Specification Management System (to obtain product process specification) and Production Management System (to obtain work order information) are implemented with the IoT framework layer of the platform (including IoT/Application Integration Gateways), through GUI-based codeless configuration. The connectivity to equipment involves various common industrial connectivity protocols that include MODBUS, OPC UA/DA, and a few proprietary protocols (e.g., reading from database from third party data collection systems)
- Digital twin system representing all the relevant equipment and products are designed and configured. Within the GUI-based digital twin designer, digital twin classes (see Fig. 3-7) (each representing a specific type of equipment or product) are created where equipment state variables and equipment process control parameters are defined as digital twin data fields, along with their metadata e.g., data type, unit, validation limits etc. Then based on the number of instances of each equipment type, the digital twin classes are instantiated to represent the actual equipment instances. The data collected from the IoT framework is mapped to data fields in each of the digital twin instances using the GUI-based or configuration file-based batch import toolset provided by the platform.

Fig. 3-7: Digital twin class (equipment type) design GUI.

- Data analytic models are at the heart of this solution. There are various analytic models tracking the process parameters to ensure they each falls within the standard range on the basis of each individual PiPs. Furthermore, because of the complexity of the production processes that involve many uncontrollable process parameters, e.g., most evidently, the quality and characteristics of raw material, controlled process parameters e.g., inlet gas flowrate to a heating furnace, and target process parameters, e.g., the heating and cooling speeds and steady temperatures of a furnace. It is not always possible to rely on single parameter range checking to determine if the process is running optimally or if certain combinations of otherwise normal ranged parameters would be conducive to product defects. The relationships among these parameters are often nonlinear, strongly coupled and some with direct or indirect feedbacks among themselves and making them difficult to assess the cause-effect for product defect or other considerations. Therefore, it is increasingly more common to combine first-principled engineering know-how and data science to create hybrid analytic models not only to detect defects and also to detect process conditions that likely to lead to the occurrence of defects. It is crucial to gather as much as possible data from the production environment for modeling, including leaving sufficient data for model validation.
- The applications, built on the digital twin API, implemented as micro-services in the Application Framework layer, support the following business functions that map to the requirements described above:
    - Quality and process real-time monitoring and alerting
    - Quality anomaly and process exception handling
    - Quality performance Lean management
    - Quality and process optimization analysis
    - Product quality traceability system

Fig. 3-8: A screenshot of the Overall Process Status View.

Fig. 3-8 shows an example of an application user interface for overall process status view.

The application is designed with a role-based user interaction model to improve user experience and enhance application security. Users of the application is divided into different roles based on their job functions and area of operations (in different processes). Each user can assume one or more roles based on the production organization details. When a user logs in, the application presents pages corresponding to the specific roles the user assumes. The application relies on the platform user authentication and fine-grain role-based authorization framework without the need to develop customer codes for the set of functions.

By leveraging the platform capability, a large portion of development work, including IoT connectivity and data collection, and digital twin design, are completed using GUI-based design and configuration toolset without the need of any coding. As result, development workload is great reduced and development cycle are substantially shortened.

For the business functions in the upper layer, though requiring code development, its scope is greatly reduced to just focus on business functions with user interaction through the UIs. In dealing with equipment and products (physical entities), the application interacts with the corresponding digital twins through their APIs, avoiding the need to deal with complex custom databases or directly with the equipment or product, it greatly simplifies the development of the application.

Since the application interaction with the physical entities is abstracted out by the digital twin layer, the application to a large degree is isolated from the physical environment in such a way that if the equipment changes, it is only needed to modify the underlying digital twin configuration without the need to change the application codes. This makes the application much more portable from one physical environment to another, for example, to the same process in another steel plant.

The end-to-end application also benefits from the security, scalability and reliability in part provided by the cloud-native distributed architecture in the platform, freeing the application developers from concerns about system security, performance and reliability in their development.

3.6 Product Digital Twin

The concept of equipment digital twin is well understood. However, the concept of product digital twin in production is relatively less reported. Both the equipment and product digital twins play important role in production process and product quality management as reported in this article.

For material product (in contract to machine or electronic device product) such a steel, the product in production subjects to various transformations, including in physical form (solid and liquid), physical properties (e.g., temperature), shape, surface properties, and micro-structure through chemical reactions, thermal and mechanical processes. Many of the transformation outcomes are not directly measurable in real time, either not easily measurable by simple sensors, or the sensors could be easily destroyed subjected to the harsh processing environment (e.g., in high temperature during heat treatment). Some properties such micro-structure requires lab tests that will take hours or days to complete. Therefore, except for a few relatively simple tests, such as hardness or surface properties, the quality of steel product relies heavily on ensuring the accuracy and stability of production process control. The process control data are properties of the production equipment and are relatively easy to collect from the equipment control systems.

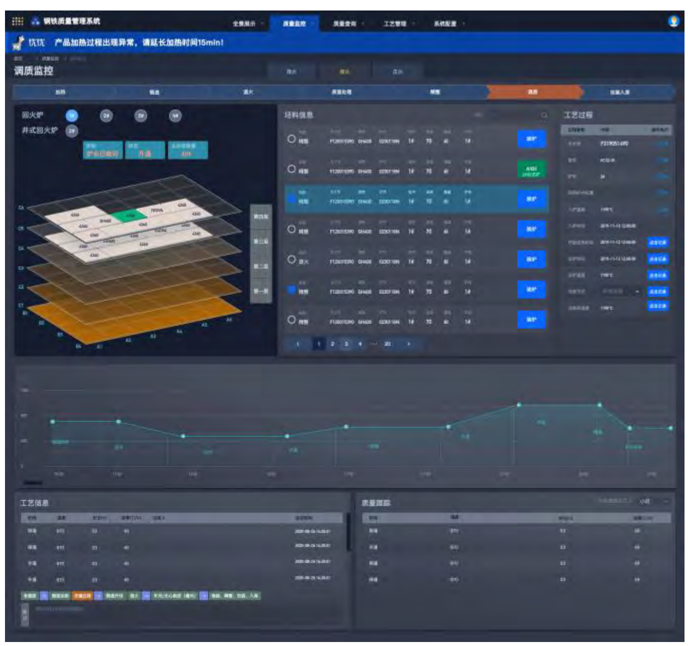


Fig. 3-9: A screenshot of a product tracking page for a heating furnace with a 3D model tracking product locations and time within the furnace, and temperature variations.

Therefore, in order ensure the product quality in every processing phase or step, it is important to track the product in production (PiP) in association with the production equipment.

In this project, the tracking of PiP is done with product digital twin, as sketched in the top-middle part of Fig. 3-6 (in red color). From information in the work orders, the system automatically creates product digital twin for each PiP. The initial product digital twin model includes information such as its ID, product type, process specification, etc. As the PiP moves through different stages in its production process, a dynamic association is made with the equipment that is processing the PiP from which key attributes such as start and end processing time and other attributes, e.g., the position of the PiP with a given piece of equipment, are recorded in the product digital twin. Any relevant actual process data from the equipment is also automatically entered into the product digital twin, and dynamically analyzed in reference to its corresponding process specification at that processing stage. Any deviation from the process specification by the actual process data can be identified, recorded and reported. Similarly, any records of quality measurement and inspection will also be entered into the product digital twin as well. When the processing of a PiP is completed, the data of the full life cycle of the product in the production process is also fully recorded and ready to be queried and analysis in the corresponding product digital twin model.

To consider the product digital twin in a different angle, products are what manufacturers made to create value for its enterprise. As the product in process moving through the production processes, it represents the value flows of the production. By tracking the product in process via product digital twin through the production processes, it in fact is tracking the *value flows* of the manufacturing enterprise. By further considering the raw material costs, energy cost, defect cost, the OEE of the equipment, labor cost, and the market or order prices of the final products, etc., it is possible to compute and visualize in near real time the value that are being created throughout the production environment. This will help the manufacturer to better manage the production in identify and reduce or eliminate cost and maximize value creation by making informed decisions in what to make and how to make them.

4. Conclusion and Learnings

By leveraging the built-in capability of Yo-i Thingswise iDOS platform, the complex project was completed from design to delivery and acceptance in about 3-4 months, less than half of the time it might need should the platform were not used. The production process and product quality management system were well received by the customer during the initial phase of trials and garnered positive praises from them. The same customer has extended this project into a different processing plant as of the writing of this article.

There are a few points of learning from this project:

- IIoT, digital twin, industrial analytics are enabling technologies (not applications by themselves), most fitting for supporting a new generation data and analytics-driven industrial operation management applications.
- IoT, digital twin and industrial analytics technologies must be fused with proven management methodologies and best practice, such as Lean management and Six Sigma, to be truly effective to provide values to the industrial operation management settings. (In other words, IoT or analytics decoupled from operation management only provides limited value.)
- IIoT, digital twin and industrial analytics technologies have great potential in advancing the digitalization of industrial operation management, in domains beyond equipment maintenance to include production process and product quality management (as demonstrated in this article), energy management (as reported in our other article, see 4, and possibly production planning and execution as well.
- Both equipment and product digital twins are effective methodology and technology to track product quality as the product being processed and to build product quality traceability records. Product digital twins in fact can be used to represent the value flows in the manufacturing floor.
- IIoT platform with a digital twin framework has been proven to greatly simplify the implementation of industrial operation management solutions that is secure, scalable and reliable.

ACKNOWLEDGEMENTS

The views expressed in the *IIC Journal of Innovation* are the contributing authors' views and do not necessarily represent the views of their respective employers nor those of the Industrial Internet Consortium.

- Return to the Beginning of this Article
- Return to the Table of Contents

Head to the IIC *Journal of Innovation* landing page for information on upcoming editions and links to past additions.