Adv Biochem Eng Biotechnol (2021) 177: 167–184
https://doi.org/10.1007/10_2020_157

Published online: 19 February 2021

Digital Twins: A General Overview of the Biopharma Industry

Michelangelo Canzoneri, Alessandro De Luca, and Jakob Harttung

Contents

M. Canzoneri (✉) and A. De Luca
Merck KGaA, Darmstadt, Germany
e-mail: michelangelo.canzoneri@merckgroup.com

J. Harttung
Industrial Affairs, Sanofi, Gentilly, France

Abstract This chapter gives an industry perspective of how digital twins are tangibly translated, implemented, and used in a biopharmaceutical environment. Technical prerequisites and components including data modeling, the lifecycle, and different skills which are required from people to be put together and collaborate efficiently with digital twins are discussed with practical examples which have been implemented in labs and in manufacturing.

Keywords 3D model, Biopharma industry, Connectivity, Contextualizing data, Critical process parameters (CPP), Critical quality attributes (CQA), Data, Digital twin, Healthcare, Holistic, Human factor, Laboratory, Lifecycle, Logistics, Manufacturing, Ontologies, Operations, People, Physical asset, Predictive, Prescriptive, Quality by design (Qbd), Research and development (R&D), Self-driving, Sensors, Simulation, Supply chain, Taxonomies, Virtual Model, VUCA

1 Introduction

A *digital twin* is the result of the convergence of two coexisting systems, the tangible and real system of a living organism or a nonliving physical entity and its virtual replica which is enabled by real-time data and underlying models through the use of digital technologies.

Digital twins have the ability to provide a holistic understanding of the system by building a network of dependencies between real-time data and their underlying meta information.

Through the use of Internet of things (IoT), advanced data analytics, artificial intelligence (e.g., machine learning, deep learning), and models (descriptive, predictive, and prescriptive) the digital twin becomes a living replicate of the physical entity which adapts to real-time information coming from its originator. In the scope of Biopharma, physical entities can range from a biological cell to a complete factory or supply chain. Models should be formalized either as mathematical models or algorithms in order to enable simulations.

So, considering a digital twin to be a real-time connected virtual-physical system where physical reality and virtual models are continuously connected, we will broadly have different levels of twins:

- The most basic digital twin establishes a one-directional relationship between a fixed physical entity and a fixed virtual model providing real-time transmission of physical parameters to the virtual environment for visualization, analysis, and experimental design. Each side of the digital twin system is fixed in the sense that only parameter values change. For instance, a digital twin of a bioreactor would be fixed in the sense that the bioreactor does not change while parameters such as cell density, temperature, pH, etc., would change.
- A more realistic approach in operational usage requires the virtual-physical system to remain operational as changes are made to the structure either of the

physical environment or the virtual environment models with bidirectional exchange of information. Typical changes could include the size of a bioreactor, variations of cell culture, and process parameter controls.
- Finally, a more advanced level provides the ability to simulate situations in the virtual environment that can then be applied to predict evolution in the physical environment and/or direct changes to the physical phenomenon. In the context of Biopharma, digital twins' potential scenarios would range from adapting control strategies of cell fermentation to optimizing supply chain policies such as stock levels of transport preferences. In this case, the digital twin is in a continuous learning state (supervised or unsupervised).

2 Technical Prerequisites and Components of Digital Twins

2.1 Context

So, in a nutshell, a digital twin consists of: The physical or biological product, the virtual manifestation, and the seamless and bidirectional connection between these two elements. Prerequisites and components will increase as we build increasingly sophisticated levels of twins.

The most essential prerequisite of a digital twin is to have an operational goal and defined value. Key examples would be higher yield for a process, lower variability and deviations, or better asset utilization. It is, however, one of the key missed prerequisites in many projects.

For the first level, static physical entity and virtual model with one-directional data connection, the main prerequisites are the following:

- Sensors on the physical system that can measure its state
- Connectivity to establish the link between the physical and virtual environment
- A virtual model of the physical assets (typically a 3D model in engineering related scenarios but could be also more abstract such as cell models)

The second level of twin where we manage lifecycle changes of both physical entity and virtual models requires the additional components:

- An asset framework to manage the relationship between sensors and the virtual model
- Configuration management of the different components

And to complete the list, for the third level of twin where the virtual model takes on a life of its own and starts being used to drive change in the physical entity, we have these additional elements:

- A dynamic virtual model (that can evolve independently from the "real" system)
- Data from executions of the system in variable conditions
- Data modeling and ontologies

There is also one additional prerequisite going across all three levels of twin with increasing complexity:

- People

For each of these prerequisites, we will provide a basic overview, insights on why they are required, and illustrations of common roadblocks to success in industrial environments. Obviously, funding, be it through private means within companies or through public subsidies, is also a prerequisite in most cases but this topic goes beyond the scope of our investigation.

3 Major Prerequisites

3.1 Sensors

The basics of a digital twin being to create a common state between a physical reality and a virtual representation, we need to have sensors that can measure the state of the physical assets and provide that data.

Such sensors exist on most modern equipment and the growth of industrial IoT is generating a lot of innovation making it easier to add additional sensors to existing equipment. There currently exists a wide range of sensors that will capture the state of a physical environment (temperature, pressure, pH, movement, flow, etc.) and we increasingly see so-called software sensors that will analyze basic physical parameters and convert them into higher level measurements [1]. This will typically be the case for measuring biological processes or using computer vision to extract complex information.

The common roadblocks in this area will usually be converting the signal obtained by the sensor into some useful information for the digital twin. Many older sensors will provide an analog signal where a digital signal is required for our twin. Another common case is where sensors are managed by a PLC that no one has the appropriate competencies to use and the sensor information remains locked within the PLC. Critical approaches to avoid these kinds of issues include defining corporate standards for PLC/sensors upon purchase and implementation that ensure such black boxes do not happen.

Beyond sensors, we shall also usually need data from standard Information Technology solutions such as MES (Manufacturing Execution System), LIMS (Laboratory Information Management System), ELN (Electronic Lab Notebook), etc. However, these are generalizable as sensors where we have a human intermediate between the measurement system and the data capture which we would generally try to avoid as we develop more sophisticated levels of digital twins. It should be noted, however, that the sensors and measurement system only need to be as good as the goal given for the twin and in many cases the required sensor quality required for observability of the model can be quite low.

3.2 Connectivity

One thing is to have measurements coming from a sensor, we also need to transmit that data from the physical asset to the virtual environment.

Traditional approaches will include SCADA and wired networks with a historian solution. Many players increasingly use variations of IoT boxes and standard mobile networks such as wi-fi and to make the connection between automation and digital environments while we are seeing mounting maturity of IoT network approaches in different variations converging towards 5G [2].

In parallel, there are a lot of emerging solutions around Industrial IoT platforms from most traditional industrial software players that make this connection between the edge and digital platforms.

Whatever the solutions, usual complexities come from the interconnection of multiple protocols originating in different worlds (automation vs IT vs telecom mainly) and the variable interpretations of standards creating confusion and inconsistencies. Another major pitfall to avoid is in the translation and volumes of data in different environments. The brutal reality of an analog sensor will usually overwhelm digital infrastructures to make the right conversions and simplifications at the right time in the process is critical. The virtual model will normally only require these condensed data sets to be operational.

3.3 Virtual Model of Physical Asset

We need to create a reality of the physical asset as a virtual model. Otherwise, we are no longer doing a digital twin, we are just doing data capture, analysis, and simulation on measurements. In other words, the virtual model needs to be self-sufficient to understand the system without having direct access to the physical reality we are creating a twin for. For instance, traditional Information Technology solutions such as ERP (Enterprise Resource Planning) have captured data on manufacturing operations but only logically without any representation of the actual physical factory and manufacturing equipment.

The representation of the physical asset as a virtual model will often take the form of a 3D model using various CAD standards [3]. These models will under usual conditions be produced by engineering teams designing the physical asset, integrating sub-models from the equipment providers contributing the various components of the physical asset. A limited number of standards exist that are quite interoperable at a basic level of modelization.

In this area, we start to see some major roadblocks appearing. Virtual models will in many cases be considered proprietary Intellectual Property of the providers, both engineering and equipment providers. It is therefore critical to ensure that initial contracts include appropriate availability and use rights of such models. Another more subtle roadblock is that these models are developed with a lack of transparency or respect of standards that basically make them unusable. Defining beforehand the standards, naming conventions, hierarchies, etc., to follow are important elements of success. The reality of today is that models are developed both by equipment providers and by engineering companies, but nobody actually uses these models beyond the short period of design and construction.

In other areas such as biological systems, 3D models are not the most efficient way to represent the system and other approaches such as computational whole-cell modeling are being used [4].

Having access to an open and programmable model where you can bring in and display the measurements provided by sensors is a last and most important constraint.

3.4 Asset Framework

As we move into more dynamic digital twins, we shall start seeing the need for an additional prerequisite. In our basic solution, we would be manually making the connection between a sensor and the corresponding objects of our virtual model. However, this quickly becomes impractical as the number of sensors increases and we start seeing evolutions of the different assets and models.

The key to overcome this challenge is through the use of an asset framework that manages the overall plant and asset hierarchy and creates an abstraction layer between the physical reality and sensors and the virtual models [5].

An asset framework is thus a hierarchical, contextualized, and digitized model to describe all physical assets (including sensors) in the system and their relationships. In this way, individual sensors and their measurements are not just data points but acquire meaning within the digital twin system and become resilient to changes to physical assets or models. A temperature sensor on a bioreactor remains as such even if we move the bioreactor to another factory or upgrade the bioreactor.

Most solutions for connectivity include asset framework capabilities. The issue is, however, double:

In many industrial environments, these asset frameworks were either not implemented upon commissioning of the equipment or they have not been maintained. Basic operations without digital twin initiatives can usually survive without a good asset framework. Digital Twins without it will not remain operational for long.

There will often be multiple asset frameworks, and these become inconsistent over time if they were not so already from the start. It is not uncommon to find asset frameworks in a historian, the ERP environment, MES, and an IoT platform solution. All at the same time with variable standards.

The most obvious way out of this dead-end is to manage the asset framework not as an element of a transactional system (Historian, ERP, etc.) but as true master data within an MDM solution that will provide a single source of reference for other systems. For most organizations, this will be a serious change.

3.5 Configuration Management

Managing change is another prerequisite for a successful digital twin. It should be obvious by now that a digital twin requires the synchronization of a lot of independently moving parts. While this is of course theoretically possible to do manually it quickly becomes complex and error prone.

Ensuring that changes are appropriately validated with the right workflows, that adequate impact analysis is done beforehand, and that consistent configurations of stable states are managed is a job by itself [5]. It will also typically require a dedicated solution. Some companies have managed to develop this as a dedicated solution but in most cases using a standard solution from the PLM world will be the best answer.

No solution is perfect and not in this case either. Leveraging PLM solutions usually developed in discrete industries has benefits but adapting these to the conditions of a Biopharma industry remains a major challenge.

3.6 Dynamic Model

Up to now, we have remained within the limited ambition of having a virtual model that continuously represents the physical environment. As we move beyond that into an area where the virtual models start to have an autonomous existence and influence on the physical environment, we start to create new prerequisites.

The virtual models need to be able to evolve by itself and influence backwards to the physical world, which essentially means that it needs to contain a model whereby it can:

- Enable simulations of system behavior without input from the physical environment. This will usually for a digital twin include a combination of mechanistic and algorithmic models to cover behaviors of the physical assets and biopharmaceutical processes.
- Drive changes in the physical environment based on measurements and predictive simulations. This will normally encompass a sophisticated process control strategy and model that defines how you can interact with the physical reality.

The dynamic model is the combination of the two. Unsurprisingly this is a much more open area where few solutions exist. Yet, having the depth of equipment and process understanding required to design such models and control strategy is a condition for building a useful digital twin. This is the moment where we start using the digital twin not just for information but to actually influence operational outcomes [6].

There are many potential pitfalls in this step. The standard process control standards in operations today are far from the level required to be useful in a digital twin context. Process characterization, understanding, and control will not be at the level of formalization to be automatically leveraged by a digital simulation. And on the models for how the system operates, we are often faced with the competing approaches of trying a bottom-up approach, component by component, that fails to capture the system dynamics that actually defines what happens; and an end-to-end heuristic model delivering insights and correlations devoid of any physical causality, amplifying noise in measurements more than modeling the real world.

Success in this area requires a convergence of practices in different disciplines which is still very emerging.

3.7 Data

Many digital twins fail due to a lack of data from previous runs of the target process and equipment. Without enough volumes of data many of the modern machine learning algorithms become unusable. While we have seen success from pure mechanistic modeling in many discrete manufacturing operations (Automotive, Aeronautics,. . .), in Biopharma we seem to still require real-world experiments to provide the context for simulations.

Capturing data is not an issue by itself but there are two challenges:

- Providing experimental data requires several actual experimentations. For new operations, this can only come from development activities. To be noted that some industries have successfully created the ability to do virtual experimentation (a case in point is autonomous driving where most major players have successfully created virtual playgrounds to increase the learning of the driving machine learning model). In Biopharma, based on the complexity of the system we shall most probably have to go in the direction of a mixed model.
- Beyond the measurement, you need to understand what it means and be able to reuse it in different context. This is an overall question of appropriately "contextualizing" data. In many cases data has been captured but without tracking the different changes that were done to the system while doing experiments. Thus, you need to "realign" those datasets to make them comparable. The different capabilities discussed above in point IV. and V. are what helps you achieve this.

Generating data will in many cases be one of the longest parts of a digital twin project or barriers to doing it.

3.8 Data Modeling and Ontologies

Good data modeling is always important for any kind of data analysis project. In cases where we are manipulating complex data with high variability, it is critical to raise the bar even more.

Developing formal ontologies for the key data domains of the twin and leveraging existing standard taxonomies is the current good practice for addressing this area [7].

The field is still emerging, so overcoming the two barriers of finding the right skill sets and building on solid initiatives that will become industry standards is a key challenge. An example of an initiative with such a potential is allotrope.org which is building a complete framework of ontologies, taxonomies, and data modeling for analytical test data.

Failing to do the appropriate effort in the early stages of a twin project might save some time but as the twin model starts to evolve the increasing load of managing changes to data model versions and realigning datasets to enable analysis across experiments quickly becomes unsustainable and may completely break the project.

3.9 People

All projects require people of course but digital twins have specific constraints. Looking at the list of prerequisites, you see that a lot of different competencies are required to be put together and collaborate efficiently. Also, many of these experts will need to work slightly outside their traditional comfort, some because the interdependencies of the digital twin create additional stress on the technologies and domains being applied.

Beyond this, there are also major mindset changes required to be successful. You need people to trust data coming from the twin to change how they operate things. This is less obvious than it might seem in many operational environments. Another change in behavior that a digital twin will require is to move from an approach of "Tests → Hypothesis → Confirmation" to an approach of "Simulation → Hypothesis → Confirmation Test."

4 Typical Lifecycle of a Twin

Building digital twins is complex, as it requires a deep understanding of the physical entity and the synchronization of the virtual and physical side of the twin. That is why digital twins generally go through iterations. If we investigate the case of digital twins for industrial assets (production equipment, factories, supply chain,etc.) the first step will generally start while designing a physical entity using CAD-type tools. A relatively limited additional investment connecting sensor data to this virtual model will then deliver a basic twin. In most cases, however, as the physical entity enters operational life the two components will get out of sync and the digital twin falls apart.

The second phase will then often be the development of a digital twin on an existing physical phenomenon. The focus then becomes much more on the design of a virtual model of this, both the 3D type modeling and more importantly dynamic modeling of the processes taking place in the physical entity. While in the first approach the focus was on the design and construction of the physical entity here the focus will usually be more on the understanding and optimization of the system with a much stronger business goal. Having the experience of how the physical entity operates and changes also makes it easier to construct the capabilities to manage change in each component of the system. This kind of digital twin can then be enriched to provide more predictive and simulation capabilities but once again it will usually start hitting limitations.

Similar dynamics will exist in digital twins of biological systems but with the added challenge of lower level of control over design and of understanding of the physical entities internal mechanisms.

The ideal lifecycle is at a higher level of maturity where in the initial design equal focus is given to the physical entity and the virtual model as well as their interactions. This also requires designing into the system from the beginning of how the virtual simulations can influence and direct the physical reality. In the case of preexisting physical phenomena such as biology, this of course requires a very deep understanding of that entity in order to formalize upfront what are the characteristics of the physical phenomenon that can be influenced or controlled and embedding in the virtual model a control strategy where outcomes of virtual model simulations can be applied to the physical entity with a continuous feedback loop.

5 Digital Twin: Potential Applications in Healthcare and Biopharma Industry

Digital Twins acting as a digital replica for the physical object, product, and/or service are the next source of competitive advantage for the Healthcare sector.

They are key to accelerate the move towards preventive and personalized medical treatments by modeling reality with advanced analytics techniques in such a way that problems can/will be predicted whether and when they occur, providing the time necessary to treat the patient in advance.

Furthermore, digital twins can provide a safe (virtual) environment for testing the impact of changes on the performance of a specific system or living body. This will enable optimal solutions, as they imply minimal capital investment, but more importantly will drastically limit risks which are obviously critical in the health sector.

A digital twin can be designed and implemented to improve "care-delivery" services, patient experience, and specific treatments as well as the overall Healthcare value chain. The potentials are unlimited as today we are just "looking" at the peak of the iceberg.

However, like in any technology, the misuse of digital twin can drive issues both on the data privacy standpoint as well as accessibility of the potential solution that, due to the high cost of entrance, may drive inequalities vs patients. Hence, the need for a strong governance body that would regulate the transparency of data usage, data privacy, and type of personalized treatment.

Let us consider what it could mean for a *single individual as a patient*: the pairing of the virtual and physical worlds allows analysis of real-time data and monitoring of responses to treatments, as well as tracking the results of behavior and lifestyle modification, to prevent any type of disease prior to their occurrence. For example, a wearable sensor (iWatch, fitbit, Oura ring, etc.) could track a patient's blood pressure, body temperature and link that information, as well as data on the patient's lifestyle and genetic characteristics, to a digital twin. Consequently, a doctor can develop and test upfront on the virtual twin the most suitable medical plan and lifestyle recommendations that, upon positive confirmation, could then be implemented on the "living" patient.

Another example, within the Healthcare industry, could come from virtualizing a hospital system to create a safe environment in which to test the impact of potential change on system performance. One of the most typical issues that hospitals face is the dilemma between immediate availability of medicines, stock management due to limited space, and the relative lifecycle of drugs that may create "expired medicines." A digital twin can simulate these complex scenarios and identify the right trade-off between inventory on-hold and availability of medicines.

Furthermore, there are many potential use cases for digital twins in *R&D and OPERATIONS* (Manufacturing and Supply Chain) including real-time monitoring, simulation, modeling, and virtual (remote) control of physical assets. Digital twin is playing a key role in manufacturing process development and many operations processes optimization, in harmonizing products with processes and in defining a holistic process control strategy and ultimately realizing the concept of "Quality by design" (QbD). The QbD is a systematic approach in pharmaceutical drug development to ensure predefined product quality by identifying and understanding the impact of all critical process parameters (CPP) on all and critical product quality attributes (CQA). The use of digital twins helps the pharmaceutical industry in aiming for the reduction of drug development and manufacturing cost as well as reducing the timelines for getting drug candidates to the patient.

The fast progress in IoT will certainly accelerate the adoption of this technology. In fact, according to a recent Gartner research, 75% of the companies using IoT in manufacturing will develop a Digital Twin within 12–18 months.

Finally, in the next section, we will make a deeper dive into two real case studies of digital twins implemented in the pharmaceutical and healthcare industry.

6 Digital Twin: Case Study from Merck KGaA Darmstadt, Germany

6.1 Overall Approach

In order to maintain competitiveness in this "new world," Merck KGaA, as a vibrant Science and Technology company, has been significantly investing since several years in digital technology for manufacturing and supply chain to pursue its vision of "Self-Driving Supply Chains."

We envisaged a Healthcare Supply Chain where the DEMAND is "automatically predicted," without any need of human intervention, through a combination of Statistical Forecasting, based on machine learning, and Predictive Forecasting, based on Real World Evidence data thanks to advanced analytics techniques able to integrate structured and unstructured data, in order to derive a forward-looking signal.

Likewise, SUPPLY is "proactively prescribed," through Control Towers (a sort of center of expertise and decision-making for supply chains), based on Digital Twin systems modeled by real-time information flows, enabling production synchronization across the overall E2E network. This enables a LOGISTICS Distribution system personalized, affordable, and agile.

Figure 1 illustrates this concept:

However, the journey towards the vision is made of a series of important steps, one after the other, that enables a progressive development of the three key elements: Technology, Process, and People.

Our recommended "journey" is made of four steps:

- *Step 1 – Integrated Supply Chain*: whereas we operate in a state of "REAL TIME" End-to-End visibility of SC performance/KPI's, gathered through online descriptive dashboarding, all this supporting the IBP (Integrated Business Planning) process towards "One Number" concept.
- *Step 2 – Predictive Supply Chain*: this is the "FORWARD LOOKING" state in which we are able to get a clear demand signal thanks to predictive capabilities (killing Bullwhip effect).

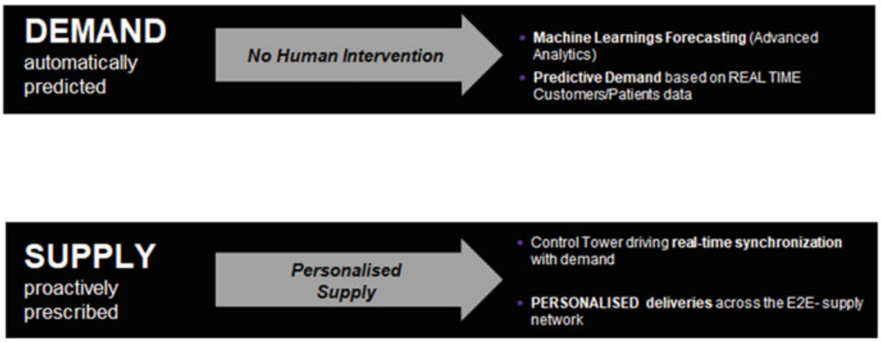

Fig. 1 Self-driving supply chain concept

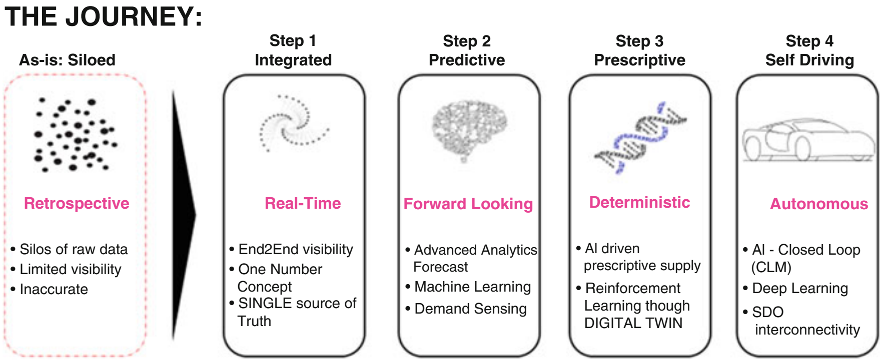

Fig. 2 The journey to implement the vision of a Self-Driving Supply Chain

- *Step 3 – Prescriptive Supply Chain*: moving towards the "DETERMINISTIC" state whereas through the creation of DIGITAL TWINS, the reinforcement learning AI solutions, we have supply planning "prescribed and recommended" by the system (upon human final decision).
- *Step 4 – Self-Driving Supply Chain:* the last step in which the overall balancing of supply vs demand as well as distribution and logistics is totally AUTONOMOUS.

Figure 2 shows the step-by-step approach:

6.2 What Is the Technological Backbone?

Nearly all Pharma companies (and not only) come from a world whereas "ERP is the king." ERP were the systems needed to reduce operational costs, to drive efficiencies through effective data gathering, and to leverage synergies by standardizing processes. While this is (was) generally true, true E2E-Supply Chain-level decision-making was never enabled. Hence the key reason why the end-to-end digital supply chain twin is an important milestone: these new systems do sit "above" any ERP or data gathering source and their analytical layer provides prescriptive insights into the interconnected decisions that are inherent in supply chains (Fig. 3).

At its core is a digital supply chain twin is:

- Connected outside-inside
- Gather and process info real-time (is always "ON")
- Autonomous
- Intelligent

Fundamentally, such a system allows the flexibility for performing simulation scenarios and modeling that can be evaluated without having to necessarily conform to the design constraints of your current supply chain nor impacting the operational activities.

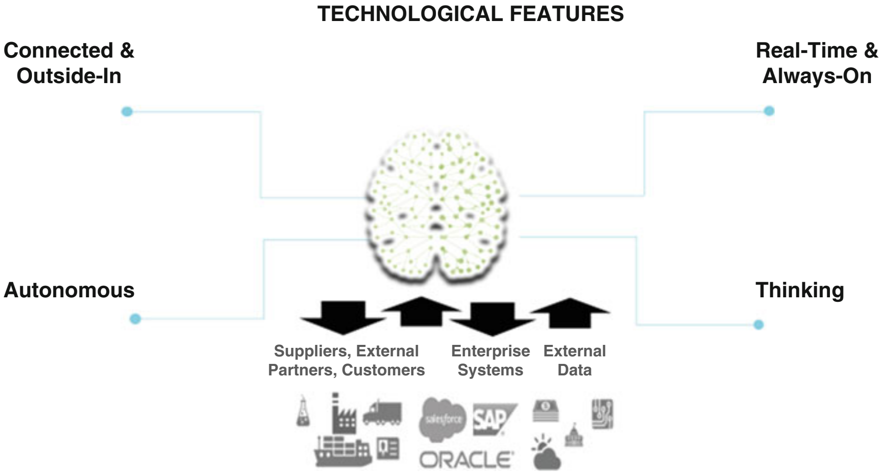

Fig. 3 The technological backbone and its features

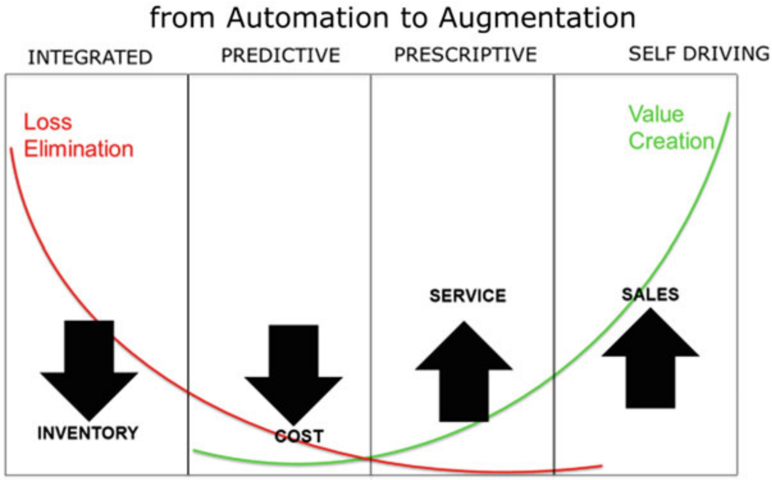

Fig. 4 The journey from Automation towards Automation

It is a journey from "Automation" where companies are increasingly automating processes and eliminating losses towards "Augmentation" enabled by new cognitive automation platforms drive value creation (Fig. 4).

However, this journey is "not exactly a walk in the park" and the two biggest challenges to overcome are:

1. Ensure a very high level of *MASTER DATA ACCURACY*
2. Empower and support a strong *CHANGE MGT PROGRAM*

The first point, Master Data Accuracy, seems trivial but it is not. Indeed, any digital solution becomes useless if data is not correct.

The second point implies the continuous "re-skilling" and "up-skilling" of the supply chain people as new skills and mindset are required.

Effective Change Management is the crossroad between the success and failure of any digital transformation.

We live in a "VUCA" (Volatility, Uncertainty, Complexity, Ambiguity) world where tons of real-time data, better processing power, and more sophisticated analytical algorithms are starting to create seismic shifts in how the supply chain will operate today and in the future.

7 Digital Twin: Case Study from Sanofi

7.1 Objective

This initiative was developed by Sanofi and partially demonstrated at the Usine Extraordinaire event took place in Paris in November 2018.

The objective was to improve operations of a cell culture production line with an endpoint goal of reducing operational incidents leading to lost batches or release delays and to improve overall productivity and yield of production.

7.2 Challenges

The two key challenges are (1) the creation of a model that could reliably simulate the biomanufacturing processes, and (2) ensuring that the operational teams integrate the use of the digital twin in daily activities and are confident with the outcome of the virtual model.

For the first challenge, the main barrier quickly became the availability and quality of data. As we started building models it became very apparent that the data being used to run an industrial process on a daily basis was insufficient to train and validate a model that could reliably replicate the process.

For the second challenge, the issue was different and caused the project to change direction somewhat. While the initial focus was very much on the simulation model, working with the operators it became clear we needed to take a much broader approach where the digital twin was not just a sophisticated technology for a few experts but a solution for the operators in their complete lifecycle experience from training to operations and continuous improvement. Creating operational impact from the digital twin required true operational familiarity and trust from the operational teams.

7.3 Digital Twin Solution

The end approach consisted of (1) building a detailed 3D CAD model of the production line facility and equipment, (2) designing a simulation model for each of the production steps based on historical runs of the same process using key production parameters and measurements and that could predict end-stage values of key outcomes (yield and failed batch) in the initial phases of the step, and (3) connecting sensor data from the process and analytical tests to the model to visualize in context of the equipment.

A first use case was then averaging this model to develop training modules where operators could learn how to operate the process and by themselves use the simulation model to understand and learn how operations parameters and decisions impacted the outcome of the biomanufacturing process.

This training experience then makes it much easier to leverage the twin in operational conditions because the environment and types of decisions are already familiar. Not only does the 3D model provide better training on how to operate an equipment but better understanding of what is happening in the bioprocess increases the right operational procedures.

The second use case was focused on providing shared real-time information on production process status. We confirmed that in many cases the key data on production status is not broadly shared, both because of basic information access and because that data is often managed in technical environments requiring very expert skills to understand its meaning. Providing easily understandable data in the context of the twin is a key enabler in creating the shared understanding of status that avoids operational issues caused by misunderstandings, lack of information, or erroneous data transmission.

The third use case was to enable a set of key decisions on the shop-floor with outcomes of twin predictions. The three main decisions targeted were:

- Decision to stop a batch early that would fail thereby enabling production to save time and quickly restart a new batch
- Decision on timing to end a batch when optimal yield/duration has been hit
- Decision to adjust process control parameters (within the specification of course) in order to optimize yield or avoid deviations and lost batches

7.4 Lessons Learnt

As stated above, integrating the human element of operations was a critical factor in the direction that the twin project took. Creating the visual element of a 3D model and training experiences was a key enabler in operations buy-in and support for the more advanced use cases.

The human factor also drives the need to clearly define and prioritize the operational decisions that the twin is aimed at enabling. It is of course possible to develop a digital twin for research investigations and broad exploration. However, in an operational environment, the possibilities of a twin can quickly just become information overload and create a lack of trust which is fatal for efficient usage. Therefore, the type of work of proactively defining which decisions are aimed at being enhanced and only focusing on them is a major success factor. If there are too many options, the decisions quickly get evaluated based on the risk one might take versus standard operations without a digital twin instead of the expected operation upside.

Finally, we mainly targeted a modeling approach focusing sequentially on each production step and using combinations of mechanistic and heuristic modeling. While this has the benefits of targeted modeling activities that remain more easily scoped and manageable, it has several caveats. The main issue is probably that as we optimize each step sub-model, we are targeting an endpoint outcome that may actually not be optimal for the end-to-end process. What we optimize for at each step is only as good as our understanding of the overall process.

We explored end-to-end model development using machine learning approaches and while promising it once again becomes apparent that higher volumes of data and better quality are absolutely required for this approach.

8 Conclusion and Outlook

The digital twin approach is still in the emerging phase of its use within the Biopharma industry. We have seen examples of use from process development to manufacturing and supply chain. However, that is only the start and fascinating opportunities also exist in other areas beyond industrial operations. A lot of ink has already been spent around the notions of quantified self and biohacking, investigating possibilities where we take twin approaches to human individuals.

While raising many ethical questions there are also areas such as clinical trials where such approaches would deliver unquestionable benefits. All levels of twins can apply from just having real-time sensors continuously updating a digital replica to the most sophisticated twin where we could do simulations, prediction of therapeutic impact.

The ultimate end goal for digital twin applications within Biopharma and Healthcare would be to no longer have any need for conducting clinical trials on living beings but already a first step would be to shift from an approach of doing experiments and then looking for a model that would explain results to an approach where we execute simulations on the twin to define a predicted best scenario and then verify whether actual therapeutic conditions follow the parameters of the predicted model.

In short, using experiments not to experiment but to validate the hypothesis produced by twin simulations. While radical, one should remember this is already the shift that has taken place in areas such as automotive where nearly all crash tests are now simulated instead of being actually conducted with a prototype vehicle.

In addition, the pairing of the virtual and physical worlds will enable a great step-change towards preventive medicine in a personalized approach. Indeed, the tracking of real-time data, the possibility of analysis and monitoring of responses to treatments in a virtual modeling twin, as well as tracking of the results of behavior and lifestyle modification, will allow the personalized prevention of any type of disease prior to their occurrence. And this is going to be a key breakthrough for science and humanity.

References

1. Chhetri SR, Al Faruque MA (2020) IoT-enabled living digital twin modeling. In: Data-driven modeling of cyber-physical systems using side-channel analysis. Springer, Cham, pp 155–182. https://doi.org/10.1007/978-3-030-37,962-9_8
2. Cheng J et al (2018) Industrial IoT in 5G environment towards smart manufacturing. J Indus Information Integr 10:10–19
3. Tao F, Zhang M (2017) Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. IEEE Access 5:20418–20,427. https://doi.org/10.1109/ACCESS.2017.2756069
4. Marucci L, Barberis M, Karr J, Ray O, Race P, Andrade M, Grierson C, Hoffmann SA, Landon S, Rech EL, Rees-Garbutt J, Seabrook R, Shaw W, Woods C (2020) Computer-aided whole-cell design: taking a holistic approach by integrating synthetic with systems biology. Front Bioeng Biotechnol:8. https://doi.org/10.3389/fbioe.2020.00942
5. Weyer S, Meyer T, Ohmer M, Gorecky D, Zühlke D (2016) Future modeling and simulation of CPS-based factories: an example from the automotive industry. IFAC-PapersOnLine 49 (31):97–102
6. Dahmen U, Rossmann J (2018) Experimentable digital twins for a modeling and simulation-based engineering approach. In: 2018 IEEE international systems engineering symposium (ISSE). IEEE, Piscataway, pp 1–8. https://doi.org/10.1109/SysEng.2018.8544383
7. Steinmetz C, Rettberg A, Ribeiro F, Schroeder G, Pereira C (2018) Internet of things ontology for digital twin in cyber physical systems. In: 2018 VIII Brazilian symposium on computing systems engineering (SBESC). IEEE, Piscataway, pp 154–159. https://doi.org/10.1109/SBESC.2018.00030