Building an organizational digital twin

Article *in* Business Horizons · August 2020

DOI: 10.1016/j.bushor.2020.08.001

3 authors:

Rashik Parmar
IBM

Aija Leiponen
Cornell University

Llewellyn D W Thomas
IESE Business School

BUILDING AN ORGANIZATIONAL DIGITAL TWIN

Rashik Parmar
IBM
Princes Exchange, Princes Square
Leeds, LS1 4HY
United Kingdom
Email: rashik_parmar@uk.ibm.com
Phone: +44 11 3280 6703

Aija Leiponen
Charles H. Dyson School of Applied Economics and Management
Warren Hall, Cornell University
Ithaca, NY 14853-6201
United States
Email: aija.leiponen@cornell.edu
Phone: +1 607 255 7588

Llewellyn D W Thomas
LaSalle Universitat Ramon Llull
Carrer de Sant Joan de la Salle, 42
08022 Barcelona, Catalunya
Spain
Email: llewellyn.thomas@salle.url.edu
Phone: +34 932 90 24 00

BUILDING AN ORGANIZATIONAL DIGITAL TWIN

ABSTRACT

The increasing digitization of the economy means that a new digital resource – a digital twin of an organization – is now feasible. We suggest five principles that assist in the construction of an organizational digital twin and show how they combine into a dynamic evolutionary process that incrementally builds and maintains the digital twin. We also discuss the organizational implications of implementing a digital twin, and how digital twins create value in an organization.

Key words: digital twin; digitization; big data; data analytics

AN EMERGING DIGITAL OPPORTUNITY

Digitization is proceeding at a fast pace. While organizational processes have been progressively digitized for decades, recently the pace of digitization has increased as more and more digital devices are brought online. For instance, there were approximately 8.4 billion connected devices in 2017, variously called the Internet of Things (IoT) and "smart" devices, up 31 percent from the previous year (Gartner, 2017a). By 2020, Intel suggests there will be a projected 200 billion smart objects, around 26 smart objects for every human on Earth (Intel, 2018). As a result, there has been an explosion of data – 90 percent of all data today was probably created in the past two years and at the current rate 2.5 quintillion bytes of data are created each day (Marr, 2018).

One of the emerging opportunities from this increasing digitization are "digital twins". Digital twins are digital replicas of assets, using 3D modelling and sensors to create digital representations (Haag & Anderl, 2018; Tao et al., 2018). Digital twins are not new. Engineering teams have used 3D renderings of computer-aided design models, asset models and process simulations for more than 30 years, and NASA has had digital simulations of spacecraft even longer (Gartner, 2017b). However, the Internet of Things (IoT) and increasing availability of real-time data has enabled a new generation of digital twins. When sensors collect data from a connected device, the sensor data can be used to update a "digital twin" copy of the device's state in real time (Haag & Anderl, 2018). The digital twin becomes an up-to-date and accurate copy of the physical object's properties and states, including shape, position, gesture, status and motion. For instance, GE, through its Predix platform, provides a software representation of physical assets such as power generation turbines, jet engines and locomotives. Digital twins enable firms to better understand, predict, and optimize the performance of an individual asset, an integrated system of assets, or a fleet of assets. As a result of these benefits, recent surveys of organization managers have shown that digital twins of physical objects are entering mainstream use, with 13 percent of organizations implementing IoT projects already using digital twins and 62 percent are either in the process of establishing digital twin use or plan to do so (Gartner, 2019).

In this paper, we suggest that the concept of digital twins can be extended from digital representations of physical objects to digital representations of whole organizations. Rather than just focusing on the data from IoT sensors, the digitization of organizational processes and the rise of internet connected consumer and industrial devices means that the data that flows from organizational assets, people, activities and their interactions can be combined into a holistic digital model of the organization – an organizational digital twin. This digital twin will be a living digital simulation model of the organization that updates and changes as the organization evolves. And upon becoming available, the digital twin will allow scenarios to be thoroughly tested to predict the performance of potential tactics and strategies.

We believe that once organizational digital twins become a reality, many of the promises of management consultants, business gurus and IT mavens of data-driven value creation can be realized (see, e.g. Davenport, Barth, & Bean, 2012; Manyika et al., 2011; Mayer-Schonberger & Cukier, 2013). Cost savings can be realized as simulations of organizational processes are used to identify more efficient ways of doing business. Improved understanding of market conditions through simulation of customer purchasing behaviors can result in competitive advantage. The ability to quickly simulate how organizational performance will respond to a particular decision can increase organizational agility through quick, even automated, decisions based upon data insights. New product development can be improved as product simulations combined with customer behavioral simulations result in products that meet the needs of customers. All these can be supercharged through the application of sophisticated analytics and artificial intelligence to data (Jarrahi, 2018).

In order to build an organizational digital twin, we argue that there are five key principles which underpin its successful construction. These principles combine into a dynamic evolutionary process which incrementally builds the digital twin. We also suggest that there are two implications, which, if not addressed, reduce the likelihood of the successful construction of a digital twin. We also outline how digital twins can create value for organizations. By applying these principles and addressing the implications of digital twins, it possible to systematically capture value from the data within an organization.

FIVE PRINCIPLES TO BUILD A DIGITAL TWIN

Fundamentally, a digital twin consists of a digital template model of an entity, data from that entity, a unique digital representation of the entity derived from the template, and the ability to monitor the entity. For a physical asset, much of the challenge is in instrumenting (such as through IoT sensors) the asset so that there is a close correspondence between the entity and the data obtained. Although not all characteristics of a physical asset may be instrumented, the nature of the asset (such as a locomotive or electric turbine) determines what characteristics best provide a representation.

To build a digital twin of a physical asset, a digital template of the asset is required. This template needs to be calibrated so that it accurately corresponds to the specific physical asset. Put differently, the digital template of a physical asset, such as a digital model of that make and model of electric turbine, needs to be adjusted so that it accurately represents the actual physical electric turbine (the specific physical asset itself). In practice, once a digital twin template has been created for a particular make and model of a digital asset, it is reused for each asset that is produced, enabling scale economies such that each subsequent digital twin is cheaper to produce than the previous. Once a unique one-to-one correspondence between a physical asset and the digital template is established, then the sensors that monitor the physical asset send data so that the unique digital twin is updated to reflect the state of the physical asset.

The process of building a digital twin for an organization is different, however, as there is no original physical design with which to build a template. Put differently, there is no physical asset which provides the contours of the initial template for the digital twin, as every organization is unique, consisting of an idiosyncratic collection of assets, processes, and interactions (Barney, 1991; Grant, 1996; Pfeffer & Salancik, 1978; Teece, Pisano, & Shuen, 1997; Thompson, 1967).

To build a digital twin of an organization, then, it is necessary to take a different, emergent approach. We suggest there are five principles that need to be considered when building a digital twin. The first principle – start with what you have – addresses production of the initial, seed, digital twin from which the full digital twin emerges. The second principle – set the data free – requires the breaking down of informational silos and allowing access to the data within the organization so that it becomes integrated into the model of digital twin. The third principle – move the digitization frontier – requires the proactive digitization of assets, processes and interactions, thereby enabling the expansion of the digital twin. The fourth principle – seek new digital opportunities – requires the leveraging of the resulting generativity that increasing digitization brings to actively seek out new digital business opportunities. The fifth principle – increment the models – requires thinking beyond the maintenance of the twin and instead consider active, incremental construction of the digital representation of the organization.

We now discuss each in turn, and show how the principles, properly executed, can result in a dynamic evolutionary process that can drive both the emergence and the growth of the digital twin.

Principle 1: Start with what you have

As there is no pre-existing template for an organization, the first digital representation is built from the existing digital artefacts, such as sensors, data, analytical models and software.

To do so, a comprehensive survey of the digital technologies in use is first undertaken, paying careful attention to consider all assets, processes, and interactions within the organization. This is not as straight forward as it seems; many data and analytical models are not explicit, but instead are embedded in existing software and hardware artefacts across the organization. For instance, many organizations use Enterprise Resource Planning software, such as that provided by SAP, which is deployed with many preconfigured data and analytical models which need to be clearly understood in order to model the organization.

There are also industry data templates that can be used. IBM, for instance, has “Industry Data Models”, consisting of a set of organizational and technical data models that are pre-designed to meet the needs of a particular industry (IBM, 2019). These often include data warehouse design models, organizational terminology and business intelligence templates. These are blueprints that provide common elements derived from best practices, government regulations, and the complex data and analytic needs of an “ideal” organization. While in many ways these are similar to the initial digital templates that are calibrated for physical assets, in practice these require extensive consulting expertise to adjust for an individual organization, as they are necessarily an abstraction of firms in a particular industry, and do not capture all the idiosyncratic assets, processes, and interactions that typify a specific organization.

Some of the data required to build a digital twin may not be from within the organization. For instance, there will be demand-side data, such as market and consumer data. Although much of this data may be collected by existing market intelligence, customer resource management, and digital marketing systems, there may be data held by third party complementors which can also provide vital market insight. There are also data from the supply-side, which are now increasingly coming from supply chain management software and from firms deep within the supply chain. At times some of these data will not be under the direct control of the organization, even though it may have access to them.

Taken together, the existing digital artefacts within an organization, extended by additional data and analytical models from suppliers and complementors, and supplemented by industry templates, provide enough digital raw material which can be shaped into the initial digital representation of the organization. These digital artefacts and the initial digital representation need to be interpreted through three models. The first, the information model, helps to understand what data artefacts are in the organization, such as sensors, process flows and analytical models. The second model is the context model, which provides insight as to the expected behaviour of these digital artefacts – such as the various states the digital artefact can be in. The third model is the impact model, which provides the interlinkages and correlations between the various digital artefacts that have been mapped. To illustrate by way of a simple example, consider a door in a lift. The information model will describe what sensors are deployed within the lift door; the context model describes what states and tolerances those sensors have, such as open, closed, opening, closing, and blocked; the impact model describes the relationship between those states and other sensors within the lift and the wider environment, such as floor level, weight, temperature, and so on, that influence the performance of the lift.

The process of shaping these existing data and analytical models into the information, context and impact models is not simple and requires specific skillsets which may not be present in the organization, a fact we discuss in more depth below. However, while it is clear this principle does not result in a complete representation of the organization, there is often a strong desire to create a complete representation. This should be resisted, as this principle demands a recognition that an organizational digital twin cannot be created fully crafted *ab initio* but will need to emerge gradually through the interaction of the other principles.

In summary, this principle requires the undertaking of a full audit of the existing sensors, connections, data and analytical models of the organization, suppliers, and complementors, and use these to create the initial entity model of the digital twin.

Principle 2: Set the data free

In contrast to physical assets which often need specific sensors installed so that data can be collected, sensors, data collection, and storage capabilities already exist within most organizations. However, much of these data are enclosed in information silos, often reflecting the broader structure of the organization. Thus, to create the digital twin, the data needs to be freed from informational silos so that it can be integrated into the digital twin. In the same way that mass attracts other masses in physics, and the greater the mass the greater the gravitational attraction, data attract complementary digital artefacts to itself (McCrory, 2010). For instance, a huge repository of data will attract software and services to leverage the data to create value. Data will also attract other data to themselves, as data becomes more valuable the more they are linked with other data (known by privacy theorists as the “aggregation effect”, see Solove, 2004). Thus, the more the data are set free, the greater the potential scope of digital twin.

To continue the metaphor, just as heavier metals have a greater gravitational attraction, so do higher quality data, a feature we call density. Firstly, dense data are normally *unique*, a characteristic which is dependent on how easy it is to infer it from other sources (Koutroumpis, Leiponen, & Thomas, 2019). The more difficult it is to recreate the data from other sources, the more unique the data become and the higher the density. For instance, data relating to an individual’s health are considered unique, as they can only be inferred with difficulty from other sources, such as social media.

Data density also depends on the extent to which is it connected to the object of measurement. We call this notion data *alienability* (Koutroumpis et al., 2019). Some types of data are inalienable from the originator and generate ongoing implications for them. For example, data about a person are not fully separable from that person in the future; in fact, for much data, that’s exactly why they are valuable to outsiders. Consider personal health records. When these data are shared or sold, those data do not just leave the originator’s control; the data buyer can implicate the originator in the future marketing campaigns or insurance policies. This makes the data of higher quality, and hence in our description, have greater gravitational attraction.

Finally, data density is related to *freshness* (Koutroumpis et al., 2019). Data that is “stale”, that is, out of date, has less gravitational attraction than current data. The level of freshness is dependent on both the useful life of data as well as how often it is sampled. The useful life of an item of data varies depending on its source: for instance, health care data has a useful life of many decades, while the useful life of personal location data may be much shorter, measured in units of hours if not days.

However, if data cannot be accessed, the power of data gravity is not given the space in which to act. If data are not accessible, they cannot be included into the information, context and impact models that underpin the digital twin. Similarly, they cannot connect to other software and analytical models to enable new services and analytics. However, giving data space does not mean sharing the data with the world. For instance, Barclays and Citigroup began opening up the data through internal hack-a-thons – bank staff were able access the previously closed off data to create new services and insights. By letting data gravity work, the banks were able to realise value from the data. So successful was this activity that both organizations subsequently opened up their data to the trusted partners to let gravity work its magic. They are now thinking of opening the data beyond the initial stakeholders.

However, given that some of the data that underpins the digital twin may be from third parties, or even subject to specific regulatory controls (such as personal data), freeing data necessitates careful consideration of legal and technical protections against unauthorized uses.

For instance, the recent General Data Protection Regulation (GDPR) of the European Union, which mandates strict personal data protection practices and allows national jurisdictions to set up additional rights to other types of data, will require firms to take particular care. Nevertheless, there are many examples of internal and external data platforms that are restricted and monitored, yet valuable and successful. ID Analytics runs one such platform for firms in several industry verticals that pool the complementary data sources to mitigate identity theft and other types of fraud.

Summarizing, to build the digital twin and to realize the value of the digital artefacts within the organization, data has to be accessible. If data cannot be accessed, then data gravity potential cannot be realized.

Principle 3: Move the digitization frontier

To create an organizational digital twin that is an accurate representation, it is necessary to actively seek out opportunities to digitize those assets and processes within the organization that are only partially digitized, or not digitized at all. Digitization is driven by the evolution of information and communication technologies (ICTs) and the mass adoption of communication devices. While digitization was occurring long before "big data", several technological factors continue to push digitization within firms. Prior to mass computerization, transformation of raw observations to digital formats was costly, often prohibitively so. These costs of data collection have shrunk with the adoption of personal computers, mobile devices, a wide variety of sensors, and the internet. Standardization of data formats, improved network interoperability, the shift to cloud computing, decreasing internet access prices, and the ever-increasing performance of the underlying computing hardware technologies, have further reduced the cost of digital transformation and transaction (cf. Nordhaus, 2002). Digitization itself guides improvement and optimization by increasing the understanding of the thing being digitized, leading to further digitization.

This recursive process of digitization creates digital artefacts – digitized collections of information. Digital artefacts can be as complex as computer operating systems, as basic as lists of numbers, or as vast as libraries of literature and film (National Research Council, 1999). The digitization of processes creates *software*; these are digital artefacts that consist of instructions that can be run on computers (Linde & Stock, 2011). Such algorithms or programs may embed a technical invention and may include operating systems, packaged software, software as a service, analytical models, or artificial intelligence. The digitization of information creates data. We can think about two types of data: *observations* and *content* (Linde & Stock, 2011). On the one hand, observations of the world can be digitized, such as laboratory data and observational data compiled by humans or IoT sensors. On the other hand, creative expression (text, sound, images, moving images, models, games, simulations) can also be digitized. These are digital artefacts that encapsulate written, aural, or visual expression, and are often called content.

While this principle requires an active search for digitization opportunities to enable a better digital representation, the digital artefacts that are created need to align and complement existing digital artifacts within the organization. A key part of this process is ensuring that the protocols and formats that result as part of the digitization process are integrated into information, context and impact models that comprise the digital twin. For example, consider elevator data collected by KONE. As part of the process of adding sensors to the lift to improve the maintenance of the lifts, significant data harmonization was also required. This was because some sensors were from Honeywell and others from GE, and they were associated with different formats and protocols. To capture and harmonize these data KONE needed to create a specific, internal standard that could be used for the data generated by all lifts within the product portfolio. Now much data are mapped to the “KONE standard” within the firm. Of course, industry standard protocols and formats may also be available, but often during the process of adding new data sources into the information, context and impact models there is no standard that exists.

In summary, this principle requires a proactive approach to organizational digitization. By finding new compatible data sources, the digital twin can be extended to better represent the organization. In many ways it is an extension of one of the patterns of data innovation first suggested by Parmar, Mackenzie, Cohn, and Gann (2014). Rather than focus on just digitizing physical assets as they suggest, this principle suggests a focus on digitizing all kinds of assets, processes, and interactions, and ensuring that the resulting data flows are compatible with the information, context and impact models that comprise the digital twin. This principle recognizes that the digital environment is not static and new opportunities to digitize constantly arise. There needs to be ongoing monitoring of the processes and assets within the organization where digitization is not yet pervasive.

Principle 4: Seek new digital opportunities

As more artefacts, processes and interactions are digitized within the organization and represented within the digital twin, then new digital opportunities become increasingly feasible. Underpinning these new digital opportunities is the notion of generativity, defined as a digital technology's capacity to enable change (Thomas & Tee, 2019; Zittrain, 2006). Generativity emerges from the digital artefacts and infrastructures, as well as from the organizational digital twin, that result from increasing digitization. Comprising of the "basic information technologies and organizational structures, along with the related services and facilities necessary for an enterprise to function" (Tilson, Lyytinen, & Sorensen, 2010, p. 748), three properties of digital infrastructure drive new digital opportunities – the ability of digital artefacts to be repurposed, the ability to use data for different purposes, and the organizational flexibility that a layered modular architecture allows.

New digital opportunities emerge through the easy repurposing of digital artefacts, or what is called the *reprogrammability*, as unlike many physical assets, digital artefacts such as software can be flexibly re-programmed to perform different functions (Tilson et al., 2010; Yoo, Henfridsson, & Lyytinen, 2010). This means that the tight coupling between form and function that characterizes physical products is loosened and organizations have more innovation opportunities. New digital opportunities are also enabled through the ability to use of data for many purposes, or what has been called the *homogeneity* of data. Data can be easily accessed by virtually any digital device, and it can be harnessed by a wide range of software for different purposes (Kallinikos, Aaltonen, & Marton, 2010). Finally the *layered modular architectures* that typify digital infrastructures (Yoo et al., 2010) provide increased organizational flexibility. There is both upward and downward flexibility: upward flexibility in the sense that the lower levels of the infrastructure support the creation of virtually any innovation that makes use of the lower level capabilities; and downward flexibility due to ability of any given innovation to draw potentially on a wide range of digital artefacts to perform desired functions (Tilson et al., 2010).

Parmar et al. (2014) suggested a number of new digital opportunities that emerge due to increasing digitization. For instance, organizations can more actively use the data that connected physical objects generate (or could generate) to *improve a product or service*. Examples of this include smart metering of energy usage that allows utilities to optimize pricing, and devices installed in automobiles that let an insurance company know how safely someone drives. Often these opportunities are directly enabled by digital twins, as organizations realize product improvements enabled by data surfaced and visualized by the digital twin. A second opportunity consists of the possibility for organizations to *combine the data within and across industries*. An example of this is a smart-city initiative like the one in Rio de Janeiro, where private utilities, transportation companies, and city agencies consolidate information so that they can deal with natural disasters more effectively. Often the impetus to combine data across and within industries is driven by a desire to add more data to a digital twin. A further opportunity is *trading data*; here, an organization whose information is valuable to another sells it, as when a cell phone service identifies traffic jams by seeing where customers in cars are slowed down and shares the information with a navigation-device company. The increasing insight into organizational data that results from the construction of a digital twin often highlights data that is non-critical and which can be usefully monetized. Finally, organizations can find new opportunities through *codifying a capabilit*y. Here, the increasing insight that digital twins enables the identification of best-in-class processes— managing travel expenses, for instance—and selling it to other organizations across the internet.

In summary, this principle consists of the application of insights from the organizational digital twin to identify and execute new digital opportunities to further extend the reach and coverage of the digital twin. While all of these opportunities may be perceived through the use of the digital twin, it may be that varied opportunities become applicable for different parts of the organization at differing times.

Principle 5: Increment the models

The final principle considers the evolution of the digital twin from the initial information, context and impact models. While initial digital representation of the organization was necessarily incomplete, this not only needs to be maintained, but it also needs to be progressively extended to describe more of the organization. Thus, there are two evolutionary processes which drive the extension of the digital representation: one process consisting of regular maintenance and improvement driven by evolution of the organization (such as organization reorganizations or new organizational activities), on one hand, and substantive improvements to broaden the scope of the digital representation itself (such as new sources of data or more sophisticated modelling), on the other. While the digital twin should fairly seamlessly adjust to regular change within the existing organization, it is less able to automatically adjust through its existing data inputs to new digital opportunities or new ways to model the organization.

The information, context and impact models that comprise the digital twin need to be constantly updated and reconfigured to ensure that new connections, new data, new processes, and new interactions are accurately modelled. As the digitization frontier is continually extended (Principle 3), the information model will need to be updated as new sensors, assets, processes and interactions are digitized. As a consequence, the context model will need to be updated to represent the various states and tolerances of the digital artefacts. And the impact model will need to be updated to better represent the relationships between the digital artefacts. In practice, the information and context layer tend to be updated incrementally, as new products and services are developed and marketed, and software and content assets within the organization are replaced and upgraded.

The impact layer tends to exhibit a different pattern. Ideally, the impact model of the digital twin will be designed so that it integrates historical data into its model and continuously learns and updates itself from the multitude of connected data sources and emerging relationships between digital artefacts to represent its near real-time status. While there is not substantive change to the information and context models, the digital twin will learn from itself, using constantly updating data that conveys various aspects of its operating condition; it will learn from human experts, such as staff with deep and relevant domain knowledge; it will learn from other connected data sources; from other similar digital twins; and from the larger systems and the environment in which it is embedded.

In practice, however, when the digital twin is first instantiated, often there is rapid change in the impact layer as the initial opportunities suggested by the information and context layers are mapped into the context layer and exposed to management. The context model then begins to stabilize as the organization generally begin to operate at the new “business-as-usual” with the digital twin. However, at some point, perhaps as a consequence of a new digital opportunity (Principle 4) or perhaps due to a new hypothesis or insight that requires new ways of thinking about the organization, or perhaps when the organization itself reaches a new state due to external conditions, there is often another rapid increase in the sophistication and insights in the impact layer as managers seek to understand newly perceived interactions in the organization. Thus, the impact layer tends to go through phases of rapid change followed by periods of stability.

In summary, this principle requires active extension of the digital twin, rather than simply maintaining and updating the information, context and impact models. Not only is active incrementing of the information and context models required, but there may be rapid change of the impact model driven by new opportunities, new insights or external change.

A dynamic evolutionary process

While these five principles provide robust and tangible guidance on the various aspects of building a digital twin, they also combine into a powerful evolutionary process which drives its emergence.

[Insert Figure 1 around here]

The evolutionary process begins with the crafting of the initial digital representation (Principle 1), as without the initial information, context and impact models that comprise an initial digital representation of the organization, it is not possible to proceed. However, once the initial, incomplete, information, context and impact models are developed, then the existing data needs to be set free (Principle 2). This will enable the flow of new data across the organization and will often suggest new interactions and insights. By setting the data free, increased organizational insight into operations enables an increment of the information and context models, and sometimes rapid improvement to the impact model (Principle 5). These improvements to the digital representation themselves result in both the identification of new assets and processes within the organization that can be digitized (Principle 3), as well as the ability to seize new digital opportunities enabled by the increased insight that the digital twin provides (Principle 4). In turn, the data created by both increased digitization of existing assets and processes and the execution of new digital opportunities will need to be set free (Principle 2), in turn leading to the necessity to increment the information, context and impact models (Principle 5). We believe that this dynamic evolutionary process is self-sustaining, and once established will enable organizations to progressively build the digital twin.

IMPLICATIONS OF DIGITAL TWINS

Before considering how a digital twin creates value for an organization, it is necessary to understand some of the implications of building an organizational digital twin. As well as benefits, a digital twin also offers risks, as evidence is mounting that many organizations are unable to grasp the opportunities offered by data (Bean & Davenport, 2019; Tabesh, Mousavidin, & Hasani, 2019). For instance, 85% of data-related projects fail as organizations are unaware of the major challenges that accompany such projects (Asay, 2017). Recent research has shown that the percentage of organizations identifying themselves as being data-driven has declined in each of the past 3 years – from 37.1% in 2017 to 32.4% in 2018 to 31.0% in 2019 (Bean & Davenport, 2019), suggesting ongoing issues for organizations dealing with digitization. Other organizations try to avoid the challenge by applying data and algorithms to legacy systems, instead acting more like a Band-Aid than a renewal and extension of capabilities (Ransbotham, 2018).

We suggest that there are the two main implications – new organizational challenges and new skill requirements – which need to be addressed in order to successfully build a digital twin.

New organizational challenges

The decision to build an organizational digital twin is not without its challenges. There are the general digital challenges, such as security and privacy, that need to be considered for all digital investments. The vast amount of confidential and inalienable data that organizations hold needs to be held securely (Alharthi, Krotov, & Bowman, 2017; Lee, 2017). Concerns as to privacy and the related risk of ethical lapses are also increasing, particularly with rise of new regulations that govern personal data, such as the new GDPR of the European Union. Specific challenges for organizational digital twins relate to ensuring organizational flexibility so that the digital twin can be extended and developed, addressing emerging social, ethical and legal issues, and maintaining technical resilience and stability. We now discuss each in turn.

A first challenge relates to the ability to effectively and efficiently extend the digital twin. A natural byproduct of the development of any digital artefact is the accrual of “technical debt” (Baldwin & MacCormack, 2011; Woodard, Ramasubbu, Tschang, & Sambamurthy, 2013). Technical debt occurs when the updating of an artefact requires the technical redesign, component renewal, upgrading, or extension, or wholesale extension or replacement of other digital artefacts. The more technical debt that is accrued, the greater the cost of subsequent incremental improvements (Brown et al., 2010). Technical debt becomes an issue if it hinders the extension of the digitization frontier (Principle 3), the leveraging of digital opportunities (Principle 4) or the extension of the digital twin itself (Principle 5). While technical debt is inevitable (Sullivan, Chalasani, Jha, & Sazawal, 1999), it vital to ensure that the dynamic evolutionary process of building a digital twin is not hindered. In essence, the generative possibilities of the digital twin should not be outweighed by the technical debt that is accrued in its construction. Paradoxically, while the goal of a digital twin is to enable organizations to model their operations and improve their agility, the actual process of implementation may actually reduce agility if technical debt becomes too high.

A second challenge relates to potential social, ethical and policy issues arising from the use of digitally driven predictions. Within the popular press there is an emerging perception that digital service offerings such as insurance, credit, or housing are influenced by data analyses that reinforce bias and unfair treatment of underrepresented populations (see for instance Ghaffray, 2019). And beyond popular concerns, there is increasing academic evidence that data analyses are indeed biased, with biases springing from data collection processes reflecting deep and hidden imbalances in institutional infrastructures and social power relations, as well as from the algorithms themselves (Zou & Schiebinger, 2018). Organizations need to be aware of emerging attempts to craft actionable assessment frameworks based upon international human rights (see for instance Mantelero, 2018; Nersessian, 2018), and follow industry leading advice (Baer & Kamalnath, 2017) to enact ethical practices, policies and procedures. Essentially, as the digital twin becomes integral to the organization, its use needs to be aligned with the emerging social, ethical and policy context, or else it could present new legal, regulatory and reputational risks.

A further challenge relates to the technical risks inherent in the successful implementation of a digital twin. As a digital twin is incrementally implemented within an organization, increasingly the organization will depend on the insights that the digital twin provides – indeed this is what the dynamic evolutionary process identified entails. While all organizations need to ensure the resilience and stability of their digital infrastructures, with digital twins this challenge becomes more significant as the digital twin becomes the operational core of the organization (Alharthi et al., 2017). Furthermore, as the digital twin will be constantly evolving due to Principles 3, 4, and 5, there is an increased risk of issues occurring due to continuous technical change. The cost of interruptions, especially when workers are prevented from completing tasks due to out-of-service infrastructure, can be huge. One Gartner study suggests that a large organization may lose as much as $540,000 per hour from a preventable technical failure, and which adds up to $647 billion a year (Lerner, 2014). While all organizations with digital systems face this challenge, when an organization has implemented a digital twin, the potential cost of these technical faults could be much higher due to the fundamental role of the digital twin in the operation of the organization. As such it is essential to ensure that the underlying infrastructure and staff which enable the digital twin are adequately maintained and supported.

New skill requirements

As the five principles above emphasize, to implement a digital twin is not an exercise of buying a technology and plugging it into an organization. The process of building a digital twin is multifaceted and doesn't just involve technology, instead representing an ongoing process of changing the way an organization operates. It requires both foundational investments in skills, projects, infrastructure, and, often, in cleaning up existing digital systems, and the mixing of people, machines, and organization processes (Davenport & Westerman, 2018). Of particular note is the vital role of people, with recent research finding that 93% of organizations identify people issues as the main obstacle to the implementation of large scale digital transformations (Bean & Davenport, 2019). Thus, in seeking to build a digital twin, not only is there a need to consider the technical infrastructure to improve the digital connections, data and interactions for the digital twin, but also the necessity to have the requisite skills in the organization to be able to execute the strategy (Davenport & Harris, 2007).

Derived from both received wisdom (Gartner, 2015; NIST; Saltz & Grady, 2017) and the practical experience of the first author in large-scale data-based digital transformation projects, we suggest that there are five roles which provide the essential skills for the successful implementation of a digital twin within an organization.[1] The *Data Engineer* understands the

[1] The roles are derived from Gartner (2015) and Saltz & Grady (2017).

creative ways that data can captured, such as through new sensors and processes, and who knows where data exists inside and outside the organization. The *Data Scientist* (sometimes called the “digital entrepreneur”) has knowledge from multiple domains to identify new, cross-cutting digital innovations, with a blend of organization and technical skills. They can describe the digital value in languages that different domains within the organization understand and are often deeply involved in identifying new digital opportunities. The *Source System Engineer* (sometimes called the “digital modeler”) understands the meaning of the digital artefacts, such as data, analytical models and software, that the firm has access to and the limitations and knows how to integrate them into the information, context and impact models so as to maintain and expand the digital twin. They can describe and design the complex models and algorithms that underpin the digital twin. The *Software Engineer* (sometimes also called the “information architect”) can create robust and dependable data and analytical models and systems that deliver the digital twin. They understand all the data roles and can develop ideas into reality. Finally, the *Digital Police* (a special type of “business expert”) understands the legal, ethical, and privacy aspects of both the underlying data and analytical models and the digital twin and can guide the team in what is viable, feasible, and sensible, as well as provides guidance on mitigating the confidentiality risks. It is necessary to ensure that each of the five roles are present and well developed in the organization to ensure both the incremental development of the digital twin, and its successful operation.

CREATING VALUE FROM DIGITAL TWINS

So far, we have identified five principles which guide in the quest to build digital twins and outlined the strategic implications of seeking to build one. We next discuss how an organization creates value from a digital twin, suggesting that value creation from digital twins occurs through three means: productivity, performance, and predictability.

Value from *productivity* results when the digital twin enables cost reductions in the execution of organizational activities. A simple example is how the digitization of internal processes, such as claims processing in insurance companies, can enable the processing of more claims per claims advisor by reducing or removing routine operations. Productivity improvements from more complex data interactions are also possible. For example, at KONE, the digitization of the elevator product line has dramatically improved the productivity of its service engineers. By allowing insights into the condition of an elevator and suggesting resolutions to potential problems, service engineers can be at the locations that most require the services, and customers can manage the equipment over its life cycle. This means less downtime, fewer faults and detailed information for maintenance crews on the performance and usage of equipment.

Value from *performance* improvements result when the insights from the digital twin permits organizations to improve and extend their offerings. Often these performance improvements are a direct consequence of the search for new digital opportunities (Principle 4). A simple example is how data collection leads to new goods and services. For instance, Netflix gathers extensive data on what movies its customers like and is able to produce new shows that meet these demands. Performance improvements can also come from more sophisticated insights derived from the digital twin. An example is how insights from a digital twin can improve the performance of the warranties that organizations often provide. The combination of IoT sensors and sophisticated analytics, integrated and considered holistically, can improve performance across four key dimensions: reducing operational costs, improving brand integrity and customer experience, identifying of new revenue streams, and supplying competitive differentiation (Gonzalez-Wertz, 2018).

Value from *predictability* arises when the digital twin leads to the removal of uncertainty from organizational processes, and hence improves organizational interactions with consumers and suppliers. A simple example at the consumer level is that of Uber, which allows customers to see the position of the taxi as it arrives to remove the uncertainty of the arrival time. The value of predictability can also be derived from much more complex models, such as the example of the Land Traffic Authority (LTA) in Singapore. Here, using historical traffic data and real-time traffic input from the LTA i-Transport system, managers were able to predict traffic flows over pre-set durations (10, 15, 30, 45 and 60 minutes), with results well above 85 percent accuracy. These predictions have not only enabled managers to better anticipate and manage the flow of traffic to prevent the build-up of congestion, but also share this data with commuters, not only providing real time data as to punctuality of public transport, but also the likelihood of getting a seat. Such technology is also being used in other cities such as London and Stockholm.

Upon becoming available to the organization, the digital twin will also facilitate decision making and control. It will allow scenarios to be assessed in detail, and repeatedly, to test potential tactics and strategies. The digital twin will provide organizations with the capacity to identify the optimal decisions and actions.

CONCLUSION

In this paper, we have suggested that the increasing digitization of the economy means that a new digital resource—an organizational digital twin—is now feasible. Digital twins describe the relationships between digitally instrumented assets and activities, thus modelling the interactions among the various sources of data within an organization. To assist in building the digital twin we suggested five principles that drive the creation of a digital twin and show how they combine into a dynamic evolutionary process that maintains and grows the digital twin. By mastering these principles and leveraging the opportunities suggested by the implications, it is possible to create value and exploit the insights from an organizational digital twin. Future leaders of the digital economy will accelerate the creation and exploitation of these digital twins.

REFERENCES

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons, 60*(3), 285-292. doi:https://doi.org/10.1016/j.bushor.2017.01.002

Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. *TechRepublic*. Retrieved from https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/

Baer, T., & Kamalnath, V. (2017). Controlling machine-learning algorithms and their biases. Retrieved from https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases

Baldwin, C. Y., & MacCormack, A. (2011). *Finding technical debt in platform and network architectures*. Paper presented at the The Mathworks, Inc.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management, 17*(1), 99-120. doi:https://doi.org/10.1177/014920639101700108

Bean, R., & Davenport, T. H. (2019). Companies Are Failing in Their Efforts to Become Data-Driven. *Harvard Business Review Digital Articles*, 2-4.

Brown, N., Cai, Y., Guo, Y., Kazman, R., Kim, M., Kruchten, P., . . . Zazworka, N. (2010). *Managing technical debt in software-reliant systems*. Paper presented at the Proceedings of the FSE/SDP workshop on Future of software engineering research, Santa Fe, New Mexico, USA.

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different. *MIT Sloan Management Review, 54*(1), 43-46.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Boston, MA: Harvard Business School Publishing.

Davenport, T. H., & Westerman, G. (2018). Why So Many High-Profile Digital Transformations Fail. *Harvard Business Review Digital Articles*, 2-5.

Gartner. (2015). Staffing data science teams. Retrieved from https://www.gartner.com/en/documents/3086717/staffing-data-science-teams

Gartner. (2017a). Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016 [Press release]. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016

Gartner. (2017b). Prepare for the impact of digital twins. Retrieved from https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins/

Gartner. (2019). Gartner Survey Reveals Digital Twins Are Entering Mainstream Use [Press release]. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2019-02-20-gartner-survey-reveals-digital-twins-are-entering-mai

Ghaffray, S. (2019). New York City wants to make sure the AI and algorithms it uses aren't biased. That's harder than it sounds. *Recode*. Retrieved from https://www.vox.com/2019/4/11/18300541/new-york-city-algorithms-ai-automated-decision-making-sytems-accountable-predictive-policing

Gonzalez-Wertz, C. (2018). Improving Warranty: IoT+AI = Found Money. *Medium*. Retrieved from https://medium.com/@hermione1/improving-warranty-iot-ai-found-money-5905ac4d16f3

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17*, 109-122-109-122.

Haag, S., & Anderl, R. (2018). Digital twin – Proof of concept. *Manufacturing Letters, 15*, 64-66. doi:https://doi.org/10.1016/j.mfglet.2018.02.006

IBM. (2019). Industry Data Models. Retrieved from https://www.ibm.com/analytics/industry-models

Intel. (2018). A Guide to the Internet of Things [Press release]. Retrieved from https://www.intel.com/content/www/us/en/internet-of-things/infographics/guide-to-iot.html

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577-586. doi:https://doi.org/10.1016/j.bushor.2018.03.007

Kallinikos, J., Aaltonen, A., & Marton, A. (2010). A theory of digital objects. *First Monday, 15*(6-7).

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2019). *Data Market Strategy*. Working Paper.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293-303. doi:https://doi.org/10.1016/j.bushor.2017.01.004

Lerner, A. (2014). The Cost of Downtime. Retrieved from https://blogs.gartner.com/andrew-lerner/2014/07/16/the-cost-of-downtime/

Linde, F., & Stock, W. G. (2011). *Information Markets. A Strategic Guideline for the I-commerce*. New York, NY: De Gruyter Saur.

Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review, 34*(4), 754-772. doi:https://doi.org/10.1016/j.clsr.2018.05.017

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*

Marr, B. (2018). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London, UK: John Murray Publishers.

McCrory, D. (2010). Data gravity - in the clouds. Retrieved from https://datagravitas.com/2010/12/07/data-gravity-in-the-clouds/

National Research Council. (1999). *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington, DC: National Academy Press.

Nersessian, D. (2018). The law and ethics of big data analytics: A new role for international human rights in the search for global standards. *Business Horizons, 61*(6), 845-854. doi:https://doi.org/10.1016/j.bushor.2018.07.006

NIST. NIST 1500 Series. Retrieved from https://bigdatawg.nist.gov/V2_output_docs.php

Nordhaus, W. D. (2002). Productivity Growth and the New Economy. *Brookings Papers on Economic Activity, 2002*(2), 211-265. doi:https://doi.org/10.1353/eca.2003.0006

Parmar, R., Mackenzie, I., Cohn, D., & Gann, D. M. (2014). The new patterns of innovation. *Harvard Business Review, 92*(1/2), 86-95.

Pfeffer, J., & Salancik, G. R. (1978). *The External Control Of Organizations*. New York, NY: Harper & Row.

Ransbotham, S. (2018). Don't let artificial intelligence supercharge bad processes Retrieved from https://sloanreview.mit.edu/article/dont-let-artificial-intelligence-supercharge-bad-processes

Saltz, J. S., & Grady, N. W. (2017, 11-14 Dec. 2017). *The ambiguity of data science team roles and the need for a data science workforce framework.* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data).

Solove, D. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York, NY: New York University Press.

Sullivan, K. J., Chalasani, P., Jha, S., & Sazawal, V. (1999). Software design as an investment activity: A real options perspective. In L. Trigeorgis (Ed.), *Real Options and Business Strategy: Applications to Decision Making*, (pp. 215-262). London, UK: Risk Books.

Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*. doi:https://doi.org/10.1016/j.bushor.2019.02.001

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, *94*(9), 3563-3576. doi:https://doi.org/10.1007/s00170-017-0233-1

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509-533. doi:https://doi.org/10.1002/(Sici)1097-0266(199708)18:7<509::Aid-Smj882>3.0.Co;2-Z

Thomas, L. D. W., & Tee, R. (2019). *Generativity: A systematic review and integrative model*. Working Paper.

Thompson, J. D. (1967). *Organizations In Action*. New York, NY: McGraw-Hill.

Tilson, D., Lyytinen, K., & Sorensen, C. (2010). Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research*, *21*(4), 748-759. doi:https://doi.org/10.1287/isre.1100.0318

Woodard, C. J., Ramasubbu, N., Tschang, F. T., & Sambamurthy, V. (2013). Design capital and design moves: The logic of digital business strategy. *MIS Quarterly*, *37*(2), 537-564. doi:https://doi.org/10.25300/misq/2013/37.2.10

Yoo, Y. J., Henfridsson, O., & Lyytinen, K. (2010). The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724-735. doi:https://doi.org/10.1287/isre.1100.0322

Zittrain, J. L. (2006). The generative Internet. *Harvard Law Review, 119*(7), 1974-2040.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist - it's time to make it fair. *Nature*, *559*(7714), 324-326. doi:https://doi.org/10.1038/d41586-018-05707-8

FIGURE 1 – An evolutionary process model of building a digital twin