Maximilian Eibl, Martin Gaedke. (Hrsg.): INFORMATIK 2017,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017

A Reference Architecture for Digitalization in the Pharmaceutical Industry

Alina M. Chircu[1], Eldar Sultanow[2] and Levent D. Sözer[3]

Abstract: Digital transformation (also called digitalization or, to a lesser extent, digitization) can enhance the customer experience, streamline operations, revolutionize existing business models, and disrupt entire industries. In this paper, we explore the opportunities for digital transformation in the pharmaceutical industry and develop a pharma-specific reference architecture for digitalization. We build on a model initially developed internally at Capgemini and show how it can be used to describe the use of several emerging digital transformation technologies: Internet of Things (IoT), Cognitive Computing (CC) and Augmented Reality (AR). Our analysis presents the architectural implications of each of these technologies for the pharma industry from the perspectives of four standard TOGAF domains (covering the technology, application, data and business architecture) and of four digitalization steps (Sense, Tag and Connect, Ingest, Analyze and Prepare, and Utilize). We also present an integrated view of the business capabilities all three technologies offer to organizations.

Keywords: Augmented Reality, Cognitive Computing, Digital Transformation, Digitalization, Internet of Things (IoT), Pharmaceutical Industry, Reference Architecture, TOGAF

1 Introduction

Digital transformation (also called digitalization or, to a lesser extent, digitization) is predicted to enhance the customer experience, streamline operations, and revolutionize existing business models through the use of new digital technologies such as social media, mobile, analytics and embedded devices [Fi13, IL14]. These technologies enable new ways for companies to create and capture value due to their unique capabilities: providing ubiquitous connectivity for people, things, and industries and replicating information without errors as many times as needed at zero marginal cost [IL14]. First movers and fast-followers in digitalization are predicted to experience significant revenue and profit growth [BLM17]. However, even as the vast majority of companies surveyed over the last few years recognize the strategic importance of digital transformation, few are adequately prepared for the disruption digital technologies will bring to their industry [Fi13; Ka16]. Barriers include, in order of importance: no sense of urgency, lack of funding, limitations of existing information technology (IT) systems, unclear roles, responsibilities, company-wide vision, and business case, silo-ed implementation, culture and leadership skills deficiencies, and regulatory concerns [Fi13].

---

[1] Bentley University, 175 Forest Street, Waltham, MA 02452-4705, USA, achircu@bentley.edu
[2] Capgemini, Bahnhofstraße 11C, 90402 Nuremberg, Germany, eldar.sultanow@capgemini.com
[3] Capgemini, Potsdamer Platz 5, 10785 Berlin, Germany, levent-deniz.soezer@capgemini.com

doi:10.18420/in2017_205

The pharmaceutical (or pharma) industry – which, according to several statistics, has already exceeded USD 1 trillion in revenues globally - is a good example of this paradox: the potential for disruption is well understood, but companies are not sure how to develop the required capabilities and capitalized on this trend [CHL15]. For example, in Germany, the pharma and healthcare industry cluster is ranked towards the bottom of a digital index scale that measures digital transformation in strategy, product, sales and customer experience, and processes [Ri16]. Current industry trends – a focus on outcome-based care, engaged patients, new competitors, availability of information, and process improvements – are creating opportunities for digital transformation through personalized, 24/7 care, omnichannel interactions, data-driven insights, and real-time responsiveness [CHL15]. Requirements for assigning unique identifiers (serializing), tagging, and tracking and tracing of medication, and, in some cases, medical equipment as well, are being incorporated in legislation around the world, with the European Union leading the way with a 2017 implementation date. This paves the way for further automation of processes and integration of data across various enterprise systems [BCS16; SC15]. Therefore, understanding these emerging digitalization trends is important not only for companies in the pharma industry, but for the software provider companies – such as ERP vendors – as well. This can support novel designs for the business models of the software vendors [Br14], which can then in turn create additional digitalization possibilities. However, the analysis of the potential opportunities often focuses on partial solutions – such as operational excellence in the supply chain [EB16] or digital customer engagement [An15], and no integrated models that can guide digital transformation for the entire industry currently exist. In this paper, we address this gap by developing a pharma-specific model – a reference architecture - for digitalization.

Organizations use architectures to represent, organize and manage complex technologies and the opportunities they offer [WAX15]. We distinguish between several levels of architectures – company-specific enterprise architectures (EA) and industry-specific reference architectures (RA). An EA is defined as the “structures, templates and methods for the overall design and description of an enterprise” [CD17] as a socio-technical system, which includes the enterprise “business and operating model, organizational structure, business processes, data, applications and technology” [ALS12]. An RA provides a templates for a technology solution, capturing both business-facing and technology-facing perspectives [Su16; WAX15], and is developed for a specific industry or function [CD17]. RAs present a high-level, organizing view for an industry, including its processes, stakeholders, organizational, informational and technology structure [CD17]. RAs are important for maintaining consistency, creating a common vocabulary, and generating efficiencies in technology development both internally and externally [Su16]. Industry-specific RAs can be viewed as industry-level templates that can be used to derive EAs for specific companies in that industry while also taking into account the industry-level interconnections [CD17]. A recent example is a telecommunications industry RA which serves as a generalized structure and EA template for telecom operators [CD17]. Company-specific EAs and industry-specific RAs are likely to become more important in the future, as companies and industries develop interconnected business platforms for customers, vendors and business partners to orchestrate the delivery of internal and external services in effective and efficient ways [SF12].

In this paper, we build on a generic RA model initially developed by several of the paper authors for Capgemini and presented in several industry forums [BCS16; SB17], and show how it can be used to describe the use of several emerging digital transformation technologies in the pharmaceutical industry.

2 Emerging Digital Transformation Technologies: Internet of Things (IoT), Cognitive Computing (CC) and Augmented Reality (AR)

Several important but not yet fully explored digital transformation technologies have emerged in recent years: Internet of Things (IoT), Cognitive Computing (CC) and Augmented Reality (AR).

IoT is significantly enabling the digitization of our society and economy. With IoT, objects and people are assigned unique identifiers, sensors, and actuators, are interconnected through wireless ad-hoc networks such Bluetooth, Zigbee or Z-Wafe, and report about their status and the surrounding environment [Ti16]. According to a recent European Commission (EU) study [Ag14], the market value of the IoT in the EU is expected to exceed one trillion euros in 2020. Globally, over 26 billion IoT devices will be deployed by 2020 [LL15], transforming industries and creating new opportunities for individuals and companies [IL14].

CC includes a variety of emerging IT systems that use artificial intelligence, natural language processing, and machine learning technologies that provide data analysis and decision-making through automated, dynamic learning based on accumulated experience, rather than manual re-programming [HKB15]. CC derives intelligence and knowledge from huge volumes of structured and unstructured data and enables real-time decisions. It relies on new approaches like brain-inspired neuromorphic computing and memcomputing, which have non-von Neumann architectures in which memory and processing logic coexist rather than being separated [Co16]. CC is emerging as an efficient, effective and flexible solution for dealing with ubiquitous data generated by digitalization, although its use, especially in areas that require full documentation and explanation of decisions, still requires further investigation [Kn17].

AR consists of technologies that increase the user’s perception and interaction with the real world by placing virtual objects in a real world scene. An AR systems possesses three main characteristics – it “combines real and virtual objects in a real environment”, “runs interactively, and in real time” and “registers (aligns) real and virtual objects with each other” [Az01]. The virtual objects’ features and behavior is derived from data that has been gathered due to digitalization as well as from information and knowledge that has been discovered from this data by CC.

3 Specifying the Structure of the Digitalization RA

The generic digitalization reference architecture has three major components: architectural domains, digitalization stages, and industry perspectives. Figure 1 presents a graphical representation of this structure.

First, the digitalization RA domains are defined based on one of the most popular open frameworks for reference architecture generation – the TOGAF framework maintained by The Open Group consortium (http://www.opengroup.org/) [BB12; CD17]. This framework has been successfully used in the past to develop industry-specific RAs – such as in the case of the telecommunications industry [CD17]. TOGAF captures architectural requirements in four essential domains: business, data, application, and technology. The business architecture describes a company's strategies, business capabilities, and processes used to meet organizational goals. The data architecture encompasses how the organizational data is stored and accessed, that is, it describes all organizational data, entities and business objects, as well as their relationships. The application architecture reflects the application landscape including the design of each application and the interactions and interfaces between the deployed software solutions. The technology architecture details the hardware and software capabilities required to support the applications, data, and business services, and their interactions [BB12].

Second, the digitalization RA recognizes that digitalization is horizontally divisible into four sequentially arranged digitalization process stages: *Tag, Sense & Connect* (1), *Ingest* (2), *Analyze & Prepare* (3) and *Utilize* (4). While many conceptualizations of process stages can be developed, the RA authors are proposing these four stages based on their collective experience. Further testing of the applicability of these stages is planned for future research. The first stage, *Tag, Sense & Connect*, involves tagging things (non-living objects or living beings) with a unique ID (coded in a Datamatrix 2D label or an RFID tag, for example) and sensors that enable identifying, locating, tracking and/or measuring their properties, movements or gestures and possibly the properties of their environment. It also includes connecting the tagged things so that they may mutually localize and recognize each other and interact with each other via wireless ad-hoc networks. Within the second stage, *Ingest*, all data captured during *Tag, Sense and Connect* is ingested (transferred and organized) into a central repository that is capable of storing large quantities of data. The third stage, *Analyse & Prepare*, includes the analytical processing steps for this data. The last stage, *Utilize*, involves using the previously developed insights in various ways (for example, to manage existing objects or processes, develop new products and services, etc.).

Last, but not least, the digitalization RA recognizes that technology components may be used not only in the target industry, but also across different interdependent industries or sectors. Applications developed and running in different industries may interact or be integrated with each other and data may be exchanged through these inter-industry connections. Business-level synergies across organizations in different industries may exist as well. For this reason, the digitalization RA needs a third dimension describing different industry perspectives. For example, the pharmaceutical industry is interconnected with the logistics and transportation industry (which supports the movement of pharma products through the supply chain) and the public health industry (which is concerned with improving the healthcare of communities, providing community education, establishing policies and engaging in disease prevention research).

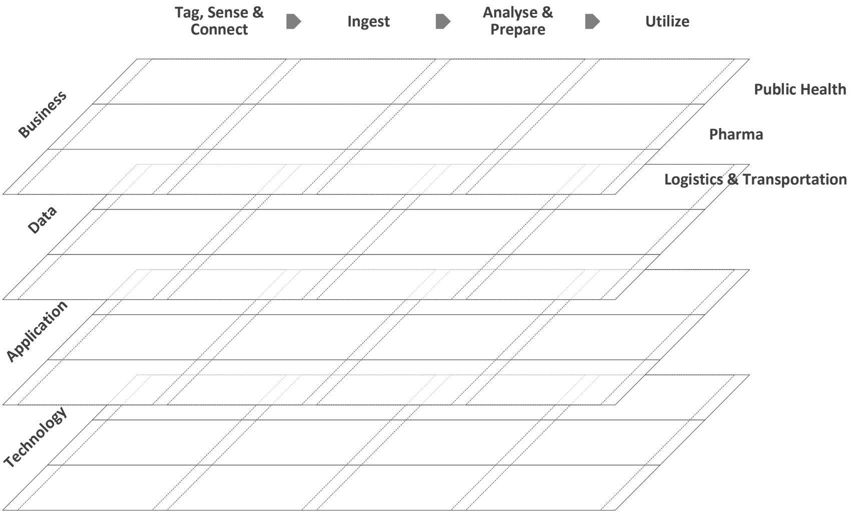

Fig. 1: Fundamental structure of the digitalization RA

4 Constructing a Digitalization RA for Pharma Industry

In the previous sections, we discussed several emerging digitalization technologies and a generic digitalization RA. To develop the specific pharma industry digitalization RA, we investigate the different architectural implications of emerging digital technologies such as IoT, CC and AR in the context of the pharmaceutical industry and related industries (such as logistics and public health).

4.1 IoT Architectural Implications for the Pharma Industry

Pharma IoT includes the digitalization of the pharmaceutical products and related processes with smart connected devices and IT services (web, mobile, apps, etc.) in drug development, clinical trials, supply chains, and patient care.

The **Technology Architecture** encompasses all IoT technologies that support the digitalization steps. The *Tag, Sense & Connect* step covers sensors (which collect data from the environment, such as location, temperature, movement, etc.), actuators and microcontrollers (which control the environment), radio-frequency (RFID) tags or other transponders which can help store and transmit data, and readers and network communication technologies that enable connections of IoT devices among themselves and with other IT systems. For example, tags enable identification of medication at the individual sale unit level, temperature sensors enable monitoring of allowable ranges during storage and transportation, and movement sensors can detect excessive shaking that may alter the properties of certain medications. This step also includes technologies for serialization (assigning unique identifiers to products), which is quickly becoming a mandate for pharmaceuticals in many countries around the world. In addition, this step may also include security technologies that ensure that the unique identifiers and other device data are not vulnerable to unauthorized reading or subject to tampering, thus preventing counterfeiting. The *Ingest* step covers technologies for managing the collected IoT data – including refreshing, aggregation, compression, archiving, or deletion. Even with extremely large data stores, keeping all IoT data in raw format indefinitely will quickly exhaust the available storage capacity. However, a sufficient level of detail needs to be available in order to support regulatory compliance such as tracking and tracing of medication, to identify and correct quality problems that may occur during manufacturing or transportation of medication, and to facilitate research for improving existing drug formulations and new drug discovery, Finally, the *Analyse & Prepare* and *Utilize* steps cover the processing and utilization of the data – either locally on the IoT device or remotely using other devices and systems.

The **Application Architecture** includes application platforms and software development kits (SDK) that facilitate application development. Examples include Amazon's AWS Greengrass and AWS IoT, which support primarily the *Ingest* and *Analyze and Prepare* domains with cloud-based services. Specific pharma applications can focus on clinical trials (monitoring drugs and subjects to ensure compliance with clinical trial rules), regular manufacturing and logistics (monitoring production, storage and transportation), and provider and patient interaction at the point of care (verifying that a drug about to be administered to a patient is genuine, confirming dosage details and administration information, etc.).

The **Data Architecture** domain covers the data model for capturing, storing and analyzing production data (including unique serial numbers and other important manufacturing data such as lot numbers, expiration dates, security codes, country-specific product codes, etc..), storage and transportation data (including changes in location over time as the product moves in the supply chain within and across manufacturer, distributors, and logistics and transportation provider facilities, and temperature and movement sensor data), patient data (such as formulation and dosage for personalized medicine administration or patient monitoring data) and health cost reimbursement data (such as reimbursement codes).

Building blocks with in the **Business Architecture** domain encompass *Treatment 2.0*, *Personalized Therapy and Health Monitoring*, and *Industrial Internet* capabilities. IoT can be used in drug development and clinical trials to create new treatment possibilities based on advanced sensors and personalized care services – a development also known as

Treatment 2.0 [KL16]. In *Personalized Therapy and Health Monitoring*, smart pills can be designed to incorporate sensors that are safe to ingest and can capture data from inside the body, or use sensors and other mechanisms that provide personalized variable doses and prevent potential abuse or unintentional overdoses. Existing medical devices such as inhalers and insulin pens can also be enhanced with sensor and connectivity technologies to collect data that can inform future care and suggest personalized therapies. Connected wearables can help monitor Parkinson's disease, multiple sclerosis, and other types of patients continuously in real-time and adjust their medication in order to improve patient outcomes and the quality of life. Last, but not least, pharma *Industrial Internet* covers the digital transformation of the product supply infrastructures from manufacturing the medicines to dispensing the medicines to patients. It encompasses manufacturing intelligence (lean and automated manufacturing), software controlled packaging execution (serialization and aggregation), integrated supply chain (traceability and collaboration), and enterprise back-end IT services utilizing supply chain data for analytics, life-cycle management and regulatory compliance reporting [Ke16].

4.2 CC Architectural Implications for the Pharma Industry

From a pharma industry perspective, CC offers a revolutionary approach to enable better use of existing medications and discovery of new drugs. CC provides automated, self-learning systems that can extract new knowledge about specific diseases, symptoms or treatment approaches from unstructured medical data, and assist doctors in making better and faster decisions.

The **Technology Architecture** encompasses all CC technologies that support the digitalization steps. To use the potential of CC those technologies must be combined with data generated by connected devices, creating "cognitive IoT" capabilities [Ma16]. The *Tag, Sense & Connect* domain covers the interfaces with IoT devices. The *Ingest* domain covers organizing unstructured data from natural language (written text or speech) input, audio signals, visual input (gestures, facial expressions, etc.) and haptic (touch-based) inputs. Within the domains of *Analyse & Prepare* and *Utilize* all this input is being processed using machine learning tools. The necessary technology architecture building blocks include signal, image, speech and natural language processors along with components for object and pattern recognition, anomaly detection and motion analysis.

The **Application Architecture** describes the process of application deployment and specific healthcare CC platforms and applications. Those applications use machine learning algorithms to process the ingested data and generate helpful outputs. According to Chappell there is a main process model describing the phases of developing CC applications. Such an application's functions are based on data – unstructured data - delivering raw input processed by the algorithm. Within the *Ingest* and *Analyse & Prepare* domains, a preprocessing analyzer structures and prepares that data so that the machine learning algorithm can be applied over the prepared data (algorithm applier) and used for deploying the application [Ch15]. One of the best known CC platforms is IBM Watson, which has “capabilities to interact in natural language, process vast amounts of Big Data for understanding patterns and insights, and learning from each interaction” [Ar15]. Similar platforms include Google’s DeepMind Health and Apixio’s Iris. An extension of the Watson platform is the Watson Health Cloud, which provides data sharing and combining for “a dynamic, aggregated view of clinical, research, and social health data that clients choose to share” [Ar15]. Watson’s algorithms can analyze any type of data – such as scientific papers, tweets or industry studies. Thus, within the *Utilize* domain, CC platforms offer solutions for different health challenges like drug discovery (i.e. discovering both new drugs and new uses for existing drugs) and precise, evidence-based treatments that can be targeted to specific patient characteristics (such as disease type, individual health record, genetic data, etc.) [La16]. Some of the most publicized applications have occurred in cancer care, where IBM has made inroads with several applications that provide expert advice to oncologists.

The input for all applications mentioned above is data. CC systems mimic the human brain which is able to process unstructured data from different sources and structure it to usable content. The **Data Architecture** comprises different data sources and types used for CC systems. In the pharmaceutical context, this may involve personal data from electronic medical records created by individual healthcare providers or from comprehensive electronic health records aggregating data from multiple providers (HealthIT, 2017). This could potentially be combined with data from genetic tests and data from other sources, including data aggregators (who collect data on employment status, social security accounts, etc.), home monitoring systems, wearables and other personal IoT devices [Rh15]. The data architecture also involves external data published or openly shared (with appropriate human subjects protection and privacy safeguards) for medical research purposes, such as lab results, medical images, tissue samples, transcripts, clinical research, and others. An example is The Cancer GenomeAtlas project to collect, analyze and provide access to data from human tissue specimens for a variety of cancers. Companies such as Amazon now provide solutions for easy sharing of such data – as with the AWS OpenData toolkit.

The **Business Architecture** encompasses various business cases describing some very beneficial outcomes of CC systems. In health care industries there is a huge potential for those systems changing treatment methods, providing diagnosis, facilitating research and automating processes. For *Medical Treatment Assistance,* self-learning CC systems like IBM Watson can be taught about cancer symptoms and treatment methods. By reading and understanding the patient’s clinical history, the system can analyze the optimal cancer treatment options and help doctors make a better informed and faster decision [Sa13; La16]. Ultimately, CC could potentially support all medical decisions with additional insights derived through self-learning.

CC can also provide *Research Improvement & Acceleration* capabilities by finding more effective new drugs faster. For example, IBM Watson can help identify DNA mutations that contribute to cancer [Ar15] and then search the medical literature to identify the drug compounds that can be useful in each situation. Scientists have uncovered 28 proteins in the last 30 years. Watson was able to identify six in one month [Sa14]. Speeding up drug discovery and better detecting toxic side effects before clinical trial stage could not just make drugs cheaper but also help find future solutions for diseases that are still incurable at the moment. To this end, IBM is collaborating with medical schools and hospitals to analyze data from various sources and accelerate the research process for drug discovery.

To help researchers and doctors doing their jobs is already a revolutionary step into a better world with less diseases and easier ways to treat them. The National Healthcare System in the UK and Google's CC DeepMind system started a project to build a smart machine which "will be able to recognize eye diseases from just digital scan" [Ta16]. Using millions of eye scans the machine can be trained to quickly detect problems to prevent any kind of eye diseases. Such *Automatic Disease Detection* could also be applied for skin scanning, x-ray analyzes and all other health relevant diagnoses. From a public health perspective, this can also help with *Health Trend Prediction* and *Illness Prevention*.

4.3 AR Architectural Implications for the Pharma Industry

AR is an important driver of innovation in the healthcare industry. There are already several healthcare areas where this technology is already being used successfully, such as enhanced patient care and better knowledge transfer during training of healthcare providers. While the use of AR in the pharmaceutical industry is not yet widespread, we believe pharma can learn from the current AR applications in order to identify future opportunities for this technology.

The **Technology Architecture** encompasses technologies and devices which describe how AR is used and how it works in technical manner. Specific hardware includes AR glasses such as Sony's SmartEyeglass, Google Glass and Microsoft Hololens, which can access data from their own sensors, camera, microphone, accelerometer and GPS (measuring movement, position, light and sound, etc.) and external sources (such as connected devices) and superimpose it on the wearer's natural field of view. It also includes gesture interaction devices such as FingerIO and Google's Soli which use miniaturized sensors and sonar or radar technology to track motions. And the next generation of AR Devices, such as Google's patented digital contact lenses, attempt to combine AR features with sensors embedded in the lenses to provide fast, secure and unobtrusive testing for diseases such as diabetes. The technology architecture also includes object recognition technologies such as object detection, 3D scanning, and optical character recognition (OCR). Last, but not least, the technology architecture includes both marker-based and marker-less AR. The former involves special cameras and visual markers applied to objects beforehand. The latter involves recognizing a variety of objects without the use of any special markers.

The **Application Architecture** encompasses AR Software Development Kits (SDKs) such as Vuforia (recently acquired by PTC) or Metaio (which is now part of Apple). Several apps for better diagnostics and treatments for patients already exist. These applications can support surgeons through the whole surgery process with pre-operative imaging analyses, real-time examinations of patients and their internal anatomy using holograms, and real-time detection of critical conditions using haptic devices [Sa08]. Applications and platforms include SurgicalTheatre's Surgical Navigation Advanced Platform, which provides multiple AR views during real surgeries and enables simulated surgeries for training, or more simple ones like AccuVein, which displays a digital map of the veins on a patient's skin in real time. For educational purposes, apps like Anatomy 4D can be used to virtualize textbooks with AR illustrations and helps students get a better understanding of organs and the human body. And for emergency situations, apps like AED4EU developed at Radbound University in the Netherlands can help save lives by navigating the user to the nearest external defibrillators.

The **Data Architecture** comprises models of virtual and real items (such as holograms), including their relationships. Holograms are not only confined to the 3D models but also comprise meta-information and links to other virtual or real items. For example, advanced CT scan images can be superimposed over the corresponding areas of a patient's body and be manipulated using AR devices, or individual CT images can be used to build 3D models. In addition, temporal data describing disease progression or effect of medication over time can also be captured.

The **Business Architecture** encompasses capabilities of *Virtual Visualization* including *Remote Disease Diagnostics/Treatments*, *Medical Treatment Assistance, Virtual Simulation,* and *Geolocation-based AR. Remote Disease Diagnostics & Treatments* allow doctors to help patients remotely. AR glasses can help instruct patients how to behave or where to look to make a remote diagnostic possible. For example, an earlier Google Glass project in Australia provided expert advice to mothers struggling with breastfeeding. The *Medical Treatment Assistance* building block describes the use of AR in its most intensive and effective way for patients' treatment – by enhancing a doctor's expertise with additional AR capabilities. Software which creates 3D models of tumors in real-time that can be displayed and manipulated using AR glasses can empower surgeons enormously and help them make better decisions before and during surgery. Surgeons can also receive advice during surgery through the AR devices. Another use case for AR is *Virtual Education* – for both patients and health care professionals in training. AR improves trainees' understanding of human anatomy and provides a realistic but cheaper and safer alternative for active learning and practice. For example, an AR system developed by ProMis enables medical students to practice laparoscopic surgery on medical manikins projected with virtual overlays of anatomical information. Finally, a very interesting and important business case is *Geolocation-based AR.* For example, detecting the nearest external defibrillator and helping a user navigate to it by using the user's smartphone may help save lives in emergency situations when someone gets a heart attack. This use case can also be extended by showing the user through the smartphone where to put the defibrillator on the body and how to use it properly.

These more general healthcare business capabilities can inform the development of more specific pharmaceutical industry business architecture capabilities. For example, AR can be used to educate both patients and healthcare providers on disease progression and effectiveness of treatments, digitalize the sales and marketing functions by creating virtual models of drugs and devices, simulate the use of medical devices in order to better prepare doctors and obtain suggestions for product improvement, test new medical products, or visualize complex drugs and devices during new product development. Thus, virtual visualization capabilities can help pharma companies design new drugs, medical devices or implants.

4.4 Bringing All Building Blocks Together

Each one of the emerging digitalization technologies we studied in this paper provides distinct but inter-connected architectural viewpoints for organizations. The complexity inherent in each technology suggests building separate RAs for each technology, to allow for updates and changes as each technology evolves. However, synergies among these technologies should also be documented by constructing an integrated reference architecture. Fig. 2 depicts such a possible integrated view for IoT of our digitalization RA for the pharmaceutical industry. It shows the horizontal slice spanning the pharma industry, as well as the adjoining industries - logistics & transportation and public healthcare – and identifies select business architecture capabilities.

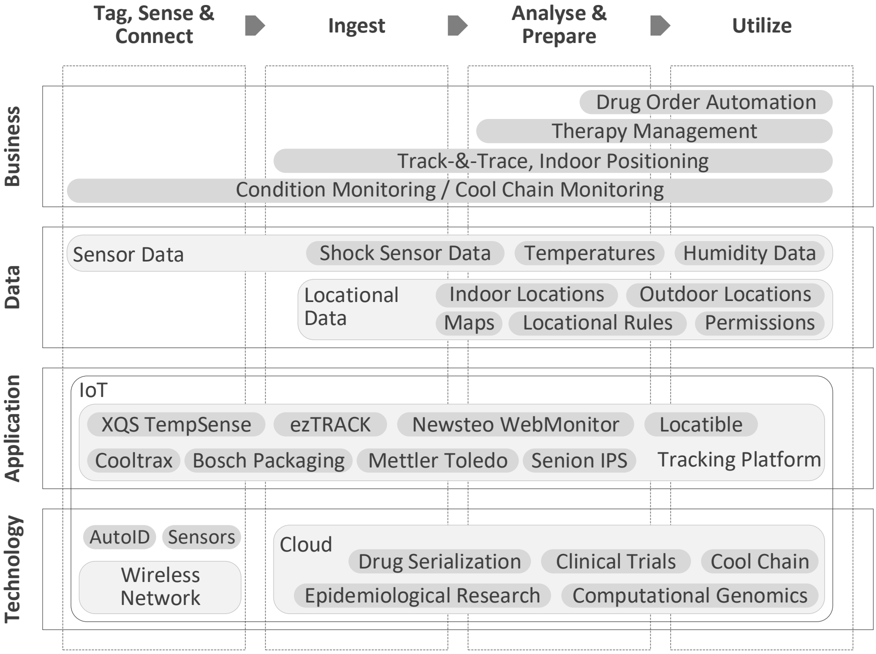

Fig. 2: IoT layer of the digitalization RA for pharma (vertical TOGAF domains slice)

5 Conclusion and Outlook

The RA presented in this paper can help pharmaceutical companies engage in digital transformation using several emerging digitalization technologies: the Internet of Things, cognitive computing, and augmented reality. Our analysis presents the architectural implications of each of these technologies for the pharma industry from the perspectives of four standard TOGAF domains (covering the technology, application, data and business architecture) and of four digitalization steps (*Sense, Tag and Connect*, *Ingest*, *Analyze and Prepare*, and *Utilize*). We also present an integrated view of the business capabilities all three technologies offer to organizations.

This paper reports on ongoing research efforts on establishing a pharma industry RA. Due to the stage of the research, the work has several limitations. Evaluation of the RA is still under way. Elements of the RA have been presented in several industry forums [BCS16; SB17] and have received very positive feedback from industry participants. We plan to further refine the RA based on this feedback and test it with more industry experts in the future. In addition, the operationalization of the digitalization steps represents the paper authors' point of view. Although this operationalization is based on the authors' extensive experience in this area, further research can explore alternative process steps. Last, but not least, the model we present in this paper could be informed by related work on RAs developed specifically for digitalization. While our literature review did not reveal any relevant high-level digitalization frameworks, we identified a few technology-specific RAs (for IoT, cloud computing, and other technologies). In future research, we plan to examine these emerging models and their implications for our generic digitalization RA. In addition, future work could continue to refine the RA model by exploring more synergies among digitalization technologies, and could show how it can be used to derive EAs for specific pharmaceutical companies.

References

[Ag14] Aguzzi, S.; Bradshaw, D.; Canning, M.; Cansfield, M.; Carter, P.; Cattaneo, G.; Gusmeroli, S.; Micheletti, G.; Rotondi, D.; Stevens, R.: Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination, 2014, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=9472, accessed 01/04/2017.

[ALS12] Ahlemann, F.; Legner, C.; Schafczuk, D.: Introduction. In (Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C., eds.): Strategic Enterprise Architecture Management. Challenges, Best Practices, and Future Developments. Springer-Verlag, Heidelberg, pp. 1-33, 2012.

[An15] Anscombe, J.; Paddison, C.; Plimley, J.; Thomas, M.; Lostaglio, K.: Time for Pharma to Dive into Digital. AT Kearney, 2015.

[Ar15] van Arnum, P.: IBM Watson, Google Forge New Models for Pharma Development, 2015, https://connect.dcat.org/blogs/patricia-van-arnum/2015/09/15/ibm-watson-google-forge-new-models-for-pharma-development#.WMKdLn--XNI, accessed

01/04/2017.

[Az01] Azuma, R.; Baillot, Y.; Behringer, R.; Feiner, S.; Julier, S.; MacIntyre, B.: Recent Advances in Augmented Reality. IEEE Computer Graphics and Applications, 21/6, pp. 34-47, 2001.

[BB12] Basten, D.; Brons, D.: EA Frameworks, Modelling and Tools. In (Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C., eds.): Strategic Enterprise Architecture Management. Challenges, Best Practices, and Future Developments. Springer-Verlag, Heidelberg, pp. 201-227, 2012.

[BCS16] Brockmann, C.; Chircu, A. M.; Sultanow, E.: Oncological Drug Track-and-Trace Systems: Implementation and Big Data Analytics. Bitkom Big Data Summit, Hanau, Germany, February 25, 2016.

[Br14] Brockmann, C.: An Approach to Design the Business Model of an ERP Vendor. Gito, Berlin, 2014.

[BLM17] Bughin, J.; LaBerge, L.; Mellbye, A.: The Case for Digital Reinvention. McKinsey Quarterly, February, 2017.

[Ch15] Chappell, D.: Introducing Azure Machine Learning, 2015, http://www.davidchappell.com/writing/white_papers/Introducing-Azure-ML-v1.0--Chappell.pdf, accessed 01/04/2017.

[CHL15] Champagne, D.; Hung, A.; Leclerc, O.: The Road to Digital Success in Pharma. McKinsey Insight, 2015, http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-road-to-digital-success-in-pharma, accessed 01/04/2017.

[Co16] CORDIS PROJECTED MEMRISTOR: A nanoscale device for cognitive computing. European Commission, 2016, http://cordis.europa.eu/project/rcn/200793_en.html, accessed 01/04/2017.

[CD17] Czarnecki, C.; Dietze, C.: Reference Architecture for the Telecommunications Industry: Transformation of Strategy, Organization, Processes, Data, and Applications. Springer, Heidelberg, 2017.

[EB16] Ehrhardt, M.; Behner, P.: Digitization in Pharma: Gaining an Edge in Operations. Strategy&. PwC, 2016, www.strategyand.pwc.com/reports/digitization-in-pharma, accessed 01/04/2017.

[Fi13] Fitzgerald, M.; Kruschwitz, N.; Bonnet, D.; Welch, M.: Embracing Digital Technology: A New Strategic Imperative. MIT Sloan Management Review Research Report, 2013.

[HKB15] Hurwitz, J.; Kaufman, M.; Bowles, A.: Cognitive Computing and Big Data Analytics. Wiley, Hoboken, NJ, 2015.

[IL14] Iansiti, M.; Lakhani, K. R.: Digital Ubiquity: How Connections, Sensors, and Data are Revolutionizing Business. Harvard Business Review, 92/11, pp. 91-99, 2014.

[Ka16] Kane, G. C.; Palmer, D.; Phillips, A. N.; Kiron, D; Buckley, N.: Aligning the Organization for Its Digital Future. MIT Sloan Management Review and Deloitte University Press, 2016, http://sloanreview.mit.edu/projects/aligning-for-digital-future/, accessed 01/04/2017.

[Ke16] Kemppainen, P.: Pharma Industrial Internet: A Reference Model Based on 5G Public Private Partnership Infrastructure, Industrial Internet Consortium Reference Architecture and Pharma Industry Standards. Journal of NBICT, 1/1, pp. 141–162, 2016.

[KL16] Kemppainen, P.; Liikkanen, S.: The Pharma Digitalization: the New Pharma Reality, Beyond the Pill, 2016, https://www.makehelsinki.com/blog/2016/04/26/the-pharma-digitalization, accessed 01/04/2017.

[Kn17] Knight, W.: The Dark Secret at the Heart of AI. MIT Technology Review, 2017.

[La16] Lau, V.: How Watson for Oncology is Advancing Cancer Care. Medical Marketing & Media, 2016, http://www.mmm-online.com/technology/how-watson-for-oncology-is-advancing-cancer-care/article/490336/, accessed 01/04/2017.

[LL15] Lee, I.; Lee, K.: The Internet of Things (IoT): Applications, Investments, and Challenges for Enterprises. Business Horizons, 58/4, pp. 431-440, 2015.

[Ma16] Matthews, S.: What is Cognitive IoT?. IBM Big Data and Analytics Hub, 2016, http://www.ibmbigdatahub.com/blog/what-cognitive-iot, accessed 01/04/2017.

[Rh15] Rhinehart, C.: The Impact of Cognitive Computing on Healthcare. Michigan Health Information Network Shared Services, 2015, https://mihin.org/wp-content/uploads/2015/06/The-Impact-of-Cognitive-Computing-on-Healthcare-Final-Version-for-Handout.pdf, accessed 01/04/2017.

[Ri16] Riemensperger, F.; Hagemeier, W.; Pfannes, P.; Wahrendorff, M.; Feldmann, M.: Demystifying Digitization: How Germany's Top500 Overcome Digital Hurdles. Accenture, 2016. https://www.accenture.com/t00010101T000000__w__/de-de/_acnmedia/PDF-5/Accenture-Top500-English-Executive-Summary.pdf, accessed 01/04/2017.

[Sa08] Samset, E.: Schmalstieg, D.; Vander Sloten, J.; Freudenthal, A.; Declerck, J.; Casciaro, S.; Rideng, Ø.; Gersak, B.: Augmented Reality in Surgical Procedures. Electronic Imaging, 2008, https://www.researchgate.net/publication/44179341_Augmented_reality_in_surgical_procedures, accessed 01/04/2017.

[Sa13] Savage, S. S.: How Watson and Cognitive Computing are Changing Healthcare and Customer Service. IBM Systems Magazine, 2013, http://www.ibmsystemsmag.com/linuxonpower/Watson/cognitive_computing/?page=2, accessed 01/04/2017.

[Sa14] Saporito, B.: IBM Watson's Startling Cancer Coup. Time Health, 2014, http://time.com/3208716/ibm-watson-cancer/, accessed 01/04/2017.

[SF12] Stettiner, E.; Fienhold, M.: EAM 2020 – the Future of the Discipline. In (Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C., eds.): Strategic Enterprise Architecture Management. Challenges, Best Practices, and Future Developments. Springer-Verlag, Heidelberg, pp. 265-286, 2012.

[SB17] Sultanow, E.; Brockmann, C.: Defining a Digitalization Reference Architecture for the Pharma Industry. In Proc. The Open Group, Berlin, Germany, April 24-27, 2017.

[Su16] Sultanow, E.; Brockmann, C.; Schroeder, K; Breithaupt, C.: Lufthansa Aviation Standard: Developing an Open Group Reference Architecture for the Aviation Industry.

In (Mayr, H. C., and Pinzger, M., eds.): INFORMATIK 2016, LNI P-259, Gesellschaft für Informatik, Bonn, pp. 825-836, 2016.

[SC15] Sultanow, E.; Chircu, A. M.: Improving Healthcare with Data-driven Track-and-trace Systems. In (Girard, J. P.; Klein, D; Berg, K., eds.): Strategic Data-Based Wisdom in the Big Data Era, IGI Global, Hershey, PA, pp. 65-82, 2015.

[Ta16] Taft, D. K.: 5 Ways Cognitive Computing Is Advancing Health Care. eWeek, 2015, http://www.eweek.com/database/slideshows/5-ways-cognitive-computing-is-advancing-health-care.html, accessed 01/04/2017.

[Ti16] The Internet of Things. European Commission, 2016, https://ec.europa.eu/digital-single-market/en/internet-things, accessed 01/04/2017.

[WAX15] Whitmore, A.; Agarwal, A.; Xu, L. D.: The Internet of Things—A Survey of Topics and Trends. Information Systems Frontiers, 17/2, pp. 261-274, 2015.