2018 IEEE 22nd International Enterprise Distributed Object Computing Conference

Enterprise Architecture 4.0 – A vision, an approach and software tool support

Adina Aldea, Maria-Eugenia Iacob
Industrial Engineering and Business Information Systems
University of Twente
Enschede, The Netherlands
{a.i.aldea, m.e.iacob}@utwente.nl

Andreas Wombacher, Marlon Hiralal, Thijs Franck
Aurelius Enterprise
Amsterdam, The Netherlands
{andreas.wombacher, marlon.hiralal, thijs.franck}
@aureliusenterprise.com

***Abstract*— Industry 4.0 has begun to shape the way organizations operate by emphasizing the need for a duality between physical machines and sensors, and the (big) data they generate, exchange and use. Manufacturing is one of several industries which is expected to be impacted by this technological revolution. Increasing the information flows and integration of systems within organizations, and along the supply chain is considered one of the main challenges that needs to be addressed by these organizations. One approach for addressing this challenge is to investigate how this abundance of (big) operational data can be used in combination with IT-driven design approaches, such as Enterprise Architecture. Therefore, in this paper we propose our vision for Enterprise Architecture 4.0, i.e. an extended Enterprise Architecture approach for the context of Industry 4.0, and we give an account of our (work-in-progress) efforts to design a model management and analytics software platform supporting this vision. The usage of the software tool is exemplified with the help of a case study with an organization that develops IT and automation systems for the husbandry sector.**

Keywords*— Enterprise Architecture 4.0; ArchiMate; Industry 4.0; Industrial Internet of Things; Big data; Advanced analytics; Manufacturing; Software tool

I. Introduction

Recently, emerging paradigms, such as Industry 4.0 (I4.0), smart manufacturing, or Industrial Internet of Things (IIOT) worldwide have put forward the idea of combining Information Technology (IT) and Operations Technology (OT). The core motivation of these initiatives is that important performance and sustainability benefits can be gained if organizations would increase their information flows, not only internally, but also throughout the supply chain. This could be achieved through a system integration starting from machines on the shop floor and reaching to business applications (e.g., ERP system). We have observed during several projects in the field that many of these integration projects fail or at the best are limited to the physical domain, in the sense they do not go beyond the scope of a specific manufacturing process, that might be enhanced with some cyber-physical capabilities. One of the reasons for this is limited understanding of the concept of digital transformation and the increasing complexity of the involved shop floor systems, processes and applications.

Another important reason is the clear divide between IT and OT, which have been traditionally designed and have evolved independently of each other, and as such, are governed by different standards, methods and tools. Due to I4.0 a subtle change can be observed on both sides, IT and OT.

For example, the field of classical Enterprise Architecture (EA) that has originally emerged from (and was ever since driven by) the need of organizations to be able to steer changes related to IT, has recently started to extend its scope with the physical domain, including for example manufacturing processes. More concretely, in the past the focus of EA models was on IT infrastructure, software applications, information flows, and business processes, designed at a rather high level of abstraction. More recently, such models and architecture modeling languages include concepts to describe entities in the physical domain. For example, ArchiMate version 3 [1] has broadened its scope with physical layer concepts to allow the extension of models to Operation Technology (OT) scenarios in manufacturing industry. In this way it would become possible for organizations to steer change processes that emerge on the work floor, while understanding their impacts on all other layers of the organization: OT, IT, business processes and business strategy and performance.

On the OT side, on the other hand, another trend can be observed related to the idea of extending the existing continuous improvement cycles by using advanced data analytics. The motivation is that established approaches, such as Lean Six Sigma, which have been applied successfully in the past decades are getting less and less effective in terms of achieved benefits, while the market demands with respect to agility, speed, and user experience are increasingly growing. Thus, new approaches and techniques are required to rapidly get new insights in the large amounts of available (operational) streaming data and to increase the benefits of a single improvement step again. However, the interpretation of analysis results in manufacturing requires a good understanding of the origin and quality of the data, of the production line, as well as of the related processes generating/using the data.

To summarize, Industry 4.0 does not only come with the vision of a bright future with respect to efficiency,

2325-6362/18/$31.00 ©2018 IEEE
DOI 10.1109/EDOC.2018.00011

sustainability and integration, but also with significant challenges yet to be overcome: lack of data standardization, heterogeneous unreliable big data, lack of common data models and semantic interoperability, scarcity of (big) data analytics expertise, complex IT-OT system and process integration, both within manufacturing organizations and throughout their respective supply chains.

Our vision is that the core problem of Industry 4.0 related to IT-OT alignment could be addressed by what we call Enterprise Architecture 4.0 (EA4.0), which essentially is an enhancement of EA with operational data, and model-based advanced analytics. The rationale behind this vision is that EA has already proved its value in providing a solution for the business-IT alignment problem and can become an effective means to address the IT-OT alignment problem as well, when a proper integration of models, data and analytics has been achieved. To achieve this vision methods and tools are needed to have these three dimensions available and seamlessly integrated in a shared environment (see Section III). We argue that each of these three aspects (models, data and analytics) is critical, and inherently dependent on the other two. More specifically, (architecture) models enriched with (operational) data would add context to and facilitate the understanding of data and would allow for associations between data and their business semantics, which, subsequently would make possible the correct interpretation of the outcomes of any type of analysis applied to such models. Conversely, data is needed to characterize entities participating in EA4.0 models (e.g., importance of some specific application component), to quantify (alternative) designs, and to eventually make possible advanced quantitative analytics (such as impact analysis). Additionally, from data architecture models might be extracted, that can be analyzed (and update in an automated fashion) and/or can participate in model transformations in which they are morphed with other existing architecture models. Finally, advanced analytics results can give an indication on the compliance of the model with the real world, and of the performance and evolution of the architecture fragment subjected to analysis. However, for the analytical techniques needed for Industry 4.0, more fine-grained models (compared with the usual detail level of ordinary EAs) are required. This is only feasible if data from existing (manufacturing) systems (and possibly IoT sensing devices) and existing models are combined and aligned, as argued earlier.

This vision of EA4.0 requires a radical change in the ways in which we have been working with models. For example, while classical EA models were perceived as relatively static artefacts, for EA4.0 models a new approach is needed in order to deal with the highly dynamic nature of these new type of artefacts, as an EA4.0 model can change due to both data associated with its architectural entities (e.g. streaming manufacturing process data, represented as time series), and due to changes in the set of architectural entities that make up the EA4.0 model (e.g., sensing devices that can be switched on and off). Besides this technical aspect concerning model management, a change in design methodology, and a mental shift towards working with the integrated concepts is required, to enable the agility, speed and user experience expected by the users. In this paper the requirements for EA4.0 are derived from several use cases we conducted within multiple projects with manufacturing organizations. These requirements are first discussed in general terms and are subsequently explained in more detail. Furthermore, as core contribution we posit the formulation of an EA4.0 development approach and the development of a software platform supporting this approach.

The remainder of this paper is organized as follows. Section II gives an overview of the current state-of-affairs in enterprise architecture, data analytics, and smart manufacturing/I4.0, with a special emphasis onto the dependency relationships between them and based on a systematic literature survey. In Section III we give an account of the methodological aspects of our EA4.0 development approach. In Section IV we identify the requirements for an EA4.0 model management and analytics platform, and we present its architecture. We validate and illustrate the usage of both the approach and the EA4.0 platform in Section V, in which a case study is presented. We conclude the paper with Section VI, where some conclusions are formulated and pointers to future work are given.

II. Background and Previous Work

In this section we briefly introduce the discipline of EA and its development over the past few years. Following this, we discuss studies which address the relationship between EA, data and analytics from several points of view. We conclude with an overview of research focused on the application of EA in the context of Industry 4.0 and the manufacturing industry.

To identify relevant studies that address the relationship between EA, data and analytics, we have performed a small systematic literature review (SLR), following the guidelines proposed by [2]. Thus, we have performed the following steps:

- Defined a research question (RQ): What is the current thinking on the relationship between EA, data, analytics, Industry 4.0 and manufacturing?;
- Selected the scientific databases: Scopus, IEEExplore and ScienceDirect;
- Defined the keyword string, linked with the help of AND and OR operators: ("business intelligence" OR "analytics" OR "big data" OR "industry 4.0" OR "internet of things" OR "cyber physical systems" OR "manufacturing") AND "enterprise architecture";
- Selected inclusion criteria: English language studies, journal or conference papers, published since 2010 (including 2010);
- Selected the exclusion criteria: duplicate studies and studies which don't help answering the RQ;
- Selected first sample of relevant studies by reading the title, abstract and conclusions;
- Read the selected studies in their entirety to confirm their contribution to answering the RQ.

As a result of the SLR, we have identified a total of 233 potential studies, of which the majority (169 studies) were found from the Scopus database. After the elimination of duplicates, 171 studies were assessed based on their title, abstract and conclusions. The final selection has provided 21 studies which are relevant for answering the RQ.

A. Enterprise architecture

EA is a discipline which focuses on the holistic management of the enterprise, based on aspects of its architecture, such as business processes, applications, information, hardware, as well as the relationships between them [3]. It is comprised of a set of frameworks, methods, models and tools to help organizations deal with emerging Business-IT capabilities and to align them to existing business processes, organizational structures, information systems, technical infrastructure [4].

Throughout the years, many EA frameworks have been developed, of which the TOGAF, followed by the DoDAF, and the Zachman framework are the most adopted in practice [4, 5]. Typically, these EA frameworks cover four interrelated domains: Business architecture (business processes of an organization), Data architecture (structure of the logical and physical data resources), Application architecture (landscape of applications, their interactions, and relationships to processes), and Technology architecture (software and hardware capabilities required to support the business processes, data, and application services of the organization).

TOGAF is a comprehensive open EA standard which contains several components, amongst which, the Architecture Development Method (ADM) stands at its core [6]. The ADM describes an iterative process for developing EA with the help of several phases (Preliminary phase, Architecture Vision, Business, Information systems, Technology, Opportunities and solutions, Migration planning, Implementation governance, Architecture change management, Requirements management). While TOGAF offers viewpoints, techniques and reference models to design EAs, it does not contain a modeling language. Therefore, the ArchiMate specification has been developed to enable organizations to model and describe EAs over time, as well as their motivation and rationale [1]. Over the years, several studies have proposed adjustments and extensions to the ArchiMate specification (e.g.: [7-11]), which have contributed to its development. The latest version of the specification (3.0) contains four core layers (Business, Application, Technology, and Physical), the Strategy layer (courses of action, capabilities and resources which can be used to model the strategy of an organization), Implementation and migration layer (programs, portfolios, project management, and plateaus that can be used in gap analysis), and several Motivation aspects (goals, requirements, stakeholders, etc. that can be used to model the motivation behind organizational change).

B. Relation to (big) data and analytics

While the discipline of EA is reaching maturity, there are still aspects which have only limited research available. One such example is the relationship between EA, data and analytics. Based on the results of our systematic literature review, we can identify several research lines.

On the one hand, there are studies which consider the integration of data and analytics into current EA practices and standards. As an example, the study of [12] identifies several phases of the TOGAF ADM (Information Systems Architecture, the Technology Architecture, Requirements management) which are most likely to be impacted by the introduction of a big data practice in an organization. Therefore, these phases of the TOGAF ADM need to be adjusted to include guidance and techniques to help organizations manage this kind of transformation. Several related studies by [13, 14] categorize and discuss the requirements brought on by big data analytics to the management of EA, while the study of [15] introduces a method to help organizations start and manage their big data practice with the help of EA management techniques.

On the other hand, there are studies which emphasize the importance of relating data and analytics to EA and have proposed approaches and frameworks which focus on this. One such example is the research of [16] which proposes an EA analytics framework which combines aspects of Service-Oriented EA and Big data in order to facilitate the discovery, analysis and optimization of EAs. Other examples include studies which propose frameworks for BI-enabled adaptive EAs [17, 18], which integrate aspects of EA frameworks with BI tools [19], or studies which propose the extension of EA modeling with BI aspects [20].

While the aforementioned studies give a good indication of the importance of combining EA, data and analytics, they do not provide guidance on how to relate data from information systems to EA models. The study of [21] addresses this limitation by proposing a method that combines EA with operational data. This is done with the help of model transformations necessary for matching concepts from an EA model to the corresponding elements from a data source, to enrich the EA model with data. Subsequently, this data can be used to perform several types of model-based analyses. We consider this study to be an inspiration for our approach, due to its similarities with our understanding of how EA, data and analytics should be related. However, we intend to go beyond the method proposed in [21] by addressing some of its limitations (e.g.: matching data with elements from an EA model), proposing several advanced analytics, and proposing tooling support designed specifically to enhance the combination of EA, data and analytics.

C. Relation to Industry 4.0 and manufacturing

The vision for Industry 4.0 is based on the duality between physical machines and sensors, and the (big) data they generate, exchange and use. Therefore, manufacturing is one of several industries which is expected to be impacted by this technological revolution. However, since the concept of Industry 4.0 and its supporting technologies can be considered fairly new, the research on this topic is rather limited, especially in relation to EA. Nonetheless, several studies have investigated its relation to EA from a few points of view.

One of the main research lines currently available revolves around investigating current paradigms and reference architectures for Industry 4.0, such as IIRA, RAMI4.0, etc., and their applications in practice [22-24]. Another interesting development comes from a series of studies originating from the same authors, which investigate the adaptation of current EA viewpoints, models, standards, frameworks and tools to cover Internet of Things (IoT) types of distributed services and devices [25-28]. Similarly, the study by [29] investigates the impact of integrating Cyber-Physical Systems (CPS) into EA, from the point of view of the TOGAF ADM. Lastly, the study of [30] focuses on the extension of EA metamodels with IoT specific concepts and layers in order to facilitate modeling of distributed devices and sensors.

From a manufacturing point of view, several studies address the suitability of the ANSI/ISA-95 (standard of the manufacturing industry) in relation to Industry 4.0 and EA [31, 32]. The study of [31] investigates the suitability of the ArchiMate 3.0 specification, and concludes that the language covers 96% of the relations needed for modelling the exchange of information between OT and IT layer. However, to cover the remaining relations, adjustments to the ArchiMate language are needed, which have also been proposed in [31]. For example, modeling a Bill of Materials (BoM) and its relations with other systems, such as ERP. Should the BoM be modeled as a business object or as a material on the physical layer? Physical objects often have a duality: they exist in the physical world, but they also have a digital representation in IT applications. Understanding and representing this duality is one of the major challenges.

III. Approach for EA 4.0

As mentioned before, our vision for EA 4.0 has as starting point the strong inherent relationship between three main aspects, namely EA models, data from enterprise information systems and IoT devices, and advanced analytics (Fig. 1).

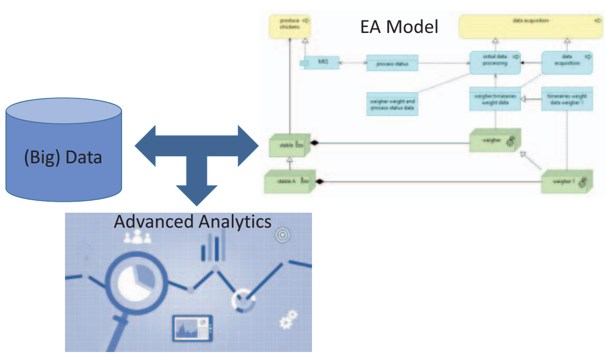

Fig. 1. The components of EA 4.0

In this section, we discuss these three aspects in more detail, and their relationships. Furthermore, we introduce our proposed approach for EA4.0 which is described by means of several workflow models (see Fig. 2 - Fig. 6).

A. EA models

As mentioned in Section II, ArchiMate is the most appropriate language for creating EA models, especially considering the recent addition of the physical layer in the latest version of the specification. For manufacturing organizations, the physical layer, (e.g.: shop floor, physical machines, raw materials, products, and finished goods) can be used to model important assets and have a strong relation to OT and IT applications and processes.

As a starting point for EA models, we do not focus on the motivation and reasoning behind designing a model (i.e. the Preliminary phase and Phase A of the TOGAF ADM), which we consider as given. We rather focus on the Phases B, C and D where the Business, Application, Data, and Technology architectures are defined and refined. However, unlike the TOGAF ADM, in the case of our approach, we propose that design and redesign of EA models is done based on advanced analytics results, thus emphasizing the iterative nature of the approach. We discuss this aspect in more detail in Section D.

Nonetheless, we consider the design and redesign of EA models as the first step of our approach. The following steps, shown in Fig. 2, include importing the model in a central repository, managing the conflicts brought on by different versions of the same model, and lastly, making the model available for analysis.

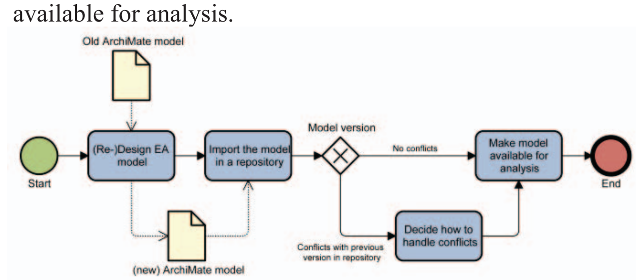

Fig. 2. Process for EA models

The central repository is essentially a database in which different types of models and data from multiple sources can be added. Its main role is to facilitate the integration of models and data with the help of several algorithms which support the necessary import and version management functions.

The import algorithm is a crucial aspect which dictates how information from models is stored in the repository. In the case of ArchiMate models and views both concepts, relationships and the direction of relationships need to be imported and their meaning needs to be preserved. Such an algorithm also needs to facilitate the relation between models and data. We discuss this aspect in more detail in Section C.

In terms of the version management algorithm, its main purpose is to ensure that when a new version of a specific model is added to the repository, the existing version does not automatically get replaced. Without this mechanism in place, the integrity of the information in the repository cannot be guaranteed. One way of facilitating versioning of models is to use the plateau concept of ArchiMate to annotate concepts and relations belonging to a version or a view. However, manually managing the relations of architecture concepts and relationships to a version, or a view is not only very cumbersome and error prone, but also results in very complex and hard to understand hierarchical plateaus. Alternatively, an implementation similar to the version management system for merging Git branches can also be used for merging the different versions of ArchiMate models added to the repository. Regardless of how the versioning is supported, the conflicts between the different model versions (e.g., when a concept has changed in both versions to be merged) need to be resolved manually, since an algorithm cannot decide which changes are intended and which need to be inverted.

The last step of our approach deals with making the models available for analysis. This implies that a new version has been added into the repository, all conflicts have been resolved and the models are considered suitable for being analyzed and/or combined with data.

B. Data

Organizations nowadays are collecting large amounts of data from a multitude of sources. This data is stored in separate systems and databases which usually have different data models and heterogenous conventions for naming concepts. Therefore, it is very important to understand what kind of data is available and where this data can be found. We consider this to be the first step of our approach concerning data, followed by the import of the selected data to the repository (Fig. 3).

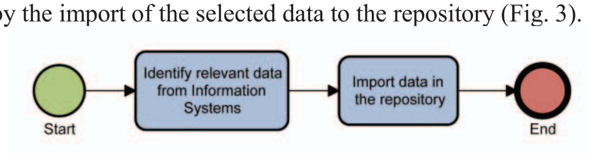

Fig. 3. Process for Data

EA can play an important role in identifying the kind of data that is available and which system or database it can be extracted from. EA models can be used to provide a good overview of the application and technology landscape of an organization. For example, ArchiMate data objects can be used to model the information which is generated and exchanged by different systems (modeled as application components) or stored in databases (modeled as system software). Alternatively, the metadata from systems or databases can be transformed into an ArchiMate model which can provide an in-depth view of the data available in each system or database (Fig. 4).

When considering the different types of data available, the following are commonly found in the manufacturing industry: time series (e.g.: a stream of temperature or pressure measurements), transactional (e.g., a relational database table), structured ad hoc (e.g., excel files), unstructured (e.g., text), binary (e.g., images), and complex data (e.g., a mix of all possible data types like the log file of an electronic microscope containing XML data describing the device settings as well as a set of images representing intermediate results of the processing in the machine). Furthermore, in terms of data ingestion methods, the following are considered as characteristic for the manufacturing industry: continuous/near real time (e.g., streaming data), batch (i.e., a bulk of data coming in with a certain frequency, e.g. once a day, or hourly), or ad hoc (e.g., applying some analysis results to a model).

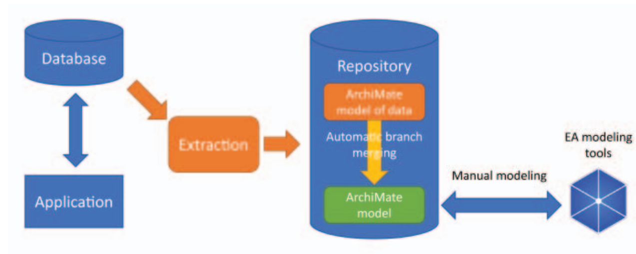

Fig. 4: Model Extraction

Finally, the data can be imported into the repository in a similar manner to the models. Therefore, this can be done automatically, with the help of an algorithm which pulls data from systems or databases on a regular basis or even real-time. Alternatively, importing data can be done manually and selectively for specific types of data.

C. Relating EA models and data

Relating EA models and data has been addressed before in [26], which proposes to have a separation between abstract model, concrete model and data. Another approach is proposed in [21], where the authors define a two-step process which starts with the selection of a subset of concepts and relationships from the EA model and, similarly, a subset of data from the available data sources. This is followed by a manual matching of the elements from these two subsets, with the results stored in a Concept Match Library for future usage. We consider a similar approach to what is proposed in [21] for matching EA models with data (Fig. 5).

For selecting the appropriate subset of EA concepts and relationships, as well as for the selection of the subset of data domain knowledge is very important. Therefore, we consider that this step requires manual processing and relies on the ability of practitioners to make an appropriate selection.

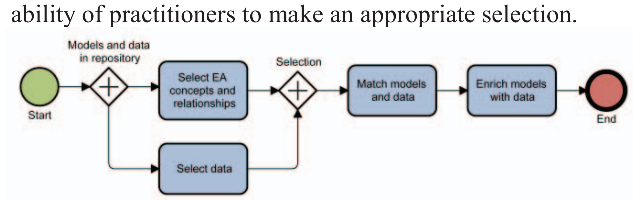

Fig. 5. Process for relating EA models and data

Elements from EA models and data can be connected in different ways. For example, an element of an EA model can be related to a single data point or a set of data points. This can be done in an automated manner by matching the label of an ArchiMate element or of its properties to the label of a column with data from a data source. However, this kind of matching heavily relies on having the same naming conventions in both the ArchiMate model and in the data source. Alternatively, this process can also be done manually, as is suggested in [21].

Furthermore, a modeling concept may be related to a (type of) query describing relevant data stored in a data storage. Based on these options, the model can be enriched with specific values which can be used for an analysis or can become a means to retrieve the data by using constraints or queries. However, the large number of concepts in a model requires strong query capabilities to support managing and editing the model. The following are a few examples of how queries can be used in relation to EA models.

1) Query 1: A user wants to see which test procedure (business process) related to a Laboratory Information Management System has produced the quality statement (business object) of a product class in the ERP. To answer this question, one has to identify the ”shortest” path between a business process and a business object for a given product;

2) Query 2: A user wants to see the changes between two versions of the model. It is easy to determine which concepts and relations have been added or removed. However, in order to understand the impact and the dependencies between additions and removals, a clustering of these changes is needed. The aim of the clustering is to provide an inherent semantic relation with respect to the modified concepts and relations. Potentially, these clusters might contain some context, that is, some unmodified concepts and relations, to make interpretation for a human easier. Thus, the underlying query would allow to identify clusters of changes between two views or plateaus which support the interpretation of changes by a human user;

3) Query 3: A user is looking for the model part describing the virtualization of a Manufacturing Execution System (MES). Thus, the query is using the MES application component in combination with the ”virtualization” model pattern, which has been defined by this company. The query result is a view containing the MES application component and the relevant concepts relevant for the virtualization pattern. The underlying query results into a sub-graph matching problem. This list of queries is not exhaustive but proposes several realistic use cases which could benefit from this kind of querying combinations of EA models and data. Furthermore, many of the above queries will not have a single result, but will provide a list of possible query results, in a similar manner as Web searches do. Thus, ranking these results, and limiting the query execution to only the “top”-k query results is another technical challenge.

The envisioned way of working is very interactive. Users would run queries to build the views they want to use. Such a way of working also assumes a high query performance. However, most of the underlying generic algorithms to answer the above queries are NP (Nondeterministic Polynomial time)-complete, and therefore have a very high computational complexity. For many of these problems heuristics are available delivering nearly optimal solutions with a much lower complexity under certain circumstances.

D. Relating EA models, data and advanced analytics

Besides query capabilities for searching or viewing data-enriched models, analytics can be performed in a model-based fashion. The difference is that analytics are not intended to be used to edit the model, but to generate new insights derived from the model. Often, analysis results produce new data or new concepts and relations, which can eventually be used redesign the EA model (Fig. 6).

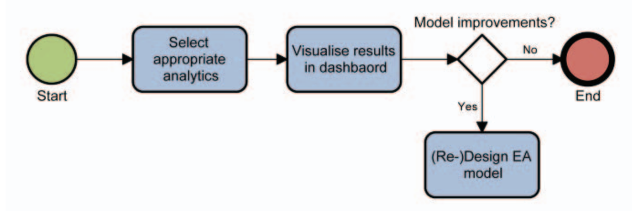

Fig. 6. Process for analytics, EA models and data

The main challenge of such advanced analytics lies in incorporating the right type of data in the EA model to make a specific type of analytics possible and in defining and implementing an efficient algorithm for that specific analytical problem. The following are a few examples of advanced analytics tools related to EA models combined with data.

1) Analytics 1: An important part of architecture governance could be to analyse a model in order to see how well architects are actually using model patterns. Thus, one can query the model for discovering ”almost correct” patterns (based on sub-graph matching techniques, and similarity measures), and draw the attention of architects to the missing or incorrect parts. Furthermore, one might want to discover new emergent model patterns, that is, model fragments, which are not considered as being patterns yet, but are used more and more by architects (i.e., model pattern mining). The underlying analytics is a graph clustering problem.

2) Analytics 2: A project manager calculates key performance indicators (KPIs) for a business transformation project (based on available usage data of underlying infrastructure and applications) to support the decision of selecting a certain design alternative. This analysis combines utilization and usage data attached to different architecture concepts in the model, and therefore needs data-enriched EA models for the analysis. The underlying technique for this type of analysis is a graph traversal algorithm.

3) Analytics 3: An architect has modeled a production process controlled by a MES from the point of view of the user interaction and of the information flow. By using process mining results he can also create a third view of the process showing the physical product viewpoint. He now wants to check whether the three views of the manufacturing process are consistent with each other. The challenge in this scenario is that the different views are not only describing different aspects of the same process, but are also imposing constraints on each other. Thus, the aim is to identify what are those constraints and whether they are fulfilled or not.

IV. REQUIREMENTS FOR EA 4.0 AND MODELS4INSIGHT

To support our approach for EA 4.0 with tooling, we have conducted several interviews with organizations from the manufacturing industry, more specifically, from the process[1]

[1] Manufacturing products continuously or in batches which potentially will be transformed into discrete objects.

and discrete[2] branches of the industry. The results of these interviews have been refined and formulated as requirements, which have been used to shape the design of an EA4.0 model management and analytics platform. This platform is currently available as a (under development) prototype, under the name of Models4Insight. In the remainder of this section we distil the results of these interviews in the form of several features the platform should support which are also reflected by the current platform architecture explained in Sub-section IV.A.

1) Importing and versioning of model parts from external applications must be made possible: many concepts in a model are managed and maintained in other applications (like e.g. customer database (CRM), products (PLM), control of the production process (MES));

2) Data import and data management capabilities in the model: a model could be related to various type of data (e.g., all model layers may be associated with some notion of cost, business transformation is related to a KPI improvement, processes are related to usage frequency, infrastructure and physical layers have related numerical configuration parameters, as well as performance metrics, such as mean time to failure, or utilization);

3) Different views of a model must be consistent with each other, but also distinguishable from each other: a model may specify the same process at different levels of details, as well as from different points of view: e.g., user interface perspective, system perspective, or product perspective;

4) New model editing capabilities to deal with the increased number of concepts in the model are necessary: due to the complexity of a model, a different way of editing and validating a model is required (e.g., it is not feasible anymore to manually search for a concept, but it should be possible to query a model on concepts and relationships);

5) Rich and fast analytics capabilities: answering the business problems requires rich analytics and query capabilities (e.g., parsing of graphs, subgraph pattern matching, and basic analytics combining data and graphs);

6) Analytics on the model requires the consistent usage of model patterns: analysis requires that certain constructs in processes, applications or infrastructure to be always modeled in the same way, otherwise it becomes very difficult to perform analysis on models, in particular when multiple models and viewpoints of the same architecture are involved.

A. About Models4Insight

Models4Insight[3] is a Software as a Service browser-based tool which means it can be used from any device with an internet connection without requiring a local installation, or updates when new versions are released. It consists of several components, such as the platform, the portal, and the underlying repository (Fig. 7). The platform allows users, after logging in, to create new projects and/or manage existing ones. A project is comprised of an ArchiMate model (created in an external EA modelling tool, such as Archi[4]), which is imported into Models4Insight.

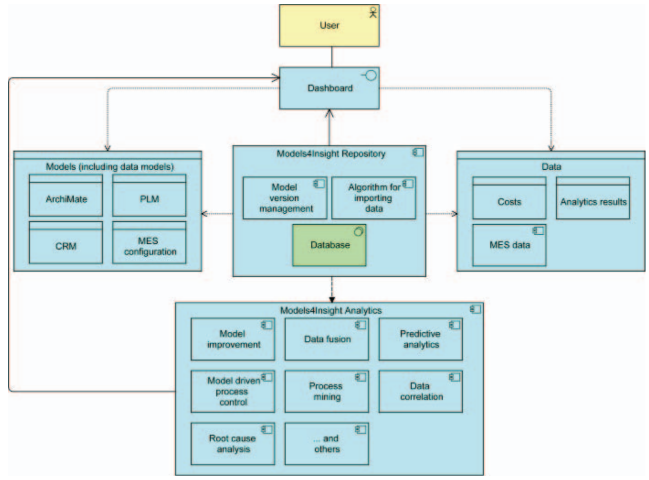

Fig. 7. Models4Insight tool architecture

Once a model is imported, it is stored in the underlying repository which, amongst others, also manages the different model versions. If a new version of a model is imported into an existing project, a pop-up window offers the possibility to manage the conflicts between the original model and its new version to determine how the two versions should be merged (Fig. 8). This process works in a similar manner to the version management system for merging Git branches.

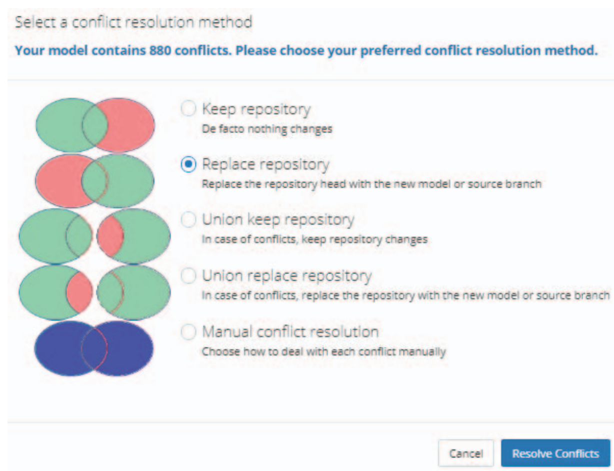

Fig. 8. Conflict management

The platform includes a project management page which contains several aspects, such as retrieving the repository version of the model, cloning or merging branches with different model versions, publishing the model to the portal for analysis, importing a new model, giving or removing access of users, navigating to different areas of the platform or portal, and seeing a history of all the changes to the project.

Once a model is imported into the platform (and repository), it can be published to the portal, where it can be

[2] Manufacturing of products assembled from multiple parts.
[3] http://www.models4insight.com
[4] https://www.archimatetool.com/

viewed and analyzed with the help of dashboards. These dashboards include both analyses based on model elements (e.g., gap analysis between two versions of the same model), and analyses based on data related to models (e.g.: consistent usage of relationships in a model, or compliance of processes described in a model with event sequences observed in application logs). In order to allow users to define their own analyses, Models4Insight supports a Python-based Analytics Library containing functions and specialized algorithms.

It is worth noting that some of the aspects mentioned in Section III are not yet implemented in the Models4Insight tool, and others still require some form of manual handling. Nevertheless, we plan to extend the automation of all features and enrich the library of analytical functionality in a future version of the tool. One example of a feature which currently requires manual input is linking data from different sources to the elements of an ArchiMate model. In future iterations, it is intended to automate this feature (possibly by using model transformations) to improve the tool's scalability and applicability to big data sources and real-time data sources.

V. Manufacturing Case study

In this section we demonstrate how our proposed approach for EA 4.0 is applied to the case study of a manufacturer of systems for the agricultural sector. Due to a confidentiality reasons, we have anonymized the name of the organization, which will be from here on referred to as AgriComp. Furthermore, in this section, we include figures containing screenshots from the Models4Insight.

A. *The case study organization and problem description*

AgriComp is an international organization which is focused on the development of IT and automation systems (climate, feeding and biometric systems) for the livestock husbandry sector. The organization provides its services to a variety of livestock types, including poultry. One of the main problems the organization is currently facing is that product failures are expensive to identify and to resolve, since new products are shipped all over the world. The intention is to collect data from the products, to provide value added services as well as reduce costs for product improvement and problem solving.

In this section, we focus on a specific instance in which the Models4Insight platform is used to give insights into the weights of chickens in several stables in order to detect anomalies. These anomalies can be either attributed to real fluctuations in the weights of the chickens or by erroneous values provided by faulty equipment.

B. *EA models*

The starting point for addressing the problem of AgriComp is the design of baseline architecture models with the help of ArchiMate 3.0, as indicated in the first step of our approach. Fig. 9 illustrates one of the baseline views of the case study model. Once these models are created, they can be imported in the Models4Insight platform. After logging into the platform, a new project can be created and the ArchiMate model can be imported. In the case of AgriComp, there is already a project in the platform with an older version of the ArchiMate model. Therefore, the conflicts between the two versions need to be handled. In our case, the new version should replace the version in the repository (Fig. 8 illustrates the version management functionality). After the conflicts are handled, the model is published in the portal, where further analysis can be done, which completes the first phase of our approach.

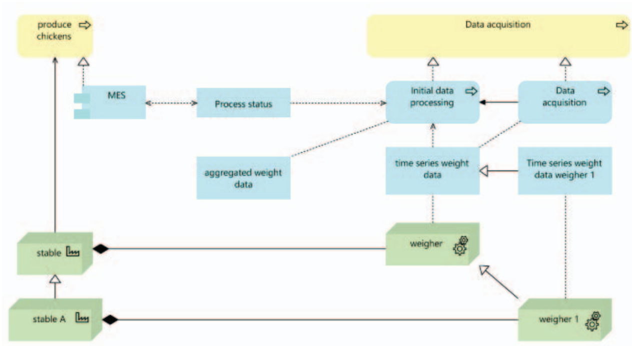

Fig. 9. Example ArchiMate model for AgriComp

These manually created models can be complemented by models maintained by other applications. One major system used by AgriComp is the MES which collects data, monitors and controls the production process. In this case, the MES stores the data in a relational database. Thus, it is possible to use the metadata from the MES to generate an ArchiMate model. Subsequently, this model can be imported into Models4Insight, where the manually created model and the automatically generated model can be merged. Fig. 10 illustrates an example ArchiMate model which has been generated based on metadata from the MES of AgriComp.

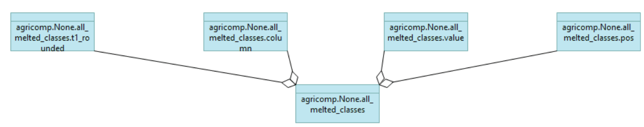

Fig. 10. Example generated model from metadata

C. *Available data*

While the MES is a very rich source of data about all the different systems that the organization is managing, in our case, we focus mainly on the data relating to the chicken weighting equipment. Thus, the data we consider comes in the form of time series weight class data and count data from each of the weighers in use. After initial processing, this is combined with data about the status of the process (Fig. 11). Identifying the system which contains the relevant data, in our case the MES, covers the first step of our approach relating to data. Selecting the right type of data represents the starting point for combining it with the ArchiMate models.

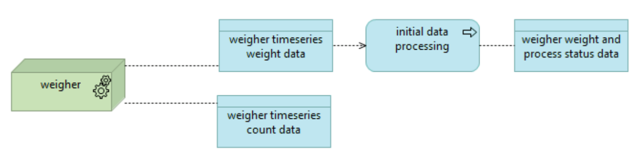

Fig. 11. Types of data related to each weigher

As mentioned before, relating the data to elements of the model is currently not automated. With the help of the generated ArchiMate model, we can manually relate the weigher instances with the corresponding data objects with the help of Access relations. Fig. 12 illustrates a snippet of the Python code used to assess the mapping of the weighers from the ArchiMate model to their corresponding data.

```python
# Load the model from the repository
model_options = {
                'projectName': 'AgriCOmp',
                'projectOwner': 'dev',
                'branchName': 'MASTER',
                'userid': 'test_user'
            }
model = ArchimateUtils.load_model_from_repository(**model_options)

# select all weighers
weighers = model.nodes[model.nodes.type==ElementType.EQUIPMENT]
weighers = weighers[weighers.name.apply(lambda x: x.startswith('weigher '))]
weighers = weighers[['name', 'id']]
weighers.columns = ['weigher', 'weigher_id']

# select and associate related Relationships
mapping = model.edges[model.edges.type==RelationshipType.ASSOCIATION]
weighers = weighers.merge(mapping, how='left',
                          left_on='weigher_id', right_on='source')
weighers = weighers[['weigher','target']]

# associate the relevant elements
data = weighers.merge(model.nodes, how='left', left_on='target', right_on='id')
data = data[['weigher','name']]
data.columns= ['weigher','data_object']
data = data.fillna('not assigned')

# prepare data and plot heatmap
data = data.groupby(by=['weigher','data_object']).size()
data = data.reset_index()
data.columns = ['weigher','data_object', 'cnt']
data = data.pivot(index='data_object', columns='weigher', values='cnt')
data = data.fillna(0)
data[data[data.index=='not assigned']==1] = -1
sns.heatmap(data, cmap="RdYlGn", center = 0)
```

Fig. 12. Python code to assess mapping model elements to corresponding data

After importing the new version of the model, the analytics library can be used to assess whether the mapping is 1-to-1 by finding all weighers in the model and checking whether they are mapped to one or more data objects.

D. *Results of analysis*

Once the data is related to the weighers from the EA model, different types of analyses can be done. This covers the last part of our approach, which is focused on analysis and visualization of results. As an example, in Fig. 13 we can see the minimum, mean and maximum weight over time for each weigher from each stable. This information is used by the farmer to assess the growing rate of the chickens and it is used by AgriComp to assess the quality of their products and identify root causes in case of errors.

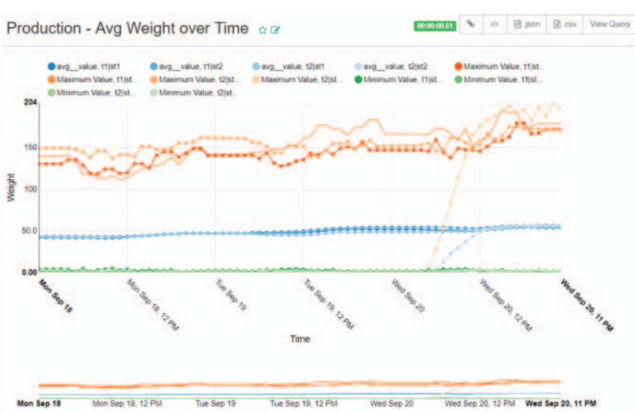

Fig. 13. Example average weight over time analysis

While this analysis can also be made without having a relation to EA models, the main advantage of our approach is that we can use the elements and relations from the model to easily aggregate data. As an example, in Fig. 14 we can use the composition relation as means to define the calculation of values for a stable based on the aggregate values of its weighers. Therefore, when the required data is spread amongst multiple systems, the ArchiMate model can take the role of a data model. This also implies that if certain ArchiMate elements or relations are removed, this will have an impact in the analysis of its related data.

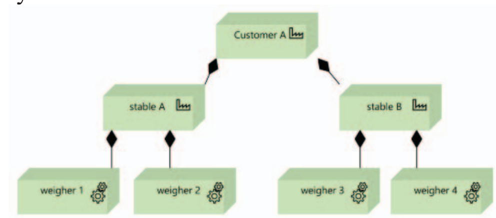

Fig. 14. Example decomposition of stables into weighers

VI. Conclusions

In this paper we have elaborated our vision on EA 4.0, that focuses on the relationship between EA models, data and advanced analytics, and we presented the design of a software platform supporting this vision. We argue that this relationship will play a crucial role in the development of organizations which operate in environments where a large amount of data is generated from many devices and sensors, such as those in the manufacturing industry. The case study we presented fits the type 2 analytics described in Section III. Due to space limitations we did not include two other case studies we carried out fitting the third type of analytics. In all case studies we have used real manufacturing data and models, which has led to several observations and conclusions, as explained below.

1) Higher level of detail for EA models: The proposed EA4.0 approach generally requires EA models to be much more detailed than most of them currently are. We observed that high level models rarely allow for a correct definition of correspondences with operational data, and hence reasoning about models is very difficult. However, having to deal with very detailed models poses serious challenges when aiming to keep models and model versions consistent.

2) New skills required: The usage of advanced analytical techniques requires that either enterprise architects have to gain new skills, or (more likely) they have to team up with other experts, such as data scientists/engineers.

3) Systematic and objective review of the models: The proposed EA4.0 approach makes possible the systematic and objective review of the models against up-to-date data from existing systems instead of through discussions between domain experts (i.e., architects and process owners). This is why it is realistic to expect that EA 4.0 models would gain importance, since they are no longer isolated static artifacts: EA 4.0 behaves much more as a living organism, reflecting the dynamic changes in all layers of the organization, and hence playing a central role in any transformation processes.

This study is a reflection of the current state of the Models4Insight platform. Future work concerns the extension of the analytics toolbox embedded in the platform, among others through validation in more Industry 4.0 case studies. Also, one part of the platform which is underdeveloped is model querying functionality, that is based on advanced graph analysis techniques. Some of these are already implemented, in particular those related to structural and topological properties of the underlying architecture models, such as the calculation of complexity, and density metrics.

References

[1] The Open Group, *ArchiMate 3.0 Specification*. Zaltbommel: van Haren Publishing, 2016.

[2] B. D. Rouhani, M. N. Mahrin, F. Nikpay, R. B. Ahmad, and P. Nikfard, "A systematic literature review on Enterprise Architecture Implementation Methodologies," *Information and Software Technology,* vol. 62, pp. 1-20, 2015.

[3] S. Buckl, F. Matthes, A. W. Schneider, and C. M. Schweda, "Pattern-based design research in enterprise architecture management," in *International Conference on Advanced Information Systems Engineering*, 2013, pp. 30-42: Springer.

[4] A. Alwadain, E. Fielt, A. Korthaus, and M. Rosemann, "A critical realist perspective of Enterprise Architecture evolution: conditioning and outcomes," *Australasian Journal of Information Systems,* vol. 18, no. 3, pp. 213-226, 2014.

[5] B. H. Cameron and E. McMillan, "Analyzing the current trends in enterprise architecture frameworks," *Journal of Enterprise Architecture,* vol. 9, no. 1, pp. 60-71, 2013.

[6] The Open Group, *TOGAF Version 9.1*. Zaltbommel: Van Haren Publishing, 2011.

[7] A. Aldea, M. E. Iacob, J. van Hillegersberg, D. Quartel, and H. Franken, "Modelling Value with ArchiMate," in *Advanced Information Systems Engineering Workshops*, Cham, 2015, pp. 375-388: Springer International Publishing.

[8] M. E. Iacob, D. Quartel, and H. Jonkers, "Capturing Business Strategy and Value in Enterprise Architecture to Support Portfolio Valuation," in *2012 IEEE 16th International Enterprise Distributed Object Computing Conference*, 2012, pp. 11-20.

[9] A. Aldea, M. E. Iacob, J. Van Hillegersberg, D. Quartel, L. Bodenstaff, and H. Franken, "Modelling strategy with ArchiMate," in *30th Annual ACM Symposium on Applied Computing, SAC 2015*, 2015, vol. 13-17-April-2015, pp. 1211-1218: Association for Computing Machinery.

[10] W. Engelsman, D. Quartel, H. Jonkers, and M. van Sinderen, "Extending enterprise architecture modelling with business goals and requirements," *Enterprise Information Systems,* vol. 5, no. 1, pp. 9-36, 2011/02/01 2011.

[11] A. Aldea, M. E. Iacob, J. Van Hillegersberg, D. Quartel, and H. Franken, "Capability-based planning with ArchiMate: Linking motivation to implementation," in *17th International Conference on Enterprise Information Systems, ICEIS 2015*, 2015, vol. 3, pp. 352-359: SciTePress.

[12] C. Kearny, A. Gerber, and A. v. d. Merwe, "Data-driven enterprise architecture and the TOGAF ADM phases," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 4603-4608.

[13] S. Kehrer, D. Jugel, and A. Zimmermann, "Categorizing Requirements for Enterprise Architecture Management in Big Data Literature," in *2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW)*, 2016, pp. 1-8.

[14] M. Lněnička, R. Máchová, J. Komárková, and I. Čermáková, "Components of big data analytics for strategic management of enterprise architecture," in *SMSIS 2017 - Proceedings of the 12th International Conference on Strategic Management and its Support by Information Systems 2017*, 2017, pp. 398-406.

[15] M. Vanauer, C. Bohle, and B. Hellingrath, "Guiding the introduction of big data in organizations: A methodology with business- and data-driven ideation and enterprise architecture management-based implementation," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2015, vol. 2015-March, pp. 908-917.

[16] R. Schmidt, M. Wißotzki, D. Jugel, M. Möhring, K. Sandkuhl, and A. Zimmermann, "Towards a Framework for Enterprise Architecture Analytics," in *2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations*, 2014, pp. 266-275.

[17] O. Akhigbe, D. Amyot, and G. Richards, "A Framework for a Business Intelligence-Enabled Adaptive Enterprise Architecture," in *Conceptual Modeling*, Cham, 2014, pp. 393-406: Springer International Publishing.

[18] E. Yu, S. Deng, and D. Sasmal, "Enterprise Architecture for the Adaptive Enterprise – A Vision Paper," in *Trends in Enterprise Architecture Research and Practice-Driven Research on Enterprise Transformation*, Berlin, Heidelberg, 2012, pp. 146-161: Springer Berlin Heidelberg.

[19] P. M. Ortega, L. G. Ávila, and J. M. Gómez, "Framework to Design a Business Intelligence Solution," in *ICT Innovations 2010*, Berlin, Heidelberg, 2011, pp. 348-357: Springer Berlin Heidelberg.

[20] R. Claes, "Expanding the discipline of enterprise architecture modeling to business intelligence with EA4BI," in *CEUR Workshop Proceedings*, 2017, vol. 2027, pp. 50-57.

[21] R. K. M. Veneberg, M. E. Iacob, M. J. Van Sinderen, and L. Bodenstaff, "Relating Business Intelligence and Enterprise Architecture - A Method for Combining Operational Data with Architectural Metadata," *International Journal of Cooperative Information Systems,* Article vol. 25, no. 2, 2016, Art. no. 1650007.

[22] M. Moghaddam *et al.*, "Next-generation enterprise architectures: Common vernacular and evolution towards service-orientation," in *Proceedings - 2017 IEEE 15th International Conference on Industrial Informatics, INDIN 2017*, 2017, pp. 32-37.

[23] M. Nardello, C. Møller, and J. Gøtze, "The industry 4.0 Journey: Start the learning journey with the reference architecture model industry 4.0," in *CEUR Workshop Proceedings*, 2017, vol. 1898.

[24] D. Schulte and A. W. Colombo, "RAMI 4.0 based digitalization of an industrial plate extruder system: Technical and infrastructural challenges," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, 2017, pp. 3506-3511.

[25] A. Zimmermann, R. Schmidt, K. Sandkuhl, M. Wißotzki, D. Jugel, and M. Möhring, "Digital Enterprise Architecture - Transformation for the Internet of Things," in *2015 IEEE 19th International Enterprise Distributed Object Computing Workshop*, 2015, pp. 130-138.

[26] A. Zimmermann, J. Bogner, R. Schmidt, D. Jugel, C. Schweda, and M. Möhring, "Digital enterprise architecture with micro-granular systems and services," in *CEUR Workshop Proceedings*, 2016, vol. 1684.

[27] A. Zimmermann, R. Schmidt, K. Sandkuhl, D. Jugel, M. Möhring, and M. Wißotzki, "Enterprise architecture management for the internet of things," in *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, 2015, vol. 244, pp. 139-150.

[28] J. Bogner and A. Zimmermann, "Towards Integrating Microservices with Adaptable Enterprise Architecture," in *2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW)*, 2016, pp. 1-6.

[29] K. Sandkuhl, "Impact of Cyber-Physical System implementation on enterprise architectures: A case study," in *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, 2015, vol. 246, pp. 855-864.

[30] I. Schirmer, P. Drews, S. Saxe, U. Baldauf, and J. Tesse, "Extending enterprise architectures for adopting the internet of things – Lessons learned from the smartPORT projects in Hamburg," in *Lecture Notes in Business Information Processing* vol. 255, ed, 2016, pp. 169-180.

[31] T. Franck, M.-E. Iacob, M. van Sinderen, and A. Wombacher, "Towards an Integrated Architecture Model of Smart Manufacturing Enterprises," in *Business Modeling and Software Design*, Cham, 2018, pp. 112-133: Springer International Publishing.

[32] E. Moones, T. Vosgien, L. Kermad, E. M. Dafaoui, A. El Mhamedi, and N. Figay, "Plm standards modelling for enterprise interoperability: A manufacturing case study for erp and mes systems integration based on isa-95," in *Lecture Notes in Business Information Processing* vol. 213, ed, 2015, pp. 157-170.