

The Stuttgart IT Architecture for Manufacturing

An Architecture for the Data-Driven Factory

Laura Kassner¹(✉), Christoph Gröger^{1,2}, Jan Königsberger¹, Eva Hoos¹,
Cornelia Kiefer¹, Christian Weber¹, Stefan Silcher^{1,3}, and Bernhard Mitschang¹

¹ Graduate School of Excellence Advanced Manufacturing Engineering,
University of Stuttgart, Nobelstraße 12, 70569 Stuttgart, Germany

{laura.kassner, jan.koenigsberger, eva.hoos,
cornelia.kiefer}@gsame.uni-stuttgart.de,

{christoph.groeger, christian.weber}@ipvs.uni-stuttgart.de

² Robert Bosch GmbH, Robert-Bosch-Platz 1, 70839 Gerlingen-Schillerhöhe,
Germany

³ eXXcellent solutions gmbh, Heßbrühlstraße 7, 70565 Stuttgart, Germany

Abstract. The global conditions for manufacturing are rapidly changing towards shorter product life cycles, more complexity and more turbulence. The manufacturing industry must meet the demands of this shifting environment and the increased global competition by ensuring high product quality, continuous improvement of processes and increasingly flexible organization. Technological developments towards smart manufacturing create big industrial data which needs to be leveraged for competitive advantages. We present a novel IT architecture for data-driven manufacturing, the Stuttgart IT Architecture for Manufacturing (SITAM). It addresses the weaknesses of traditional manufacturing IT by providing IT systems integration, holistic data analytics and mobile information provisioning. The SITAM surpasses competing reference architectures for smart manufacturing because it has a strong focus on analytics and mobile integration of human workers into the smart production environment and because it includes concrete recommendations for technologies to implement it, thus filling a granularity gap between conceptual and case-based architectures. To illustrate the benefits of the SITAM's prototypical implementation, we present an application scenario for value-added services in the automotive industry.

Keywords: IT architecture · Data analytics · Big data · Smart manufacturing · Industrie 4.0

1 Introduction

The manufacturing industry is changing under the influence of increased global competition: product life cycles become shorter, products and processes become more complex, production conditions become more turbulent. Manufacturing

companies can only succeed in this shifting environment if they ensure high product quality, continuous improvement of processes and flexible organizational structures [1].

Initiatives such as Industrie 4.0 [2] and Smart Manufacturing [3] promote the digitalization of manufacturing operations and the use of cyber-physical systems (CPS) [4] to enable the vision of decentralized, self-controlling, self-optimizing products and processes [5]. These developments are supported especially by the rise of the internet of things. Increasingly, large amounts of heterogeneous industrial data, that is, *big industrial data* [6], are created across the entire product life cycle. These data include both structured and unstructured portions, for instance, machine sensor data on the shop floor, product usage data, customer complaints data from social networks or failure reports written by service technicians. One central challenge in Industrie 4.0 is the exploitation of these data to extract valuable business insights and knowledge from them [7]. Sample fields of application for the exploitation of big industrial data are product design optimization, manufacturing execution and quality management.

The predominant manufacturing IT architecture in practice is the information pyramid of manufacturing [8] (see Fig. 1). It fails to enable comprehensive data exploitation because it has several limitations, as reported in [9]: (1) complex point-to-point integration of heterogeneous IT systems limits a flexible integration of new data sources; (2) strictly hierarchical aggregation of information prevents a holistic view for knowledge extraction; (3) isolated information provisioning for the manufacturing control level and the enterprise control level impedes employee integration on the factory shop floor.

This work is an extension of [9]. We build on the concept of the *data-driven factory* developed therein, which is recapitulated in Sect. 2. In this work, we put a stronger focus on the industry-near, use-case-driven IT architecture for the data-driven factory, the *Stuttgart IT Architecture for Manufacturing (SITAM)* which overcomes the insufficiencies of the traditional information pyramid of manufacturing, presented in Sect. 4. The SITAM enables *service-oriented integration*, *advanced analytics* as well as *mobile information provisioning*, which are central requirements of the data-driven factory in order to exploit big industrial data for competitive advantages. The general introduction of the architecture (Sect. 4) and the description of the prototype and application scenario (Sect. 6) are presented as in [9].

In extension of [9], we have added the following new contributions:

1. A detailed analysis of existing reference IT architectures for smart manufacturing and Industrie 4.0 in Sect. 3.
2. An analysis of existing technologies for the core layers and components of SITAM in Sect. 5.
3. A more elaborate evaluation of the SITAM architecture in comparison with existing reference architectures and with respect to available technologies as well as the use case in Sect. 7.

2 Motivation: A Data-Driven Factory for Leveraging Big Industrial Data

In this section, we first analyze the limitations of the traditional information pyramid of manufacturing with respect to big industrial data in Sect. 2.1, then present the concept of the data-driven factory [9] in Sect. 2.2. Further details on the data-driven factory can be found in [9].

2.1 Limitations of the Information Pyramid of Manufacturing

The information pyramid of manufacturing, also called the hierarchy model of manufacturing, represents the prevailing manufacturing IT architecture in practice [10]. It is used to structure data processing and IT systems in manufacturing companies and it is standardized in ISA 95 [8]. In a simplified version, the information pyramid is comprised of three hierarchical levels (see Fig. 1): the *enterprise control level* refers to all business-related activities and IT systems, such as enterprise resource planning (ERP) systems, the *manufacturing control level* focuses on manufacturing operations management especially with manufacturing execution systems (MES) and the *manufacturing level* refers to the machines and automation systems on the factory shop floor.

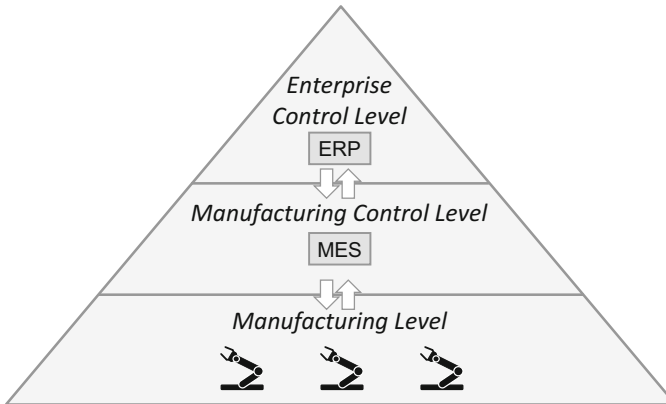


Fig. 1. Information pyramid of manufacturing [9].

Data processing in the information pyramid is based on three fundamental principles [10]:

- *Central automation* to control all activities top-down starting from the enterprise control level
- *Information aggregation* to condense all data bottom-up starting from the manufacturing level

- *System separation* to allow only IT systems at adjacent levels to directly communicate with each other.

The digitalization of manufacturing operations as well as the massive use of CPS lead to big industrial data, i.e., enormous amounts of heterogeneous industrial data at all levels of the information pyramid and across the entire product life cycle [6]. For instance, besides huge amounts of structured machine data and sensor data resulting from the shop floor, there are unstructured data on service reports and customer opinions in social networks. Exploiting these data, that is, extracting valuable business insights and knowledge, enables comprehensive optimization of products and processes [7]. For instance, customer satisfaction can be correlated with product design parameters using CAD data and CRM data or root causes of process quality issues can be analyzed using machine data and ERP data.

However, data processing according to the information pyramid of manufacturing prevents comprehensive data exploitation due to the following major technical limitations (L_i):

- L_1 : Central automation and system separation lead to a *complex and proprietary point-to-point integration of IT systems*, which significantly limits a flexible integration of new data sources across all hierarchy levels [11].
- L_2 : Strictly hierarchical information aggregation leads to *separated data islands* preventing a holistic view and strong analytics for knowledge extraction [6].
- L_3 : Central control and information aggregation lead to *isolated information provisioning* focusing on the manufacturing control level and the enterprise control level and thus impede employee integration through information provisioning on the manufacturing level [12].

To conclude, the function-oriented and strictly hierarchical levels of the information pyramid of manufacturing support a clear separation of concerns for the development and management of IT systems. However, the information pyramid lacks flexibility, holistic data integration and cross-hierarchical information provisioning. These factors significantly limit the exploitation of big industrial data and necessitate new manufacturing IT architectures, which are discussed in the following section.

2.2 The Data-Driven Factory

The data-driven factory [9] is a holistic concept to exploit big industrial data for competitive advantages of manufacturing companies. For this purpose, the data-driven factory addresses central economic challenges of today's manufacturing (Westkämper [1]), particularly agility, learning ability and employee orientation.

The data-driven factory takes a holistic view on all data generated across the entire product life cycle, including both structured data and unstructured data, i.e. data with a relational schema as well as text, audio, video and image

data without such a schema. In contrast to earlier integration approaches, especially Computer Integrated Manufacturing [13], the data-driven factory does *not* aim at totally automating all operations and decision processes but explicitly integrates employees in order to benefit from their knowledge, creativity and problem-solving skills.

From a manufacturing point of view, the data-driven factory is defined by the following core characteristics (see Fig. 2):

- The data-driven factory enables *agile manufacturing* (Westkämper [1]) by exploiting big industrial data for proactive optimization and agile adaption of activities.
- The data driven factory enables *learning manufacturing* [14] by exploiting big industrial data for continuous knowledge extraction.
- The data driven factory enables *human-centric manufacturing* [15] by exploiting big industrial data for context-aware information provisioning as well as knowledge integration of employees to keep the human in the loop.

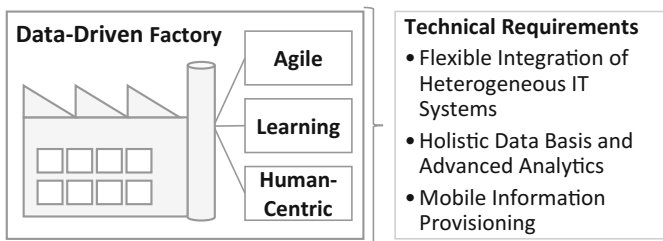


Fig. 2. Characteristics and technical requirements of the data-driven factory [9].

Based on the above characteristics and taking into account the limitations of the information pyramid of manufacturing (see Sect. 2.1), we have derived the following technical core requirements (R_i) for the realization of the data-driven factory (see Fig. 2):

- R_1 : *Flexible integration of heterogeneous IT systems* to rapidly include new data sources for *agile manufacturing*, e.g., when setting up a new machine
- R_2 : *Holistic data basis and advanced analytics* for knowledge extraction in *learning manufacturing*, e.g., to prescriptively extract action recommendation from both structured and unstructured data
- R_3 : *Mobile information provisioning* to ubiquitously integrate employees across all hierarchy levels for *human-centric manufacturing*, e.g., including service technicians in the field as well as product designers

In order to realize these requirements, a variety of IT concepts and technologies has to be systematically combined in an overall IT architecture. Since

the information pyramid of manufacturing lacks flexibility, holistic data integration and cross-hierarchical information provisioning (\mathbf{R}_1 - \mathbf{R}_3), we develop a novel manufacturing IT architecture that enables the data-driven factory.

The data-driven factory leverages big industrial data for agile, learning and human-centric manufacturing. In this way, it creates new potentials for competitive advantages for manufacturing companies, especially with respect to efficient and simultaneously agile processes, continuous and proactive improvement as well as the integration of knowledge and creativity of employees across the entire product life cycle.

3 Reference Architectures for Smart Manufacturing and Industrie 4.0

We did a comprehensive literature analysis on recent architectural approaches for IT-based manufacturing. An overview of recent reference architectures can be found in [16,17]. As result, we have identified three major groups of work:

- *Abstract frameworks for Industrie 4.0 and Smart Manufacturing*, which represent meta models and roadmaps for standardization issues, especially the Reference Architectural Model Industrie 4.0 (RAMI, [18]) as well as the SMLC framework for Smart Manufacturing [3].
- *Cross-domain-spanning reference architectures*, which also target the manufacturing industry, e.g. the Industrial Internet Reference Architecture (IIRA) [19] and the Industrial Data Space (IDS) [20].
- *Concrete manufacturing IT architectures*, which structure IT components and their relations in and across manufacturing companies on a conceptual level, especially Vogel-Heuser et al. [10], Minguez et al. [11], Holtewert et al. [21], Papazoglou et al. [22].

In the following we discuss the identified types of reference architectures and analyze them with respect to the technical core requirements identified in Sect. 2.2.

R_1 : *Flexible Integration of Heterogeneous IT Systems*

The above frameworks are defined on a significantly higher abstraction level than the information pyramid of manufacturing. Decomposed to its full structure, the pyramid contains a lot of additional hierarchical layers condensed to a reasonable minimum in the examined reference architectures [8]. The RAMI includes just a few layers of the equipment hierarchy model as dimensional perspective, which are extended by the layers product and connected world [18]. In [19], the IIC defines an Industrial Internet System by its technologies like manufacturing execution systems and programmable logical controllers which in turn are functionally related to the layers of the information pyramid of manufacturing. To solve data integration issues, they suggest a service-oriented architecture which includes a flexible method to combine services via metadata references at run-time to allow for dynamical composition in order to provide a real-time

response to changes in the environment [19]. The ZVEI [18] defines the concept of an administrative shell for I4.0 components, which contains a resource manager to expose services like OPC-UA¹ to other components. This administrative shell serves as digital representation for real-world assets on the shopfloor and allows integration with other administrative shells through a service-oriented architecture. To conclude, the common core of the above IT architectures is a service-oriented architecture (SOA) [23] in order to enable a flexible integration of IT systems – i.e. IT services – across all hierarchy levels [11, 21].

R₂: Holistic Data Basis and Advanced Analytics

In [10], the need for a common data model standardizing the interfaces and the data of the IT services is underlined. In [22], a knowledge repository is part of the architecture. In contrast, the IIC defines complex event processing and advanced analytics as part of multiple hierarchical layers of the industrial internet system to meet different processing requirements, e.g. edge analytics in close proximity to the place where the data is required for realtime processing. Integration techniques like syntactical and domain transformation are addressed, but not discussed in detail. [20] propose an industrial data space for the industry to exchange and integrate data across enterprise borders in a secure manner. For the integration, vocabulary and schema matching are used, as well as knowledge database management. However, the integration concepts of these reference architectures are very abstract and don't provide further details. A holistic data model or technique to integrate data is still missing.

R₃: Mobile Information Provisioning

In [20–22], a marketplace with IT services is proposed in addition. These services are offered via apps in [20, 21]. However, concrete approaches for displaying tailored information from integrated data sources on mobile devices to support the information needs of workers are not discussed, nor are the particular challenges of mobile data provisioning addressed.

Table 1 shows an overview of the existing IT architectures for Smart Manufacturing and Industrie 4.0 evaluated against the requirements of the data-driven factory. All in all, these existing manufacturing IT architectures mainly address the limitation of a complex and proprietary point-to-point integration of IT systems in the information pyramid of manufacturing and enable the flexible integration of heterogeneous IT systems (R₁) by defining a service-oriented architecture. At the moment, only the IIRA includes a holistic data basis and advanced analytics (R₂) to allow for knowledge extraction in learning manufacturing. However, they still lack mobile information provisioning (R₃) to address isolated information provisioning. Our concept of the data-driven factory and the SITAM architecture address all three limitations. The SITAM provides a detailed structure in order to serve as an implementation guideline and describes a holistic approach as detailed in the following sections.

¹ <https://opcfoundation.org/about/opc-technologies/opc-ua/>.

Table 1. Evaluation of IT architectures for smart manufacturing and industrie 4.0 against the requirements of the data-driven factory. (● fulfilled; ◐ partly fulfilled; ○ not fulfilled.)

	R ₁ integration	R ₂ analytics	R ₃ mobile
RAMI [18]	●	○	○
SMLC [3]	●	○	○
IIRA [19]	●	●	○
IDS [20]	◐	◐	◐
Vogel-Heuser et al. [10]	●	◐	○
Minguez et al. [11]	●	○	○
Holtewert et al. [21]	●	○	◐
Papazoglou et al. [22]	●	◐	○

4 SITAM: Stuttgart IT Architecture for Manufacturing

The SITAM architecture [9] is a conceptual IT architecture enabling manufacturing companies to realize and implement the data-driven factory. The architecture is based on the results and insights of several research projects we have undertaken in cooperation with various industry partners, particularly from the automotive and the machine construction industry.

In the following, we present an overview of the SITAM architecture in Sect. 4.1 and detail its components in Sects. 4.2–4.6.

4.1 Overview

The SITAM architecture (see Fig. 3) encompasses the entire product life cycle: *Processes*, *physical resources*, e.g., CPS and machines, *IT systems* as well as *web data sources* provide the foundation for several layers of abstracting and value-adding IT.

The *integration middleware* (see Sect. 4.2) encapsulates these foundations into services and provides corresponding data exchange formats as well as mediation and orchestration functionalities.

The *analytics middleware* (see Sect. 4.3) and the *mobile middleware* (see Sect. 4.4) build upon the integration middleware to provide predictive and prescriptive analytics for structured and unstructured data around the product life cycle and mobile interfaces for information provisioning.

Together, the three middlewares enable the *composition of value-added services* for both human users and machines (see Sect. 4.5). In particular, services can be composed ad-hoc and offered as mobile or desktop apps on an *app marketplace* to integrate human users, e.g., by a mobile manufacturing dashboard with prescriptive analytics for workers. The added value from these services feeds back into the product life cycle for continuous proactive improvement and adaptation.

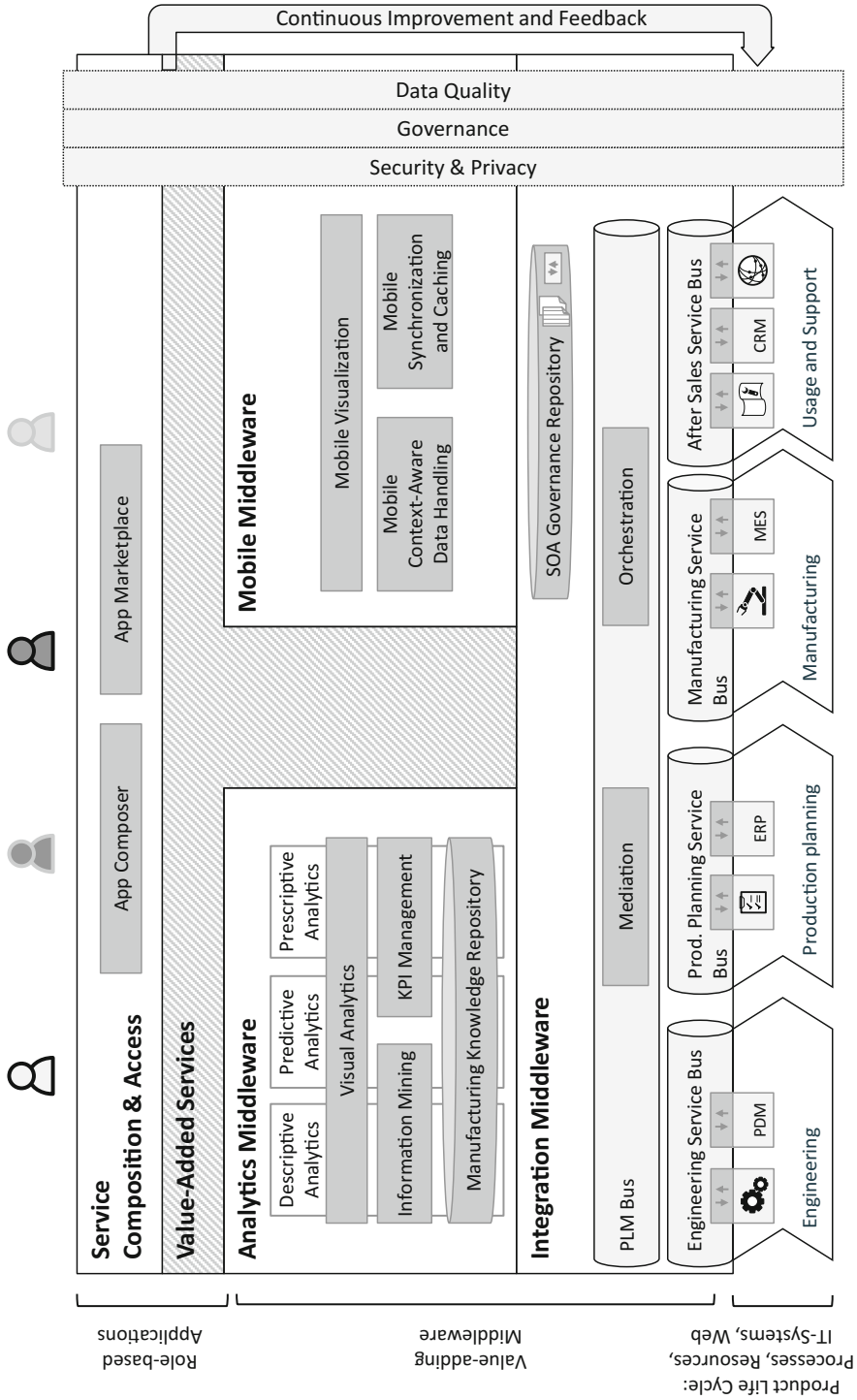


Fig. 3. Overview of the Stuttgart IT Architecture for Manufacturing (SITAM) [9].

Cross-architectural topics (see Sect. 4.6) represent overarching issues relevant for all components and comprise *data quality, governance* as well as *security and privacy*.

In the following, the components of the SITAM architecture are described in greater detail.

4.2 Integration Middleware: Service-Oriented Integration

The SITAM's *integration middleware* represents a changeable and adaptable integration approach which is based on the SOA paradigm [23]. The integration middleware is specifically tailored to manufacturing companies, providing the much needed flexibility and adaptability required in today's turbulent environment with a permanent need of change.

To enable those benefits, it builds on a concept of hierarchically arranged *Enterprise Service Buses* (ESBs) following [24]. Each one of these ESBs is responsible for the integration of all applications and services of a specific phase of the product life cycle.

All phase-specific ESBs are connected via a *superordinate Product-Lifecycle-Management-Bus* (PLM Bus). The PLM Bus is responsible for communication and mediation between phase-specific busses as well as for the orchestration of services.

This concept enables, for example, the easier integration of external suppliers without opening up too much of a company's internal IT systems to them by just "plugging" their own ESB into the PLM Bus. Besides, it also reduces the complexity by abstraction over the introduced integration hierarchy.

A dedicated sub-component providing real-time capabilities is used in the manufacturing phase to connect CPS and other real-time machine interfaces to the overall ESB compound.

The ESB hierarchy effectively abstracts and decouples technical systems and their services into a more business-oriented view, which we call *value-added services*. Value-added services use the basic services providing access to application data, orchestrate and combine them.

This decoupling also evens out different speeds in the development and change of applications or services. Companies often face the problem of having to integrate, e.g., legacy mainframe applications with modern mobile apps, which inherently have very different development speeds. By decoupling business-oriented services from the technical systems/services, each application can be developed separately and at its own pace, while the integration middleware handles all transformations and mediations that might be necessary to maintain compatibility.

Each phase-specific ESB also utilizes its own *phase-specific data exchange format* to handle the different requirements of each phase. For example, engineering has to be able to exchange large amounts of data, e.g., CAD models, whereas manufacturing requires the quick exchange of a large amount of smaller data chunks, e.g., MES production data. Aftersales on the other hand needs to

handle both large CAD data as well as small, lightweight data structures, e.g., live car data.

The separation into different phase-specific ESBs allows each department or business unit to make use of specialized data exchange formats tailored to phase-specific needs.

To sum up, the hierarchical composition of phase-specific ESBs across the entire product life cycle and the changeable service-oriented abstraction of IT systems address requirement R_1 (flexible integration of heterogeneous IT systems) of the data-driven factory.

4.3 Analytics Middleware: Advanced Analytics

The analytics middleware is service-oriented and comprises several manufacturing-specific analytics components which are crucial for a data-driven factory: The *manufacturing knowledge repository* for storing source data and analytics-derived insights, *information mining* on structured and unstructured data, *management of key performance indicators* (KPIs), and *visual analytics*. The analytics middleware includes functionalities for descriptive, predictive and prescriptive analytics, with prescriptive analytics being a novel introduction which provides actionable problem solutions or preventative measures before critical conditions lead to losses [25]. In providing *integrative, holistic* and *near-real time analytics* on big industrial data of all data types, the SITAM analytics middleware transcends the analytics capabilities of existing approaches (see Sect. 2). This significantly contributes to the learning and agile characteristics of the data-driven factory.

Source data are extracted using predefined ETL functions from the integration middleware. Integrated data of structured and unstructured type from around the product life cycle are stored in the *manufacturing knowledge repository* along the lines of [26] for maximum integration, minimum information loss and flexible access. Over the course of the product life cycle, this repository is enriched with various knowledge artefacts, e.g., analytics results like data mining models, business rules and free-form documents such as improvement suggestions. To store structured and unstructured source data in a scalable manner, the repository combines SQL and NoSQL storage concepts. It also includes the functionality for flexibly creating semantic links between source data and knowledge artefacts to support reasoning and knowledge management (see [26]).

The *information mining* component can be subdivided into classical data mining and machine learning tools for structured data on the one hand, and tools for various types of unstructured data – text, audio, video – on the other hand.

We will discuss text analytics [27] in more detail since its use in a framework for integrative data analytics is novel and since text data harbor a wealth of hitherto untapped knowledge. Typically, text analytics applications have been focused on one isolated unstructured data source and one analytical purpose, without integrating the results with analytics on structured data and with the disadvantage of information loss along the processing chain [28].

To secure flexibility of analytics and easy integration of data from different sources, we propose a set of basic and custom text analytics toolboxes, including domain-specific resources for the manufacturing and engineering domains and on an individual product domain level. This type of toolbox is similar to the generic and specific text analytics concepts proposed in [28]. Value-added applications of these text analytics tools fall into two main categories: (1) information extraction tasks and (2) direct support of human labor through partial automation. For example, presenting the top ten errors for a specific time span based on text in shop floor documentation is an information extraction task which helps workers gain insights into weaknesses of the production setup. Using features of text reports, for example occurrences of particular domain-specific keywords, to predict the likelihood of certain error codes which a human expert must manually assign to these text reports, constitutes an example of a direct support analytics task (see [29] for an implementation and proof of concept of this use case within the SITAM architecture).

Information mining can then be applied to discover knowledge, which is currently hidden in a combination of structured data and extracts from unstructured data. For example, process and machine data from the shop floor can be matched up with timestamps and extracted topics or relations from unstructured error reports to discover root causes for problems which have occurred. Real-time process data from the shop floor can be compared to historical data to discover indicators for problematic situations and prescribe measures for handling them, for example speeding up a machine when a delayed process has been discovered.

In order to constitute the backbone of a truly data-driven factory, information mining has to be conducted near real-time, on a variety of data sources as-needed, and manufacturing processes, sales, delivery, logistics and marketing campaigns have to adjust to meet the prescriptions derived from analytics results.

The *management of key performance indicators* is another important component and can be greatly improved by readily available and flexible analytics on a multitude of data sources. Instead of being an off-line process conducted by the executive layer based on aggregated reporting data, KPI management can become a continuous and pervasive process, as data analytics feedback loops are in place for all processes around the product life cycle and at any level of the process hierarchy.

Finally, the analytics middleware also includes *visual analytics* for data exploration through human analysts: This type of analytics mainly combines information mining and visualization techniques to present large data sets to human observers in an intuitive way, allowing them to make sense of the data beyond the capabilities of analytics algorithms. Thereby, visual analytics keep the human in the loop according to human-centric manufacturing.

Thus, the analytics capabilities of our reference architecture for the data-driven factory transcend those of related conceptual work in several aspects: (1) They include prescriptive, not just predictive or descriptive analytics, (2) they fully integrate structured and unstructured data beyond the manufacturing process, (3) they stretch across the entire product life cycle and provide a holistic

view as well as holistic data storage, and (4) they are decentralized yet integrative, since analytics services are combined as needed to answer questions or supervise processes and keep the human in the loop. Advanced analytics mostly contribute to the fulfillment of requirement R_2 , but also R_3 and R_1 of the data-driven factory.

4.4 Mobile Middleware: Mobile Information Provisioning

The mobile middleware enables mobile information provisioning and mobile data acquisition by facilitating the development and integration of manufacturing-specific mobile apps. Mobile apps [30] are running on smart mobile devices, such as smartphones, tablets, and wearables, and integrate humans into the data-driven factory. Due to their high mobility, workers on the shop floor have to have access to the services of the factory *anywhere and anytime*, e.g., viewing near real-time information or creating failure reports on-the-go, supported by the mobile devices' cameras and sensors. Workers can also actively participate in the manufacturing process, e.g., they can control the order in which products are produced. Furthermore, mobile apps offer an intuitive *task-oriented touch-based* design and enable users to consume only relevant data. Mobile devices also allow for the collection of new kinds of data, e.g., position data or photos. This enables new kinds of services such as context-aware apps and augmented-reality apps [31].

However, the development of mobile apps differs from the development of stationary applications due to screen sizes, varying mobile platforms, unstable network connections and other factors. In addition, manufacturing-specific challenges arise [31], e.g., due to the complex data structures as well as the high volume of data. In contrast to existing approaches (see Sect. 3), the mobile middleware addresses these manufacturing-specific needs.

The mobile middleware comprises three components: (1) mobile context-aware data handling, (2) mobile synchronization and caching as well as (3) mobile visualization.

The *mobile context-aware data handling* component provides manufacturing-specific context models describing context elements and relations, e.g., on the shop floor, as well as efficient data transfer mechanism so that only relevant data in the current context is transmitted to the mobile device. For instance, a shop floor worker specifically needs information on the current machine he is working at.

The *mobile synchronization and caching* component supports offline usage of mobile apps. This is important because a network connection cannot always be guaranteed, particularly on the factory shop floor. The component offers mechanisms to determine which data should be cached using context information provided by the context models.

The *mobile visualization* component provides tailored visualization schemas for manufacturing data, e.g., for CAD product models. For example, it provides a visualization schema to represent a hierarchical product structure and to browse

it via touch gestures. Various screen sizes and touch-based interaction styles are considered.

To sum up, the mobile middleware enables the integration of the human by supporting the development and integration of mobile apps. This is done by offering manufacturing-specific services for data handling and visualization. Thus, by addressing requirement R_3 (mobile information provisioning), the mobile middleware contributes to the human-centric characteristic of the data-driven factory, i.e., keeping the human in the loop.

4.5 Service Composition and Value-Added Services

The service-based and integrative nature of the SITAM architecture allows it to provide value-added services in several ways. We define *value-added services* as services which provide novel uses and thus create value by transcending the limitations of the information pyramid of manufacturing (see Sect. 2.1): By providing flexible interfaces for data and service provisioning (addressing limitation L_1), by integrating, analyzing and presenting data from several phases around the product life cycle (addressing limitation L_2) and by providing access to information in all the contexts in which it is needed and in which the traditional model may fail to do so (addressing limitation L_3). The value-added services offered in the SITAM architecture cut across the architectural layers, packaging and combining functionalities of the integration middleware, the analytics middleware and the mobile middleware.

In the SITAM architecture, services are composed and adapted *on the basis of user roles* and the information needs and permissions associated with them. For example, a shop floor worker receives detailed alerts related to the process step he is responsible for, whereas his production supervisor is concerned with the aggregated state of the entire manufacturing process across all process steps.

Ad-hoc *service composition* is enabled by the *app composer*. The app composer offers this functionality for users in all roles, regardless of their educational background or their ability to code. For example, data sources and analytics services can be mashed up and composed via drag-and-drop in a graphic user interface. Atomic or composed services can then be offered and distributed as apps in the *app marketplace* for all types of devices, both stationary and mobile.

Since there is very little in the way of dedicated service composition frameworks to build on, we are in the process of developing an implementation for clean and easy service composition, including a graphic user interface for non-technical users. We take inspiration from mashup platforms, such as [32], and app generator tools, such as [33].

To sum up, flexible service composition contributes to the fulfillment of requirement R_1 (flexible integration of heterogeneous IT systems) and the provisioning of composed services as mobile apps helps to fulfill requirement R_3 (mobile information provisioning) of the data-driven factory.

4.6 Cross-Architectural Topics

Security and privacy, governance and data quality are overarching topics which must be considered at all layers of the architecture: at the data sources, in analytics and mobile middleware as well as in the applications. In the following, we focus on *SOA governance* and *data quality* as they require specific concepts for the data-driven factory. For general security and privacy issues in data management, we refer the reader to [34].

The governance of complex service-oriented architectures is often neglected in existing manufacturing IT architectures, such as [22], even though a lack of governance is one of the main reasons for failing SOA initiatives [35].

SOA governance covers a wide range of aspects (a list of key aspects can be found in [36]). With more and more systems being integrated – especially CPS, but also for example social media services – it is becoming difficult to keep track of planned changes to those systems and services. For this reason, service change management and service life cycle management governance processes track and report those changes to service consumers and providers, governed for example via consumer and stakeholder management processes.

When setting up those governance processes, it is important to keep them as lightweight and unobtrusive as possible in order to minimize complexity and managerial effort. To support this, the SITAM architecture contains a central *SOA Governance Repository*, which is built on a specific SOA governance meta model described in [36]. The SOA Governance Repository contains service data as well as operations data, spanning and providing support during all phases of the service life cycle, and therefore also supporting novel software development concepts like DevOps.

Apart from SOA governance, the *need for high quality data* is a direct consequence of the concept of the data-driven factory. A data quality framework for the data-driven factory needs to enable data quality measurement and improvement (1) near-real-time (2) at all analysis steps from data source to user (3) for all types of data accumulating in the product life cycle, especially structured data as well as unstructured textual, video, audio and image data.

Existing data quality frameworks, e.g., [37,38], fail to satisfy these requirements. Hence, we translate these requirements into an extended data quality framework, which allows a flexible composition of data quality dimensions (e.g., timeliness, accuracy, relevance and interpretability) at all levels of the SITAM architecture (see [38] for an example list of data quality dimensions). Furthermore, we define sets of concrete indicators considering data consumers at all levels, from data source to user, and we allow for near real-time calculation of data quality (e.g., the confidence or accuracy of machine learning algorithms, language of text and speech, author of data sources and the distribution of data points on a timeline). This makes the quality of data and of resulting analytics results transparent at all levels and therefore enables holistic data quality improvement.

To sum up, we have seen that SOA governance and data quality are crucial factors across all layers of the SITAM architecture. A flexible composition of

IT systems and services can be offered using service-oriented architectures. But complex service-oriented architectures are prone to fail without systematic SOA governance. Besides, a holistic data quality framework forms the basis to measure and improve data quality from data source to user, including the generated analytics results.

5 Technologies for SITAM

In the following, we review technologies suited for the implementation of the SITAM architecture. We focus on the middleware components which provide the core functionalities of the SITAM and on data quality controls as an important prerequisite for data-driven manufacturing with strong analytics. We first address technologies for each of the middlewares in Sect. 5.1 for integration, Sect. 5.2 for analytics and Sect. 5.3 for mobile. We then discuss technologies for assessing data quality as a central cross-architectural topic in Sect. 5.4.

5.1 Integration

The *integration middleware* layer of the SITAM consists of several components: (1) The service bus hierarchy, (2) mediation components for communication between the different life cycle phases, (3) an orchestration component and (4) the SOA Governance Repository. This chapter presents possible technologies to implement these components.

The goal of the service bus hierarchy is to structure the services into multiple, phase-specific integration environments, tailored to the respective needs of the phase (cf. Sect. 4.2). To realize this, basically any off-the-shelf Enterprise Service Bus can be used as they all provide the necessary functionalities. Options range from proprietary products such as IBM's Integration Bus² or the Oracle Service Bus³ to open source alternatives such as the WSO2 Service Bus⁴).

Communication between the services as well as the different phases can be realized with a number of standardized technologies and protocols like SOAP over HTTP, SOAP over Message Queue or REST. For machine-to-machine communication, protocols like OPC-UA⁵ and MQTT⁶ exist, which also support real-time data exchange.

The phase-specific data exchange formats introduced in Sect. 4.2 can be defined in a protocol-independent format which then can be translated into different representations, e.g. XML Schema⁷ or JSON Schema⁸. These exchange

² <http://www-03.ibm.com/software/products/en/ibm-integration-bus>.

³ <http://www.oracle.com/technetwork/middleware/service-bus/overview/index.html>.

⁴ <http://wso2.com/products/enterprise-service-bus/>.

⁵ <https://opcfoundation.org/developer-tools/specifications-unified-architecture>.

⁶ <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html>.

⁷ <https://www.w3.org/TR/xmlschema11-1/>.

⁸ <http://tools.ietf.org/html/draft-zyp-json-schema-04>.

formats can also rely on existing definition formats like STEP [39] or JT Open [40], which are both used for the exchange of CAD data. The mediation component guaranteeing the communication between different life cycle phases can be realized programmatically as deployable Java artifacts or, in case XML-based formats are used, via Extensible Stylesheet Language Transformations (XSLT⁹).

To orchestrate atomic services into value-added services, workflows described using the Business Process Execution Language (BPEL¹⁰) can be used. If the composite service requires advanced logic, a workflow can be combined with additional program logic.

The SOA repository helps to manage all services and SOA artifacts across the complete product life cycle. There are several products available, among others IBM's WebSphere Service Registry and Repository¹¹ or WSO2's Governance Registry¹². Unfortunately, existing products don't fulfill the requirements for a comprehensive SOA governance approach [36], which necessitates the development of a custom-tailored solution. As database backend a traditional relational database system can be used as well as a NewSQL database [41] or a triple store. The repository could be implemented either as a client/server or as a web-based system.

5.2 Analytics

The analytics layer of the SITAM requires a number of technical components. These components are typically organized in an integrated analytical tool stack. For such stacks, the Lambda Architecture [42] is becoming the de-facto standard in industry practice for scalable and robust analytical tool stacks and therefore represents the basis for an implementation of the analytics layer of SITAM. The Lambda Architecture mainly differentiates between components for batch data processing to store and analyze historic data in-depth with rather high latency and components for stream data processing for near-real-time data analysis of current data. In the following, we briefly describe the application of the Lambda Architecture to realize SITAMs analytics layer and highlight major tools.

With respect to batch data processing, the basis is a data lake approach on top of Hadoop¹³ to implement the manufacturing knowledge repository, with structured, unstructured and semi-structured portions and semantic linking between related data. An alternative option would be a combination of a relational database and a NoSQL system, e.g. a content management system, for scenarios which do not need a massive scale-out. In both cases, semantic relations can be implemented either as relational or as NoSQL links, e.g., using a graph-based approach (see [26]).

Considering the SITAM's components for information mining, KPI management and visual analytics, the Apache data processing family also provides

⁹ <https://www.w3.org/TR/xslt>.

¹⁰ <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>.

¹¹ <http://www-03.ibm.com/software/products/en/wsrr>.

¹² <http://wso2.com/products/governance-registry/>.

¹³ <https://hadoop.apache.org/>.

libraries for scalable batch machine learning and data mining, e.g., with Apache Mahout¹⁴ and SparkR¹⁵. Further, there are a number of free and commercial data mining toolkits and libraries for structured data analytics and reporting, some of which also include libraries for preprocessing unstructured text data. Toolkits such as WEKA¹⁶, KNIME¹⁷ or RapidMiner¹⁸ offer graphical interfaces for rapid data exploration and prototypical analytics design. Only some of them also allow integration into custom applications.

Apart from various linguistic preprocessing tasks which are already integrated into structured data mining libraries, there exist several dedicated frameworks for text analytics. GATE, the General Architecture for Text Engineering¹⁹, and Apache UIMA²⁰, the Unstructured Information Management Architecture are both widely used in research and industry projects and provide capabilities for building full text processing pipelines with all processing steps, from reading in data sources, through standard preprocessing steps and custom-built analytics components, to outputting results in various data formats.

These batch components are complemented by stream processing components for near-real-time analytics. Tools for this include various options from the open source world focusing on massively scalable processing of data streams, e.g., Apache Spark Streaming²¹ or Apache Storm²². In addition, there are classical commercial stream data processing platforms, e.g. IBM InfoSphere Streams²³ or Oracle Streams²⁴, which typically provide more enhanced functions for analyzing data streams but lack scalability in comparison with their open source counterparts.

5.3 Mobile

Different technologies are available for mobile visualization, context-aware data provisioning and mobile synchronization.

With respect to mobile visualization, we distinguish between native and web app development of mobile apps. Native apps are developed for a specific mobile platform such as iOS²⁵ or Android²⁶. There are libraries and frameworks for native apps to support and facilitate the development of user interfaces. The

¹⁴ <http://mahout.apache.org/>.

¹⁵ <https://spark.apache.org/docs/latest/sparkr.html>.

¹⁶ <http://www.cs.waikato.ac.nz/ml/weka/>.

¹⁷ <https://www.knime.org/>.

¹⁸ <http://rapidminer.com/>.

¹⁹ <https://gate.ac.uk/>.

²⁰ <http://uima.apache.org/>.

²¹ <https://spark.apache.org/streaming/>.

²² <https://storm.apache.org/>.

²³ <http://www-03.ibm.com/software/products/en/ibm-streams>.

²⁴ <http://www.oracle.com/technetwork/testcontent/streams-oracle-streams-twp-128298.pdf>.

²⁵ <https://developer.apple.com/>.

²⁶ <https://developer.android.com/index.html>.

main purpose is to provide uniform user interface and interaction design according to their style guide. However, they are often limited to standard visualization such as lists and menu bars. More complex visualizations have to be developed individually for the respective use cases. There are also lots of frameworks supporting the development of web apps. They are not restricted to any style guide and can be used to develop responsive design which fits multiple devices. Popular frameworks, especially for mobile usage, are angular.js²⁷ and jQuery²⁸. Complex visualization for web apps can be supported by dedicated frameworks for complex visualization such as d3.js²⁹ or rapid.js³⁰.

Context-aware data provisioning requires the management of context data and store them into a context model. There are several different approaches to model context based on key-value, logic-based, ontology, rule-based, or graphical model [43]. A review of context models to support context-aware provisioning can be found in [44].

Mobile synchronization requires local storage on the mobile devices. For native apps, light-weight databases such as SQLITE³¹ can be used. For web apps, HTML5 provides local storage in key-value format³². For example, the chrome browser provides the Index db API to manage local offline storage³³.

5.4 Data Quality

For the implementation of the data quality layer, technologies which allow the measurement and improvement of structured as well as unstructured data are needed. Many commercial toolkits dedicated to the quality of structured data exist, but open source toolkits are rare (e.g. the DuDe tool³⁴ for duplicate detection and OpenRefine³⁵, a tool for cleaning and transforming structured data). Neither open source nor commercial toolkits are available for unstructured data. In Sect. 4.6 we mentioned concrete indicators for the quality of structured and unstructured data, such as the confidence of machine learning algorithms, language of text and speech and the distribution of data points on a timeline. Here, we provide concrete technologies which can be used to measure data quality based on these indicators. The *confidence* of the tools in the natural language processing library OpenNLP³⁶ can be retrieved for each classification decision. For automatic detection of the *language of texts*, e.g., LibTextCat³⁷, which is

²⁷ <https://angularjs.org/>.

²⁸ <https://jquery.com/>.

²⁹ <https://d3js.org/>.

³⁰ <http://jointjs.com/>.

³¹ <https://www.sqlite.org/>.

³² http://www.w3schools.com/html/html5_webstorage.asp.

³³ https://developer.chrome.com/apps/offline_storage.

³⁴ <http://hpi.de/naumann/projects/data-quality-and-cleansing/dude-duplicate-detection.html>.

³⁵ <http://openrefine.org/>.

³⁶ <https://opennlp.apache.org/>.

³⁷ <http://software.wise-guys.nl/libtextcat/>.

a C library, or the associated versions in other programming languages such as Java or Python can be used. *Outliers* are a well-studied task in the field of data quality and can be detected for example using the programming language R and the Rlof package³⁸.

6 Prototype and Application

In the following, we present current work on the realization of the SITAM architecture in a prototypical implementation in Sect. 6.1. Moreover, we introduce a real-world application scenario from the automotive industry using the SITAM architecture in Sect. 6.2 in order to illustrate its benefits for a number of value-added services.

6.1 Prototypical Implementation

Our current prototype covers core components in every layer of the SITAM architecture, in particular with respect to analytics, governance, mobile and repository aspects. In the following, we sketch major solution details and technologies we utilized. The latter were chosen from the large available pool of free and open source software to underline the broad applicability of the SITAM architecture and make the implementation easily adaptable to various industrial real-world settings.

The *integration middleware* relies on WSO2's Enterprise Service Bus, to realize the hierarchical ESB structure as well as the orchestration of basic services and mediation between phase-specific ESBs as described in [24]. As all interfaces are based on standards, the ESB hierarchy can also be heterogeneous, allowing to select different products from different vendors that might better support certain phase-specific requirements. Services within the prototype are implemented as either conventional SOAP web services or REST services. Data exchange formats are described as XSD documents and stored in the SOA Governance Repository. The repository itself relies on semantic web technologies, mainly the Resource Description Framework (RDF³⁹), and provides a web-accessible as well as a Web Service interface as described in [45]. The use of those technologies allows for example the use of semantic reasoning to detect new dependencies or missing information within the repository.

In the *analytics middleware*, the manufacturing knowledge repository is implemented as a federation of a relational database and a NoSQL system – we used the content management system Alfresco CMS⁴⁰ – to store structured and unstructured data. These systems are integrated by a specific link store using a graph database such as Neo4j⁴¹. The information mining component includes

³⁸ <https://cran.r-project.org/web/packages/Rlof/>.

³⁹ <https://www.w3.org/TR/rdf11-concepts/>.

⁴⁰ <https://www.alfresco.com/>.

⁴¹ <https://neo4j.com/>.

tools from the Apache UIMA framework⁴² for unstructured data analytics, with the uimaFit extension⁴³ for on-the-fly analytics service composition. Structured data mining capabilities are taken from the WEKA data mining workbench⁴⁴. On this basis, manufacturing-specific predictive and prescriptive analytics are realized using various data mining techniques, especially decision tree induction and text categorization, as described in [26, 29, 46], respectively.

Regarding the *mobile middleware*, we implemented several *mobile apps*, e.g., a mobile analytics dashboard for shop floor workers [26] and a mobile product structure visualizer for engineers. We have implemented native apps for Android and for Windows as well as platform independent web apps using standardized web technology such as HTML5.

An *app marketplace* and a graphical interface for intuitive access to the *app composer* are currently under development.

6.2 Use Case: Quality Management and Process Optimization in the Automotive Industry

To demonstrate the concept of the data-driven factory as well as the SITAM architecture, we have cooperated with an OEM to develop a real-world application scenario for the automotive industry. The scenario focuses on quality management and process optimization as critical success factors for OEMs especially in the automotive premium segment. An overview of all involved components and participants can be seen in Fig. 4.

An automotive manufacturer collects *big industrial data*, including structured sales and machine data, sensor and text data around the product life cycle. These data originally reside in isolated databases; for instance, text reports about product and part quality from development, production and aftersales are all gathered via different IT systems. To ensure a realistic representation of source data and processes, on the one hand, we take advantage of publicly available data sources, such as the records of automotive complaints covering the US market and maintained by the NHTSA⁴⁵. On the other hand, we make use of anonymized data and internal knowledge resources of our industry partner.

On this basis, the SITAM architecture is applied to exploit these data for quality management and process optimization. In the following, we give an overview of representative value-added services and role-based apps across the product life cycle which are enabled by the SITAM architecture (see Fig. 4). We focus on car paint quality as a recurring example (all data samples in the following are fictitious for reasons of confidentiality).

During product development and testing, quality data are collected through the mobile *dev Q app* by engineers and test drivers on the go, including text reports and image material. The *aftersales Q app* is used to collect aftersales

⁴² <http://uima.apache.org/>.

⁴³ <https://code.google.com/archive/p/uimafit/>.

⁴⁴ <http://www.cs.waikato.ac.nz/ml/weka/>.

⁴⁵ <http://www.nhtsa.gov/NCSA>.

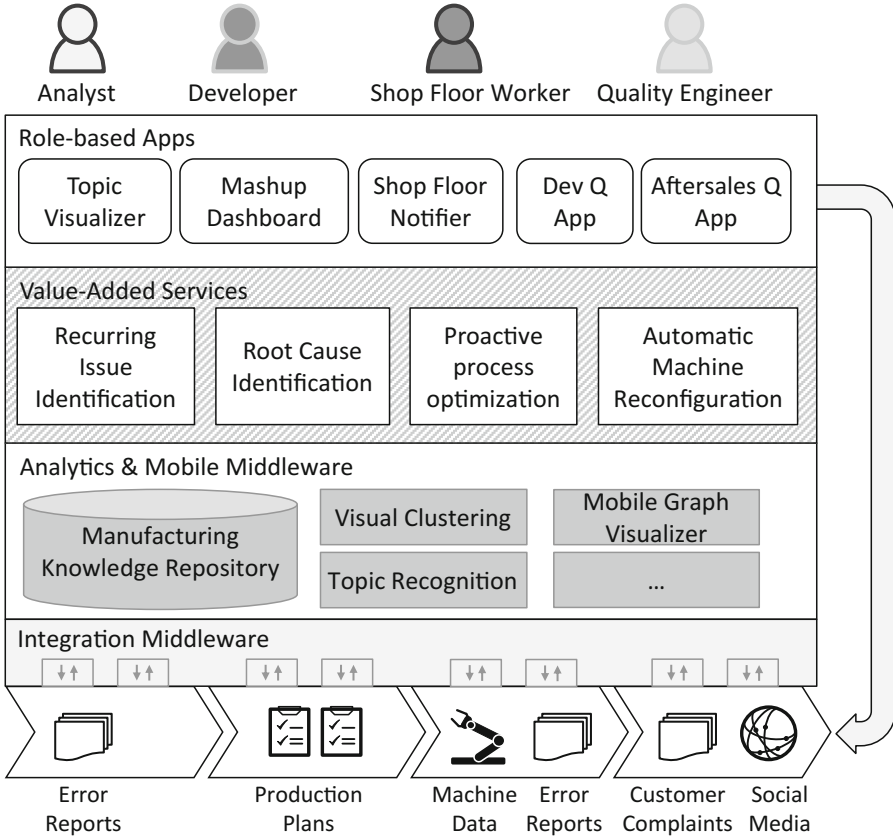


Fig. 4. Value-added services and role-based apps in the application scenario [9].

quality data for the warranty and recovery process of damaged car parts in the form of unstructured text reports (e.g., “customer states that car paint is coming off after washing”, “flaking paint on fender during extreme summer heat”). It has different profiles for quality engineers (whose primary task is the definition of new error codes), for quality expert workers (whose task it is to assign error codes to damaged parts) and for executives (who are interested in comparing aggregated error code data over time). In addition, quality data come in the form of *customer complaints* and via *social media* crawling services.

After aggregating these data into the manufacturing knowledge repository via the integration middleware, *topic recognition* on the text data is performed as an information mining step. The topics (e.g., “paint flaking – heat”, “paint damage – washing”) are presented to a human *analyst* via *visual clustering* to pick the most pressing ones or perform minor reclassification. This constitutes a value-added service of *recurring issue identification* and is performed via the

topic visualizer app, which makes use of the *mobile graph visualizer* from the mobile middleware.

Next, the problem topics are combined with historical data from the production phase, especially machine data, shop floor environment data, and structured error counts for *root cause identification* (e.g., elevated humidity in the paint shop leading to a lower quality of paint and a higher risk of flaking when exposed to harsh environmental conditions). This analytics step is executed in an analytics and data *mashup dashboard* app, where data sources and analytics algorithms are combined ad-hoc, but can also be stored for recurring use.

Identified root causes and condition patterns serve as input for *proactive process optimization*. It makes use of prescriptive analytics to automatically identify potentially problematic situations (e.g., critical humidity in paint shops) during process execution and recommend actions to on-duty workers through a *shop floor notifier* app (e.g., to air the paint shops to decrease humidity) or trigger *automatic machine reconfiguration* (e.g., increasing air conditioning and heating to decrease humidity).

7 Evaluation and Benefits

This section evaluates the benefits of the SITAM architecture with respect to the requirements of the data-driven factory as well as in contrast to the reference architectures described in Sect. 3.

The application scenario from Sect. 6.2 allows us to analyze the fulfillment of the technical requirements of the data-driven factory and contrast it with the traditional information pyramid of manufacturing.

In the scenario, diverse systems across the product life cycle, such as machines, social media sources as well as sensors, are encapsulated as services and are uniformly represented in the SOA governance repository to ease integration and access in the integration middleware. By this service-oriented abstraction, the SITAM architecture enables a flexible integration of heterogeneous data sources as well as a flexible service composition fulfilling requirement R_1 . This enables *agile manufacturing*, the first characteristic of the data-driven factory. Accessible service-based and role-based information provisioning also works towards keeping the human in the loop (*human-centric manufacturing*).

To merge structured and unstructured data from different life cycle phases, e.g., aftersales quality data and machine data in the application scenario, all data are integrated in the manufacturing knowledge repository of the analytics middleware. Moreover, predictive and prescriptive analytics are used to derive action recommendations for process optimization according to the application scenario. Thus, the SITAM architecture provides a holistic data basis encompassing the product life cycle as well as advanced analytics for knowledge extraction fulfilling requirement R_2 . This analytics capability provides functionalities for *learning manufacturing*, such as learned improvements for the quality-optimal design of both processes and products. It also is a prerequisite for agile process adaptations (*agile manufacturing*), such as the near real-time adaptation of production conditions to prevent known product quality issues.

In the application scenario, various mobile apps support seamless integration of employees, e.g., for data acquisition by test drivers using the dev Q app or for notifications of shop floor workers using the shop floor notifier. The mobile middleware facilitates the development of such manufacturing-specific apps using predefined manufacturing context models as well as specific visualization components, especially for product models. These apps can be easily deployed on various devices using the app marketplace. In this way, the SITAM architecture enables mobile information provisioning and fulfills requirement R_3 of the data-driven factory to ubiquitously integrate employees across all hierarchy levels. Thus, it provides the framework for *human-centric manufacturing* in keeping the human expert in the loop through data provisioning and data gathering.

The SITAM architecture thus enables flexible system and data integration, advanced analytics and mobile information provisioning and thus fulfills all technical requirements (R_1 – R_3) of the data-driven factory.

Table 2. Comparison of the SITAM to types of IT architectures for smart manufacturing and Industrie 4.0 (● fulfilled; ◐ partly fulfilled; ○ not fulfilled.)

	Abstract frameworks for Industrie 4.0 and Smart Manufacturing	Cross-domain-spanning reference architectures	Stuttgart IT Architecture for Manufacturing	Concrete manufacturing IT architectures
Integration (R_1)	●	◐	●	●
Analytics (R_2)	○	◐	●	◐
Mobile (R_3)	○	◐	●	◐
Granularity	—————→			
Reference architecture	●	●	●	○
Concrete implementation	○	○	●	●

Table 2 shows the evaluation of the three groups of architectures described in Sect. 3 and the SITAM against the three requirements of the data-driven factory as well as in terms of granularity and concreteness. We find that the SITAM fills an important granularity gap: It provides both a full reference architecture and concrete recommendations for implementation. We have also included a discussion of technologies suited for the realization of its individual components in Sect. 5 where we point out which technologies already exist and which need to be further developed before they can be used in an industry context. In contrast, the abstract frameworks and the cross-domain-spanning reference architecture provide only the reference architecture and the concrete architectures provide only implementation details. None of the other architectures for smart manufacturing fulfills all requirements of the data-driven factory or addresses all

the limitations of the information pyramid of manufacturing. Most notably, the SITAM excels over other architectures in its capability to keep the human in the loop, particularly in three areas: (1) data integration, where the hierarchy of ESBs provides maximum flexibility for including and accessing data sources as needed; (2) analytics with its particular focus on including unstructured data sources and visualizing intermediate results; and (3) mobile with its enormous impact on tailored data provisioning and active human participation.

8 Conclusion and Future Work

In this article, we have presented in detail the Stuttgart IT Architecture for Manufacturing (SITAM) [9] which (1) flexibly integrates heterogeneous IT systems, (2) provides holistic data storage and advanced analytics covering the entire product life cycle, and (3) enables mobile information provisioning to empower human workers as active participants in manufacturing. We have given an overview of technologies which are required for the implementation of the SITAM and pointed out concrete examples of infrastructures and toolboxes which can be used, as well as identified gaps in the technology landscape where more work is needed. We have compared the SITAM against major reference architectures for smart manufacturing and Industrie 4.0 and found that it surpasses them in several points, the most important ones being the integration of the human worker and the concrete technological recommendations.

We have prototypically implemented core components of the SITAM architecture in the context of a real-world application scenario concerned with quality and process management in the automotive industry. Our conceptual evaluation shows that the SITAM architecture enables the realization of the data-driven factory and the exploitation of big industrial data across the entire product life cycle.

Acknowledgements. The authors would like to thank the German Research Foundation (DFG) as well as Daimler AG for financial support of this project as part of the Graduate School of Excellence advanced Manufacturing Engineering (GSaME) at the University of Stuttgart.

References

1. Westkämper, E.: Towards the Re-industrialization of Europe: A Concept for Manufacturing for 2030. Springer Science & Business Media, Heidelberg (2013)
2. MacDougall, W.: Industrie 4.0: Smart Manufacturing for the Future (2014)
3. Davis, J., Edgar, T., Porter, J., Bernaden, J., Sarli, M.: Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Comput. Chem. Eng.* **47**, 145–156 (2012)
4. Shi, J., Wan, J., Yan, H., Suo, H.: A survey of cyber-physical systems. In: 2011 International Conference on Wireless Communications and Signal Processing, Piscataway, NJ, pp. 1–6. IEEE (2011)

5. Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. *Int. J. Sci. Eng. Technol.* **8**, 37–44 (2014)
6. Kemper, H.G., Baars, H., Lasi, H.: An integrated business intelligence framework. closing the gap between IT support for management and for production. In: Rausch, P., Sheta, A.F., Ayesh, A. (eds.) *Business Intelligence and Performance Management. Advanced Information and Knowledge Processing*, pp. 13–26. Springer, London (2013). doi:[10.1007/978-1-4471-4866-1_2](https://doi.org/10.1007/978-1-4471-4866-1_2)
7. Gölzer, P., Cato, P., Amberg, M.: Data processing requirements of industry 4.0 - use cases for big data applications. In: *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, Paper 61 (2015)
8. ISA: Enterprise-Control System Integration. ANSI/ISA 95.00.01-2000, Instrument Society of America (2000)
9. Gröger, C., Kassner, L., Hoos, E., Königsberger, J., Kiefer, C., Silcher, S., Mitschang, B.: The data-driven factory. Agile, learning and human-centric manufacturing. In: *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS)*, Scitepress (2016)
10. Vogel-Heuser, B., Kegel, G., Bender, K., Wucherer, K.: Global information architecture for industrial automation. *Automatisierungstechnische Praxis* **51**, 108–115 (2009)
11. Minguez, J., Lucke, D., Jakob, M., Constantinescu, C., Mitschang, B.: Introducing SOA into production environments - the manufacturing service bus. In: Sihn, W., Becker, T., Kolev, I. (eds.) *Proceedings of the 43rd CIRP International Conference on Manufacturing Systems (CMS)*, pp. 1117–1124. Neuer Wissenschaftlicher Verlag, Wien (2010)
12. Bracht, U., Hackenberg, W., Bierwirth, T.: A monitoring approach for the operative CKD logistics. *wt Werkstattstechnik Online* **101**, 122–127 (2011)
13. Groover, M.P.: *Automation, Production Systems, and Computer-Integrated Manufacturing*, 3rd edn. Prentice Hall, Upper Saddle River (2008)
14. Hjelmervik, O.R., Wang, K.: Knowledge management in manufacturing: the soft side of knowledge systems. In: Wang, K., Kovacs, G.L., Wozny, M., Fang, M. (eds.) *PROLAMAT 2006. IIFIP*, vol. 207, pp. 89–94. Springer, Boston (2006). doi:[10.1007/0-387-34403-9_10](https://doi.org/10.1007/0-387-34403-9_10)
15. Zuehlke, D.: Smart factory - towards a factory-of-things. *Ann. Rev. Control* **34**, 129–138 (2010)
16. Weyrich, M., Ebert, C.: Reference architectures for the internet of things. *IEEE Softw.* **33**, 112–116 (2016)
17. EFFRA: Platforms for connected factories of the future. Technical report, Communications Networks, Content and Technology Directorate-General DG CONNECT, A3 and European Factories of the Future Research Association (EFFRA) (2015)
18. VDI/VDE, ZVEI: Reference Architecture Model Industrie 4.0 (RAMI4.0). Technical report, Plattform Industrie 4.0 (2015)
19. Lin, S.W., Miller, B., Durand, J., Joshi, R., Didier, P., Chigani, A., Torenbeek, R., Duggal, D., Martin, R., Bleakley, G., King, A., Molina, J., Schrecker, S., Lembree, R., Soroush, H., Garbis, J., Crawford, M., Harper, E., Raman, K., Witten, B.: Industrial internet reference architecture. Technical report 1.7, Industrial Internet Consortium (2015)
20. Otto, B., Auer, S., Cirullies, J., Jürjens, J., Menz, N., Schon, J., Wenzel, S.: Industrial Data Space - Digitale Souveränität für Daten. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Technical report (2016)

21. Holtewert, P., Wutzke, R., Seidelmann, J., Bauernhansl, T.: Virtual fort knox - federative, secure and cloud-based platform for manufacturing. In: Cunha, P.F.D.C. (ed.) *Economic Development and Wealth through Globally Competitive Manufacturing Systems*, Procedia CIRP., Red Hook, NY, Curran, vol. 7, pp. 527–532 (2014)
22. Papazoglou, M.P., van den Heuvel, W.J., Mascolo, J.E.: A reference architecture and knowledge-based structures for smart manufacturing networks. *IEEE Softw.* **32**, 61–69 (2015)
23. Erl, T.: *Service Oriented Architecture: Principles of Service Design*. The Prentice Hall Service-oriented Computing Series from Thomas Erl. Prentice Hall, Upper Saddle River (2008)
24. Silcher, S., Dinkelmann, M., Minguez, J., Mitschang, B.: Advanced product life-cycle management by introducing domain-specific service buses. In: Cordeiro, J., Maciaszek, L.A., Filipe, J. (eds.) *ICEIS 2012. LNBIP*, vol. 141, pp. 92–107. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40654-6_6](https://doi.org/10.1007/978-3-642-40654-6_6)
25. Evans, J.R., Lindner, C.H.: Business analytics: the next frontier for decision sciences. *Decis. Line* **43**, 4–6 (2012)
26. Gröger, C., Schwarz, H., Mitschang, B.: The manufacturing knowledge repository. consolidating knowledge to enable holistic process knowledge management in manufacturing. In: Hammoudi, S. (ed.) *Proceedings of the 16th International Conference on Enterprise Information Systems*, Lisbon, Portugal, 39–51 April 2014, pp. 27–30. [S.l.], SciTePress (2014)
27. Aggarwal, C.C., Zhai, C.: An introduction to text mining. In: Aggarwal, C.C., Zhai, C. (eds.) *Mining Text Data*, pp. 1–10. Springer, New York (2012)
28. Kassner, L., Gröger, C., Mitschang, B., Westkämper, E.: Product life cycle analytics - next generation data analytics on structured and unstructured data. In: Teti, R. (ed.) *9th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRPICME 14*, Procedia CIRP., Red Hook, NY, Curran, vol. 33, pp. 35–40 (2015)
29. Kassner, L., Mitschang, B.: Exploring text classification for messy data: an industry use case for domain-specific analytics. In: Pitoura, E., Maabout, S., Koutrika, G., Marian, A., Tanca, L., Manolescu, I., Stefanidis, K. (eds.) *Proceedings of the 19th International Conference on Extending Database Technology (EDBT)*, Open-Proceedings.org, pp. 491–502 (2016)
30. Clevenger, N.C.: *IPad in the Enterprise: Developing and Deploying Business Applications*. Wiley, Indianapolis (2011)
31. Hoos, E., Gröger, C., Mitschang, B.: Mobile apps in engineering: a process-driven analysis of business potentials and technical challenges. In: Teti, R. (ed.) *9th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRPICME 14*, Procedia CIRP., Red Hook, NY, Curran, vol. 33 (2015)
32. Daniel, F., Matera, M.: *Mashups: Concepts, Models and Architectures*. Data-Centric Systems and Applications. Springer, Heidelberg (2014)
33. Francese, R., Risi, M., Tortora, G., Tucci, M.: Visual mobile computing for mobile end-users. *IEEE Trans. Mob. Comput.* **15**, 1033–1046 (2015)
34. Whitman, M.E., Mattord, H.J.: *Principles of Information Security*, 3rd edn. Thomson Course Technology, Boston (2007)
35. Meehan, M.: SOA adoption marked by broad failure and wild success (2008)
36. Königsberger, J., Silcher, S., Mitschang, B.: SOA-GovMM: a meta model for a comprehensive SOA governance repository. In: Joshi, J.B.D. (ed.) *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration*, Piscataway, NJ, pp. 187–194. IEEE (2014)

37. Sebastian-Coleman, L.: *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework*. Elsevier Science, Burlington (2013)
38. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. *J. Manag. Inf. Syst.* **12**, 5–33 (1996)
39. International Organization for Standardization: *Industrial Automation Systems and Integration* (1994)
40. International Organization for Standardization: *Industrial Automation Systems and Integration - JT File Format Specification for 3D Visualization* (2016)
41. Stonebraker, M.: Newsq: an alternative to NoSql and old SQL for new OLTP apps. *Commun. ACM* (2011)
42. Marz, N., Warren, J.: *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Manning Publications Co., Greenwich (2015)
43. Nalepa, G., Bobek, S.: Rule-based solution for context-aware reasoning on mobile devices. *Comput. Sci. Inf. Syst.* **11**, 171–193 (2014)
44. Bolchini, C., Curino, C.A., Quintarelli, E., Schreiber, F.A., Tanca, L.: A data-oriented survey of context models. *ACM SIGMOD Rec.* **36**, 19 (2007)
45. Königsberger, J., Mitschang, B.: A semantically-enabled SOA governance repository. In: *Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration*. IEEE (2016)
46. Gröger, C., Schwarz, H., Mitschang, B.: Prescriptive analytics for recommendation-based business process optimization. In: Abramowicz, W., Kokkinaki, A. (eds.) *BIS 2014. LNBIP*, vol. 176, pp. 25–37. Springer, Cham (2014). doi:[10.1007/978-3-319-06695-0_3](https://doi.org/10.1007/978-3-319-06695-0_3)