Closing the Gap between Smart Manufacturing Applications and Data Management

Emanuel Marx, Matthias Stierle, Sven Weinzierl, Martin Matzner

Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Digital Industrial Service Systems, Nürnberg, Germany
{emanuel.marx, matthias.stierle, sven.weinzierl, martin.matzner}@fau.de

Abstract. Smart manufacturing refers to the intensified collaboration of machines, products, and people throughout the manufacturing and the supply chain. This facilitates innovative products, services, business models, and processes. Smart manufacturing is premised on emerging technologies such as cloud computing, mobile computing, the Internet of Things, data analytics, and artificial intelligence. A plethora of companies struggles with the implementation of corresponding applications. In research and practice, we see general data management approaches with primary attention on building architectures that are not tailored to fit a particular domain/ application scenario. However, a robust data management concept is vital, as smart manufacturing decisively depends on data. To address this substantial deficit, we conduct a comprehensive literature review, an expert workshop, and semi-structured expert interviews with one of the leading German automotive manufacturers. The result is a catalog of requirements and a framework for data management that fosters the implementation of smart manufacturing applications.

Keywords: smart manufacturing, smart factory, data management, data analytics, expert interview

1 Introduction

The manufacturing industry is undergoing a paradigm shift, in which machines, products, and people tightly collaborate and are self-organized, enabling innovative products, services, business models, and processes [1–3]. New technologies such as cloud computing, mobile computing, internet of things, data analytics, and increasingly artificial intelligence (AI) facilitate this transition [4]. Aside from smart manufacturing, various synonymous terms such as smart factory, Industry 4.0, or additive manufacturing are popular [5]. Smart manufacturing applications, e.g. collaborative robots for safe human-machine interaction [6] combine increasingly available data spawned by manufacturing systems to create new value-added potentials [7]. As data (analytics) are vital resources for smart manufacturing, a robust data management concept is critical to foster smart manufacturing applications [8].

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

https://doi.org/10.30844/wi_2020_u1-marx

Researchers have developed various concepts for data management, mainly under the term of data warehouse [9] or data lakes [10]. Current publications also consider the ascending influence of big data [10] and propose specially designed data architectures [11].

However, practical applications of actual smart manufacturing have been observed to be rather rare [12, 13], which can be traced back to inadequate data management concepts. A sustainable data management concept depends on the specific use and setting [14] and requires a practical evaluation [11]. In general, scientific publications either prioritize the elaboration of a particular smart manufacturing application scenario [15] or the development of a reference architecture for data management [16]. Bridging both worlds has been mainly neglected by research [17, 18]. Smart manufacturing brings along specific challenges, which must be considered by a sufficient data management concept. A tremendous number of IT systems and software [2, 19], a lack of central integration across various databases [20], a diverse composition and use of systems [21], and a shifting customer demand towards a highly flexible and customizable product, affect the manufacturing process [22]. Isolated, these properties are not exceptional, but in combination, the unique complexity hinders the theory transfer between application domains [23].

Against this background, the research question of this paper is: "*how to develop a data management concept for smart manufacturing applications?*". To answer this question, we develop a catalog of requirements and a framework for data management. Accordingly, we adopt the design science research process (DSR) idea of iteratively building and evaluating an artifact [26]. We first developed a catalog of requirements based on a systematic literature review and a workshop with experts in the domain of manufacturing. In line with the identified requirements, we designed a data management framework which we then evaluated by conducting interviews with a new set of information systems and data management experts. Workshop and interviews were conducted with one of the leading German automotive manufacturers, allowing us to extract valuable knowledge and feedback on our artifact [27].

The contribution of this work is a catalog of requirements and a framework, (1) ensuring the usability of all relevant data for smart manufacturing application scenarios and (2) creating awareness about necessary management tasks. With convenient access to all relevant data and the resulting analytics findings, decision support for employees or independent decision-making systems are possible. This enhancement should lead to greater efficiency and productivity in manufacturing processes as employees are supported in executing their tasks, and transparency along the value chain is enhanced.

This paper is structured as follows: Section 2 presents theoretical foundations regarding data management concepts. Section 3 describes the applied research method. In sections 4, we present our artifacts: a list of potential smart manufacturing applications, a catalog of requirements, and a framework. In section 5, we evaluate the artifacts. Finally, we discuss our finding in section 6 and conclude with a summary.

https://doi.org/10.30844/wi_2020_u1-marx

2 Research Background

By "*providing integrated access to multiple, distributed, heterogeneous databases and other information sources*" [9] data warehousing traditionally was the pacemaker for the development of data management approaches. Data warehouses were built for data extraction to fulfill mainly static and continuous reporting needs to support decision making [24]. Initially, data warehouse research focused on technical challenges arising with querying information from various data sources [9, 25], with a primary focus on data modeling [26, 27].

With the increasing prevalence of data-driven technologies – powerful drivers for smart manufacturing – the requirements for data management have changed drastically [28, 29]. The quantity of sources has become highly diverse and unstructured data from, e.g., social media has become the basis for many AI application [20]. Applications are now profoundly linked to processes (e.g., creating a product) and resources (e.g., employees) in the organization [28]. Consequently, traditional management aspects, such as governance and compliance, must be included in data management to fulfill the changing needs for data management, many enterprises have proposed so-called "data lakes" architectures for combining diverse data sources and structures [10, 30].

Instead of preprocessing data for specific use cases into a pre-defined data model, data lakes combine data sources on a raw data level enabling a wide range of applications and assuring agility of data analytics [10]. Therefore, the traditional process of extract, transform, load (ETL) has been adapted [31]. While the extraction used to be done exclusively as a batch process, e.g., following a pull approach, some sources push data to data lakes as a data stream which needs to be processed in real-time [32]. The transformation phase is still dealing with data cleansing and standardization, but a preliminary calculation of measures and aggregation of data is not required anymore. Instead, data is physically stored as raw data, and merged and transformed into a virtual layer that delivers the desired result to the analytics application for visualization [33].

Various scholarly works proposed data lake architectures of which some limit their view onto the technical level [20], but several also address managerial aspects [10, 34]. However, none of them offers an integrated view of data management that includes application scenarios, managerial aspects such as change management, and the integration of the data lake into the business processes of the organization.

3 Research Method

For the research design, we follow a design science research (DSR) approach [35, 36]. DSR alternates between "building" and "evaluating" [36]. To substantiate this iterative process, we carried out a workshop and interviews. According to Benbasat et al. [37], expert interviews are appropriate to observe the utilization and development of an artifact in a specific context. **Table 1** displays an overview of both the workshop participants and the interviewees. Workshop and interviews were part of a case study

https://doi.org/10.30844/wi_2020_u1-marx

based on the recommendations of [38] and [39] with one of the leading German automotive manufacturers. Case studies are sufficient for theory elaboration, which combines building a general theory and evaluating an empirical context [40]. The company already deploys numerous smart manufacturing applications to accomplish and optimize products, processes, and services and generates a reasonable amount of transferable knowledge. While we observed extent applications, we gained valuable insights into challenges and opportunities for implementing smart manufacturing.

Table 1. Overview of the interviewees

#	*Position*	*Stage*	*Type*
1	Manager IE (Industrial Engineering)	1	Workshop
2	Employee IE - Data Analytics applications	1	Workshop
3	Employee IE - Time Analysis	1	Workshop
4	Employee IE - Staff Planning	1	Workshop
5	Manager Productivity Controlling	1	Workshop
6	Project Manager IT Controlling	1	Workshop
7	System Administration	1	Workshop
8	Developer Application Systems Production - Planning	2	Interview
9	Developer Application Systems Production - Assembly	2	Interview
10	Specialist Database Architectures	2	Interview
11-13	Data Scientist - Implementation of data analytics projects	2	Interview
14	Roll-out Expert Application Systems Production	2	Interview
15	Specialist Production Platforms	2	Interview

We initiated our research with a systematic literature review, according to the scheme proposed by Webster and Watson [41], consisting of three steps. First, for the identification of papers, the databases "ScienceDirect", "IEEE Xplore" and "EBSCO host" were considered and the searching statement "("data analytics" OR "big data architecture" OR "central database" OR "information system design") AND requirement AND manufacturing" was used. The first step resulted in 2,748 papers. In the second step, 102 out of the 2,748 identified papers were prioritized and elaborated as to be relevant based on their abstract, title, keywords and journal-ranking. Third, we conducted a forward- and backward review with the search engine "Google Scholar". Finally, based on the resulting 148 papers, we extracted universal technical, organizational, and procedural requirements for creating a data management concept.

We simultaneously carried out a seven-hour workshop with domain experts from the case company on the 28th of November 2018. The workshop's objective was to develop a compilation of smart manufacturing applications through a keyword-based discussion. The group of participating experts consisted primarily of experts from the field of industrial engineering in the automotive industry, whereby both operatively and strategically acting persons were involved in order to develop a multitude of application scenarios from different perspectives (cf. **Table 1**, upper half). To attune the understanding of the overarching topic, first, the fundamentals of smart manufacturing and data analytics were introduced. Then, the manufacturing context was characterized in the form of typical issues and extent solutions. Based on this, smart manufacturing applications were discussed from which requirements for successful implementation were derived. The results of the literature review and workshop were consolidated to identify implementation requirements from a data

https://doi.org/10.30844/wi_2020_u1-marx

perspective which were compiled in a catalog. Based on this catalog, we designed a framework with relevant elements for data management from a technical, procedural, and organizational perspective.

To evaluate and further develop our framework, we conducted interviews with a new set of experts. We invited experts from different departments, which manufacture a large number of variants in mass production, who have substantial knowledge in information systems, and experience with the deployment of data analytics, the development of reference architectures or the application of a digital production system (cf. **Table 1**, lower half). We chose semi-structured interviews as they allow improvisation and exploration of the underlying phenomenon [42]. The questionnaire involved a technical, an organizational, and a procedural section. For each section, we discussed completeness, consistency, traceability, and transferability of both catalog of requirements, and framework. In total, we conducted eight face-to-face interviews in January 2019, which lasted between 67 and 123 minutes. We recorded, transcribed, anonymized and sent back the transcriptions to the interviewees to provide additional comments. The final transcripts were used for our analysis.

4 Closing the Gap between Smart Manufacturing Applications and Data Management

4.1 Smart Manufacturing Applications

A suitable data management framework depends on the specific use and setting [14]. Thus, we established an exhaustive list of smart manufacturing applications in an expert workshop. To support the ideation process and to position the outcome systematically, we first defined an ancillary framework with two dimensions (**Table 2**).

The first dimension constitutes the capabilities of data analytics as a vital foundation for smart manufacturing, which is well suited to categorize the essential outcome of smart manufacturing applications [43, 44]. Data analytics consists of four successive levels: *descriptive, diagnostic, predictive, and prescriptive analytics* [43–45]. What happened? Descriptive analytics evaluates historical data with especially statistical methods and visualizes the results in the form of dashboards for monitoring or controlling [46]. Why did it happen? Diagnostic analytics uncover reasons and causes of past states with merely statistical methods by inspecting dependencies and correlations of parameters [45]. What could happen? Predictive analytics uses primarily statistical and machine-learning methods for predicting future states or trends [44, 45]. What should happen? Prescriptive analytics intends to determine optimal actions with respect on grounding conditions (e.g., cost-related) through the use of especially machine-learning, optimization, and simulation-based methods [47].

The second dimension characterizes four primary task areas within manufacturing: *production management, work organization, work design and management, and profitability analysis* [48–50]. Production management describes the optimization of production systems and the management and control of required resources. The affiliated area of work organization is characterized by tasks such as work structuring and operating time organization. Work design and management considers the individual workplace by, e.g., carrying out time studies and evaluating the procedure. In the profitability analysis, key figures are recorded, evaluated, and visualized for the person responsible for the workshop or management level.

We identified fifteen applications: (A1) cause analysis of production errors, (A2) prediction of production times, (A3) personnel control with predictive maintenance, (A4) container management optimization, (A5) material flow optimization, (A6) workforce management, (A7) factory layout planning, (A8) rework optimization, (A9) line clocking testing, (A10) direct ergonomics feedback, (A11) production time calculation for new starts, (A12) work process design, (A13) productivity controlling, (A14) task-based absence analysis, and (A15) productivity increase in the series.

Table 2. Smart manufacturing application scenarios

	Descriptive Analytics	*Diagnostic Analytics*	*Predictive Analytics*	*Prescriptive Analytics*
Production management		A1	A2, A3	A4, A5
Work organization			A6	A7, A8
Work design and management	A9, A10		A11	A12
Profitability analysis	A13	A14	A15	

4.2 A Catalog of Requirements for a Data Management Concept

Based on the literature review and the smart manufacturing application scenarios identified in the expert workshop, we derived 22 requirements for the implementation of smart manufacturing (cf., **Figure 1**). As we iteratively built and evaluated a catalog of requirements and framework, we here present the final iteration of our artifact. The requirements are grouped into four categories. Requirements that arise through using different systems and data sources compose the category *system landscape (A)*. *Data analytics (B)* considers factors to realize a qualitative and expedient use of the stored data. While data analytics covers general requirements derived from theory, smart manufacturing applications, as seen in the previous chapter, entail unique requirements represented by application *scenario (C)*. We also must pay attention to the implementation and monitoring of information systems in a company. *Data engineering (D)* describes operational, legal, and technical basics regarding data [51].

A	B	C	D
System Landscape	**Data Analytics**	**Application Scenario**	**Data Engineering**
Compatibility (A1)	Interoperability (B1)	Real-time data refresh (C1)	Governance concept (D1)
Continuous information integration (A2)	Data quality standards (B2)	Reasonable data extraction (C2)	Operation management (D2)
Processing heterogeneous data types (A3)	Standardized approach (B3)	Fair performance (C3)	Fair infrastructure (D3)
Cloud computing (A4)	Process-related integration (B4)	Consistent data handling(C4)	Security concept (D4)
	Responsibility assignement (B5)	High availability (C5)	
	Specification of employee qualification (B6)	Scalability (C6)	
	Communication strategy (B7)	Flexibility (C7)	

Figure 1. Overview of the catalog of requirements for a data management concept

Compatibility (**A1)** describes the interaction and the consistency of new applications with established information systems [52]. This mainly refers to interfaces for and arrangements of the data exchange and the underlying data structure. Semi-automatic or manual mapping (merging of different data models) potentially entails a high error rate [53]. For this reason, processes and architectures should be prepared for **continuous information integration (A2)** between the data source and processing level. To process and evaluate all available data, **processing heterogeneous data types (A3)** is necessary [54]**.** This includes both structured data from ERP systems, MES and PPC systems and unstructured data such as texts or images from knowledge management or similar systems. Future IT infrastructures should **support cloud computing (A4)** [55] to ensure an immediate or later integration into the cloud. This also allows a service-oriented architecture in manufacturing [56].

Interoperability (B1) is the ability to collaborate on information systems [57]. With a broad spectrum of software solutions in practice [58], the architecture has to integrate various platforms and tools [59]. Also, the system requires data interpretation capabilities, including a general understanding of the data basis [60]. Fundamental **data quality standards (B2)** and continuous data quality checks must be established [61]. Intrinsic data quality describes the completeness, accuracy, validity, and consistency of a dataset and contextual data quality refers to the value and relevance of the data-dependent on the given facts [62]. Data quality is ensured by specifying data schemes or data models [53, 54]. A **standardized approach (B3)** with elementary steps is necessary to reduce complexity and planning efforts and to facilitate cross-sectoral implementation [63]. Still, a dynamic business environment calls for a reasonable amount of flexibility and adaptability [53]. To enable sustainable applicability and continuous improvement, a **process-related integration (B4)** needs to be specified. This includes determining the use, timing, and objectives of applications within business processes. An important aspect is the identification and verification of the added value of an implemented application [64]. Existing databases and the pool of applications already deployed in the enterprise have to be controlled, maintained, and managed by **responsibilities assignment (B5)** to keep up efficiency and performance [61]. Besides, specialists can exchange expertise in the enterprise. To ensure the successful implementation of the application scenarios, the **specification of employee qualifications (B6)**, including a recruitment strategy, is necessary [65]. The processed applications must also be intelligible to be reusable [66]. Thus, planning and implementing a **communication strategy (B7)** is vital.

An essential requirement of manufacturing is **real-time data refresh (C1)**. Especially for time-critical tasks, contemporary situation analysis, and notification to respond is indispensable. This is particularly important for applications scenarios such as productivity controlling (A13), whose usefulness or output is strongly dependent on always up-to-date data. Some applications, e.g. prediction of production times (A2), benefit from a wide variety of (context) information, which requires the implementation of various analytics methods. To prevent data from being misinterpreted or unnecessarily integrated, **reasonable data extraction (C2)** must be assured, which disconnects the data stored from the data processed. In manufacturing, reading, writing, retrieving, processing and displaying data is typically executed at a high frequency resulting in an infrastructure capable of high data throughput (**decent performance (C3)**) [67]. Parallel data processing and executing in different systems/ modules simultaneously [68] and continuous data updates or peak loads should not affect the performance of the system [69]. Data should have a uniform granularity level and a consistent structure. However, there is often a discrepancy between interacting systems in practice [70]. Source and processing systems need to be matched, and norms for data import must be specified with **consistent data handling (C4)** [70]. Applications such as material flow optimization (A5) need to surveil the production process constantly and if necessary, notify the operators in cases of deviations. This leads to the requirement that the system must feature **high availability (C5)** to ensure a steady execution of requests [69]. Closely related is reliability. To avoid single points of failure, bridging, and sustaining performance has to be ensured [69]. Vertical ("scale-up") and horizontal ("scale-out") **scalability (C6)** describes the ability to spatially and technologically extend or reduce objects [68, 71]. Vertical refers to capacity expansion through additional hardware or license extensions, whereas horizontal represents the addition of database servers or cloud instances [71]. This requirement arose from the experts' need to test smart manufacturing applications on a small scale and then expand them quickly and efficiently. Another criterion for the design of information systems is **flexibility (C7)**. This allows changes and extensions of data storage and analysis. Crucial influencing factors are the data structure or the data model since these are the basis for [46].

Governance concept (D1) describes the inevitable administration and control of data [61]. This embraces designing authorization structure (e.g., the evaluation of data across domain boundaries), describing and classifying data with meta data management, establishing guidelines and standards, and implementing risk management and compliance [20, 30]. **Operation management (D2)** analyzes the estimated cost structure and ensures controllability during operation [72]. Costs are expenses for introduction/ administration and operating costs for the development/ maintenance of an information system. Strategic aspects such as the expected acceptance and the coordination of comparable development activities result in

https://doi.org/10.30844/wi_2020_u1-marx

synergy effects [66]. With the evaluation criterion, **decent infrastructure (D3)**, technical and organizational factors are taken into consideration [73]. To avoid a limited capacity of data storing and analysis, expandability is desirable. Flexibility in dealing with changes in the environment and overall infrastructure and general adaptability/ reconfigurability are also essential requirements [74]. Stability in the operation of the overall architecture and maintainability ensure a smooth process. Additional aspects are fault tolerance and automatic recovery [68, 74]. To prevent the misuse and loss of data, a **security concept (D4)** is relevant. Protective devices or mechanisms and measures to safeguard the data in case of system failures must be available [75], including the protection of personal data in line with the country-specific legal situation [66].

4.3 A Data Management Framework for Smart Manufacturing Applications

Based on the identified requirements, we developed a data management framework to support smart manufacturing applications, consisting of three layers (cf., **Figure 2**).

The central layer represents the technical implementation of smart manufacturing applications. This layer displays a typical "data lifecycle," beginning with data integration, which represents the system's capability to gather data of various, heterogeneous data sources through appropriate interfaces and methods. Then the data has to be stored in a structured manner, e.g., by pre-defined data models. Unstructured data, such as text data, requires an ex-ante pre-processing layer. Along with data storing occurs data processing, which combines activities to ensure the integrity and usability of the data. This includes data cleaning, transforming, and standardizing. Next, necessary data has to be extracted from the database and to be converted to an exploitable form without changing the initial data with data virtualization. Finally, data analytics is performed with five fundamental types of outcome: microservices, applications, ad-hoc analysis, ad-hoc reporting, and standard reporting. To ensure the technical operability, data governance and security are requisite, and a cloud-compatible architecture is advisable. The applications need to be embedded in the ongoing business, which is portrayed by the business layer on top. This layer includes the integration of the application into the affected business processes and management tasks to ensure implementation and operation. These tasks are communication policy, change management, stakeholder management, and employee qualification. The bottom layer depicts relevant management tasks with a focus on data handling. Besides the determination of necessary data for an application and therefore, the data source, this includes the tasks compliance management, access control, data quality/ management and standards for data models and meta data.

https://doi.org/10.30844/wi_2020_u1-marx

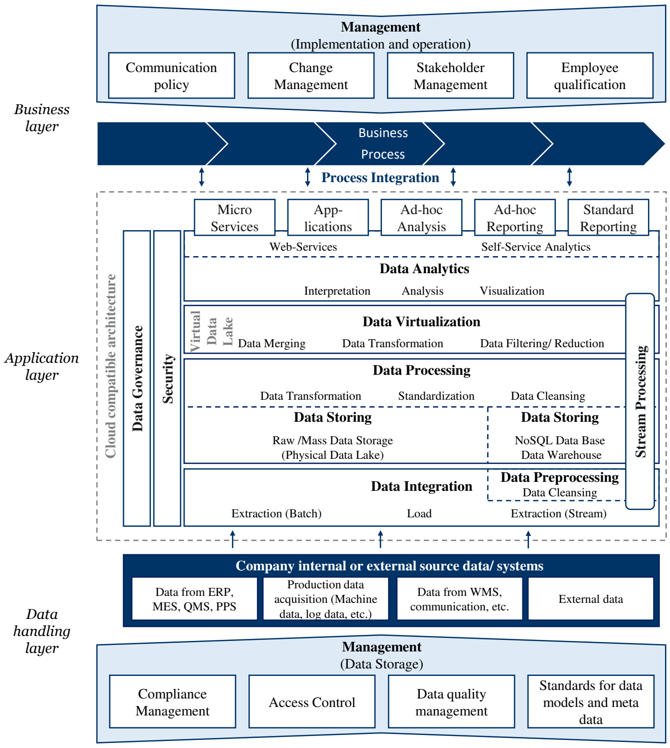

Figure 2. A data management framework for smart manufacturing applications

5 Evaluation

The feedback and gained knowledge of the conducted expert interviews were adopted to evaluate and enhance both artifacts - the catalog of requirements and the data management framework - iteratively. Beginning with the initial research motivation, all experts confirmed the importance of aligning specific use and setting with data management. They also unanimously emphasized the insufficient treatment of organizational and procedural aspects in current technical implementation in companies, which corresponds with our observation in recent research.

Concerning the technical perspective, the experts agree upon a model of logical layers, especially transferability and the often given high heterogeneity of systems. All experts have considered combining data warehouse, NoSQL database, and a distributed file system in the data storing layer appropriate. For an expert, meta data management is the *"basic obligation"* for a database architecture to operate. Despite having convenient access to data, there is a risk of a missing possibility of identifying data, to transfer information, and to generate knowledge. Regarding the technical solution of this issue, the expert opinions differ. On the one hand, the identification keys must be standardized across all systems. On the other hand, this would increase the complexity in the operational systems and recommend implementing mapping.

The experts recommended web-services as a basis on which smart manufacturing applications can be attached. Furthermore, several experts highlighted the need to differentiate between application and reporting, while some additionally subdivided reporting into standard and ad-hoc reporting. One expert placed self-service analytics on the same level as web-service, which was considered appropriate. The expert's inputs lead to some adjustments in the data analytics layer of the model. Instead of equating applications and reporting - based on the literature - web services and self-service analytics are the basis of said in the final model. Besides, microservices and ad-hoc analysis were added and reporting was divided into standard and ad-hoc reporting.

Concerning the organizational perspective, the tasks of stakeholder management, access control, employee qualification, change management, communication guidelines, and data quality management were confirmed. Additionally, the experts recommended including compliance management and the standardization of data models and meta data, which were added to the management perspective of the data handling layer. An expert advocates the creation of an organizational unit that deals with issues of data governance in a structured way to make sustainable decisions. This includes the use of key indicators, the specification of data models, and commonly the standardization of the data landscape. In the context of the latter, the costs for standardization should be evaluated economically. Data governance was added as an overarching function of the application layer. Regarding current developments and challenges, the handling of personal data was also discussed. Due to the complexity of the topic, the creation of an independent category of tasks for the handling of legal issues on the organizational level was discussed. Because of a lack of an agreeable result, such an element was not included in the final artifact.

All experts agreed on the sine qua non of a cultural change across all levels and a shift of the basic attitude regarding the economic justification for smart manufacturing projects. It is hardly possible to capture the added value of a smart manufacturing application in specific financial terms at an early stage of the project. An understanding of the relationship between effort and results of smart manufacturing use cases must be created. Often only great efforts in preparing and executing analyses uncover easy-to-implement solutions. Particular attention should rest on change management. Creating a problem awareness and even a sense of urgency, sensitizing and informing executives about potential applications and providing management support and appropriate governance structures should be elementary components. To enable a firm integration into the company processes, the motivation of the employee must be created by showing the added value for the employees themselves.

Concerning the procedural perspective, all experts expressed that, despite the diversification in the application scenarios, a basic orientation makes sense. Guidelines should be established, including a description of the execution, the appropriate type of process integration, and the required tools or qualifications of a particular use case type. One suggestion for anchoring smart manufacturing in the department's processes was the transfer of general practices in manuals or similar documents.

https://doi.org/10.30844/wi_2020_u1-marx

6 Discussion and Conclusion

In this paper, we presented a data management approach to foster the development and implementation of smart manufacturing applications. As we discussed, an appropriate data management concept depends on the specific use of data [14]. While most publications focus either on data management or smart manufacturing applications, bridging both is rather scarce in research [17, 18]. We address this deficit by providing a catalog of requirements and a design of a framework for data management that fosters the implementation of smart manufacturing applications. The objective was to provide a methodical basis for the technical implementation and guidelines for sustainable integration of applications in organization and business processes. The present work includes the identification of potential application scenarios with a generic description of properties to enable further use in other research projects.

To develop the catalog of requirements, we conducted a systematic literature review on data management concepts to ensure rigor and conducted a workshop with seven industry experts from one of the leading German automotive manufacturers to identify specific use and setting requirements. We established 15 company independent applications with a high potential in manufacturing. As a result, 22 requirements cataloged with the categories *system landscape, data analytics, application scenario, or data engineering* were specified. With these requirements, we defined a data management framework, which combines technical, procedural, and organizational aspects. To evaluate both artifacts, we conducted interviews with eight data experts to check for completeness, consistency, traceability, and transferability.

The contribution of this paper to the body of knowledge of IT management is as follows: The overall objective in smart manufacturing is to increase efficiency and productivity by exploiting data analytics for continuous process optimizations in production and along the entire value stream. To realize an optimal design of the processes, various information must be included and processed. This multiplicity of information results in a high degree of decision complexity since different factors have to be considered simultaneously. Practitioners must not only reach out for implementing smart manufacturing applications, but also consider the fundament in the form of sufficient data management.

Nevertheless, the study design is subject to some limitations, which in return lay the foundation for future research. First, while the single case research allowed us an in-depth evaluation of the artifact in a representative company, conducting a multiple case study could offer more valuable insights. Second, we did not consider company-specific requirements in the development of the concept, which displays a limitation of our artifact. Each company has different conditions for and characteristics of manufacturing systems, due to various processes and a diverse system landscape. For a possible implementation with the associated technologies, a reconciliation of our concept regarding the respective corporate environment is required. Still, the layered design of the system with a flexible choice of technology deployment ensures transferability of potential applications.

https://doi.org/10.30844/wi_2020_u1-marx

References

1. Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D.: Deep learning for smart manufacturing. Methods and applications. Journal of Manufacturing Systems 48, 144–156 (2018)
2. Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., Apon, A. (eds.): Automotive big data. Applications, workloads and infrastructures (2015)
3. Urbach, N., Drews, P., Ross, J.: Digital business transformation and the changing role of the IT function. MIS Quarterly Executive 16, 2–4 (2017)
4. Urbach, N., Ahlemann, F., Böhmann, T., Drews, P., Brenner, W., Schaudel, F., Schütte, R.: The impact of digitalization on the IT department. Business & Information Systems Engineering 61, 123–131 (2019)
5. Mittal, S., Khan, M.A., Romero, D., Wuest, T.: Smart manufacturing: Characteristics, technologies and enabling factors. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 233, 1342–1361 (2019)
6. Surdilovic, D., Schreck, G., Schmidt, U. (eds.): Development of collaborative robots (COBOTS) for flexible human-integrated assembly automation (2010)
7. Bullinger, H.-J., Meiren, T., Nägele, R.: Smart services in manufacturing companies. Unpublished
8. Tao, F., Qi, Q., Liu, A., Kusiak, A.: Data-driven smart manufacturing. Journal of Manufacturing Systems 48, 157–169 (2018)
9. Widom, J.: Research Problems in Data Warehousing. In: Fourth International Conference on Information and Knowledge Management (CIKM 1995) (1995)
10. Fang, H.: Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. In: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 820–824. IEEE (2015 - 2015)
11. Sang, G.M., Xu, L., Vrieze, P.T.d. (eds.): A reference architecture for big data systems (2016)
12. Groggert, S., Wenking, M., Schmitt, R.H., Friedli, T. (eds.): Status quo and future potential of manufacturing data analytics. An empirical study (2017)
13. Li, J., Tao, F., Cheng, Y., Zhao, L.: Big data in product lifecycle management. The International Journal of Advanced Manufacturing Technology 81, 667–684 (2015)
14. Tariq, R.S., Thabet, N.: Big data challenges. Computer Engineering & Information Technology 4 (2015)
15. Ji, W., Wang, L.: Big data analytics based fault prediction for shop floor scheduling. Journal of Manufacturing Systems 43, 187–194 (2017)
16. Santos, M.Y., Oliveira e Sá, J., Costa, C., Galvão, J., Andrade, C., Martinho, B., Lima, F.V., Costa, E. (eds.): A big data analytics architecture for industry 4.0 (2017)
17. Riggins, F.J., Wamba, S.F. (eds.): Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics (2015)
18. Li, S., Peng, G.C., Xing, F.: Barriers of embedding big data solutions in smart factories. Insights from SAP consultants. Industrial Management & Data Systems 119, 1147–1164 (2019)
19. Gröger, C., Kassner, L., Hoos, E., Königsberger, J., Kiefer, C., Silcher, S., Mitschang, B. (eds.): The data-driven factory. Leveraging big industrial data for agile, learning and human-centric manufacturing (2016)
20. Hai, R., Geisler, S., Quix, C. (eds.): Constance. An intelligent data lake system (2016)
21. Cuzzocrea, A., Song, I.-Y., Davis, K.C.: Analytics over large-scale multidimensional data. In: International Workshop on Data Warehousing, pp. 101–103

https://doi.org/10.30844/wi_2020_u1-marx

22. Bauernhansl, T., Hompel, M. ten, Vogel-Heuser, B.: Industrie 4.0 in Produktion, Automatisierung und Logistik. Anwendung - Technologien - Migration. Springer Fachmedien Wiesbaden, Wiesbaden (2014)
23. Wixom, B.H., Watson, H.J.: An empirical investigation of the factors affecting data warehousing success. MIS Quarterly 25, 17 (2001)
24. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology. ACM Sigmod Record 26, 65–74 (1997)
25. Gupta, H.: Selection of views to materialize in a data warehouse. IEEE Transactions on Knowledge and Data Engineering 17, 98–112 (2005)
26. Tryfona, N., Busborg, F., Borch Christiansen, J.G. (eds.): starER. A conceptual model for data warehouse design (1999)
27. Kimball, R., Ross, M.: The data warehouse toolkit. The complete guide to dimensional modeling. John Wiley & Sons, Indianapolis, IN (2011)
28. Davenport, T.H., Barth, P., Bean, R.: How 'big data' is different. MIT Sloan Management Review 54 (2012)
29. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data. The next frontier for innovation, competition, and productivity. New York, NY (2011)
30. Mathis, C.: Data lakes. Datenbank Spektrum 17, 289–293 (2017)
31. Bruckner, R.M., List, B., Schiefer, J. (eds.): Striving towards near real-time data integration for data warehouses (2002)
32. Miloslavskaya, N., Tolstoy, A.: Big data, fast data and data lake concepts. Procedia Computer Science 88, 300–305 (2016)
33. van der Lans, Rick F.: Architecting the multi-purpose data lake with data virtualization. Lisse (2018)
34. Madera, C., Laurent, A. (eds.): The next information architecture evolution. The data lake wave (2016)
35. Hevner, A.R., Ram, S., March, S.T.: Design science in information systems research. MIS Quarterly 28, 75–105 (2004)
36. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decision Support Systems 15, 251–266 (1995)
37. Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. MIS Quarterly 11, 369 (1987)
38. Eisenhardt, K.M.: Building theories from case study research. The Academy of Management Review 14, 532–550 (1989)
39. Yin, R.K.: Case study research and applications. Design and methods. Sage, Los Angeles, CA (2018)
40. Ketokivi, M., Choi, T.: Renaissance of case research as a scientific method. Journal of Operations Management 32, 232–240 (2014)
41. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future. Writing a literature review. MIS Quarterly 26, 13–23 (2002)
42. Cooper, D.R., Schindler, P.S.: Business research methods. McGraw-Hill Irwin, New York, NY (2014)
43. Bauer, D., Maurer, T., Henkel, C., Bildstein, A.: Big-Data-Analytik. Datenbasierte Optimierung produzierender Unternehmen. Zenodo, Stuttgart (2017)
44. Delen, D., Demirkan, H.: Data, information and analytics as services. Decision Support Systems 55, 359–363 (2013)
45. Watson, H.J.: Tutorial: big data analytics. Concepts, technologies, and applications. Communications of the Association for Information Systems 34, 1247–1268 (2014)

https://doi.org/10.30844/wi_2020_u1-marx

46. Hu, H., Wen, Y., Chua, T.-S., Li, X.: Toward scalable systems for big data analytics. A technology tutorial. IEEE Access 2, 652–687 (2014)
47. Freitag, M., Kück, M., Alla, A.A., Lütjen, M.: Potenziale von Data Science in Produktion und Logistik. Teil 1 - Eine Einführung in aktuelle Ansätze der Data Science. Industrie Management 31, 22–26 (2015)
48. Bokranz, R., Landau, K.: Handbuch Industrial Engineering. Produktivitätsmanagement mit MTM. Schäffer-Poeschel, Stuttgart (2012)
49. Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung: Industrial Engineering. Standardmethoden zur Produktivitätssteigerung und Prozessoptimierung. Hanser, München (2015)
50. Dorner, M., Stowasser, S.: Das Produktivitätsmanagement des Industrial Engineering. • Modellentwicklung • Führungssystem • Produktivitätskennzahl • Verbesserung der Arbeitsproduktivität • Produktivitätscontrolling. Zeitschrift für Arbeitswissenschaft 66, 212–225 (2012)
51. Scheffler, A., Wirths, C.P.: Data innovation @ AXA Germany. Journey towards a data-driven insurer. In: Urbach, N., Röglinger, M. (eds.) Digitalization cases, pp. 363–378. Springer International Publishing, Cham (2019)
52. Nyhuis, P., Reinhart, G., Abele, E. (eds.): Wandlungsfähige Produktionssysteme. Heute die Industrie von morgen gestalten. Produktionstechnisches Zentrum Hannover, Garbsen (2008)
53. Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F.: Guide to intelligent data analysis. How to intelligently make sense of real data. Springer London, London (2010)
54. Anagnostopoulos, I., Zeadally, S., Exposito, E.: Handling big data. Research challenges and future directions. The Journal of Supercomputing 72, 1494–1516 (2016)
55. Balachandran, B.M., Prasad, S.: Challenges and benefits of deploying big data analytics in the cloud for business intelligence. Procedia Computer Science 112, 1112–1122 (2017)
56. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyya, R.: Big data computing and clouds. Trends and future directions. Journal of Parallel and Distributed Computing 79-80, 3–15 (2015)
57. Sheth, A.P.: Changing focus on interoperability in information systems. From system, syntax, structure to semantics. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (eds.) Interoperating geographic information systems, 495, pp. 5–29. Springer US, Boston, MA (1999)
58. Gerloff, C., Cleophas, C. (eds.): Excavating the treasure of IoT data. An architecture to empower rapid data analytics for predictive maintenance of connected vehicles (2017)
59. Jacoby, M., Antonić, A., Kreiner, K., Łapacz, R., Pielorz, J. (eds.): Semantic interoperability as key to IoT platform federation (2017)
60. Wunder, M., Grosche, J.: Verteilte Führungsinformationssysteme. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
61. Khatri, V., Brown, C.V.: Designing data governance. Communications of the ACM 53, 148/152 (2010)
62. Hazen, B.T., Boone, C.A., Ezell, J.D., Jones-Farmer, L.A.: Data quality for data science, predictive analytics, and big data in supply chain management. An introduction to the problem and suggestions for research and applications. International Journal of Production Economics 154, 72–80 (2014)
63. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. Big Data 1, 51–59 (2013)
64. Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., Zhao, J.L.: Transformational issues of big data and analytics in networked business. MIS Quarterly 40, 807–818 (2016)

65. Ghasemaghaei, M., Ebrahimi, S., Hassanein, K.: Data analytics competency for improving firm decision making performance. The Journal of Strategic Information Systems 27, 101–113 (2018)
66. Bose, R.: Advanced analytics. Opportunities and challenges. Industrial Management & Data Systems 109, 155–172 (2009)
67. Divate, R., Sah, S., Singh, M.: High performance computing and big data. In: Srinivasan, S. (ed.) Guide to big data applications, 26, pp. 125–147. Springer International Publishing, Cham (2018)
68. Singh, D., Reddy, C.K.: A survey on platforms for big data analytics. Journal of Big Data 2, 1–20 (2015)
69. Kimball, R., Ross, M.: The data warehouse toolkit. The definitive guide to dimensional modeling. John Wiley & Sons, Indianapolis, IN (2013)
70. van der Aalst, W.: Process mining. Data science in action. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
71. Vermeulen, A.F.: Practical data science. A guide to building the technology stack for turning data lakes into business assets. Apress, Berkeley, CA (2018)
72. Periasamy, M., Chelliah, P.R.: Big data analytics. Enabling technologies and tools. In: Mahmood (Ed.) 2016 - data science and big data, pp. 221–243
73. Diemer, J.: Sichere Industrie-4.0-Plattformen auf Basis von Community-Clouds. In: Vogel-Heuser, B., Bauernhansl, T., Hompel, M. ten (eds.) Handbuch Industrie 4.0 Bd.1. Produktion, pp. 177–204. Springer Berlin Heidelberg, Berlin, Heidelberg (2017)
74. Krishnan, K.: Data warehousing in the age of big data. Morgan Kaufmann, Waltham, MA (2013)
75. D. Sudarsan, S., Jetley, R., Ramaswamy, S.: Security and privacy of big data. In: Big Data, pp. 121–136

https://doi.org/10.30844/wi_2020_u1-marx