DATA PROCESSING REQUIREMENTS OF INDUSTRY 4.0 – USE CASES FOR BIG DATA APPLICATIONS

Research in Progress

Gölzer, Philipp, University of Erlangen-Nuremberg, Nuremberg, Germany, philipp.goelzer@fau.de

Patrick, Cato, University of Erlangen-Nuremberg, Nuremberg, Germany, patrick.cato@fau.de

Michael, Amberg, University of Erlangen-Nuremberg, Nuremberg, Germany, michael.amberg@fau.de

Abstract

Industry 4.0 stands for the 4th Industrial revolution and the new paradigm of autonomous and decentralized control in production. Products and production systems are enhanced to Cyber Physical Systems which have the capability to communicate with each other, to build ad-hoc networks and for self-control and self-optimization. From the IT-perspective this involves a new level of networking, data integration and data processing in production. Established technologies like Internet of Things, Cloud or Big Data are propagated solution-components of Industry 4.0. So far, there is no founded elaboration of IT-requirements and no differentiated discussion on how solution-components fulfil these requirements. This research uses the method of content analysis to extract requirements of Industry 4.0 from current research publications. Objective of analysis is a structured compilation of requirements regarding data processing. The resulting category scheme enables further development of solution-components in the application domain of Industry 4.0. Furthermore, this paper shows how the requirements can be matched to the capabilities of Big Data software solutions. As a result, two general use cases for Big Data applications in Industry 4.0 were identified and characterized.

Keywords: Big Data, Production, Industry 4.0.

1 Introduction

The continuous evolution in the field of information systems has the potential to enhance established solutions in almost every field of human creation (Lee, 2008). In the field of production, this development enhances the 4th Industrial Revolution, frequently noted as Industry 4.0. Global mega trends, such as globalization, individualization or demography lead to shortened product life cycles, higher product variety and increase dynamic and complexity on the market side. To remain competitive under these conditions, a company's production has to be highly flexible and adaptable (Westkämper, 2011).

Industry 4.0 propagates the objective of an autonomous and decentralized production (Bauernhansl et al., 2014) to address the increasing dynamic and complexity on the market. Products and production systems have to be enhanced to Cyber Physical Systems (CPS). CPS are able to communicate with each other, to detect their environment, to interpret available data and to act on the physical world (Lee, 2008). These characteristics of CPS enable autonomous and decentralized production networks which have the capability for self-control and self-optimization. This new production paradigm of Industry 4.0 fundamentally improves industrial processes and leads to a higher efficiency in production. Autonomous and decentralized production has specific requirements for IT-support. Established technologies like Internet of Things, Cloud or Big Data are propagated solution-components of Industry 4.0 (Kagermann et al., 2013).

So far, there is no founded elaboration of IT-requirements and no detailed discussion on how solution-components fulfil these requirements. The description of requirements can mostly be found in an implicit form. Publications either describe requirements on a high level of abstraction, declared as challenges or by artefacts resulting from design science research. Solution-components are named but not specified. Detailing of IT-requirements and matching with capabilities of solution-components is a necessary step for further research, development and evaluation.

This paper analyses the requirements of Industry 4.0 with the focus on data processing and matches the requirements on the capabilities of the solution-component Big Data. It is structured in 5 chapters. Chapter 1 introduces to topic and motivation. Chapter 2 explains the background of Industry 4.0 and the capabilities of Big Data solutions. Chapter 3 describes the method used to extract data processing requirements from current research publications. Chapter 4 presents and explains results. Chapter 5 closes with discussion of requirements and matching with capabilities of Big Data solutions.

2 Background

2.1 Industry 4.0

In Industry 4.0 products and production systems like machines, warehouses and operating resources are enhanced to CPS and are connected to global production networks (Kagermann et al., 2013). These intelligent entities in production are able to communicate with each other, interpret available data, trigger actions and together have the capability for autonomous self-control and self-optimization (Lee, 2008). Intelligent products can be clearly identified, located at all times, know their history, status and alternative ways to completion. Intelligent production systems are connected to company's business processes, IT-systems and to the entire value chain in the production network. This enables real-time control and optimization of the value chain, starting with an order to the final delivery of the product (Kagermann et al., 2013). The convergence of the physical world and the digital world with CPS enables the new paradigm of autonomous and decentralized production (Brettel et al., 2014; Monostori, 2014).

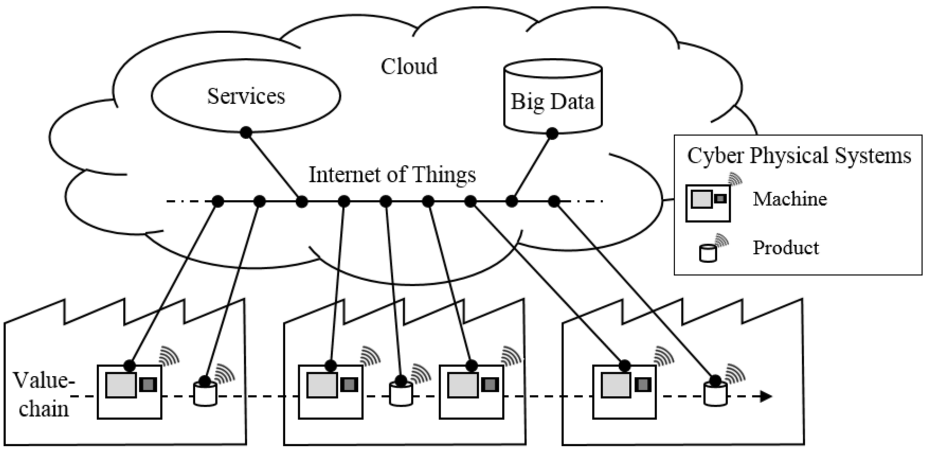

Figure 1. Solution-components of Industry 4.0

To fulfil the vision of Industry 4.0, a variety of concepts and solution-components are drawn and discussed in current research (Figure 1). This includes, but is not limited to (1) Cyber Physical Systems as intelligent entities in production (Sztipanovits et al., 2013), (2) Internet of Things as communication platform for CPS (Madisetti and Bahga, 2014), (3) Cloud solutions for decentralized services (Verl et al., 2013) and (4) Big Data solutions for high-performance processing of large data amounts in production (Kagermann et al., 2013; Lee et al., 2013).

A typical application scenario in Industry 4.0 are intelligent products, carrying all necessary information about their production processes. These products communicate with available and appropriate production resources in the value chain and decide on their own which steps to take next. Another scenario are intelligent machines who are able to predict breakdowns or quality problems and organize maintenance in time. They use patterns and knowledge in historical data from the production network to enhance their decisions processes and take actions.

2.2 Big Data software solutions

Big Data software solutions are an important component of Industry 4.0 (Kagermann et al., 2013). A closer look on Big Data software solutions brings a wide variety in characteristics, capabilities and applications (Schmarzo, 2013). Apache Hadoop is one of the most wellknown Big Data software solution, however, there is a great variety of others solutions e. g. Redis, SimpleDB, CouchDB, MongoDB, Terrastore, HBase or Cassandra (Cattell, 2011).

A common denominator is the usage of NoSQL data bases. Several taxonomies have been proposed to classify the different NoSQL data bases (Cattell, 2011; Pokorny, 2013). Pokorny (2013) for instance uses the criteria of the data model and identified three kind of models: Column-oriented (e. g. Cassandra), key-value (e. g. SimpleDB) and document-based (e.g. Mongo DB). Another criteria for classification of Big Data solution is the principle of data processing (Agrawal et al., 2011). The first principle is batch processing and distributed computing of data (Gupta et al., 2012). Large and complex data is split into small subsets and then processed concurrently. A common algorithm is MapReduce which is tuned for a specific use cases. A representative software solution for this principle is Hadoop HDFS with MapReduce (White, 2012). The second principle is to store data in a semi-structured data model which is adapted to the specific access pattern of a use case (Kaur and Rani, 2013). This enables real-time queries and random access on data without time-consuming operations and data joins. The software solutions Cassandra (Hewitt, 2010), SimpleDB (Chaganti and Helms, 2010) or MongoDB (Chodorow, 2013) are representatives of this principle.

Both classification criteria (data model, principle of processing) are important characteristics, when selecting a proper solution for the data processing requirements of a use case. Table 1 compares four different Big Data software solution regarding general capabilities and characteristics. Each solution is a representative for the classification described above. A final selection of an appropriate Big Data software solution depends on use case, existing infrastructure and application scenario.

BIG DATA SOFTWARE SOLUTIONS				
Capabilities /Characteristics	Hadoop HDFS & MapReduce	Cassandra	MongoDB	SimpleDB
Data model	File system	Column	Document	Key-Value
Batch processing / distributed computing	Yes	No	No	No
Real-time queries	No	Yes	Yes	Yes
Random access	No	Yes	Yes	Yes
Horizontal scaling	Yes	Yes	Yes	Yes
Strength	Data processing	Write	Read	Full Indexing
Architecture type	Master-Slave	Peer-to-Peer	Master-Slave	Web Service / Cloud Computing
CAP theorem	Consistency, Partition Tolerance	Availability, Partition Tolerance	Consistency, Partition Tolerance	Availability, Partition Tolerance

Table 1. Capabilities and characteristics of representative Big Data software solutions

3 Method

A widely accepted method in IS research to make valid interference from text can be found in the scientific technique of content analysis (Myers, 1997). Content analysis uses clear rules and systematic procedures for analysis and interpretation of text (Klenke, 2008; Krippendorff, 2004; Mayring, 2000). Compliance of rules and procedures delivers rigorous and replicable results (Krippendorff, 2004). Core of the content analysis is a category scheme, achieving the objectives of analysis. Categories can be developed by deduction from theory or by induction using the analysed material (Mayring, 2000). As relevant theory in the field of Industry 4.0 is widespread in various scientific disciplines (production, logistic, IT, AI, mathematics…), this study uses an inductive category development approach based on scientific publications in the field of Industry 4.0. This allows a grounded interpretation of material without pre-assumptions.

3.1 Process of content analysis

The analysis of requirements in the field of Industry 4.0 regarding data processing was conducted using the process of content analysis according to Mayring (2000, 2008). Vision, objectives and concepts behind the idea of Industry 4.0 are object of analysis. Objective of analysis is the structured compilation of explicit and implicit requirements for data processing in the Industry 4.0 concept. The research questions are: (1) What are the requirements regarding the data that need to be processed? (2) What are the requirements regarding the processing of the data?

Scientific publications (material) in the field of Industry 4.0 constitute the basis for this analysis. The search was conducted on journal papers, conference papers and white papers from the following scientific sources: general databases (ScienceDirect, IEEE Explore), German journal data bases (ZWF, WT, IM) and institutions/federations (Fraunhofer, VDI, Acatech, NIST). Publications were filtered using the keywords ‘industry 4.0’, ‘cyber physical system’, ‘internet of things’, ‘autonomy’, ‘decentralized’, ‘self-control’ in combination with the keyword ‘production’, ‘manufacturing’ or ‘logistic’. The filter was applied on title, abstract and keywords of publications. The search resulted in 87 publications in the period 2005 to 2014 (see appendix). Thereof, 55 research publications are written in German.

After selection of the material, rules for the process of category development were defined. Therefore, unit of analysis, selection criteria and level of abstraction were specified. The unit of analysis defines rules for the amount of text which is the basis for interpretation. As requirements are mostly described in implicit form, we choose ‘phrase’ as minimum and ‘section’ as maximum unit of analysis. This allows understanding and interpretation of requirements in the individual context of the paper. The selection criteria defines decision rules whether a unit of analysis contributes to the research question and objectives of analysis. In case of data processing, we analysed requirements regarding (1) Data: types, structure, format and sources and (2) Processing of data: operations, performance and conditions. The level of abstraction defines the rules to build a category for a unit of analysis which fulfils the selection criteria. In case of data processing requirements of industry 4.0, the rule is to choose the level of abstract in a way that categories for requirements are not specific to any approach or solution, but are applicable to the idea of Industry 4.0 in general.

As a next step, one researcher reviewed parts of the material applying the defined rules. The review was conducted using the software MAXQDA 11. The categories derived in this first loop were built closely from material. After the review of 30 publications, 31 row categories for requirements were derived. At this point, a revision of rules and category scheme was conducted using a cluster analysis to merge related aspects and to receive distinct categories. Each content of 22 resulting categories was then checked for reliability and its fit in the category. Therefore, a code description and anchor examples for each category have been defined. Based on the resulting categories and codes the final examination of the material was conducted. A summative check of reliability was performed upon the coding of a second researcher. Therefore, the second researcher was instructed in object, objectives, research question and rules and then analysed the complete material. Mayring (2008) proposes a reliability of at least 70% for acceptable results of a content analysis. The summative reliability was calculated according to Holsti (1969) and proves the reliability of this analysis. Table 2 shows the aggregated reliabilities for each main category.

Description / Category	C10	C20	C30	C40	C50	C60	Total
Codings of Coder 1	27	34	23	73	27	29	213
Codings of Coder 2	25	30	28	81	30	27	221
Matching Codings	22	26	19	54	24	22	167
Summative Reliability	85%	81%	75%	70%	84%	79%	77%

Table 2. Number of codings and summative reliability by main category

4 Results

The result of the content analysis is a structured compilation of Industry 4.0 requirements regarding data processing. It provides a comprehensive view on the data that need to be processed and on the processing of that data in an Industry 4.0 production. Analysed publications only address certain aspects or solution-components of Industry 4.0 and rarely describe requirements or structures for requirements. Table 3 shows the resulting category scheme with 6 main categories and 22 subcategories describing object, subject and conditions of data processing. According to our two research questions, it is grouped in requirements for data and for processing of data.

The first main category '*Data Model'* (C10) shows requirements for characteristics of data, structure and sources to integrate in the context of Industry 4.0. The subcategory '*Unify semantics'* contains requirements for a unified description of information and meanings in production. The unification of interfaces between entities in production are content of the next subcategory '*Unify interfaces*'. Together, semantics and interfaces address communication and data exchange among CPS in a comprehensive production network which various systems and objects. The second main category '*Data Integration'* (C20) refers to different perspectives of data integration within an enterprise and beyond. The first subcategory '*Integrate life cycle'* contains requirements to integrate life cycle data of CPS in engineering and operation processes. The next subcategory '*Integrate horizontally'* focuses on requirements to integrate data along the value chain in an entire production network. The third subcategory '*Integrate vertically*' contains requirements to integrate data from the automation pyramid (enterprise-, control-, device- and sensor-level). The third main category '*Data content*' (C30) shows requirements for necessary data to be processed in Industry 4.0. Necessary data comprises product data, production process data, business data and data from sensors and actors.

The fourth main category '*Decision processing'* (C40) refers to requirements for autonomous, decentralized self-control and self-optimization performed in CPS networks. Requirements to build CPS networks dependent on environment and current situation are dedicated to subcategory '*Ad-hoc networking*'. The next subcategory '*Optimize network*' focuses on requirements for overall system goals and optimization when local decisions are made by CPS. The subcategory '*Admit autonomy*' contains requirements for autonomy and freedom in decision processes in CPS networks. Requirements for utilization of models of the current production are part of the subcategory '*Utilize models'*. The next sub-category '*Monitor conditions*' contains requirements for CPS to monitor, diagnose and perform actions online. The fifth main category '*Knowledge processing*' (C50) refers to requirements for processing of actual and past data to generate additional value for decisions. The subcategory '*Detect patterns*' contains requirements to use data analytics to identify optimization potentials in production. Requirements to prepare, filter and compile data to support decision-making are part of the subcategory '*Prepare data'*. The next subcategory '*Transform know-how*' contains requirements to transform expert knowledge and experience in information models. Requirements to use past data for predictive analysis are dedicated to the subcategory '*Predict parameters*'. The sixth main category

'*Real-time*' (C60) focuses on requirements for processing performance. Requirements for access the status of entities in real-time are part of the subcategory '*Access status*'. The next subcategory '*Access description*' contains requirements for access the description of an entity, in real-time. Real-time requirements to build CPS networks are dedicated to the subcategory '*Build networks*'. The subcategory '*Control production*' contains requirements for real-time operative production control.

DATA REQUIREMENTS			
Main category	Subcategory	Freq.	Requirement Description
Data model (C10)	Unify semantics (C11)	15	Unify information models and meanings
	Unify interfaces (C12)	12	Unify interfaces and communication
Data integration (C20)	Integrate life cycle (C21)	10	Integrate data along the life cycle of CPS
	Integrate horizontally (C22)	13	Integrate data along the value chain and network
	Integrate vertically (C23)	11	Integrate data of the automation pyramid
Data content (C30)	Include product data (C31)	3	Include product data and description
	Include process data (C32)	7	Include production processes data and description
	Include business data (C33)	1	Include business data and parameters
	Include sensor data (C34)	12	Include sensor and actor data from CPS
PROCESSING REQUIREMENTS			
Main category	Subcategory	Freq.	Requirement Description
Decision processing (C40)	Ad-hoc networking (C41)	21	Build networks depending on situation
	Optimize network (C42)	18	Optimize network in local decision-making
	Admit autonomy (C43)	8	Admit autonomy in decision-making of CPS
	Utilize models (C44)	13	Utilize comprehensive models of real production
	Monitor conditions (C45)	13	Monitor, diagnose and perform actions online
Knowledge processing (C50)	Detect patterns (C51)	7	Detect patterns and similarities in production
	Prepare data (C52)	7	Prepare, compile and filter data
	Transform know-how (C53)	6	Transform know-how and expert knowledge
	Predict parameters (C54)	7	Predict decision parameters based on past data
Real-time processing (C60)	Access status (C61)	14	Access the status of CPS in real-time
	Access description (C62)	2	Access the description of CPS in real-time
	Build networks (C63)	4	Build CPS networks in real-time
	Control production (C64)	9	Control operative production in real-time

Table 3. Resulting Categories for data processing requirements of Industry 4.0.

5 Discussion

5.1 Data processing requirements of Industry 4.0

Industry 4.0 seeks to improve production processes and efficiency. This improvement requires a comprehensive integration of data (C20) and standardized semantics and interfaces (C10) to enable efficient communication and data exchange. Requirements for data integration include the horizontal, vertical and life cycle perspective. Many publications address the requirements regarding integration (Brettel et al., 2014; Vogel-Heuser et al., 2009), but only a few deal with the necessary data content. What is the content of the data? Requirements found, often address sensor data (C34) (Jatzkowski and Kleinjohann, 2014) or rarely data to describe the production process of a product (C32) (Denkena et al., 2014). Only the publication of Mertens (2014) focuses on business data (C33) as enabler for monetary evaluations and claims the reuse of concepts from the CIM era.

Efficient processing of this comprehensive data is another requirement of Industry 4.0. Decision processes in Industry 4.0 are performed by a random group of CPS that build an ad-hoc network (C41). The group builds upon individual situations and condition of environment in production, e. g. failure of a machine (Schuh et al., 2014). This requires random access on status and description of entities in production. Autonomous and decentralized control of CPS in production also require goals for decision-making. In Industry 4.0, the requirement for overall system goals ranks high in frequency. The goal is to optimize the overall value chain or rather the overall production network when decisions are made by decentralized entities (C42) (Brettel et al., 2014). This implies the requirement to access and process of data of the entire production network within decision processes. Further enhancement of decision process in Industry 4.0 require the usage of a wide range of extracted and formalized knowledge (C50) generated from historical data (Lee et al., 2014). This knowledge can be used to determine parameters or predictions for decision processes. Sensors in the operative production deliver data in cycles of milliseconds (Bauernhansl et al., 2014). As a result, there is a huge and continuously growing amount of historical data which need to be processed. The requirement for real time access on status data of CPS (C61) ranks high in frequency, but a differentiated specification of real-time is missing in this main category. Consensus in hard real-time requirements can be found in the operative production control (C64) (Lee, 2008).

5.2 Implications for Big Data solutions in Industry 4.0

The results cover a wide range of Industry 4.0 data processing requirements. We see four issues requiring high performance processing of large data volumes and appropriate Big Data approaches.

(1) *Large and continuously growing amount of data*: The comprehensive data integration (vertical, horizontal, lifecycle) (C20) generates Big Data to access and process. The data comprises of active data (e.g. status and description of entities in the operative production) and growing passive or past data from the life cycle of entities (e.g. sensor data from machines).

(2) *Knowledge processing*: Processing of a large and complex amount of past data (e.g. sensor data) for analytics, mining and prognosis (C50) to enhance decision-making and for optimization.

(3) *Random access on data*: Individual situations and events in production are trigger for CPS to build ad-hoc networks for decision-making (C41). This requires random access on status and description of all entities within the overall production network (C42).

(4) *Real-time access and processing*: Operative production requires real-time control and decision-making (C60). Decision processes have to consider overall system goals and optimization (C42). This requires processing of comprehensive models (C44) and accessing network data, in real-time.

Comparing these four issues with the capabilities of Big Data solutions (Table 1) leads to the finding, that two unique Big Data approaches are necessary. The issues form two general Big Data use cases with fundamental differences regarding access and processing patterns and underlying data. The first use case *Data Mining* handles time consuming data analytics, mining and prognosis on large amounts of passive data (C50). Real-time queries and random access on data are not crucial in this case. This requires Big Data solutions that support batch processing and distributed computing e.g. Hadoop HDFS with MapReduce. The second use case *Entity Access* performs ad-hoc queries on entity data from the overall network for operative decision-making (C40, C60). This requires Big Data solutions that support real time queries and random access. Depending on infrastructure and application scenario Cassandra, MongoDB or SimpleDB could be a relevant software solution. Both use cases have a general character and require individual adaption to the context of application.

Use Case	Data Mining	Entity Access
Description	Analytics, mining and prognosis on past data for enhanced decision-making and for optimization.	Ad-hoc access and processing of actual entity data from the overall network for operative decision-making.
Related category	C50	C40, C60
Data	Passive (e.g. sensor data)	Active (e.g. status, description of entities)
Data amount	Growing	Constant
Real-time queries	No	Yes
Random access	No	Yes
Software solution	e.g. Hadoop HDFS + MapReduce	e.g. Cassandra, MongoDB or SimpleDB

Table 4. Characteristics of identified use cases for Big Data applications in Industry 4.0.

6 Limitations

There are possible limitations that need to be considered when interpreting the results. The analysed publications are limited to specified sources and search terms. Further sources and terms might be relevant, e.g. from neighbouring scientific disciplines. Not all aspects of Industry 4.0 which might contribute to data processing requirements were considered, e.g. data security or IT-architecture. Additionally the building of categories might be biased by the selected method and the author´s background and insights. The calculation of reliability following Holsti (1969), neglects effects of aggregation and number of categories. Only a brief overview is given on characteristics of representative Big Data software solutions and the high level assignment might not suit to individual use cases.

7 Conclusions

Industry 4.0 stands for the 4th Industrial revolution and the new production paradigm of autonomous and decentralized control. It involves a new level of data integration and data processing in industrial production. The structured elaboration of data processing requirements is essential for appropriate application of solution-components (e. g. Internet of Things, Cloud or Big Data) and for the further development and implementation of Industry 4.0. Surprisingly, data processing requirements or general requirements of Industry 4.0 are rarely subject of current research publications. The majority of publications addresses Industry 4.0 from a technological point of view, highlighting new possibilities without elaboration of requirements.

This research provides a category scheme for data processing requirements of Industry 4.0. The scheme can be used to match requirements and capabilities of solution-components to get new insights of necessary application. For Big Data and high performance processing of large data volumes in Industry 4.0, we identified the two general use cases *Data Mining* and *Entity Access*, requiring unique Big Data approaches. The latter has not yet been described in the context of Industry 4.0.

The identified Big Data use cases are representatives that need to be adapted to the individual context of application. Further research is needed for a complete and detailed description of data processing requirements of Industry 4.0, to fully understand implications on IT-infrastructure and IT-solution-components. This could result in a framework for data processing requirements of Industry 4.0.

References

Agrawal, D., Das, S. and Abbadi, A. El. (2011), “Big data and cloud computing: current state and future opportunities”, *14th International Conference on Extending Database Technology*, pp. 1–4.

Bauernhansl, T., Hompel, M. ten and Vogel-Heuser, B. (2014), *Industry 4.0 in Produktion, Automatisierung und Logistik*, Springer Vieweg Wiesbaden.

Brettel, M., Friederichsen, N., Keller, M. and Rosenberg, M. (2014), "How Virtualization , Decentralization and Network Building Change the Manufacturing Landscape : An Industry 4.0 Perspective", *International Journal of Mechanical, Aerospace, Industrial and Mechatronics Engineering*, Vol. 8 No. 1, pp. 37–44.

Cattell, R. (2011), "Scalable SQL and NoSQL Data Stores", *ACM SIGMOD Record*, Vol. 39 No. December, pp. 12–27.

Chaganti, P. and Helms, R. (2010), *Amanzon Simple DB Developer Guide*, Packt Publishing, p. 252.

Chodorow, K. (2013), *MongoDB: the definitive guide*, O'Reily Media. Inc.

Denkena, B., Schmidt, J. and Krüger, M. (2014), "Data Mining Approach for Knowledge-based Process Planning", *International Conference on System-Integrated Intelligence*, Vol. 15, pp. 406–415.

Gupta, R., Gupta, H. and Mohania, M. (2012), "Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?", in Srinivasa, S. (Ed.),*Big data analytics : first international conference, BDA 2012, New Delhi, India, December 24 - 26, 2012 ; proceedings*, Springer, Heidelberg, pp. 42–61.

Hewitt, E. (2010), *Cassandra: the definitive guide*, O'Reily Media. Inc.

Holsti, O.R. (1969), *Content Analysis for the Social Science and Humanities*, Addison-Wesley, Reading.

Jatzkowski, J. and Kleinjohann, B. (2014), "Towards Self-reconfiguration of Real-time Communication within Cyber-physical Systems", *International Conference on System-Integrated Intelligence*, Vol. 15, pp. 54–61.

Kagermann, H., Wahlster, W. and Helbig, J. (2013), *Recommendations for implementing the strategic initiative Industrie 4.0*, Frankfurt/Main.

Kaur, K. and Rani, R. (2013), "Modeling and Querying Data in NoSQL Databases", *IEEE International Conference on Big Data*, pp. 1–7.

Klenke, K. (2008), *Qualitative research in the study of leadership*, Emerald group publishing, Bingley.

Krippendorff, K. (2004), *Content Analysis: An Introduction to Its Methodology*, Sage Publications.

Lee, E.A. (2008), "Cyber Physical Systems: Design Challenges", *IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, pp. 363–369.

Lee, J., Kao, H.-A. and Yang, S. (2014), "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment", *CIRP Conference on Industrial Product-Service Systems*, Vol. 16, pp. 3–8.

Lee, J., Lapira, E., Bagheri, B. and Kao, H. (2013), "Recent advances and trends in predictive manufacturing systems in big data environment", *Manufacturing Letters*, Society of Manufacturing Engineers (SME), Vol. 1 No. 1, pp. 38–41.

Madisetti, V. and Bahga, A. (2014), *Internet of Things - A Hands-On Approach*, VTP, 1. Edition.

Mayring, P. (2000), "Qualitative Content Analysis", *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol. 1 No. 2.

Mayring, P. (2008), *Qualitative Inhaltsanalyse*, Beltz Verlag, Weinheim Basel.

Mertens, P. (2014), "Industrie 4.0 = CIM 2.0", *Industrie Management*, Vol. 30 No. 1, pp. 28–30.

Monostori, L. (2014), "Cyber-physical Production Systems: Roots, Expectations and R&D Challenges", *CIRP Conference on Manufacturing Systems*, Vol. 17, pp. 9–13.

Myers, M.D. (1997), "Qualitative Research in Information Systems", *MISQ Quaterly*, Vol. 21 No. 2, pp. 241–242.

Pokorny, J. (2013), "NoSQL databases: a step to database scalability in web environment", *International Journal of Web Information Systems*, Vol. 9 No. 1, pp. 69–82.

Schmarzo, B. (2013), *Big Data: Understanding How Data Powers Big Business*, John Wiley & Sons, Indianapolis.

Schuh, G., Potente, T., Varandani, R., Hausberg, C. and Fränken, B. (2014), "Collaboration Moves Productivity to the Next Level", *CIRP Conference on Manufacturing Systems*, Vol. 17, pp. 3–8.

Sztipanovits, J., Ying, S., Corman, D., Davis, J., Mosterman, P.J., Prasad, V. and Stormo, L. (2013), *Strategic R&D Opportunities for 21st Century Cyber-Physical Systems*.

Verl, A., Lechler, A., Wesner, S., Kirstädter, A., Schlechtendahl, J., Schubert, L. and Meier, S. (2013), "An Approach for a Cloud-based Machine Tool Control", *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 7, pp. 682–687.

Vogel-Heuser, B., Kegel, G., Bender, K. and Wucherer, K. (2009), "Global Information Architecture for Industrial Automation", *ATP*, Vol. 1 No. 2, pp. 108–115.

Westkämper, E. (2011), "New Paradigms of Advanced Manufacturing Engineering", *17th International QFD Symposium, ISQFD, Stuttart*.

White, T. (2012), *Hadoop: the definitive guide*, O'Reily Media. Inc.

Appendix (References for content analysis)

Anderl, R., Strang, D., Picard, A. and Christ, A. (2014), "Integriertes Bauteildatenmodell - Informationsträger für cyber-physische Produktionssysteme", *ZWF*, Vol. 109 No. 1-2, pp. 64–69.

Aruväli, T., Maass, W. and Otto, T. (2014), "Digital Object Memory Based Monitoring Solutions in Manufacturing Processes", *DAAAM International Symposium on Intelligent Manufacturing and Automation*, Elsevier B.V., Vol. 69, pp. 449–458.

Atmosudiro, A., Faller, M. and Verl, A. (2014), "Durchgängige Datenintegration in die Cloud", *WT*, Vol. 104 No. H.3, pp. 151–155.

Auerbach, T., Kamps, S., Veselovac, D. and Klocke, F. (2013), "Digitales Technologiewissen durch intelligente Fertigungssysteme", *ZWF*, Vol. 108 No. 7-8, pp. 7–8.

Aurich, J.C. and Meissner, H. (2014), "Entwicklung cybertronischer Produktionssysteme", *ZWF*, Vol. 109 No. 1-2, pp. 70–73.

Bauer, F., Gausemeier, J., Pruschek, P. and Rehage, G. (2013), "Arbeitsvorbereitung 4.0 - Cloud-basierte Nutzung virtueller Werkzeugmaschinen", *WT*, Vol. 103 No. H2, pp. 118–123.

Bauer, W., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T. and Schlund, S. (2014), "Industrie 4.0 – flexiblere und reaktionsfähigere Produktionsarbeit", *WT*, Vol. 104 No. H.3, pp. 134–138.

Bauernhansl, T. (2014), "Revolution geht weiter", *WT*, Vol. 104 No. H.3, p. 70569.

Bauernhansl, T., Schatz, A. and Jäger, J. (2014), "Komplexität bewirtschaften – Industrie 4 . 0 und die Folgen", *ZWF*, Vol. 109 No. Bild 1, pp. 347–350.

Blunck, H. and Windt, K. (2013), "Komplexität schafft Spielraum für Selbststeuerung", *WT*, Vol. 103 No. 2, pp. 109–113.

Brecher, C., Özedemir, D., Ecker, C., Eilers, J. and Lohse, W. (2013), "Modellbasierte Rekonfigurierbarkeit", *WT*, Vol. 103 No. H.2, pp. 157–161.

Brettel, M., Friederichsen, N., Keller, M. and Rosenberg, M. (2014), "How Virtualization , Decentralization and Network Building Change the Manufacturing Landscape : An Industry 4.0 Perspective", *International Journal of Mechanical, Aerospace, Industrial and Mechatronics Engineering*, Vol. 8 No. 1, pp. 37–44.

Brikhold, M. and Lechler, A. (2014), "Modellierung von Automatisierungssystemen nach Vorgaben des BSI", *WT*, Vol. 104 No. H.5, pp. 301–306.

Dangelmaier, W., Franke, H., Pape, U. and Rüther, M. (2002), "Dezentrale, intelligente Multiagentensysteme", *ZWF*, Vol. 97 No. 12, pp. 640–643.

Denkena, B., Boujnah, H., Henning, H., Schmidt, J. and Meyer, J. (2013), "Anwendungspotentiale gentelligenter Bauteile", *WT*, Vol. 103 No. H2, pp. 152–156.

Denkena, B., Litwinski, K.M., Dahlmann, D., Haythem, B., Kiesner, J., Krüger, M., Schmidt, J., et al. (2014), "Industrie 4.0 in der Zerspanung", *ZWF*, Vol. 109 No. 7-8, pp. 537–541.

Denkena, B., Schmidt, J. and Krüger, M. (2014), "Data Mining Approach for Knowledge-based Process Planning", *International Conference on System-Integrated Intelligence*, Vol. 15, pp. 406–415.

Dombrowski, U. and Wagner, T. (2014a), “Arbeitsbedingungen im Wandel der Industrie 4.0”, *ZWF*, Vol. 109 No. 5, pp. 351–355.

Dombrowski, U. and Wagner, T. (2014b), “Mental Strain as Field of Action in the 4th Industrial Revolution”, *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 17, pp. 100–105.

Elsweier, M., Nyhuis, P. and Nickel, R. (2010), “Assistenzsystem zur Diagnose in der Produktionslogistik”, *ZWF*, Vol. 105 No. 6, pp. 562–568.

Engelhardt, P., Weidner, M., Einsank, W. and Reinhart, G. (2013), “Situationsbasierte Steuerung modularisierter Produktionsabläufe”, *ZWF*, Vol. 108 No. 12, pp. 987–991.

Fleischer, J., Lanza, G., Appel, D. and Stricker, N. (2013), “Life Cycle Performance 4 . 0 Strategische und technische Lösungen für den intelligenten Betrieb von Maschinen und Anlagen”, *WT*, Vol. 103 No. H.2, pp. 124–129.

Franke, J., Merhof, J., Fischer, C. and Risch, F. (2010), “Intelligente Steuerungskonzepte für wandlungsfähige Produktionssysteme”, *Industrie Management*, Vol. 26 No. 2, pp. 61–64.

Franke, J., Merhof, J. and Hopfensitz, S. (2010), “Einsatz von dezentralen Multiagentensystemen”, *ZWF*, Vol. 105, pp. S.1075–1078.

Frazzon, E.M., Hartmann, J., Makuschewitz, T. and Scholz-Reiter, B. (2013), “Towards Socio-Cyber-Physical Systems in Production Networks”, *CIRP Conference on Manufacturing Systems*, Vol. 7, pp. 49–54.

Gausemeier, J., Dumitrescu, R., Jasperneite, J., Kühn, A. and Trsek, H. (2014), “Der Spitzencluster it ´ s OWL auf dem Weg zu Industrie 4 . 0”, *ZWF*, Vol. 109 No. 5, pp. 336–346.

Gausemeier, J., Ebbesmeyer, P. and Steffen, D. (2005), “Selbstoptimierung als Perspektive für die Gestaltung intelligenter Fertigungssysteme”, *ZWF*, Vol. 100 No. 9, pp. 495–501.

Gronau, N. (2013), “Bionic Manufacturing - Steuerung der Fabrik mit Mitteln der Natur im Zeitalter von Industrie 4.0”, *Industrie Management*, Vol. 29 No. 6, pp. 12–16.

Grundstein, S., Schukraft, S., Görges, M. and Scholz-reiter, B. (2013), “Autonome Produktionssteuerung bei zentraler Produktionsplanung”, *ZWF*, Vol. 108 No. 12, pp. 992–994.

Günthner, W., Hompel, M. ten, Chisu, R., Nettsträter, A. and Roidl, M. (2010), “Das Internet der Dinge in der Intralogistik”, *WT*, Vol. 100 No. H3, pp. 146–151.

Håkansson, A. and Hartung, R. (2014), “An Infrastructure for Individualised and Intelligent Decision-making and Negotiation in Cyber-physical Systems”, *International Conference on Knowledge-Based and Intelligent*, Elsevier Masson SAS, Vol. 35, pp. 822–831.

Han, C., Jornet, J.M., Fadel, E. and Akyildiz, I.F. (2013), “A cross-layer communication module for the Internet of Things”, *Computer Networks*, Vol. 57 No. 3, pp. 622–633.

Harjes, F. and Scholz-Reiter, B. (2014), “Autonomous Control in Closed Dynamic Logistic Systems”, *International Conference on System-Integrated Intelligence*, Elsevier B.V., Vol. 15, pp. 313–322.

Hering, N., Brandenburg, U. and Kropp, S. (2013), “Energieeffiziente Produktions- planung und -regelung”, *ZWF*, Vol. 108 No. 10, pp. 783–787.

Herkommer, O. and Hieble, K. (2014), “Ist Industrie 4.0 die nächste Revolution in der Fertigung ?”, *Industrie Management*, Vol. 30 No. 1, pp. 42–46.

Herkommer, O., Kauffmann, M. and Hieble, K. (2014), “Erfolg der Produktion von morgen”, *ZWF*, Vol. 109 No. 3, pp. 153–158.

Holtewert, P., Wutzke, R., Seidelmann, J. and Bauernhansl, T. (2013), “Virtual Fort Knox Federative, Secure and Cloud-based Platform for Manufacturing”, *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 7, pp. 527–532.

Hutle, M. (2013), “Zukunft der Industriesysteme”, *Industrie Management*, Vol. 29 No. 1, pp. 62–64.

Jatzkowski, J. and Kleinjohann, B. (2014), “Towards Self-reconfiguration of Real-time Communication within Cyber-physical Systems”, *International Conference on System-Integrated Intelligence*, Vol. 15, pp. 54–61.

Jentsch, D., Riedel, R., Jäntsch, A. and Müller, E. (2013), “Fabrikaudit Industrie 4.0”, *ZWF*, Vol. 108 No. 9, pp. 678–681.

Kagermann, H., Wahlster, W. and Helbig, J. (2013), *Recommendations for implementing the strategic initiative Industrie 4.0*, Frankfurt/Main.

Klocke, F., Veselovac, D. and Keitzel, G. (2013), "Cloudbasierte Informationssysteme", *WT*, Vol. 103 No. H.2, pp. 90–95.

Krenz, P. and Wulfsberg, J.P. (2012a), "Unfold Collective Intelligence ! (Teil 1)", *ZWF*, Vol. 107, pp. 152–157.

Krenz, P. and Wulfsberg, J.P. (2012b), "Unfold Collective Intelligence ! (Teil 2)", *ZWF*, Vol. 107 No. 5, pp. 349–354.

Krückhans, B. and Meier, H. (2013), "Industrie 4.0 – Handlungsfelder der Digitalen Fabrik zur Optimierung der Ressourceneffizienz in der Produktion Stand der Technik", *HNI Verlagsschriftenreihe*, pp. 31–40.

Landherr, M., Holtewert, P., Kuhlmann, T., Lucke, D. and Bauernhansl, T. (2013), "Virtual Fort Knox", *WT*, Vol. 103 No. H.2, pp. 146–151.

Lanza, G., Stricker, N. and Peters, S. (2013), "Ad-hoc Rescheduling and Innovative Business Models for Shock- robust Production Systems", *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 7, pp. 121–126.

Lappe, D., Veigt, M., Franke, M., Kolberg, D. and Schlick, J. (2014), "Vernetzte Steuerung einer schlanken Intralogistik", *WT*, Vol. 104 No. H.3, pp. 112–117.

Lee, E.A. (2008), "Cyber Physical Systems: Design Challenges", *IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, pp. 363–369.

Lee, J., Kao, H.-A. and Yang, S. (2014), "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment", *CIRP Conference on Industrial Product-Service Systems*, Vol. 16, pp. 3–8.

Lee, J., Lapira, E., Bagheri, B. and Kao, H. (2013), "Recent advances and trends in predictive manufacturing systems in big data environment", *Manufacturing Letters*, Society of Manufacturing Engineers (SME), Vol. 1 No. 1, pp. 38–41.

Lentes, J., Eckstein, H. and Zimmermann, N. (2014), "amePLM – eine Plattform zur Informationsbereit - stellung in der Produkt- und Produktionsentstehung", *WT*, Vol. 104 No. H.3, pp. 146–150.

Letmathe, P., Maslan, N. and Zeise, P. (2013), "Ressourceneffizienz in Produktionsprozessen", *ZWF*, Vol. 108 No. 6, pp. 449–452.

Lin, K.-J. and Panahi, M. (2010), "A real-time service-oriented framework to support sustainable cyber-physical systems", *IEEE International Conference on Industrial Informatics*, pp. 15–21.

Lütjen, M. and Pattanashetti, P. (2014), "Remote Manufacturing", *Industrie Management*, Vol. 30 No. 4, pp. 16–20.

Marczinski, G. (2014), "Fabrikleistung wirksam steigern Schlank Automatisiert Autonom", *ZWF*, Vol. 109 No. 5, pp. 356–360.

Mertens, P. (2014), "Industrie 4.0 = CIM 2.0", *Industrie Management*, Vol. 30 No. 1, pp. 28–30.

Monostori, L. (2014), "Cyber-physical Production Systems: Roots, Expectations and R&D Challenges", *CIRP Conference on Manufacturing Systems*, Vol. 17, pp. 9–13.

Morlock, F. and Dorka, T. (2014), "Automatisierungspyramide für die integrierte Sach- und Dienstleistungserbringung", *WT*, Vol. 104 No. H.7/8, pp. 448–452.

Müller, R., Vette, M., Leenhard, H. and Speicher, C. (2014), "Cyber-physisch geprägte Montageplanungsmethoden", *WT*, Vol. 104 No. H.3, pp. 124–128.

Neumann, M. and Westkämper, E. (2014), "Method for a Situation-based Adaptation and Validation of the Manufacturing Capability of Assembly Systems", *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 17, pp. 118–123.

Otto, J., Henning, S. and Niggemann, O. (2014), "Why Cyber-physical Production Systems Need a Descriptive Engineering Approach – A Case Study in Plug & Produce", *International Conference on System-Integrated Intelligence*, Elsevier B.V., Vol. 15, pp. 295–302.

Overmeyer, L., Kleinert, S., Radosavac, M., Krühn, T., Heißmeyer, S. and Bentlage, A. (2013), "Kognitive Transportsysteme durch optische Technologien", *WT*, Vol. 103 No. H.2, pp. 139–145.

Pal, K. and Karakostas, B. (2014), “A Multi Agent-based Service Framework for Supply Chain Management”, *International Conference on Ambient Systems, Networks and Technologies*, Elsevier Masson SAS, Vol. 32, pp. 53–60.

Plorin, D., Jentsch, D. and Riedel, R. (2013), “Ambient Assisted Production”, *WT*, Vol. 103 No. H2, pp. 135–138.

Rehder, E. and Schatz, A. (2014), “Reifegradbasierte Kopplung – koordinierte Selbststeuerung für dezentrale Produktionssysteme”, *WT*, Vol. 104 No. H.3, pp. 129–133.

Reinhart, G., Engelhardt, P., Geiger, F. and Philipp, T.R. (2013), “Cyber-Physische Produktionssysteme”, *WT*, Vol. 103 No. H.2, pp. 84–89.

Reinhart, G., Engelhardt, P. and Genc, E. (2013), “RFID-basierte Steuerung von Wertschöpfungsketten”, *WT*, Vol. 103 No. H.2, pp. 104–108.

Reinhart, G., Meis, J.-F., Reisen, K. and Schindler, S. (2013), “Bionik in der Produktionsorganisation”, *ZWF*, Vol. 108 No. 9, pp. 608–612.

Rekersbrink, H., Ludwig, B. and Scholz-reiter, B. (2007), “Entscheidungen selbst- steuernder logistischer Objekte”, *Industrie Management*, Vol. 23 No. 4, pp. 25–30.

Rüsing, E. (2014), “Herausforderungen in der inter- nationalen Fertigungsindustrie”, *Industrie Management*, Vol. 30 No. 4, pp. 57–61.

Sauer, O. (2013), “Informationstechnik für die Fabrik der Zukunft”, *Industrie Management*, Vol. 29 No. 1, pp. 11–14.

Scheifele, S., Friedrich, J., Lechler, A. and Verl, A. (2014), “Flexible, Self-configuring Control System for a Modular Production System”, *2nd International Conference on System-Integrated Intelligence*, Elsevier B.V., Vol. 15, pp. 398–405.

Schlechtendahl, J., Kretschmer, F., Lechler, A. and Verl, A. (2014), “Communication Mechanisms for Cloud based Machine Controls”, *CIRP Conference on Manufacturing Systems*, Vol. 17, pp. 830–834.

Schmiedgen, P., Noennig, J.R. and Nyhuis, P. (2014), “Resilienzsteigerung durch Wissenstransfer in CPPS”, *WT*, Vol. 104 No. H.3, pp. 164–168.

Schmitt, R. and Beaujean, P. (2007), “Selbstoptimierende Produktionssysteme”, *ZWF*, Vol. 102 No. 9, pp. 520–524.

Scholz-reiter, B., Böse, F., Jagalski, T. and Windt, K. (2007), “Selbststeuerung in der betrieblichen Praxis”, *Industrie Management*, Vol. 23 No. 3, pp. 7–10.

Schuh, G., Potente, T., Thomas, C. and Hauptvogel, A. (2013), “High Resolution Production Management”, *WT*, Vol. 103 No. H.2, pp. 96–99.

Schuh, G., Potente, T., Varandani, R., Hausberg, C. and Fränken, B. (2014), “Collaboration Moves Productivity to the Next Level”, *CIRP Conference on Manufacturing Systems*, Vol. 17, pp. 3–8.

Schuh, G., Potente, T., Wesch-Potente, C., Weber, A.R. and Prote, J.-P. (2014), “Collaboration Mechanisms to Increase Productivity in the Context of Industrie 4.0”, *CIRP Robust Manufacturing Conference*, Elsevier B.V., Vol. 19, pp. 51–56.

Shafiq, S.I., Sanin, C., Szczerbicki, E. and Toro, C. (2014), “Implementing Virtual Engineering Objects (VEO) with the Set of Experience Knowledge Structure (SOEKS)”, *Procedia Computer Science*, Elsevier Masson SAS, Vol. 35, pp. 644–652.

Sztipanovits, J., Ying, S., Corman, D., Davis, J., Mosterman, P.J., Prasad, V. and Stormo, L. (2013), *Strategic R&D Opportunities for 21st Century Cyber-Physical Systems*.

VDI/VDE. (2013), *Thesen und Handlungsfelder Cyber-Physical Systems : Chancen und Nutzen aus Sicht der Automation*, Düsseldorf.

VDI/VDE. (2014a), *Industrie 4.0 Statusreport: Gegenstände, Entitäten, Komponenten*, Düsseldorf.

VDI/VDE. (2014b), *Industrie 4.0 Statusreport: Wertschöpfungsketten*, Düsseldorf.

VDI/VDE. (2014c), *Industrie 4.0 Statusreport: Auf dem Weg zu einem Referenzmodell*, Düsseldorf.

Verl, A., Lechler, A., Wesner, S., Kirstädter, A., Schlechtendahl, J., Schubert, L. and Meier, S. (2013), “An Approach for a Cloud-based Machine Tool Control”, *CIRP Conference on Manufacturing Systems*, Elsevier B.V., Vol. 7, pp. 682–687.