International Journal of Computer Integrated Manufacturing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tcim20

Advancing manufacturing systems with big-data analytics: A conceptual framework

Dominik Kozjek, Rok Vrabič, Borut Rihtaršič, Nada Lavrač & Peter Butala

To cite this article: Dominik Kozjek, Rok Vrabič, Borut Rihtaršič, Nada Lavrač & Peter Butala (2020) Advancing manufacturing systems with big-data analytics: A conceptual framework, International Journal of Computer Integrated Manufacturing, 33:2, 169-188, DOI: 10.1080/0951192X.2020.1718765

To link to this article: https://doi.org/10.1080/0951192X.2020.1718765

INTERNATIONAL JOURNAL OF COMPUTER INTEGRATED MANUFACTURING
2020, VOL. 33, NO. 2, 169–188
https://doi.org/10.1080/0951192X.2020.1718765

Taylor & Francis
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Advancing manufacturing systems with big-data analytics: A conceptual framework

Dominik Kozjek[a], Rok Vrabič[a], Borut Rihtaršič[b], Nada Lavrač[c] and Peter Butala[a,d]

[a]Department of Control and Manufacturing Systems, University of Ljubljana, Ljubljana, Slovenia; [b]Litostroj Power, Ljubljana, Slovenia; [c]Jožef Stefan Institute, Ljubljana, Slovenia; [d]Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa

ABSTRACT

With the intensive development and implementation of information and communication technologies in manufacturing, large amounts of heterogeneous data are now being generated, gathered and stored. Handling large amounts of complex data – often referred to as big data – represents a challenge as there are many new approaches, methods, techniques, and tools for data analytics that open up new possibilities for exploiting data by converting them into useful information and/or knowledge.

However, the application of advanced data analytics in manufacturing lags behind in terms of penetration and diversity in comparison with other domains such as marketing, healthcare and business, meaning that the available data often remain unexploited. This paper proposes a new conceptual framework for systematically introducing big-data analytics into manufacturing systems. To this end, the paper defines a new stepwise procedure that identifies what knowledge and skills, and which reference models, software and hardware tools, are needed for the development, implementation and operation of data-analytics solutions in manufacturing systems. The feasibility of the proposed conceptual framework is demonstrated in a case study from an engineer-to-order company and by mapping the framework to several previous data-analytics projects.

ARTICLE HISTORY
Received 28 February 2019
Accepted 14 January 2020

KEYWORDS
Manufacturing system; data analytics; big data; conceptual framework

1. Introduction

The open question in manufacturing is how to produce products better, faster and more efficiently. One of the obstacles to effectively addressing of this question is the increased complexity of manufacturing systems, making them hard to manage. Market globalization, increased competition, and economic and political fluctuations are some of the factors that affect the complexity of manufacturing and call for an improved responsiveness, flexibility, robustness, resilience and adaptability of manufacturing systems.

Two major sources of complexity in manufacturing systems are the incompleteness of information (Peklenik 1995) and the incompleteness of knowledge (Suh 2005). While until recently the main problem was the lack of information sources, today industry is already being faced with the inverse problem. The intensive developments in information and communication technologies in the past decade and their implementation in manufacturing systems has resulted in the emergence of many new sources, which generate large volumes of various types of data at an increasing rate. However, due to the lack of knowledge and understanding, these data often remain unexploited as people simply do not know how to extract useful information and/or knowledge from the data, or they do not recognize the potential that is hidden in all this collected data.

It is now time to upgrade manufacturing systems by utilizing methods that will help to exploit the data, with the goal not only being to reduce the incompleteness of the information, but also to discover new insights and knowledge about manufacturing systems. This will enable better management of the complexity associated with manufacturing systems, and consequently, their better performance.

From the early beginnings of the intensified collection of manufacturing data a decade or two ago, until now, large amounts of digital data have been generated and stored. These data represent a tremendous potential for analyzing and discovering new knowledge that could contribute to the more efficient operation of manufacturing systems and their elements, such as processes, tools, other equipment, and of course people, but on the other hand represent a challenge in terms of their exploitation (Esmaeilian, Behdad, and Wang 2016;

CONTACT Dominik Kozjek dominik.kozjek@fs.uni-lj.si Department of Control and Manufacturing Systems, University of Ljubljana, Ljubljana, Slovenia

Rihtaršič and Sluga 2017). In parallel, computer scientists have developed various concepts, methods and data-analytics tools that enable the effective processing of data. Therefore, in manufacturing, new possibilities for analyzing the collected data, for discovering new knowledge, and for upgrading existing IT systems with new data-analytic tools, have emerged.

Data-driven analytics solutions have proved to be a powerful tool in a large number of studies. However, the performance and reliability of data-driven solutions often demands large amounts of training data. There already exist numerous approaches, methods, techniques and tools that can tackle the large size, dimensionality, dynamics, and other complex properties of data, but their customization for the manufacturing domain, new integration architectures and control algorithms, together with the willingness of the manufacturing stakeholders to use them, are needed (Babiceanu and Seker 2016).

Recently, the problem of the management and use of large and complex data has led to the development of the emerging *big-data* paradigm. The term *big data* denotes data whose effective management and use are not possible with conventional approaches, due to their size and/or other characteristics, such as a lack of structure, variability, speed, distributivity, diversity, incompleteness, un-credibility, un-verifiability, etc. (Babiceanu and Seker 2016; Boyd and Crawford 2012; Esmaeilian, Behdad, and Wang 2016; Gartner 2016; Hitzler and Janowicz 2013; Hurwitz et al. 2013; Laney 2001; Villars and Olofson 2011; Wang, Törngren, and Onori 2015). *Big-data analytics* can be perceived as a wide framework for extracting the value from such large and complex data. It provides approaches, methods, techniques and tools that together form efficient data-analytics systems. To successfully apply big-data analytics in manufacturing systems, skills and knowledge of information and communication technologies, and in particular data science, as well as the engineering know-how of manufacturing systems, and expert knowledge of manufacturing processes, need to be integrated.

However, the application of big-data analytics in the manufacturing domain lags behind in terms of penetration and diversity in comparison with other domains, such as marketing, healthcare, and business (Babiceanu and Seker 2016). In our view, the reason for this situation is mainly due to the problem of linking information and communication technologies and data science know-how with the engineering and expert knowledge of manufacturing systems and processes. The importance of the big-data paradigm for manufacturing is often highlighted, but in both industry and academia this paradigm is not defined concisely and there is a lack of general and practical reference models that would show how this paradigm can support the operation of manufacturing systems.

This paper proposes a new conceptual framework for introducing big-data analytics into manufacturing systems. The objective is to clarify the relation between the big-data paradigm and the manufacturing systems, and to practically and systematically show how to develop and implement data-analytics solutions in manufacturing systems.

1.1. Section summary and content overview

The paper addresses a general problem in production, i.e. managing the complexity of manufacturing systems. Two important sources of complexity that arise from the problem of managing and using large amounts of data are highlighted: (1) the incompleteness of information and (2) the incompleteness of knowledge. This is followed by a reasoning, explaining how this type of complexity can be better managed through the introduction of advanced data-analytics. The problem of the lack of knowledge about information and communication technologies and information science in the manufacturing domain, and vice-versa, is identified. As a solution to this problem, the use of tools and methods, introduced by the *big-data* paradigm, is proposed. It is explained how big-data analytics can help improve the complexity management in production and how new concepts derived from this paradigm will help introduce data analytics into manufacturing systems. The rest of the paper is structured as follows. First, a review of the literature and related concepts is given. The gap, addressed by the newly proposed concepts in this paper, is identified. The central part of the paper (Section 3) introduces a new conceptual framework that facilitates and accelerates the deployment of advanced data-analytics solutions in manufacturing systems. The feasibility and wide applicability of the framework are further demonstrated by mapping several existing data-analytics projects into the proposed framework. Finally, the paper concludes with a summary.

2. Big-data analytics in manufacturing

Numerous papers related to the application of big-data analytics in manufacturing were published in recent years. This clearly indicates the strong interest of researchers as well as the relevance of the topic. Research and applications of big-data analytics in manufacturing can be divided into (1) theoretical research on general models for introducing big-data analytics into manufacturing, examinations of the existing situations in industry and the development of conceptual solutions, and (2) the applied research and development of specialized dedicated solutions. Table 1 shows a possible classification of the related works found in the literature.

A common feature of these research studies is the use of basic concepts, such as the identification of new, potentially useful data sources, data integration and the innovative use of data in order to improve the performance of the observed system. The data used in such projects and the data that are intended for use with the developed data-analytics solutions originate from the manufacturing environment and from elsewhere, for example, from the internet, sensor networks at places of public events, etc. In several studies and projects, typical big-data technologies, such as NoSQL databases and the Hadoop software framework, are used.

The use of intelligent heuristic approaches and data-analysis techniques (e.g., machine learning) are often dictated by the speed and automation requirements. These methods generally prove to be effective for fault-diagnosis problems in complex manufacturing processes, where the process states are described by non-trivial patterns of a large number of parameters (Precup et al. 2015). Intelligent heuristic methods of analysis enable a high degree of automation and the recognition of complex patterns that go beyond human capabilities. Deep learning gives outstanding results in this field (Wang et al. 2018). However, these methods are still rarely used in practice. The reason for this is the lack of studies on holistic reference data-analytics solutions that would describe practical ways of presenting the results of analyses and their actual use in a real manufacturing environment, and which would provide a sufficient degree of confidence to the end user.

Various conceptual models as tools for assisting the introduction of data analytics in manufacturing systems have been developed. For example, Lechevalier, Narayanan, and Rachuri (2014) propose a domain-specific framework for the applications of predictive analytics in production. The main contributions by O'Donovan et al. (2015b) are a set of data and system requirements for implementing equipment-maintenance applications in industrial environments, and an information system model that provides a scalable and fault-tolerant big-data pipeline for integrating, processing and analyzing industrial equipment data. A framework for the conceptualization, planning and implementation of big-data projects in companies is presented by Dutta and Bose (2015). Zhang et al. (2017a) propose an overall architecture for big-data analytics for the purpose of making better product-lifecycle-management and cleaner-production decisions based on big data. Zhang et al. (2017b) propose a framework for big-data-driven product-lifecycle management to address challenges such as the lack of reliable data and valuable knowledge that can be employed to support the optimized decision making of product-lifecycle management. Tao et al. (2018) propose a data-driven smart-manufacturing framework that consists of four modules: the manufacturing module, the data-driver module, the real-time monitor module, and the problem-processing module. Jun, Lee, and Kim (2019) propose a cloud-based big-data analytics platform for manufacturing industry.

For introducing big-data analytics into manufacturing systems, besides models focused on the manufacturing domain, other more general reference models and concepts from other domains, and general well-known data-analytics reference models such as CRISP-DM (Chapman et al. 1999, 2000), KDD (Fayyad, Piatetsky-Shapiro, and Smyth 1996) and SEMMA (developed by the SAS Institute), can also be used.

The gap in the existing literature on big-data analytics in manufacturing is that the proposed concepts and solutions are either useful only for certain types of manufacturing problems, or are not sufficiently specific and do not describe in sufficient detail the data-analysis procedures and the elements that are needed for the development of specific data-analysis solutions in manufacturing systems. This gap is addressed by the conceptual framework proposed in the following section.

3. Conceptual framework for data analytics in manufacturing systems

In industry, the awareness of the potential and of the hidden value of manufacturing data is on the

Table 1. Related work.

Reference	General introd. model/concept. solut./ examination of existing situation in the industry	Specialized solution/ case study	Topic or/and Application field
Arnold 2016		✓	fault diagnostics
Hu et al. 2015		✓	fault diagnostics
Ing et al. 2017		✓	fault diagnostics
Kozjek et al. 2017b		✓	fault diagnostics
Kumar et al. 2016		✓	fault diagnostics
Lei et al. 2016		✓	fault diagnostics
Mangal and Kumar 2016		✓	fault diagnostics
Mohanty et al. 2015		✓	fault diagnostics
Vrabič, Kozjek, and Butala 2017		✓	fault diagnostics
Yin and Zhao 2016		✓	fault diagnostics
Yu 2016		✓	fault diagnostics
Bastani 2016		✓	predictive analytics
Butte and Patil 2016		✓	predictive analytics
Grolinger et al. 2016		✓	predictive analytics
Han and Chi 2016		✓	predictive analytics
Kozjek et al. 2017a		✓	predictive analytics
Wang, Du, and Xi 2015		✓	predictive analytics
Wang and Zhang 2016		✓	predictive analytics
D'Oca and Hong 2015		✓	system state monitoring
Fan et al. 2015		✓	system state monitoring
Green 2015		✓	system state monitoring
Kaewunruen 2014		✓	system state monitoring
Phillips et al. 2015		✓	system state monitoring
Hammer et al. 2017	✓		system state control
Kohlert and König 2016		✓	system state control
Koo, Piratla, and Matthews 2015		✓	system state control
Liu and Jiang 2016		✓	system state control
Thompson and Kadiyala 2014	✓	✓	system state control
Tsuda et al. 2015		✓	system state control, fault diagnostics
Kozjek et al. 2018b		✓	operations management
Afshari and Peng 2015		✓	product design support
Hazen et al. 2014	✓	✓	supply chain analysis
Zhong et al. 2016b	✓		Big Data for supply-chain management
Chan et al. 2018	✓		big data and machine-learning technologies for design
Liu et al. 2019b		✓	cloud processing of the traffic big data, logistics
Zhong et al. 2015b		✓	logistics trajectory discovery
Liu et al. 2019a		✓	optimal routes planning
Gölzer et al. 2015	✓	✓	modeling of production data
Lenz, Wuest, and Westkämper 2018	✓		machine tool data analytics
Modoni et al. 2017	✓		data integration
Marini and Bianchini 2016	✓	✓	modeling of production data
Zhong et al. 2016a		✓	data visualization
Shin, Woo., and Rachuri 2014	✓	✓	modeling of production data, forecasting energy consumption
Ansari, Glawar, and Nemeth 2019	✓		prescriptive maintenance
Chien, Liu, and Chuang 2015		✓	identifying causes for improving system performance
Crespino et al. 2016	✓		anomaly detection in aerospace product manufacturing
Kozjek et al. 2018a		✓	identifying the business and social networks in the domain of production
Papacharalampopoulos et al. 2016		✓	efficiency of data-acquisition and data-storage systems
Stark et al. 2014	✓		Life-cycle engineering
Xu, Li, and Lu 2016		✓	feature selection
Zhong et al. 2015a		✓	identifying causes for improving system performance
O'Donovan et al. 2015b	✓		information system model for scalable and fault-tolerant big-data analytics
Ismail, Truong, and Kastner 2019	✓		manufacturing process data analysis pipelines
Huber, Voigt, and Ngomo 2016	✓	✓	system architecture, fault diagnostics
Krumeich et al. 2014	✓		architecture for implementing predictive systems in companies
Yang et al. 2014	✓		architecture for a data-analysis system
Zhang et al. 2017a	✓		big-data analytics architecture for cleaner manufacturing and maintenance processes
Bilal et al. 2016a	✓		big-data architecture for construction-waste analytics

(*Continued*)

Table 1. (Continued).

Reference	General introd. model/concept. solut./ examination of existing situation in the industry	Specialized solution/ case study	Topic or/and Application field
Jun, Lee, and Kim 2019	✓	✓	Cloud-based big-data analytics platform using algorithm templates for manufacturing industry
Kang, Chien, and Yang 2016		✓	big-data platform for semiconductor manufacturing
Babiceanu and Seker 2015	✓		framework for manufacturing cyber-physical systems
Dutta and Bose 2015	✓		framework for managing a Big-Data project
Kong et al. 2014	✓		framework for network manufacturing in the big-data environment
Lechevalier, Narayanan, and Rachuri 2014	✓		framework for predictive analytics in manufacturing
Shao et al. 2012	✓		framework for interoperable sustainable manufacturing process analysis applications development
Tao et al. 2018	✓		framework for smart production
Wu et al. 2017	✓		fog computing-based framework for process monitoring and prognosis in cyber-manufacturing
Wang, Zhang, and Li 2016	✓		cloud-based and big-data centric framework for a smart factory
Zhang et al. 2017b	✓		framework for Big-Data-driven product-lifecycle management
Babiceanu and Seker 2016	✓		Big Data and virtualization for manufacturing cyber-physical systems
Lee et al. 2013	✓		predictive manufacturing systems in big-data environment
O'Donovan et al. 2015a	✓		big data in manufacturing
Wang and Wang 2016	✓		Big Data in cyber-physical systems
Rüßmann et al. 2015	✓		big data and Industry 4.0.
Lee, Kao, and Yang 2014	✓		service innovation and smart analytics for industry 4.0 and big-data environment
Xiang, Chen, and Jiang 2016	✓		manufacturing resources integration and sharing modes in big-data environment
Wang and Alexander 2016	✓		Big Data in additive manufacturing
Adhikari et al. 2016	✓		trust issues for big data
Dubey et al. 2016	✓		the impact of big data on world-class sustainable manufacturing
Bilal et al. 2016b	✓		Big Data in the construction industry
Lee et al. 2015a	✓		industrial big-data analytics and cyber-physical systems for future maintenance and service innovation
Lee et al. 2015b	✓		intelligent factory agents with predictive analytics for asset management
Kazuyuki 2017	✓		Big-Data use and innovation in Japanese manufacturing companies

increase. But the problem at hand is how to extract that value and from which data, as there are numerous methods and tools for managing and using complex and large volumes of data.

The situation calls for an interdisciplinary systemic approach. The following questions arise: (1) How to combine the data-analytics tools and the large amounts of generated manufacturing data, and (2) How to associate various experts in order to maximize the value gained from the generated data. These are the questions addressed in this paper and systematically answered by the proposed conceptual framework.

3.1. General description

The proposed conceptual framework is based on the findings published in the literature about (1) the approaches and solutions for the development and implementation of data-analytics solutions in the manufacturing domain, and (2) other general approaches and solutions for introducing data analytics (not only in the manufacturing domain), as well as on (3) the findings and experiences from several projects the authors have conducted in recent years, which include various experiments and developments of data-analytics solutions for manufacturing systems.

The definitions of two key terms used within the conceptual framework are given in Table 2.

To extract value from the data with the aim to boost the performance of manufacturing systems, data-analytics solutions must be developed and implemented in manufacturing systems. So, the questions are (1) how can these data-analytics solutions be developed and implemented, and (2) which elements are needed for the development and implementation. The proposed conceptual framework includes two core conceptual tools, i.e., two abstractions that show how data-analytics solutions can be developed and implemented in manufacturing systems and what is needed for that. These two core tools are (1) the structured scheme of elements (shown in Figure 1(a)) and (2) the reference procedure for the development and implementation of data-analytics solutions (shown in Figure 1(b) in the form of a functional diagram).

The framework elements are in the scheme (illustrated in Figure 1(a)) (1) in the first dimension (shown on the x-axis) arranged according to how strongly they are connected to the basic domains, i.e., on the left-hand side *big-data analytics*, on the right-hand side *manufacturing systems*, and the interdisciplinary domain in the middle, and (2) in the second dimension (the y-axis), the elements are classified into three levels of abstraction, i.e., from the most abstract level *knowledge and skills* on the top, the *reference models* level in the middle, down to the *implementational level* at the bottom.

The reference procedure, shown in Figure 1(b) in the form of a functional diagram, proposes a phase-by-phase procedure for the development and implementation of data-analytics solutions. For each phase (in Figure 1(b) illustrated as a square with the name of the phase in the middle), its inputs and outputs (input and output arrows on the left- and the right-hand sides of the square), and the most likely required controls (downward arrows on the upper side of the square) and mechanisms/resources (upward arrows on the underside of the square) are marked. This reference procedure is actually a central element of the conceptual framework as it is connected to and it interconnects all the other elements in the scheme, as indicated in Figure 1.

The following subsection describes the individual elements of the conceptual framework (listed in Figure 1(a)).

3.2. Framework elements

3.2.1. Implementational level

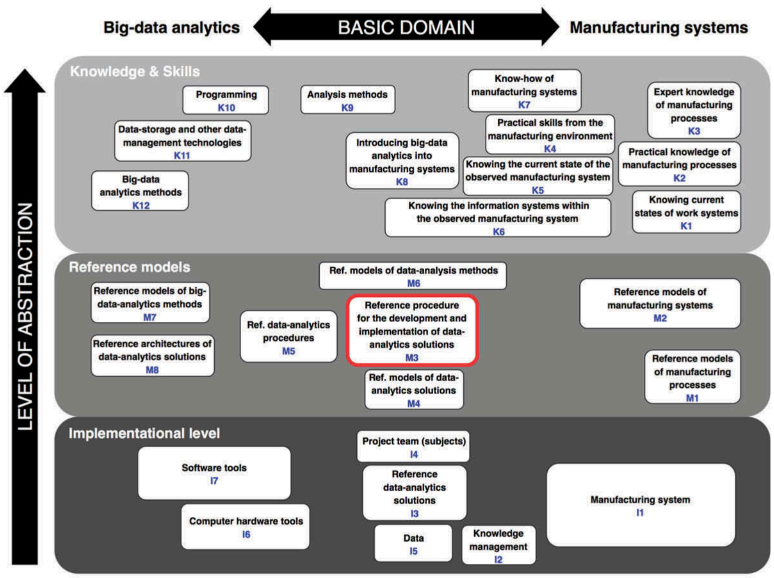
Manufacturing system. The basic element on the implementational level is the observed manufacturing system in which various processes, which might be supported by newly developed and implemented data-analytics solutions, are being carried out. It has to be pointed out that there are a huge variety of different processes that are running in manufacturing systems – from engineering processes (e.g., product design, process design), managerial processes (e.g., production management, project management, operation management), to production processes (e.g., machining, assembly). Each of these processes is specific in terms of the execution mechanisms, involved individuals and data, and are interconnected to other processes via inputs and outputs. So, no generalized analytic solutions can be developed, but reference models and possibilities for 'ad-hoc' development are needed.

Knowledge management. Knowledge management is used to manage the existing knowledge and new knowledge gained through the development, implementation and operation of data-analytics solutions.

Data. Various business, engineering and manufacturing IT systems have been implemented and integrated into manufacturing companies with the goal being to reduce the incompleteness of the information, ensure better decision making, improve traceability, etc. All these IT systems generate and store the data about processes, products, manufacturing resources, quality, energy management, maintenance

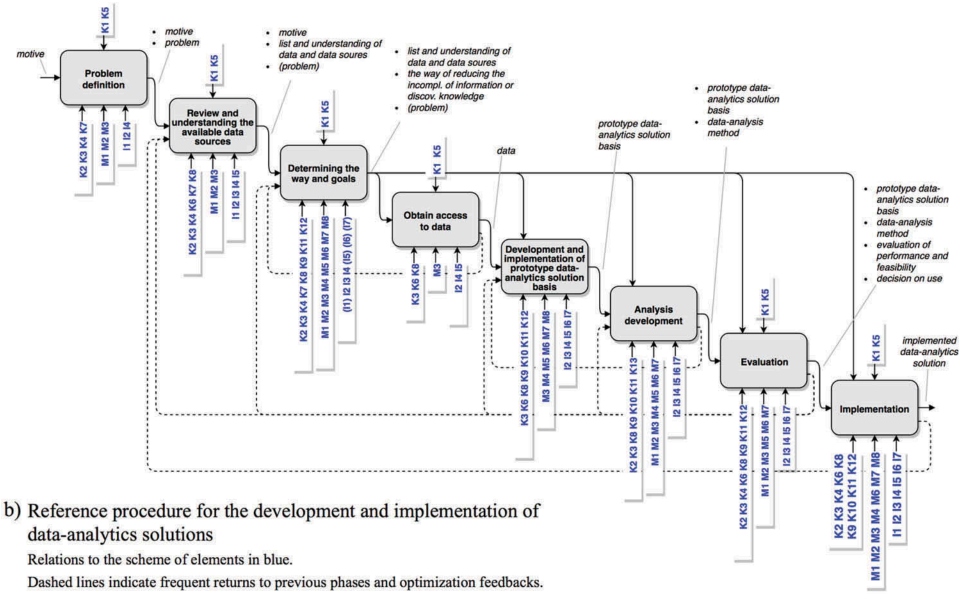
Table 2. Key terms.

Term	Definition
Conceptual framework	The conceptual framework is defined as an environment that includes relevant elements for a given purpose, which in this case is the development, implementation and operation of big-data analytics solutions in manufacturing systems.
Data analytics solution	A data-analytics solution is either in the form of an implemented repetitive data-analysis process (a tool) or in the form of discovered new knowledge. Its purpose is to extract value from the data.



a) Scheme of elements

The reference procedure is emphasized in red.



b) Reference procedure for the development and implementation of data-analytics solutions

Relations to the scheme of elements in blue.

Dashed lines indicate frequent returns to previous phases and optimization feedbacks.

Figure 1. Two core abstract tools of the proposed conceptual framework: a) Structured scheme of elements and b) Reference procedure for the development and implementation of data-analytics solutions.

and the production environment. In addition, a range of product-engineering data, such as CAD models, CNC programs, quality-control results, test measurements, etc., as well as business data on orders, supplies, deadlines, prices, etc., is available in digital form. Data in the manufacturing system can originate from process databases, resources and knowledge bases, they can be generated by machine controllers, they can be temporarily available in a machine controllers' memory unit, they can be collected via sensor networks, etc. Other data are, for example, the data from the products that are already in use, and internet data could also be available.

Project team. A project team is a group of individuals that manage and implement the development and implementation of a process for data-analytics solutions. On the one hand, due to the complexity of manufacturing systems and, on the other hand, due to the demanding field of information and communication technologies, the project team must ensure distinctive interdisciplinarity, which is difficult to achieve with a single individual. To ensure the required interdisciplinarity, project team members are likely to originate from several segments of a manufacturing system and from elsewhere, e.g., from a research institution that collaborates with a manufacturing company. Good communication and cooperation between the members of the project team are crucial for the successful and efficient development and implementation of data-analytics solutions.

Hardware and software tools. Another group of elements on the implementational level are the hardware and software tools required for the development and implementation of data-analytics solutions. Software tools can be divided into (1) tools for storing and managing big data, such as NoSQL databases (MongoDB, BigTable, Dynamo, etc.), the Hadoop software framework, etc., (2) tools for the analysis and mining of data, e.g., the programming language R, software tools and libraries: Rapidminer, Weka, ClowdFlows, Orange, Scikit-learn, Keras, Matlab, etc., (3) programming languages, e.g., Python, Java, and C, and (4) other tools, such as tools for visualization, data acquisition, etc.

Reference data-analytics solutions. When a data-analytics solution is developed and implemented, and if this solution can also be used in other projects, such a solution is referred to as a *reference data-analytics solution*.

3.2.2. Knowledge and skills

The key knowledge and skills from the manufacturing domain are (1) the engineering know-how of manufacturing systems, i.e., the knowledge of the functioning of manufacturing systems and their elements, building blocks, resources, processes, etc., (2) knowing the current state of the observed manufacturing system, (3) expert knowledge, (4) practical knowledge of manufacturing processes, (5) practical skills from the manufacturing environment, (6) knowing current states of work systems within the observed manufacturing system, and (7) knowing the information systems within the observed manufacturing system. This know-how is usually derived from individuals (or is accessible to individuals) that operate on different levels and segments of the manufacturing system.

The knowledge and skills from the domain of big-data analytics can be divided, according to Chen, Mao, and Liu (2014) and Grobelnik and Jaklič (2017), into: (1) programming, (2) knowing the data-storage and other data-management technologies, (3) typical big-data analytics methods, e.g., Bloom Filtering, Hashing, Indexing, Triel and parallel processing, and (4) analysis methods, e.g., clustering analysis, factor analysis, correlation analysis, regression analysis, A/B testing, statistical analysis and data-mining algorithms.

Interdisciplinary know-how in the middle between the basic domains is the knowledge and skills of introducing big-data analytics into manufacturing systems. The knowledge and skills are also one of the outputs of carrying out the processes of development and implementation of data-analytics solutions into manufacturing systems. They must be properly managed and used in subsequent projects.

3.2.3. Reference models

Reference models of manufacturing systems. Data analytics should be properly linked to understanding the structures and operation of manufacturing systems and their elements. Models of manufacturing systems can help with this. Reference models must enable the identification of manufacturing processes with the potential to improve their efficiency by reducing the incompleteness of information and by discovering new knowledge, and the identification of information flows and data sources.

Reference models of manufacturing processes. A reference model of a manufacturing process defines how a manufacturing process is carried out, which are the process steps or phases, what tools (resources) are needed, etc. While in the development of data-analytics solutions, reference models of manufacturing systems mainly assist in identifying the manufacturing processes, reference models of manufacturing processes are used for identifying the segments of manufacturing processes where the efficiency of these processes can be improved, and in searching for practical ways of reducing the incompleteness of the information and discovering new knowledge.

Reference architectures of data-analytics solutions. The purpose of reference architectures is to facilitate planning of the structure and behavior of the data-analytic system. The reference architecture supports the understanding of the structures, behaviors and interrelationships between the elements of the data-analytic system, which can be hardware and software tools, data models, data-management methods, data-analysis methods, visualization tools, etc.

An example of the reference architecture of the data-analytics solution is the so-called *value chain* (Chen, Mao, and Liu 2014; Hu et al. 2014). In the *value chain* concept, according to the system-engineering approach, a typical big-data analytics system is structured into four phases: (1) data generation, (2) data acquisition, (3) data storage, and (4) data analysis. In the case of using this reference architecture in the development of a data-analytics solution, each of these phases defines the necessary tools, appropriate methods, data flows within each phase and between phases, etc.

The data-analytics system can be divided into the level structure as proposed in (Hu et al. 2014). The level structure is composed of three levels: (1) the infrastructure layer, (2) the computing layer, and (3) the application layer.

Another example of the reference architecture is a technology-independent reference architecture, presented by Pääkkönen and Pakkala (2015). This reference architecture is based on the analysis of several big-data-analytics-system implementations. It is designed to facilitate the design of the architecture and the choice of technologies or commercial solutions in the development of data-analytics systems.

Reference models of big-data-analytics methods. The ways and approaches of the typical big-data-analytics methods (e.g., Bloom Filtering, Hashing, Indexing, Triel and parallel processing) are described by reference models in the form of pseudo codes, sequences of steps/operations, etc. These reference models are implemented in the hardware and software tools (implementational level), or directly in reference models of data-analytics solutions or in data-analytics solutions.

Reference models of data-analysis methods. In addition to big-data-analytics methods, the ways and approaches of data-analysis methods (e.g., clustering analysis, factor analysis, correlation analysis, regression analysis, A/B testing, statistical analysis and data-mining algorithms) are described by the reference models and implemented in the elements on the implementational level or in the data-analytics solutions.

Reference data-analytics procedures. Other general concepts, such as KDD (Knowledge Discovery in Databases) (Fayyad, Piatetsky-Shapiro, and Smyth 1996) and SEMMA (Sample, Explore, Modify, Model, Assess; developed by the SAS Institute), can be used within individual or in between the phases of the data-analytics solution's development and implementation procedure. The decision on the use of these models depends on the specific problem and the type of data-analytics solution.

Reference models of data-analytics solutions. In the development and implementation of data-analytics solutions, in some cases it is possible to use models of data-analytics solutions that are useful not only for a specific manufacturing system, process or data, they can be applied to other manufacturing systems, processes or data. Such a model is a reference data-analytics solution model.

Reference procedure for the development and implementation of data-analytics solutions. The reference procedure for the development and implementation of data-analytics solutions, shown in Figure 1(b), is the central element of the conceptual framework. It is presented in more detail in Section 3.3.

3.3. Reference procedure for the development and implementation of data-analytics solutions

In the paper, a new reference procedure for the development and implementation of data-analytics solutions is proposed. The proposed reference procedure is shown in Figure 1(b). It originates from the Cross Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al. 1999, 2000). It modifies and adjusts the CRISP-DM model, taking into account the following facts and requirements:

- Current situation in manufacturing systems: (1) the large size and variety of generated data, which most people do not even know exist; (2) the people that are managing IT systems in the company are usually not experts in other

informatics fields, e.g., in advanced approaches such as machine learning and artificial intelligence; (3) the people that would finally use data-analytics solutions and have a direct interaction with the data-analytics systems, usually do not fully understand the structure and operation of IT systems, advanced data-analysis methods, and the abilities and potential of data analytics; (4) the people who are well acquainted with data analytics usually do not fully understand the functioning of manufacturing systems and their processes; (5) the set of knowledge needed to understand the operation of manufacturing systems, their elements and manufacturing processes is too wide to be easily and rapidly conquered in practice by an individual data analyst in the phase of understanding the target domain, etc.
- Interdisciplinarity in the development and implementation of data-analytics solutions. The procedure should enable the integration of diverse knowledge, which is most often sourced from several people from various domains having limited communication possibilities.
- In a case of big-data analytics, unlike more conventional data analytics, the choice of technology and storage and management techniques often strongly depends on the subsequently used data-analysis methods, and on the properties of the data under consideration.
- The size and complexity of the data disable quick and easy modification of the data-storage and management parts of a data-analytics system.
- In manufacturing systems, in addition to the fundamental and/or self-evident problems related to the incompleteness of the information, often more specifically defined problems and potentially useful data-analytics solutions cannot be determined before the understanding of the available data and integrating this knowledge with knowing the possibilities offered by the advances in data analytics.

The proposed procedure is begun by the phases *problem definition, review and understanding of the available data sources*, and *determination of the way and goals*. In the phases of the *review and understanding of the available data sources*, and *determining the way and goals*, by integrating heterogeneous knowledge, innovative ways of reducing the incompleteness of the information and discovering new knowledge, and/or problems that are related to the incompleteness of the information and the lack of knowledge, and for which it was not previously known that they even exist, or it was previously believed that such solutions are impossible, might be discovered.

Due to the data's size, heterogeneity, security, and other reasons the phases of the development of data-analytics solutions often cannot be performed using data directly from locations where the data in the manufacturing system are stored, or they originate from; therefore, in the phase *obtain access to data* it is necessary to enable access to data, or if possible, to export a representative sample of the data, which is needed for the development of the data-analytics solution.

A prototype data-analytics solution plays an important role in development. The development of the prototype data-analytics solution is divided into two phases: (1) *development and implementation of a prototype data-analytics solution basis*, and (2) *analysis development*. In the phase of *development and implementation of prototype data-analytics solution basis*, the focus is mainly on the technology and techniques for data storage and management taking into account the potential analysis methods that will be used in the *analysis development* phase. In the *analysis development* phase, the focus is on finding the final analysis models.

The *evaluation* and *implementation* phases are performed after the *analysis development* phase.

Each phase of the reference procedure model is presented in more detail in Table 3. To better show the feasibility of the proposed reference procedure, in Table 3, for each phase, the description of its implementation in a case study from a typical engineer-to-order (ETO) company that manufactures industrial and energy equipment is given. The case of the ETO company is chosen here as it is the most complex example of a manufacturing system where the incompleteness of information is extremely high and additional knowledge derived from the data might significantly contribute to an improved performance.

Table 3. Reference-procedure phases.

Phase name	General description & Case study
Problem definition	**General description**: The *problem-definition* phase is intended to determine the specific problem to be addressed by the data-analytics solution. Examples of specific problems are the occurrence of unplanned machine stops during the observed manufacturing process, inaccurate estimation of the project's duration, time-consuming conventional search and insight into documentation of past engineering projects with the purpose of using the acquired experiences and knowledge from these past projects, etc. The knowledge and reference models that are needed at this initial stage are derived mostly from the manufacturing domain. **Case study**: In the observed company, for production, projects are structured according to the principle of work breakdown into smaller components, i.e., parts, modules, subassemblies, or higher-level tasks. To each of these components there corresponds one or more work orders. Each work order defines a sequence of operations that must be executed on different workstations or systems. Typically, several dozen projects and approximately one thousand work orders are in the production process at any given moment. One issue related to the operations management is that an actual sequence of work order's operations is not always the same as the planned one due to the dynamically changing situation in production, external and internal disturbances, requests for changes induced by customers, etc., meaning that the information on the planned sequence of operations is incomplete. A specific problem that needs to be addressed by a data-analytics solution is in this case the following: A planned sequences of work orders' operations sometimes does not match the actual ones.
Review and understanding the available data sources	**General description**: There is (in most cases) no easy and quick way to know and understand all the data sources and structures. Therefore, in practice, in addition to an important data-understanding phase, a comprehensive, systematic and effective overview of all the available data sources is needed. In this context, it is necessary to include both groups of subjects, i.e., the people that know and manage the IT systems in the company every day, as well as people who are experts and practically acquainted with the manufacturing system, its resources, and processes. The output of this phase, i.e., the list and understanding of the available data sources, needs to be effectively presented, and key findings and understandings need to be communicated to the data-analytics experts, i.e., people with knowledge and skills of data storage, management and analysis technologies, methods and techniques. **Case study**: The company operates with several data sources, i.e., an enterprise resource planning (ERP) system, manufacturing execution system (MES), Supervisory Control And Data Acquisition (SCADA) system, etc. The MES was developed and implemented in the company about a decade ago with the purpose, of establishing an independent information/control system and to introduce feedback loops based on real-time data acquisition in production. The MES includes numerical, categorical, and textual descriptions of manufacturing operations, work orders, work systems, etc. Until now, large amounts of data have been generated and stored. There are several real-world issues that need to be considered before further use of these data, i.e., missing records of operations, unintentional and intentional error entries, security issues related to privacy and the nondisclosure of business know-how, etc.
Determining the way and goals	**General description**: In this phase, the way to reduce the incompleteness of information or how to discover new knowledge needs to be roughly determined, e.g., which are potentially useful data sources, an estimation of the dimensionality and size of the data under consideration, the potential methods of analysis, which are the methods of evaluation and validation, the technologies and methods for data management, etc. At this stage it is necessary to use most of the knowledge and skills from both domains. Teamwork is crucial at this point, and a key role would be played by so-called data scientists (Davenport and Patil 2012; Grobelnik and Jaklič 2017), which would facilitate the cooperation and communication between individuals from various domains, and thus enable the integration of heterogeneous knowledge, experiences, and ideas. For the purpose of searching and demonstrating ideas, a demo data-analytics solution based on smaller data excerpts can also be developed at this stage. Knowing the conditions and possibilities, the goals of this development and implementation cycle need to be determined in relation to the problem addressed. **Case study**: In the observed company, large amounts of data, which includes information about work orders' planned and actually performed sequences of operations, and which holds descriptions of work orders and operations, have been collected and stored over the past few years. It is assumed that patterns which indicate whether a sequence of operations will change or not might be revealed from the numerical and textual descriptions of the work orders and operations (MES data). However, these patterns cannot be easily revealed by simple lookups into the database due to the size, dimensionality and variety of the data. Machine-learning techniques for prediction tasks seem to be a promising solution. The final application could be realized in the form of a plug-in program, e.g., when a planner is determining the sequence of operations of some work orders, the plug-in program informs the planner whether the actual work-order's sequence of operations is expected to be the same as the planned one, or not. The goals in this development and implementation cycle are to find out whether changes in the work orders' sequences of operations can be predicted, and what is the possible data-analysis method.
Obtain access to data	**General description**: In this phase, with the support of the people who know the IT and data-management systems in the observed manufacturing system and/or those who are familiar with the sources of the observed data, it is necessary to provide access to the data, based on which the prototype data-analytics solution will be developed in the following phases. At this point, issues and constraints that are related to data privacy, security, etc., could be faced. A good presentation and justification of the way and purpose of the use of the data might be necessary in order to obtain access to the data.

(*Continued*)

Table 3. (Continued).

Phase name	General description & Case study
	Case study: Not all the members of the project team are employees of the observed company. A part of the project team is working for the research institution that occasionally collaborates with this company. Due to the issues related to privacy and the non-disclosure of business know-how, access to the data is not possible without the consent of the responsible people in the company. On the basis of the presentation of the research plans, competences and references of the project team, the access to the data is obtained after the negotiations. A backup of the MES database for a time period of 18 months (starting from January 2010) is obtained. This dataset includes numerical, categorical and textual descriptions of approximately 60,000 manufacturing operations, 14,000 work orders, and 352 work systems. For the purpose of better understanding the data, meetings with key people responsible for the data and IT systems within the company are held, which clarify the uncertainties associated with the data (this data clarification step can be seen as a transition back to the phase *review and understanding the available data sources*).
Development of architecture and implementation of prototype data-analytics solution basis	**General description**: In practice, before the implementation of such data-analytics solutions is approved, it is necessary, for example, to have a demonstration of the feasibility of the solution, a demonstration of use, a performance evaluation, an estimation of implementation and maintenance costs, security risks, etc. For this reason, a prototype data-analytics solution is developed and implemented. In this phase, using the knowledge that arises predominantly from the domain of big-data analytics, it is necessary to determine the appropriate architecture of the prototype data-analytics solution and then to implement it. When determining the architecture of the prototype data-analytics solution's basis, the potential analysis methods that will be tested in the next phase and used in the evaluation of data-analytics solutions, must be taken into account. The choice of technologies and methods for storage and management of the data could depend on the choice of analysis methods. A prototype data-analytics solution basis is used in the next phase for testing the analysis methods and the feasibility of the solution. In the next phase, it enables the efficient development of the analysis part. It should enable quick and easy access to the data, and easy testing of different analysis methods and parameter configurations. The development of the more efficient architecture of the prototype data-analytics solution can be achieved by integrating knowledge from the manufacturing domain, data modelling, and typical big-data analytics methods. **Case study**: For the development of the architecture and implementation of the prototype data-analytics solution basis, the value chain (Chen, Mao, and Liu 2014; Hu et al. 2014) reference architecture is selected. Due to security and for practical reasons, the MES data excerpt for the experiments is stored and managed with a prototype data-analytic system that is separated from the company's IT system. The size and dimensionality of the data and the amount of time necessary to conduct the research allows the implementation of a prototype data-analytics solution in the form of a console application on a regular PC. According to the format, the structure and size of the data under consideration, the MySQL relational database-management system is selected to store and manage the data. The Python programming language is selected as the central tool of the data-analytics system. The hash method is frequently used in order to efficiently manage the data. The Scikit-learn software tool (Pedregosa et al. 2011) is used to implement a machine-learning approach.
Analysis development	**General description**: In the phase of analysis development, a prototype data-analytics solution basis, and knowledge and skills on data-analysis methods and typical big-data analytics methods, are used to develop the analysis part of the solution. The results of this phase are analysis models, i.e., selected, implemented and integrated analysis methods, and the corresponding parameter settings. **Case study**: For a work order it is necessary to predict whether the actual sequence of operations will be different to the planned one or not. This is a binary classification problem. Machine-learning techniques for prediction tasks are used. The observed data is stored in the form of a relational database, consisting of three tables: 1. *Work orders*. (an individual row corresponds to an individual work order) Attributes: ○ Work order ID, ○ Corresponding parts ID, ○ WBS code, ○ Quantity, ○ etc. 2. *Operations*. (an individual row corresponds to an individual operation) Attributes: ○ Operation ID, ○ Work system, ○ Corresponding work order ID, ○ Planned start date, ○ Actual start date and time, ○ Textual description of work, ○ etc. 3. *Parts*. (an individual row corresponds to a part or a group of parts) Attributes: ○ ID, ○ Name or short textual description, ○ etc.

(*Continued*)

Table 3. (Continued).

Phase name	General description & Case study
	According to the problem type, the structure and the properties of the data, the technique called *wordification* seems to be the promising solution. The general idea of the *wordification* approach is a transformation from a relational database representation into a Bag-Of-Words feature vector representation (Perovšek et al. 2015). The input is a relational database, and the output is a set of feature vectors, which can be viewed as a corpus of text documents, and each text document represents an individual entry of the main data table. The basic idea of the *wordification* approach was used for the experiments in this study. The main table is in this case the *work orders* table. To each work order corresponds a set of words that describes the work order and the corresponding operations. The associated target class denotes whether the work order's actual sequence of operations differs from the planned one or not. The analysis is developed and implemented in such a way that features (types of words) can be easily added or removed, and thus to enable testing different combinations of words describing a work order. Attributes, on the basis of which words describing an individual work order are formed, are: work order ID, textual description of operation's work, work system of operation, information about corresponding parts, etc. An example of a work order's text that corresponds to one instance in a data-mining table is shown below:

Abbreviations of work systems' names

Textual descriptions of operations' work

Words describing the sequence of operations

Names of needed components

```
ZK6 d 110x71 NSU5 turning to a dimension NRU3 make all holes KCO clean
threaded holes chamfering edges and mark according to the instructions of
the control PL1 clean and oil surfaces NrOfOperationsInSequence=5 workSys-
temOfOperation_ZK6      workSystemFirstOperation_ZK6      workSystemOfOpera-
tion_ZK6_workSystemNextOperation_NSU5     workSystemOfOperation_idxPl=0_ZK6
workSystemOfOperation_NSU5  workSystemPreviousOperation_ZK6_workSystemOfOp-
eration_NSU5            workSystemPreviousOperation_ZK6_workSystemOfOpera-
tion_NSU5_workSystemNextOperation_NRU3  workSystemOfOperation_NSU5_workSys-
temNextOperation_NRU3    workSystemPreviousOperation_ZK6_workSystemOfOpera-
tion_NSU5_workSystemNextOperation_NRU3_workSystemNextNextOperation_KCO
workSystemPreviousOperation_ZK6_workSystemNextOperation_NRU3   workSystemO-
fOperation_idxPl=1_NSU5 workSystemOfOperation_NRU3 workSystemPreviousOpera-
tion_NSU5_workSystemOfOperation_NRU3                 workSystemPreviousOpera-
tion_NSU5_workSystemOfOperation_NRU3_workSystemNextOperation_KCO   workSys-
temOfOperation_NRU3_workSystemNextOperation_KCO    workSystemPreviousOpera-
tion_NSU5_workSystemOfOperation_NRU3_workSystemNextOperation_KCO_workSys-
temNextNextOperation_PL1   workSystemPreviousPreviousOperation_ZK6_workSys-
temPreviousOperation_NSU5_workSystemOfOperation_NRU3_workSystemNextOpera-
tion_KCO   workSystemPreviousPreviousOperation_ZK6_workSystemPreviousOpera-
tion_NSU5_workSystemOfOperation_NRU3_workSystemNextOperation_KCO_workSys-
temNextNextOperation_PL1 workSystemPreviousOperation_NSU5_workSystemNextOp-
eration_KCO  workSystemOfOperation_idxPl=2_NRU3  workSystemOfOperation_KCO
workSystemPreviousOperation_NRU3_workSystemOfOperation_KCO workSystemPrevi-
ousOperation_NRU3_workSystemOfOperation_KCO_workSystemNextOperation_PL1
workSystemOfOperation_KCO_workSystemNextOperation_PL1      workSystemPrevi-
ousPreviousOperation_NSU5_workSystemPreviousOperation_NRU3_workSystemOfOp-
eration_KCO_workSystemNextOperation_PL1              workSystemPreviousOpera-
tion_NRU3_workSystemNextOperation_PL1     workSystemOfOperation_idxPl=3_KCO
workSystemOfOperation_PL1 workSystemLastOperation_PL1 workSystemPreviousOp-
eration_KCO workSystemOfOperation_PL1     workSystemOfOperation_idxPl=4_PL1
wordComponentsName_560  wordComponentsName_nut  wordComponentsName_1232828
wordComponentsName_34crnimo6+qt wordComponentsName_m64
```

	The multi-nomial naive Bayes classifier, which is well suited for text classification, and the tf-idf transformation are used to generate a classifier.
Evaluation	**General description**: In this phase it is necessary to determine the appropriate methods for a comprehensive evaluation of the data-analytics solution. It is necessary to carefully evaluate the data-analytics solution from (1) the theoretical point of view of methods of analysis, and (2) from the point of the practical application and feasibility of its implementation in the real manufacturing environment. In this phase it is necessary to include most of the participating people from the project team, and to use knowledge from both domains. It is necessary to carefully examine and critically assess all the actions carried out in the development phases. A decision on the use of the developed data-analytics solution should be formed. If the main purpose is to implement a data-analytics solution in a real manufacturing system in the form of, e.g., a repetitive data-analysis process and, if the properties of the data for analysis and the technology and infrastructure of the final data-analytics solution will differ from the technology and infrastructure of the prototype data-analytics solution, particular attention should be paid to the possible performance differentiation of the storage and other data-management subsystems. **Case study**: To evaluate the performance of generated classifiers, a *separate training and test set* method is used. The training set includes 7053 work orders, of which the last operation started before 1st-October -2010, and the test set includes 6317 work orders, of which the first operation started after 1st-October -2010. In the set of work orders for the observed time period of 18 months, the ratio of work orders having the same planned sequence of operations as the actual one to work orders, of which the planned sequence of operations differs from the actual one, is 3.77:1. Due to this imbalance, precision, recall and F-measure are selected as the performance measures to evaluate the prediction model's performance. The resulting precision, recall and F-measure for the best combination of words corresponding to a work order are 0.75, 0.42 and 0.54, respectively. The resulting confusion matrix is:

(*Continued*)

Table 3. (Continued).

<table>
<tr><th>Phase name</th><th>General description & Case study</th></tr>
<tr><td></td><td>
<table>
<tr><th></th><th>Predicted as the same sequence of operations</th><th>Predicted as changed sequence of operations</th></tr>
<tr><td>The same sequence of operations</td><td>4,968</td><td>167</td></tr>
<tr><td>Changed sequence of operations</td><td>683</td><td>499</td></tr>
</table>
This result indicates that for 42% of work orders, of which actual sequence of operations will differ from the planned one, the change in the sequence of operations can be predicted with 75% precision. Looking at this result, several things need to be considered. The properties of the observed data might change over time. The results on the basis of this experimental data are not necessarily the same as the results on the most current data. Due to the loss of information when the data was exported from the MES database (that is, the loss of the information about accurate planned times), some of the planned operation start times in the work order's sequence of operations could be exactly the same, although in the source data they might not be so, etc.
</td></tr>
<tr><td><i>Implementation</i></td><td>General description: The result of the implementation phase can be in the form of an implementation of a repetitive data-analysis process or in the form of a report on discovered new knowledge. In the case of the implementation of a data-analytics system, i.e., in the form of additional software (and hardware) tools implemented in the manufacturing system, it is economical to aim for the use of components of information systems that are already implemented in the manufacturing system.
Case study: In this case study, the result is in the form of a report on discovered new knowledge. It is revealed that changes in the work orders' sequence of operations can be predicted to some extent. This could be useful, for example, as a warning system in the process of planning work orders.</td></tr>
</table>

3.4. Demonstrating the use of a conceptual framework on selected studies

This section demonstrates the use of a conceptual framework on the selected studies of introducing data analytics in manufacturing systems. Five existing case studies of developing data-analytics solutions in manufacturing systems, i.e., (Kozjek et al. 2017a, 2017b, 2018a, 2018b; Vrabič, Kozjek, and Butala 2017), are selected. Data-analytics solutions developed within these projects are either innovative ways of reducing the incompleteness of information and discovering new knowledge through additional use of data or they enable the more efficient reduction of information incompleteness than the conventional approaches.

In Table 4 the selected projects are described and positioned into the three basic manufacturing-system levels, i.e., the (1) operational, (2) coordination and (3) strategic levels, according to the concept of the Adaptive Distributed Manufacturing System (ADMS), presented in (Sluga, Butala, and Peklenik 2005) and (Peklenik 1995). Each project is classified into the ADMS concept depending on the purpose of the data-analytics solution that was developed or studied within the individual project. The concept of *value chain* (Chen, Mao, and Liu 2014; Hu et al. 2014) is used to classify the projects according to the structure of a typical data-analytics system with the purpose to show the focus of the developed and investigated solutions within these projects according to the structure of the typical big-data analytics system. For each project some of the associations to the framework elements are highlighted.

The mapping of the existing projects into the framework on the basis of the ADMS and *value-chain* concepts in Table 4 shows the applicability of the framework on all the basic levels of a manufacturing system and for all the segments of a typical big-data analytics system.

4. Conclusions

The paper explains what big-data analytics is and how it can help to improve the performance of manufacturing systems. A new conceptual framework that facilitates the introduction of big-data analytics in manufacturing systems is proposed. The framework systematically presents the relations of the big-data paradigm with the manufacturing systems and describes the elements that are needed for the development, implementation, and operation of data-analytics solutions that gain the value from large and complex data. The reference procedure for the development and implementation of data-analytics solutions, and the so-called structured scheme of required elements are the main framework tools that show the step-by-step procedure of introducing data analytics into manufacturing systems and what are the required tools, reference models, knowledge and skills.

Table 4. Mapping between the framework and the case studies.

<table>
<tr><th>No.</th><th>Title, Description, and Mapping</th></tr>
<tr><td>1</td><td>
<p>Title: Interpretative identification of the faulty conditions in a cyclic manufacturing process (Kozjek et al. 2017b)</p>
<p>Description: The proposed data-analysis method integrates well-known heuristic algorithms, i.e., decision trees and clustering, with the purpose of identifying types of faulty operating conditions in a cyclic manufacturing process. The result of the analysis is an interpretable model for decision support that can be used for fault identification, to search for root causes, and to develop prognostic systems. The use of the method is demonstrated in the case study in which the proposed method is applied to real industrial data from a plastic injection-moulding process (PIM) with the aim of revealing the types and the root causes of the complex faults, i.e., the unplanned machine stops.</p>
<p>Mapping: The data-analysis method developed within this project represents a reference data-analysis method. When determining the sequence of data-analysis workflow steps, besides informatics, expert and practical knowledge of typical cyclic manufacturing processes, such as PIM and die-casting, were used. To better manage the large size of the data, a method based on indexing (a typical big-data analytics method) was proposed. When applying the developed data-analysis method to the real industrial data of the PIM process and interpreting and evaluating the data-analysis results, practical knowledge of the structure and operation of the observed manufacturing system and knowledge of the actual state of the manufacturing system were intensively used throughout the experiments, etc.</p>
<table>
<tr><th colspan="3">Positioning into levels of manufacturing system according to the ADMS concept</th><th colspan="4">Positioning according to the <i>value-chain</i> concept</th></tr>
<tr><td>Operational</td><td>Coordination</td><td>Strategic</td><td>Generation</td><td>Acquisition</td><td>Storage</td><td>Analysis</td></tr>
<tr><td>✓</td><td>✓</td><td></td><td></td><td></td><td></td><td>✓</td></tr>
</table>
</td></tr>
<tr><td>2</td><td>
<p>Title: A data-driven holistic approach to fault prognostics in a cyclic manufacturing process (Kozjek et al. 2017a)</p>
<p>Description: A data-driven holistic approach, which includes data generation, acquisition, storage, processing, and prognostics, is shown in the case of a typical cyclic manufacturing process, i.e., plastic injection moulding (PIM). The approach is able to tackle the high dimensionality and the large size of the data to create and evaluate prediction models for prognostics of the unplanned machine stops.</p>
<p>Mapping: Within this project the development of a prototype data-analytics solution architecture is the focus. The data-analytic system is developed based on the reference architecture <i>value chain</i>. Important parts of the data-analytic system are the document-oriented NoSQL database and the proposed structure of storing the data. To implement the data-analytic system, among others, the programming language Python, the NoSQL database MongoDB and the machine-learning programming library Scikit-learn were used. The steps to induce and evaluate the prediction model use the following data-analysis methods: machine-learning techniques for classification and the techniques for the classification of imbalanced datasets. Knowing the actual states of the PIM work systems and the reference model of the observed manufacturing process (i.e., PIM) are needed for the identification of faulty PIM cycles and determining the features of feature vectors for the machine-learning process, etc.</p>
<table>
<tr><th colspan="3">Positioning into levels of manufacturing system according to the ADMS concept</th><th colspan="4">Positioning according to the <i>value-chain</i> concept</th></tr>
<tr><td>Operational</td><td>Coordination</td><td>Strategic</td><td>Generation</td><td>Acquisition</td><td>Storage</td><td>Analysis</td></tr>
<tr><td>✓</td><td></td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
</table>
</td></tr>
<tr><td>3</td><td>
<p>Title: Knowledge elicitation for fault diagnostics in plastic injection moulding: A case for machine-to-machine communication (Vrabič, Kozjek, and Butala 2017)</p>
<p>Description: The machine-to-machine (M2M) approach is explored, where several work systems share the process data to improve the accuracy of the fault-detection model. The model is based on machine learning and it was applied to industrial data from approximately two million process cycles performed on several plastic injection moulding (PIM) work systems. The results showed that the fault-prediction model can be improved by sharing the data among work systems, and that it is possible to generalise the process knowledge and apply it to a different work system (without prior knowledge), to some extent.</p>
<p>Mapping: Inspired by the data-integration concept used in many big-data analytics applications, this project investigates whether sharing the manufacturing process data among several PIM work systems contributes to a better prediction of the faults. The result of carrying out data-analytics solution development and the implementation procedure is the discovery of new knowledge. The knowledge gained through this project should be properly managed and stored in the knowledge base and be used by the manufacturing company or by the PIM machine producer in future projects, etc.</p>
<table>
<tr><th colspan="3">Positioning into levels of manufacturing system according to the ADMS concept</th><th colspan="4">Positioning according to the <i>value-chain</i> concept</th></tr>
<tr><td>Operational</td><td>Coordination</td><td>Strategic</td><td>Generation</td><td>Acquisition</td><td>Storage</td><td>Analysis</td></tr>
<tr><td>✓</td><td></td><td></td><td>✓</td><td></td><td></td><td>✓</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>
<p>Title: Big-data analytics for operations management in engineer-to-order (ETO) manufacturing (Kozjek et al. 2018b)</p>
<p>Description: The objective of the research is to investigate manufacturing data which are collected by a MES during operations and to develop data-driven tools for supporting operations management in ETO manufacturing. The developed tools can be used for the simulation of production and the forecasting of potential resource overloads. Machine-learning techniques are used. The visualization principles to facilitate decision making for operations management and to efficiently present the aggregated information are proposed.</p>
<p>Mapping: Since no existing software tools for simulating production (which simulate production in the way that was needed for this project) were available, it was necessary to develop and implement a dedicated software solution, which required considerable knowledge and experience in the programming and use of various software-management and data-analysis tools. The experiments and solutions are based on the large amounts of real manufacturing data collected in the observed company. Especially in making the assumptions that had to be made for the purpose of generating a valid simulation, knowing the structure and operation of the observed manufacturing system, and knowing the structure and operating of information system within the observed company turned out to be absolutely necessary. Key people in the project team were those who had a lot of practical experience in managing the production process in the observed company, and a good understanding and knowledge of the information systems within the company, etc.</p>
</td></tr>
</table>

(*Continued*)

Table 4. (Continued).

No.	Title, Description, and Mapping

Positioning into levels of manufacturing system according to the ADMS concept			Positioning according to the *value-chain* concept			
Operational	Coordination	Strategic	Generation	Acquisition	Storage	Analysis
	✓					✓

5 **Title**: Identifying the business and social networks in the domain of production by merging the data from heterogeneous internet sources (Kozjek et al. 2018a)

Description: It is demonstrated that relevant information about the structures of networks that connect individuals, companies, and institutions for the domain of production can be obtained by merging publicly available internet data using a combination of advanced computational methods including web crawling, machine learning, and creating mash-ups of publicly available services. The feasibility and the applicability of the approach are shown for a case in the automotive domain and a case of the scientific community consisting of people from industry and academia. The proposed approach can be used for the modelling and analysis of various forms of collaboration between and within businesses.

Mapping: The reference data-analytics solution model, i.e., a method for efficient data acquisition is developed and implemented within this project. Crawling (a typical big-data acquisition method), network analysis, and machine-learning techniques are the main components of the proposed method. The development and implementation of the method is inspired by practical experiences from production, i.e., people, businesses and institutions form networks as part of their technical, economic and social activities, and these networks have an impact on business operations, etc.

Positioning into levels of manufacturing system according to the ADMS concept			Positioning according to the *value-chain* concept			
Operational	Coordination	Strategic	Generation	Acquisition	Storage	Analysis
		✓		✓	✓	✓

The advantage and the difference of the proposed conceptual framework compared to other existing concepts are that the proposed concept is widely applicable, as it is useful for solving a wide variety of problems that relate to the introduction of big-data analytics into manufacturing systems. At the same time, it clearly presents the sequence of steps that should be taken for the development and implementation of a data-analytics solution, the likely needed tools, reference models, knowledge, and skills.

The feasibility of the proposed framework is shown in the case of the development and implementation of a data-analytics solution for predicting the changes in work orders' sequences of operations in a typical engineer-to-order manufacturing company. The mapping of several existing projects about the introduction of big-data analytics in manufacturing systems with the framework further shows the feasibility and wide applicability of this framework.

Data-analytic tools, i.e., hardware and software tools, algorithms, data-analytics approaches, etc., come primarily from the field of informatics and computer science. But for their successful implementation in manufacturing systems, in addition to IT experts, experts from the manufacturing domain need to be involved. The proposed conceptual framework facilitates cooperation between these two groups of experts. The results of this research will thus contribute to the integration of knowledge in the field of production with knowledge in the fields of IT and data science.

Future work will include conducting projects that involve the introduction of advanced data analytics into manufacturing systems, and the presented conceptual framework will be used as a backbone to support conducting and coordinating the projects' activities. This will be a further validation of the proposed conceptual framework.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Ministry of Higher Education, Science and Technology of the Republic of Slovenia [1000-15-0510]; and by the Slovenian Research Agency [P2-0103,P2-0270].

References

Adhikari, A., A. Hojjati, J. Shen, J. T. Hsu, W. P. King, and M. Winslett. 2016. "Trust Issues for Big Data about High-Value Manufactured Parts". 2016 IEEE 2nd International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing in IEEE International Conference on Intelligent Data and Security, New York, USA, 24–29. doi:10.1109/BigDataSecurity-HPSC-IDS.2016.50.

Afshari, H., and Q. Peng. 2015. “Using Big Data to Minimize Uncertainty Effects in Adaptable Product Design.” ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Boston, Massachusetts, USA.

Ansari, F., R. Glawar, and T. Nemeth. 2019. “PriMa: A Prescriptive Maintenance Model for Cyber-physical Production Systems.” *International Journal of Computer Integrated Manufacturing* 32 (4–5): 482–503. doi:10.1080/0951192X.2019.1571236.

Arnold, N. 2016. *Wafer Defect Prediction with Statistical Machine Learning*. Massachusetts Institute of Technology, Cambridge, USA.

Babiceanu, R. F., and R. Seker. 2015. “Manufacturing Cyber-physical Systems Enabled by Complex Event Processing and Big Data Environments: A Framework for Development.” *Studies in Computational Intelligence* 594: 165–173. doi:10.1007/978-3-319-15159-5_16.

Babiceanu, R. F., and R. Seker. 2016. “Big Data and Virtualization for Manufacturing Cyber-physical Systems: A Survey of the Current Status and Future Outlook.” *Computers in Industry* 81: 128–137. doi:10.1016/j.compind.2016.02.004.

Bastani, K. 2016. “Compressive Sensing Approaches for Sensor based Predictive Analytics in Manufacturing and Service Systems.” PhD diss., Virginia Tech.

Bilal, M., L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha. 2016b. “Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends..” *Advanced Engineering Informatics* 30 (3): 500–521. doi:10.1016/j.aei.2016.07.001.

Bilal, M., L. O. Oyedele, O. O. Akinade, S. O. Ajayi, H. A. Alaka, H. A. Owolabi, J. Qadir, M. Pasha, and S. A. Bello. 2016a. “Big Data Architecture for Construction Waste Analytics (CWA): A Conceptual Framework.” *Journal of Building Engineering* 6: 144–156. doi:10.1016/j.jobe.2016.03.002.

Boyd, D., and K. Crawford. 2012. “Critical Questions for Big Data.” *Information, Communication & Society* 15 (5): 662–679. doi:10.1080/1369118X.2012.678878.

Butte, S., and S. Patil. 2016. “Big Data and Predictive Analytics Methods for Modeling and Analysis of Semiconductor Manufacturing Processes.” 2016 IEEE Workshop on Microelectronics and Electron Devices (WMED), Boise, ID, USA. doi:10.1109/WMED.2016.7458273.

Chan, K. Y., C. K. Kwong, P. Wongthongtham, H. Jiang, C. K. Y. Fung, B. Abu-Salih, Z. Liu, T. C. Wong, and P. Jain. 2018. “Affective Design Using Machine Learning: A Survey and Its Prospect of Conjoining Big Data.” *International Journal of Computer Integrated Manufacturing* 1–25. doi:10.1080/0951192X.2018.1526412.

Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. “CRISP-DM 1. 0: Step-by-Step Data Mining Guide” The CRISP-DM Consortium.

Chapman, P., R. Kerber, J. Clinton, T. Khabaza, T. Reinartz, and R. Wirth. 1999. “The CRISP-DM Process Model.” Discussion paper. The CRISP–DM Consortium: NCR Systems Engineering Copenhagen (Denmark), DaimlerChrysler AG (Germany), Integral Solutions Ltd. (England) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands). DOI:10.1046/j.1469-1809.1999.6320101.x.

Chen, M., S. Mao, and Y. Liu. 2014. “Big Data: A Survey.” *Mobile Networks and Applications* 19 (2): 171–209. doi:10.1007/s11036-013-0489-0.

Chien, C. F., C. W. Liu, and S. C. Chuang. 2015. “Analysing Semiconductor Manufacturing Big Data for Root Cause Detection of Excursion for Yield Enhancement.” *International Journal of Production Research* 55 (17): 5095–5107. doi:10.1080/00207543.2015.1109153.

Crespino, A. M., A. Corallo, M. Lazoi, D. Barbagallo, A. Appice, and D. Malerba. 2016. “Anomaly Detection in Aerospace Product Manufacturing: Initial Remarks.” 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), IEEE, Bologna, Italy. doi:10.1109/RTSI.2016.7740644.

D’Oca, S., and T. Hong. 2015. “Occupancy Schedules Learning Process through a Data Mining Framework.” *Energy and Buildings* 88: 395–408. doi:10.1016/j.enbuild.2014.11.065.

Davenport, T. H., and D. J. Patil. 2012. “Data Scientist.” *Harvard Business Review* 90: 70–76.

Dubey, R., A. Gunasekaran, S. J. Childe, S. F. Wamba, and T. Papadopoulos. 2016. “The Impact of Big Data on World-class Sustainable Manufacturing.” *International Journal of Advanced Manufacturing Technology* 84 (1–4): 631–645. doi:10.1007/s00170-015-7674-1.

Dutta, D., and I. Bose. 2015. “Managing a Big Data Project: The Case of Ramco Cements Limited.” *International Journal of Production Economics* 165: 293–306. doi:10.1016/j.ijpe.2014.12.032.

Esmaeilian, B., S. Behdad, and B. Wang. 2016. “The Evolution and Future of Manufacturing: A Review.” *Journal of Manufacturing Systems* 39: 79–100. doi:10.1016/j.jmsy.2016.03.001.

Fan, C., F. Xiao, H. Madsen, and D. Wang. 2015. “Temporal Knowledge Discovery in Big BAS Data for Building Energy Management.” *Energy and Buildings* 109: 75–89. doi:10.1016/j.enbuild.2015.09.060.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. 1996. “From Data Mining to Knowledge Discovery in Databases.” *AI Magazine* 17 (3): 37–54.

Gölzer, P., L. Simon, P. Cato, and M. Amberg. 2015. “Designing Global Manufacturing Networks Using Big Data.” *Procedia CIRP* 33: 191–196. doi:10.1016/j.procir.2015.06.035.

Green, P. L. 2015. “Bayesian System Identification of Dynamical Systems Using Large Sets of Training Data: A MCMC Solution.” *Probabilistic Engineering Mechanics* 42: 54–63. doi:10.1016/j.probengmech.2015.09.010.

Grobelnik, M., and J. Jaklič. 2017. ““Knowledge and Skills of Data Scientists: Overview and Analysis of Current Situation in Slovenia.” [Znanja in Sposobnosti Podatkovnih Znanstvenikov: Pregled in Analiza Stanja V Sloveniji.].” *Uporabna Informatika* 25 (1): 17–44.

Grolinger, K., A. L’Heureux, M. A. M. Capretz, and L. Seewald. 2016. “Energy Forecasting for Event Venues: Big Data and Prediction Accuracy.” *Energy and Buildings* 112: 222–233. doi:10.1016/j.enbuild.2015.12.010.

Hammer, M., K. Somers, H. Karre, and C. Ramsauer. 2017. “Profit per Hour as a Target Process Control Parameter for

Manufacturing Systems Enabled by Big Data Analytics and Industry 4.0 Infrastructure." *Procedia CIRP* 63: 715–720. doi:10.1016/j.procir.2017.03.094.

Han, J. H., and S. Y. Chi. 2016. "Consideration of Manufacturing Data to Apply Machine Learning Methods for Predictive Manufacturing." 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, Austria, 109–113.

Hazen, B. T., C. A. Boon, J. D. Ezell, and L. A. Jones-Farmer. 2014. "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. doi:10.1016/j.ijpe.2014.04.018.

Hitzler, P., and K. Janowicz. 2013. "Linked Data, Big Data, and the 4th Paradigm." *Semantic Web* 4 (3): 233–235. doi:10.3233/SW-130117.

Hu, H., Y. Wen, T. S. Chua, and X. Li. 2014. "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial." *IEEE Access* 2: 652–687. doi:10.1109/ACCESS.2014.2332453.

Hu, T., M. Zheng, J. Tan, L. Zhu, and W. Miao. 2015. "Intelligent Photovoltaic Monitoring Based on Solar Irradiance Big Data and Wireless Sensor Networks." *Ad Hoc Networks* 35: 127–136. doi:10.1016/j.adhoc.2015.07.004.

Huber, M. F., M. Voigt, and A. C. N. Ngomo. 2016. *Big Data Architecture for the Semantic Analysis of Complex Events in Manufacturing*, 353–360. Bonn, Germany: Informatik, Lecture Notes in Informatics (LNI).

Hurwitz, J., A. Nugent, F. Halper, and M. Kaufman. 2013. *Big Data for Dummies*. New Jersey, USA: John Wiley & Sons.

Inc, G. 2016. *"Big Data. "* Gartner IT Glossary. https://www.gartner.com/it-glossary/big-data.

Ing, C. K., T. L. Lai, M. Shen, K. Tsang, and S. H. Yu. 2017. "Multiple Testing in Regression Models with Applications to Fault Diagnosis in Big Data Era." *Technometrics* 59 (3): 351–360. doi:10.1080/00401706.2016.1236755.

Ismail, A., H. L. Truong, and W. Kastner. 2019. "Manufacturing Process Data Analysis Pipelines: A Requirements Analysis and Survey." *Journal of Big Data* 6. doi:10.1186/s40537-018-0162-3.

Jun, C., J. Y. Lee, and B. H. Kim. 2019. "Cloud-based Big Data Analytics Platform Using Algorithm Templates for the Manufacturing Industry." *International Journal of Computer Integrated Manufacturing* 32: 723–738. doi:10.1080/0951192X.2019.1610578.

Kaewunruen, S. 2014. "Monitoring Structural Deterioration of Railway Turnout Systems via Dynamic Wheel/rail Interaction." *Case Studies in Nondestructive Testing and Evaluation* 1: 19–24. doi:10.1016/j.csndt.2014.03.004.

Kang, S., W. T. K. Chien, and J. G. Yang. 2016. "A Study for Big-data (Hadoop) Application in Semiconductor Manufacturing." 2016 IEEE International Conference on Industrial Engineering and Engineering Management, Bali, Indonesia, 1893–1897. doi:10.1109/IEEM.2016.7798207.

Kazuyuki, M. 2017. "Survey of Big Data Use and Innovation in Japanese Manufacturing Firms." RIETI Policy Discussion Paper Series.

Kohlert, M., and A. König. 2016. "Advanced Multi-sensory Process Data Analysis and On-line Evaluation by Innovative Human-machine-based Process Monitoring and Control for Yield Optimization in Polymer Film Industry." *Technisches Messen* 83 (9): 474–483. doi:10.1515/teme-2015-0120.

Kong, W., L. Li, F. Qiao, and Q. Wu. 2014. "Network Manufacturing in the Big Data Environment." 2014 International Conference on System Science and Engineering (ICSSE), 13–17. doi:10.1109/ICSSE.2014.6887895.

Koo, D., K. Piratla, and C. J. Matthews. 2015. "Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (Iot)." *Procedia Engineering* 118: 489–497. doi:10.1016/j.proeng.2015.08.465.

Kozjek, D., R. Vrabič, B. Rihtaršič, and P. Butala. 2018b. "Big Data Analytics for Operations Management in Engineer-to-order Manufacturing." *Procedia CIRP* 72: 209–214. doi:10.1016/j.procir.2018.03.098.

Kozjek, D., R. Vrabič, D. Kralj, and P. Butala. 2017a. "A Data-Driven Holistic Approach to Fault Prognostics in A Cyclic Manufacturing Process." *Procedia CIRP* 63: 664–669. doi:10.1016/j.procir.2017.03.109.

Kozjek, D., R. Vrabič, D. Kralj, and P. Butala. 2017b. "Interpretative Identification of the Faulty Conditions in a Cyclic Manufacturing Process." *Journal of Manufacturing Systems* 43 (Part 2): 214–224. doi:10.1016/j.jmsy.2017.03.001.

Kozjek, D., R. Vrabič, G. Eržen, and P. Butala. 2018a. "Identifying the Business and Social Networks in the Domain of Production by Merging the Data from Heterogeneous Internet Sources." *International Journal of Production Economics* 200: 181–191. doi:10.1016/j.ijpe.2018.03.026.

Krumeich, J., J. Schimmelpfennig, D. Werth, and P. Loos. 2014. *Realizing the Predictive Enterprise through Intelligent Process Predictions Based on Big Data Analytics: A Case Study and Architecture Proposal*, 1253–1264. Shanghai, China: Informatik 2014, Gesellschaft für Informatik (GI).

Kumar, A., R. Shankar, A. Choudhary, and L. S. Thakur. 2016. "A Big Data MapReduce Framework for Fault Diagnosis in Cloud-based Manufacturing." *International Journal of Production Research* 54 (23): 7060–7073. doi:10.1080/00207543.2016.1153166.

Laney, D. 2001. "3D Data Management: Controlling Data Volume, Velocity and Variety." *META Group Research Note* (6 February).

Lechevalier, D., A. Narayanan, and S. Rachuri. 2014. "Towards a Domain-specific Framework for Predictive Analytics in Manufacturing." Proceedings - 2014 IEEE International Conference on Big Data, Washington, DC, USA, 987–995. doi:10.1109/BigData.2014.7004332.

Lee, J., E. Lapira, B. Bagheri, and H. A. Kao. 2013. "Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment." *Manufacturing Letters* 1 (1): 38–41. doi:10.1016/j.mfglet.2013.09.005.

Lee, J., H. A. Kao, H. D. Ardakani, and D. Siegel. 2015b. "Intelligent Factory Agents with Predictive Analytics for Asset Management." In *Industrial Agents: Emerging Applications of Software Agents in Industry*, Morgan

Kaufmann, Boston, USA, 341–360. Chap. 19. DOI: 10.1016/B978-0-12-800341-1.00019-X.

Lee, J., H. A. Kao, and S. Yang. 2014. “Service Innovation and Smart Analytics for Industry 4.0 And Big Data Environment.” *Procedia CIRP* 16: 3–8. doi:10.1016/j.procir.2014.02.001.

Lee, J., H. D. Ardakani, S. Yang, and B. Bagheri. 2015a. “Industrial Big Data Analytics and Cyber-physical Systems for Future Maintenance & Service Innovation.” *Procedia CIRP* 38: 3–7. doi:10.1016/j.procir.2015.08.026.

Lei, Y., F. Jia, J. Lin, S. Xing, and S. X. Ding. 2016. “An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data.” *IEEE Transactions on Industrial Electronics* 63 (5): 3137–3147. doi:10.1109/TIE.2016.2519325.

Lenz, J., T. Wuest, and E. Westkämper. 2018. “Holistic Approach to Machine Tool Data Analytics.” *Journal of Manufacturing Systems* 48 (Part C): 189–191. doi:10.1016/j.jmsy.2018.03.003.

Liu, C., H. Li, Y. Tang, D. Lin, and J. Liu. 2019a. “Next Generation Integrated Smart Manufacturing Based on Big Data Analytics, Reinforced Learning, and Optimal Routes Planning Methods.” *International Journal of Computer Integrated Manufacturing* 32: 820–831. doi:10.1080/0951192X.2019.1636412.

Liu, C., and P. Jiang. 2016. “A Cyber-physical System Architecture in Shop Floor for Intelligent Manufacturing.” *Procedia CIRP* 56: 372–377. doi:10.1016/j.procir.2016.10.059.

Liu, C., Y. Zhou, Y. Cen, and D. Lin. 2019b. “Integrated Application in Intelligent Production and Logistics Management: Technical Architectures Concepts and Business Model Analyses for the Customised Facial Masks Manufacturing.” *International Journal of Computer Integrated Manufacturing* 32 (4–5): 522–532. doi:10.1080/0951192X.2019.1599434.

Mangal, A., and N. Kumar. 2016. “Using Big Data to Enhance the Bosch Production Line Performance: A Kaggle Challenge.” 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2029–2035. doi:10.1109/BigData.2016.7840826.

Marini, A., and D. Bianchini. 2016. “Big Data as a Service for Monitoring Cyber-Physical Production Systems.” 30th European Conference on Modelling and Simulation, Regensburg, Germany.

Modoni, G. E., M. Doukas, W. Terkaj, M. Sacco, and D. Mourtzis. 2017. “Enhancing Factory Data Integration through the Development of an Ontology: From the Reference Models Reuse to the Semantic Conversion of the Legacy Models.” *International Journal of Computer Integrated Manufacturing* 30 (10): 1043–1059. doi:10.1080/0951192X.2016.1268720.

Mohanty, S., B. Jagielo, C. B. Bhan, S. Majumdar, and K. Natesan. 2015. “Online Stress Corrosion Crack Monitoring in Nuclear Reactor Components Using Active Ultrasonic Sensor Networks and Nonlinear System Identification - Data Fusion Based Big Data Analytics Approach.” ASME 2015 Pressure Vessels and Piping Conference, Boston, Massachusetts, USA. doi:10.1115/PVP2015-45849.

O’Donovan, P., K. Leahy, K. Bruton, and D. T. O’Sullivan. 2015b. “An Industrial Big Data Pipeline for Data-driven Analytics Maintenance Applications in Large-scale Smart Manufacturing Facilities.” *Journal of Big Data* 2. doi:10.1186/s40537-015-0034-z.

O’Donovan, P., K. Leahy, K. Bruton, and D. T. J. O’Sullivan. 2015a. “Big Data in Manufacturing: A Systematic Mapping Study.” *Journal of Big Data* 2. doi:10.1186/s40537-015-0028-x.

Pääkkönen, P., and D. Pakkala. 2015. “Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems.” *Big Data Research* 2 (4): 166–186. doi:10.1016/j.bdr.2015.01.001.

Papacharalampopoulos, A., J. Stavridis, P. Stavropoulos, and G. Chryssolouris. 2016. “Cloud-based Control of Thermal Based Manufacturing Processes.” *Procedia CIRP* 55: 254–259. doi:10.1016/j.procir.2016.09.036.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel et al. 2011. “Scikit-learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12:2825–2830.

Peklenik, J. 1995. “Complexity in Manufacturing Systems.” *CIRP Journal of Manufacturing Systems* 24 (1): 12–25.

Perovšek, M., A. Vavpetič, J. Kranjc, B. Cestnik, and N. Lavrač. 2015. “Wordification: Propositionalization by Unfolding Relational Data into Bags of Words.” *Expert Systems with Applications* 42 (17–18): 6442–6456. doi:10.1016/j.eswa.2015.04.017.

Phillips, J., E. Cripps, J. W. Lau., and M. R. Hodkiewicz. 2015. “Classifying Machinery Condition Using Oil Samples and Binary Logistic Regression.” *Mechanical Systems and Signal Processing* 60–61: 316–325. doi:10.1016/j.ymssp.2014.12.020.

Precup, R. E., P. Angelov, B. S. J. Costa, and M. Sayed-Mouchaweh. 2015. “An Overview on Fault Diagnosis and Nature-inspired Optimal Control of Industrial Process Applications.” *Computers in Industry* 74: 75–94. doi:10.1016/j.compind.2015.03.001.

Rihtaršič, B., and A. Sluga. 2017. “Quality Management in the Era of IoT & Big Data: A Case Study in ETO Company.” 61. EOQ congress. 2017, Bled, Slovenia.

Rüßmann, M., M. Lorenz, P. Gerbert, M. Waldner, J. Justus, P. Engel, and M. Harnisch. 2015. “Industry 4.0. The Future of Productivity and Growth in Manufacturing Industries.” *Boston Consulting Group*.

Shao, G., F. Riddick, J. Y. Lee, D. B. Kim, Y. T. T. Lee, and M. Campanelli. 2012. “A Framework for Interoperable Sustainable Manufacturing Process Analysis Applications Development.” 2012 Winter Simulation Conference (WSC), Berlin, Germany.

Shin, S. J., J. Woo., and S. Rachuri. 2014. “Predictive Analytics Model for Power Consumption in Manufacturing.” *Procedia CIRP* 15: 153–158. doi:10.1016/j.procir.2014.06.036.

Sluga, A., P. Butala, and J. Peklenik. 2005. “A Conceptual Framework for Collaborative Design and Operations of Manufacturing Work Systems.” *CIRP Annals - Manufacturing Technology* 54 (1): 437–440. doi:10.1016/S0007-8506(07)60139-5.

Stark, R., H. Grosser, B. Beckmann-Dobrev, S. Kind, and I. N. P. I. K. O. Collaboration. 2014. “Advanced Technologies in Life Cycle Engineering.” *Procedia CIRP* 22: 3–14. doi:10.1016/j.procir.2014.07.118.

Suh, N. P. 2005. "Complexity in Engineering." *CIRP - Annals Manufacturing Technology* 54 (2): 46–63. doi:10.1016/S0007-8506(07)60019-5.

Tao, F., Q. Qi, A. Liu, and A. Kusiak. 2018. "Data-driven Smart Manufacturing." *Journal of Manufacturing Systems* 48 (Part C): 157–169. doi:10.1016/j.jmsy.2018.01.006.

Thompson, K., and R. Kadiyala. 2014. "Making "Water Systems Smarter Using M2M Technology"." *Procedia Engineering* 89: 437–443. doi:10.1016/j.proeng.2014.11.209.

Tsuda, T., S. Inoue, A. Kayahara, S. Imai, T. Tanaka, N. Sato, and S. Yasuda. 2015. "Advanced Semiconductor Manufacturing Using Big Data." *IEEE Transactions on Semiconductor Manufacturing* 28 (3): 229–235. doi:10.1109/TSM.2015.2445320.

Villars, R. L., and C. W. Olofson. 2011. "Big Data: What It Is and Why You Should Care." White Paper, IDC.

Vrabič, R., D. Kozjek, and P. Butala. 2017. "Knowledge Elicitation for Fault Diagnostics in Plastic Injection Moulding: A Case for Machine-to-machine Communication." *CIRP Annals - Manufacturing Technology* 66 (1): 433–436. doi:10.1016/j.cirp.2017.04.001.

Wang, J., and J. Zhang. 2016. "Big Data Analytics for Forecasting Cycle Time in Semiconductor Wafer Fabrication System." *International Journal of Production Research* 54 (23): 7231–7244. doi:10.1080/00207543.2016.1174789.

Wang, J., Y. Ma, L. Zhang, R. X. Gao, and D. Wu. 2018. "Deep Learning for Smart Manufacturing: Methods and Applications." *Journal of Manufacturing Systems* 48 (Part C): 144–156. doi:10.1016/j.jmsy.2018.01.003.

Wang, L., and C. A. Alexander. 2016. "Additive Manufacturing and Big Data." *International Journal of Mathematical, Engineering and Management Sciences* 1 (3): 107–121. doi:10.33889/IJMEMS.2016.1.3-012.

Wang, L., and G. Wang. 2016. "Big Data in Cyber-Physical Systems, Digital Manufacturing and Industry 4.0." *International Journal of Engineering and Manufacturing* 6 (4): 1–8. doi:10.5815/ijem.2016.04.01.

Wang, L., M. Törngren, and M. Onori. 2015. "Current Status and Advancement of Cyber-physical Systems in Manufacturing." *Journal of Manufacturing Systems* 37 (Part 2): 517–527. doi:10.1016/j.jmsy.2015.04.008.

Wang, M., S. Du, and L. Xi. 2015. "Predicting Machined Surface Topography Based on High Definition Metrology." *IFAC-PapersOnLine* 48 (3): 1013–1017. doi:10.1016/j.ifacol.2015.06.216.

Wang, S., C. Zhang, and D. Li. 2016. "A Big Data Centric Integrated Framework and Typical System Configurations for Smart Factory." International Conference on Industrial IoT Technologies and Applications. Industrial IoT 2016. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 173, GuangZhou, China. doi:10.1007/978-3-319-44350-8_2.

Wu, D., S. Liu, L. Zhang, J. Terpenny, R. X. Gao, T. Kurfess, and J. A. Guzzo. 2017. "A Fog Computing-based Framework for Process Monitoring and Prognosis in Cyber-manufacturing." *Journal of Manufacturing Systems* 43 (Part 1): 25–34. doi:10.1016/j.jmsy.2017.02.011.

Xiang, F., X. Chen, and G. Jiang. 2016. "A New Manufacturing Resources Integration and Sharing Modes in Big Data Environment." 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, China, 1984–1987. doi:10.1109/ICIEA.2016.7603914.

Xu, S., X. Li, and W. F. Lu. 2016. "Randomized K-d Tree ReliefF Algorithm for Feature Selection in Handling High Dimensional Process Parameter Data." 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Hefei, China. doi:10.1109/ETFA.2016.7733508.

Yang, H., M. Park, M. Cho, M. Song, and S. Kim. 2014. "A System Architecture for Manufacturing Process Analysis Based on Big Data and Process Mining Techniques." 2014 IEEE International Conference on Big Data, Washington, DC, USA, 1024–1029. doi:10.1109/BigData.2014.7004336.

Yin, J., and W. Zhao. 2016. "Fault Diagnosis Network Design for Vehicle On-board Equipments of High-speed Railway: A Deep Learning Approach." *Engineering Applications of Artificial Intelligence* 56: 250–259. doi:10.1016/j.engappai.2016.10.002.

Yu, J. 2016. "Machinery Fault Diagnosis Using Joint Global and Local/nonlocal Discriminant Analysis with Selective Ensemble Learning." *Journal of Sound and Vibration* 382: 340–356. doi:10.1016/j.jsv.2016.06.046.

Zhang, Y., S. Ren, Y. Liu, and S. Si. 2017a. "A Big Data Analytics Architecture for Cleaner Manufacturing and Maintenance Processes of Complex Products." *Journal of Cleaner Production* 142 (Part 2): 626–641. doi:10.1016/j.jclepro.2016.07.123.

Zhang, Y., S. Ren, Y. Liu, T. Sakao, and D. Huisingh. 2017b. "A Framework for Big Data Driven Product Lifecycle Management." *Journal of Cleaner Production* 159: 229–240. doi:10.1016/j.jclepro.2017.04.172.

Zhong, R. Y., C. Xu, C. Chen, and G. Q. Huang. 2015a. "Big Data Analytics for Physical Internet-based Intelligent Manufacturing Shop Floors." *International Journal of Production Research* 55 (9): 2610–2621. doi:10.1080/00207543.2015.1086037.

Zhong, R. Y., G. Q. Huang, S. Lan, Q. Y. Dai, X. Chen, and T. Zhang. 2015b. "A Big Data Approach for Logistics Trajectory Discovery from RFID-enabled Production Data." *International Journal of Production Economics* 165: 260–272. doi:10.1016/j.ijpe.2015.02.014.

Zhong, R. Y., S. Lan, C. Xu, Q. Dai, and G. Q. Huang. 2016a. "Visualization of RFID-enabled Shopfloor Logistics Big Data in Cloud Manufacturing." *International Journal of Advanced Manufacturing Technology* 84 (1–4): 5–16. doi:10.1007/s00170-015-7702-1.

Zhong, R. Y., S. T. Newman, G. Q. Huang, and S. Lan. 2016b. "Big Data for Supply Chain Management in the Service and Manufacturing Sectors: Challenges, Opportunities, and Future Perspectives." *Computers & Industrial Engineering* 101: 572–591. doi:10.1016/j.cie.2016.07.013.