A TAXONOMY OF RECURRING DATA ANALYSIS PROBLEMS IN MAINTENANCE ANALYTICS

Research paper

Zschech, Patrick, TU Dresden, Dresden, Germany, patrick.zschech@tu-dresden.de

Abstract

Modern maintenance strategies increasingly focus on vast amounts of diverse data and multifaceted analytical approaches in order to make efficient use of given resources and unveil hidden potentials. While there is often no universal solution approach to a specific case at hand, it is still possible to observe recurring problem classes for which generic solution templates can be applied and thus the establishment of a reusable knowledge base appears beneficial. To this end, we apply a taxonomy development approach to identify and systematize dimensions and characteristics of recurring data analysis problems in data-driven maintenance scenarios. Our research method integrates findings from a systematic literature review and expert interviews with data scientists from industry. Thus, we add descriptive theory to the field of maintenance analytics and propose a taxonomy that distinguishes between analytical maintenance objectives, data characteristics and analytical techniques.

Keywords: Big Data Analytics, Data Analysis Pattern, Condition-Based Maintenance, Predictive Maintenance.

1 Introduction

In recent years, utilizing big data has been established as a core topic in information systems (IS) research and practice, where the term refers to a situation of expanding availability of large amounts of data that are generated with increasing frequency from multiple business activities and heterogenous systems (Abbasi et al., 2016; Watson, 2014; Chen et al., 2012). From an organizational perspective, this ubiquitously generated data can be seen as a primary asset that changes the way how corporate decision-making is carried out (Constantiou and Kallinikos, 2015; Sharma et al., 2014; Zschech et al., 2017). However, big data cannot be regarded as self-explanatory, as it requires analytical techniques in order to identify valuable insights from vast amounts of noise-affected data (Müller et al., 2016). In this context, analytics as a multidisciplinary concept can be defined as "*(...) the process of introspecting data to discover hidden patterns, meaningful relationships, and interesting associations which can be converted into actionable insights"* (Ramannavar and Sidnal, 2016, p.294). Depending on the question to be answered and the data given at hand, the complexity of analytical techniques may range from simple tasks, e.g. the summarization of univariate measures, up to more sophisticated tasks, e.g. the identification of non-linear and complex high-level interactions between variables (Ramannavar and Sidnal, 2016). Hence, the field of big data analytics (BDA) comprises manifold techniques from various converging disciplines, including statistical analysis, mathematical modelling, data mining or machine learning, which allow access to the diversity of data from multiple perspectives (Chen et al., 2012; Kaisler et al., 2014; Manyika et al., 2011).

A promising area for the application of BDA is the manufacturing sector. According to Manyika et al. (2011), more data are generated in this sector than elsewhere. In 2010 alone, they estimated an amount of two exabytes of newly generated data, which provides a fundamental basis for improving various areas of interest, such as process performance, quality control, production scheduling, energy efficiency or maintenance (Brodsky et al., 2015; Flath and Stein, 2018; Manyika et al., 2011). The area of maintenance is of particular interest, since today's industry is characterized by increasingly complex production systems and machinery, which require sophisticated maintenance systems to guarantee low environmental risks, high reliability and human safety. Thus, it is necessary to establish a maintenance strategy that makes efficient use of given resources and avoids redundant expenditures (Bousdekis et al., 2015; Elattar et al., 2016; Peng et al., 2010; Heng et al., 2009). At this point, the amount and the variety of data is a vital asset, since the ubiquitous use of IT facilitates the generation of multifaceted data, such as sensor values from condition monitoring, machine settings from operation profiles, event logs from process executions or status messages from machine applications. This is an ideal starting point for the discovery of unknown potentials and the realization of various benefits, including higher transparency, better understanding of technical processes for health assessment and root cause analysis, replacement of subjective decision-making and faster response to faults due to minimal human intervention (Accorsi et al., 2017; Karim et al., 2016; Manyika et al., 2011; Meeker and Hong, 2014). Thus, the application of BDA in maintenance (hereinafter referred to as ‘maintenance analytics’, MA) offers great opportunities to extract hidden knowledge and make better use of given resources.

However, applying BDA to unveil hidden potentials can often be a complex and challenging task, since it requires multidisciplinary expertise like solid domain understanding, experiences with different data sources or analytical modelling skills (e.g. Agarwal and Dhal, 2014; Debortoli et al., 2014; Schumann et al., 2016). Even if the focus is placed on a specific problem, such as performance prediction, there is often no ‘silver bullet approach’ that conquers the multidisciplinary problem with a universal solution (Stein and Flath, 2017). Nevertheless, on a certain level of abstraction, it is still possible to observe recurring problem classes for which generic solution templates can be applied and thus the establishment of a reusable knowledge base appears beneficial (e.g. Brodsky et al., 2015; Bousdekis et al., 2015; Eckert and Ehmke, 2017). In this context, Russo (2016) introduces the vision of so-called ‘data analysis patterns’ in analogy to design patterns in software engineering. Such patterns can be considered as guiding models or templates to instruct users how to apply an intentional solution design for recurring data analysis problems (RDAP) based on accumulated experiences instead of rediscovering a problem solution every time again from scratch. Thus, solution designs for RDAP could help to convey analytical expertise to domain-experts without deep analytical skills. This is particularly helpful against the backdrop of BDA, where data of different types and structures need to be accessed via multiple analytical techniques. Once being developed, such models could, for example, be integrated into superior data analysis platforms based on reusable model repositories for solving recurring analytics tasks (e.g. Brodsky et al., 2015). Towards this vision, the scope of this research paper is on the establishment of a reusable knowledge base, focusing on RDAP in data-driven maintenance environments from a BDA perspective and thus we pose the following research question:

RQ: *What are dimensions and characteristics of recurring data analysis problems in maintenance analytics?*

Despite the fact, that there is a huge body of knowledge considering the existent literature on (data-driven) maintenance, the establishment of RDAP from a big data perspective can be considered as a relatively young field. For this reason, we choose a taxonomy development approach to answer our research question and add to the field by providing descriptive theory in terms of analyzing and summarizing salient dimensions and characteristics of the phenomenon under consideration (Gregor, 2006). Following this approach, this paper is structured as follows: In Section 2, we provide a conceptual background and refer to related work. In Section 3, we describe the taxonomy development process in detail. In Section 4, we present the results and subsequently discuss our contribution and its limitations in Section 5. Finally, we draw a conclusion and give an outlook for further research in Section 6.

2 Conceptual Background and Related Work

Data-driven approaches have been used in maintenance for a long time, since data can be considered as a vital resource in order to obtain a machine’s health and examine deviations from the expected behaviour (Schwabacher, 2005). From a technical perspective, this kind of maintenance is often referred to as ‘condition-based maintenance’ (CBM). In CBM, comprehensive data collections are gathered and processed by a condition monitoring (CM) system to assess the current state of the equipment and derive recommendations for the optimal time of maintenance (Jardine et al., 2006; Peng et al., 2010; Veldman et al., 2011). Thus, in contrast to reactive maintenance strategies, divergent machine behaviour can be detected and classified at an early stage by means of diagnostic techniques to avoid unnecessary work. Furthermore, by using suitable indicators and prognostic techniques, it is possible to determine the future state of a machine or its remaining useful lifetime, which is often also referred to as 'predictive maintenance' (PdM) (Hashemian and Bean, 2011; Peng et al., 2010; Elattar et al., 2016). Due to the ubiquitous use of IT, the amount and the variety of maintenance-related data has increased considerably, including for example high-dimensional time series data, machine states, image data or textual error messages. This leads to novel BDA approaches to unveil valuable insights, such as demonstrated by Sipos et al. (2014), where the authors use event logs for equipment failure prediction.

When searching for explicit concepts dealing with MA from a BDA perspective, only a few contributions could be identified (see Section 3.2 for details on the search process): Karim et al. (2016) introduce an MA concept highlighting the importance of big data and the process of knowledge discovery in maintenance. The concept distinguishes between four types of analytical perspectives, which can be arranged as interconnected phases. Kans and Galar (2017) adopt this approach without any significant modifications and discuss its relevance in a broader scope with regard to the digital transformation in industry. Similarly, Famurewa et al. (2017) refer to the existing MA concept, but apply it to railway infrastructure maintenance as a specific domain and thus extend it by additional aspects. The MA concept proposed and discussed within all three contributions stays at an abstract level though, without specifying more profound characteristics of RDAP. Nevertheless, it can be considered as a valid starting point, providing sufficient space and flexibility towards the achievement of our research goal.

Moreover, by taking into account a broader scope on (data-driven) maintenance research, it can be seen that the body of knowledge is extensive, as hundreds of papers are published every year in academic journals, conferences or technical reports (see Section 3.3 for details on the search process). They address diverse topics due to a broad variety of systems and components and range from theories to practical applications. Due to the plethora of maintenance research, there are also numerous review papers structuring the field with an emphasis on different technologies, models and algorithms for data processing and decision-making. Bousdekis et al. (2015), for example, examined patterns of characteristics for the combination of different analytical methods. However, this approach suffers from a narrow scope, since it is limited to prognostic-based decision support. As such, to our best knowledge, there is no single contribution yet that structures the field accordingly from a BDA perspective, addressing the diversity of objectives, data types and analytical techniques in a holistic manner.

3 Research Method

Taxonomies play an important role in IS research, since they provide a structure to organize knowledge of a specific field, help to understand and analyze complex domains and enable researchers to study the relationship among concepts (Nickerson et al., 2013). Moreover, taxonomies help to describe and classify characteristics and dimensions of objects by summarizing commonalities found in discrete observations and thus serve as a viable research approach for manifesting descriptive theories (Gregor, 2006). In our particular case, we aimed at identifying, describing and structuring dimensions and characteristics of RDAP within the field of maintenance.

To carry out the taxonomy development process, we applied the method proposed by Nickerson et al. (2013), as it provides systematic guidance for taxonomy development. As such, we followed the example of recently developed taxonomies in both analytics research (e.g. Nadj and Schieder, 2017) and other IS domains (e.g. Jöhnk et al., 2017; Püschel et al., 2016; Tilly et al. 2017). The method basically consists of the following main elements: (i) determining a meta-characteristic, (ii) specifying ending conditions, and (iii) identifying dimensions and characteristics towards the taxonomy creation. The actual step of the taxonomy creation can be carried out either with an 'empirical-to-conceptual' or a 'conceptual-to-empirical' path. We applied a combination of both by running several iterations, which can be summarized with three major phases: First, we created a baseline taxonomy using initial literature on MA. Second, we extended the initial taxonomy using survey papers derived by a review on

CBM and PdM. Third, we revised and modified the taxonomy conducting qualitative, semi-structured interviews with data science experts from industry. In the following, we describe the development process in detail referring to the meta-characteristic, ending conditions and all three phases applied.

3.1 Meta-characteristic and Ending Conditions

The meta-characteristic is the root element within the taxonomy development process, as it is the most comprehensive one and serves as a basis for the choice of all other characteristics (Nickerson et al., 2013). Thus, our meta-characteristic was defined in accordance with the research question: *characteristics of recurring data analysis problems*. However, since RDAP can be considered as an abstract concept, it required a more precise manifestation to derive appropriate characteristics. While Russo (2016) leaves the design of constituent elements unspecified, a useful approach could be identified within the contribution of Tsai et al. (2014). The authors discuss the analysis of multifaceted data types in 'internet of things' environments and state that any data analysis problem can be broken down into the following three elements: (i) **objectives** - defining the problem to be solved in terms of contextual information such as assumptions and limitations, (ii) **data** - describing specific data characteristics, and (iii) mining algorithms or **analytical techniques** - specifying the actual step of data processing to fulfil the requirements given by the objectives. This structure was adopted as a tripartite meta-characteristic, as it could help to distinguish between outputs, inputs and throughputs of RDAP.

Due to the iterative method character, it also required the specification of ending conditions, which can be either subjective (SEC) or objective (OEC) (Nickerson et al., 2013). For SEC, we adopted the criteria of a taxonomy being sufficiently *robust* (i.e. it contains enough dimensions and characteristics to clearly differentiate the objects of interest, SEC1), *concise* (i.e. it is not overloaded with too many dimensions and characteristics to not exceed the cognitive load of the taxonomy user, SEC2) and *comprehensive* (it can classify all known objects within the domain under consideration, SEC3). For OEC, we adopted the following four criteria: *All objects have been examined* (OEC1), *at least one object is classified under every characteristic of every dimension* (OEC2), *no new dimensions or characteristics were added in the last iteration* (OEC3) and *no dimensions or characteristics were modified in the last iteration* (OEC4). Moreover, Nickerson et al. (2013) postulate the fundamental objective requirements of characteristics being mutually exclusive and having non-redundant dimensions. However, due to hierarchical and combinatorial relationships, both criteria led to an inflated set of characteristics within individual dimensions. For this reason, we followed the example of other recently developed taxonomies (e.g. Jöhnk et al., 2017; Püschel et al., 2016) and allowed non-exclusive characteristics and the creation of sub-dimensions to guarantee transparency and parsimony within our taxonomy.

3.2 Phase 1: Initial Literature on Maintenance Analytics

In a first iteration, we started with a conceptual-to-empirical approach using initial literature on MA. To identify relevant contributions, we carried out a literature review, since reviews help to identify sources being relevant to the topic and support the effective use of existing knowledge (vom Brocke et al., 2009; Webster and Watson, 2002). In particular, we applied a database search using the following digital libraries: EBSCOhost, ScienceDirect, ACM Digital Library, JSTOR, IEEE Xplore Digital Library and AIS Electronic Library. Since we explicitly focussed on the MA concept in this first step, the search term was limited to the keyword 'maintenance analytics' using the search fields 'title', 'abstract' and 'keywords' whenever such filter mechanisms were provided by the digital libraries. Moreover, the search was restricted to articles that address concepts or frameworks from a BDA perspective as introduced in Section 1 (i.e. considering multiple analytical objectives, techniques and data sources). Based on this search, it was possible to identify one paper which deals with a conceptual discussion of MA (Karim et al., 2016) and another two relevant articles could be identified which refer to this approach by conducting a forward search (Kans and Galar, 2017; Famurewa et al., 2017; day of search: 2017-08-06). Further details on their respective research focus are outlined in Section 2.

Examining this literature, a first draft of a taxonomy could be derived. In line with the tripartite meta-characteristic, we extracted initial dimensions and related characteristics to gather distinct features of

RDAP. Of particular interest was the distinction between descriptive, diagnostic, predictive and prescriptive analytics tasks, as proposed by Karim et al. (2016) in terms of constitutive pillars for MA. This elementary distinction could also be confirmed by Kans and Galar (2017) and Famurewa et al. (2017), who adopted the conceptual proposal, but also by Brodsky et al. (2015), where the authors applied a similar structure to the reusable knowledge base for manufacturing analytics. Moreover, it was possible to derive a simple data type classification and a distinction between different classes of analytical techniques. At this stage, however, the resulting taxonomy was neither sufficiently robust (SEC1) nor was it comprehensive enough (SEC3) due to missing dimensions and characteristics which are able to adequately differentiate among RDAP. For this reason, we carried out a second iteration by making use of the plethora of research on (data-driven) maintenance.

3.3 Phase 2: Literature Review on CBM and PdM

To make use of the large body of knowledge on maintenance research, a second literature review iteration was initiated, thus following again a conceptual-to-empirical approach. However, since the concept of MA is not widely recognized yet within the literature, alternative search terms had to be selected. Thus, by screening taxonomies for maintenance philosophies (e.g. Kothamasu et al., 2006, p.1014), we identified the keywords ‘predictive maintenance’ and ‘condition based maintenance’ as representative approaches for data-driven maintenance. Both approaches can be considered as solid maintenance concepts, where data utilization plays an important role as previously introduced in Section 2. Considering only the results of the database ScienceDirect, both terms led to a high number of search results (2.103 hits for CBM, 3.063 hits for PdM, day of search: 2017-08-08), illustrating the extent of the existing knowledge base. At a closer look, these results could be roughly divided into four groups: (i) development of context-specific maintenance solutions, (ii) comparison and evaluation of different maintenance strategies, (iii) conceptual discussion of CBM and PdM programs and components, and (iv) survey/review papers summarizing the field with an emphasis on different technologies, models and algorithms for data processing and decision-making. Due to the quantity and the type of the search results, we decided to focus only on contributions that represent the current state of research and summarize the field. For this reason, we combined the previously derived concept names with the keywords ‘review’, ‘survey’, ‘overview’ and ‘state of the art’, using the same databases and search fields as in the first phase (see Section 3.2). Thus, after the removal of non-review articles, duplicates and contributions with limited access, a total number of 94 survey papers could be identified. This amount had to be further reduced due to a narrowly limited focus of 79 articles, which address too specific components, techniques or other particular aspects. Moreover, another five review papers could be identified conducting a backward search. Hence, a total number of 20 survey articles served as basis for the second iteration of the taxonomy development process (Ahmad and Kamaruddin, 2012; Ahmadzadeh and Lundberg, 2014; An et al., 2015; Ao, 2011; Bousdekis et al., 2015; Dragomir et al., 2009; Elattar et al., 2016; Goyal and Pabla, 2015; Hashemian and Bean, 2011; Heng et al., 2009; Jardine et al., 2006; Kothamasu et al., 2006; Lee et al., 2014; Peng et al., 2010; Prajapati et al., 2012; Schwabacher, 2005; Si et al., 2011; Veldman et al., 2011; Vogl et al., 2016; Zhu et al., 2016).

Using this knowledge base, it was possible to extend the initial taxonomy draft. While only a few but essential modifications could be derived for the meta-characteristic 'analytical maintenance objectives', we obtained extensive and crucial input for the distinction of data characteristics and analytical techniques based on existing classification schemas from the literature. Thus, more than 80 characteristics were identified that could be grouped and organized within more than 25 dimensions. As such, the resulting taxonomy was robust enough considering the amount of characteristics and dimensions (SEC1). However, now it lacked of being concise (SEC2) and comprehensive (SEC3), as it consisted of too many unstructured, partially overlapping dimensions and thus could not be applied to real-life objects. A major issue could be observed within the meta-characteristic ‘analytical techniques’ due to a high degree of complexity, since techniques can often be classified in several ways as exemplified by Elattar et al. (2016, p.132): *“Sometimes, the classification is based on the type of available data and knowledge about the system. Another time (...) to the type of the used methodology.”*. Hence, more iterations were planned using insights from the empirical practice in order to meet all SEC.

3.4 Phase 3: Interviews with Data Science Experts

In a third phase, we applied an empirical-to-conceptual approach to integrate knowledge from industrial practice and see how real world scenarios could be mapped to our previously derived taxonomy. In particular, we followed the example of Jöhnk et al. (2017) conducting multiple interviews with experts from industry. The interviewees were selected based on experiences with data analysis problems in maintenance. Since it was planned to carry out several interview iterations incorporating the findings from all other interviewees, which required short contact paths and fast response times, we selected experts working for the same company. To avoid biases, however, we chose a network partner operating as a service provider, who deals with a wide range of client projects. As such, it was possible to survey a total of seven experts from a medium-sized IT service provider offering data-driven solutions to clients from various industrial branches, such as plant engineering, automotive or semiconductor industry. All interviewees had several years of data science (DS) working experience and successfully carried out various analytical maintenance projects for industrial clients (cf. Table 1). Moreover, we took care that the experts chosen were not involved in too many identical client projects to preserve the diversity of maintenance projects and prevent potential biases. Hence, by taking all distinct projects into account, we were able to incorporate the experience from a total of at least 19 maintenance projects, while each project, moreover, consisted of several individual data analysis problems.

ID	Position of the interviewee	Yrs. of DS experience	Maint. projects	Feedback in sequence I	Feedback in sequence II	Feedback in sequence III
1	System Analyst	6	7	01: MJC	08: MNC	15: NC
2	Project Manager	4	5	02: MJC	09: NC	16: NC
3	System Analyst	9	8	03: MNC	10: MNC	-
4	Executive System Architect	6	6	04: MNC	11: NC	-
5	System Analyst	3	5	05: MJC	12: NC	-
6	Division Manager Analytics	10	7	06: NC	13: NC	-
7	Head of Industry Division	12	7	07: MNC	14: NC	-
MJC: major change - MNC: minor change - NC: no change						

Table 1. Descriptive details on the semi-structured expert interviews

To carry out the interviews, we applied a qualitative, semi-structured approach. Qualitative interviews have the advantage of taking account for the respondents' specific context and thus are suitable for explorative studies (Schultze and Avital, 2011). As such, we included closed- and open-ended questions within personal face-to-face interviews, where it was possible to discuss specific taxonomy elements by means of additional material and modification proposals. Moreover, each interview was designed to be a single empirical-to-conceptual iteration within the overall development process, i.e. the feedback of each interview was directly incorporated in the taxonomy in order to be available for all other interviewees. Thus, due to the given OEC, any modification of characteristics and/or dimensions (OEC3, OEC4) had to be revised by all other experts (OEC1). However, to keep the number of iterations low, the interviews were organized in sequences, where the overall process stopped after all experts revised major and minor changes without proposing any further modifications. As a result, a total of 16 empirical-to-conceptual iterations were carried out (cf. Table 1). Within the interviews, we applied a semi-structured interview guide (Myers and Newman, 2007) and addressed the following three main aspects: First, we provided information about the research project and introduced the scope of the taxonomy development approach. Second, the interviewees delivered contextual information about their maintenance projects and described recurring commonalities of data analysis problems. Third, the taxonomies were discussed and the expert feedback was incorporated. This tripartite structure was applied to the first seven interviews, using the taxonomy results from the second literature-driven phase as well as the modified version from the previous interview iterations. In the remaining nine interviews only potential modification were discussed until no more changes were necessary.

As a result of this third major phase, the findings of both conceptual-to-empirical iterations could not only be evaluated, but also be enriched with experiences and insights from the industrial practice.

Thus, it was possible, one the one hand, to reduce the degree of complexity to make the taxonomy more precise (SEC2) and, on the other hand, to add more necessary characteristics to make the taxonomy more comprehensive (SEC3). As a result, we were confident that, after the conduction of all three phases, our defined OEC and SEC were met and no more iterations were required. The final taxonomy consists of 67 characteristics organized in 21 dimensions, which are distributed among the tripartite meta-characteristics as follows: analytical maintenance objectives (18/7), data characteristics (27/9), analytical techniques (22/5). When presenting the detailed results in the following section, we point out whether the selection and inclusion of the taxonomy elements are rather motivated by the findings of the literature review or the results of the expert interviews.

4 A Taxonomy of Recurring Data Analysis Problems

In this section, we present the final version of our resulting taxonomy, distinguishing between various dimensions and characteristics of RDAP in MA (cf. Figure 1). The following subsections are organized in accordance to the tripartite meta-characteristic.

Analytical Maintenance Objectives					
Analytical Type	Descriptive	Diagnostic	Predictive/Prognostic	Prescriptive	
Descriptive	Measures	Visualizations			
Diagnostic	Fault Detection	Fault Isolation	Fault Identification		
Predictive/Prognostic	System Health State	Remaining Useful Life			
Prescriptive	Optimal Time of Maintenance	Optimal Action of Maintenance			
Maintenance Paradigm	Breakdown Maintenance	Time-Based Preventive Maint.	Condition-Based Maintenance		
Degree of Maintenance	Perfect Maintenance	Imperfect Maintenance			
Data Characteristics					
Data Type	Condition Monitoring Data	Event Data	Metadata	Business Data	
Condit. Monitoring Type	Single Value	Time Waveform	Multidimensional		
Monitoring Frequency	Continous Records	Regular Records	Irregular Records		
Variety of Sensors	Single Sensor	Multiple Homogeneous Sensors	Multiple Heterogeneous Sensors		
Physical Relation	Direct Data	Indirect Data			
Event Type	Machine State	Operating Step	Machine Configurat.	Malfunction	Maint. Action
Malfunction Type	Continous Degradation	Sudden Change of State	Sudden Incident		
Data Labeling	Labeled Data	Unlabeled Data			
Data Censoring	Censored Data	Uncensored Data			
Analytical Techniques					
Knowledge Integration	Empirical Observations	Physical Models	Expert Knowledge		
Descriptive & Diagnostic Approach	Summary Statistics	Hypothesis Testing	Clustering	Classification	
	Anomaly Detection	Frequent Pattern Mining	Process Mining		
Predictive/Prognostic Approach	Machine Learning Models	Trend Projection Models	Reliability & Hazard Rate Models	Stochastic Filters	Graphical Models
Decision-Making Appr.	Evidence-Based	Optimization	Simulation		
Pre-processing	Signal Processing	Image Processing	Natural Language Processing	Single Value Processing	

Figure 1. Taxonomy for recurring data analysis problems in maintenance analytics

4.1 Analytical Maintenance Objectives

From an organizational point of view, the maintenance function pursues various objectives, such as ensuring availability, reliability and product quality or preserving plant and environmental safety

(Muchiri et al., 2011). From a data-driven perspective, however, it is rather of interest how given data assets can be analytically exploited in order to support the achievement of such superior objectives. As such, Karim et al. (2016) basically distinguish within their original MA concept between descriptive, diagnostics, predictive and prescriptive tasks, defining the core of analytical objectives. All four types are related to maintenance-relevant questions and can be differentiated as follows: **Descriptive MA** primarily deals with questions like *"What happened/is happening?"*, since it summarizes collected data from various maintenance sources and provides summary statistics in terms of measures (e.g. number of faults) and visualizations (e.g. failure rate chart). **Diagnostic MA** tries to answer questions like *"Why did/does it happen?"* by discovering patterns and delivering explanations for abnormal behaviour. It can basically be divided into fault detection (i.e. indicating faults and malfunctions), fault isolation (i.e. determining the cause and the related component) and fault identification (i.e. designating the type and the nature of the fault) (Jardine et al., 2006; Kothamasu et al., 2006; Vogl et al., 2016). While descriptive and diagnostic MA are rather focused on the past, **predictive** or **prognostic MA** is concerned with a more forward-looking perspective to answer the question *"What is likely to happen?"*. The goal of this approach is to take current machine conditions and past operation profiles into account and either predict the probability that a machine operates without a failure up to a time in the future (i.e. system health state estimation) or calculate the remaining useful life (RUL) as the time left before a failure occurs (Peng et al., 2010; Elattar et al., 2016; Bousdekis et al., 2015; Jardine et al., 2006; Dragomir et al., 2009). **Prescriptive MA** goes one step further and supports the process of maintenance decision-making to answer the question *"What should be done?"*. For this purpose, it builds on the results of the previously described types, integrates additional data (e.g. costs and resource information), and transforms them into actionable maintenance recommendations to identify the optimal actions and/or the optimal time of actions (Bousdekis et al., 2015; Jardine et al., 2006).

In addition to the distinction between analytical objectives, it also needs to be distinguished between different maintenance strategies which have been applied in the past and which in return should be carried out in the future. Besides the existence of more fine grain taxonomies for maintenance paradigms (e.g. Veldman et al., 2011), it can basically be differentiated between **breakdown maintenance** (i.e. after a failure occurs), **time-based preventive maintenance** (i.e. after a defined period of time) and **CBM** (i.e. consideration of current state) (Bousdekis et al., 2015; Goyal and Pabla, 2015; Prajapati et al., 2012). This consideration is necessary, since the choice of analytical approaches depends on individual requirements, such as maintenance costs, quality aspects or safety issues. For example, it may not always be advisable to switch from a breakdown strategy to a CBM approach if the costs of failure and replacement of a component are lower than the costs for measuring its condition. Similarly, it should be divided between **perfect** and **imperfect** maintenance actions. The first type implies a situation of full restoration, while the latter allows minor corrections, since it might not always be desirable to restore a unit to its original health condition (Heng et al., 2009; Kothamasu et al., 2006).

4.2 Data Characteristics

According to the literature, maintenance-relevant data can be grouped into two main categories: event data and condition monitoring data (Jardine et al., 2006; An et al., 2015; Si et al., 2011). **CM data** are measurements related to the health conditions of a physical asset, such as vibration data, temperature, pressure or humidity. **Event data**, on the other hand, refer to information about what happens on a physical asset (e.g. faults, failures, operations) and which actions are taken (e.g. repair, configurations) (Jardine et al., 2006). Furthermore, two additional categories could be determined within the expert interviews: metadata and business data. **Metadata** provide further context details of the physical asset (e.g. machine type, location, machine manufacturer), whereas **business data** are used to describe the environmental context, ranging from quality measures and performance indicators over different maintenance-related cost and resource information up to comprehensive productions plans and scheduling specifications. At a next level of granularity, the CM data and event data can be further classified, where the specifications lead to different pre-processing and analysis techniques. CM data can be divided into single values, time waveforms and multidimensional data (Jardine et al., 2006). A **single value** is a CM variable which is measured at a particular time or aggregated for a certain period of time (e.g. average, min, max). **Time waveforms**, on the other hand, refer to a collection of time series data that spans over a certain period of time with single values at each point of time. **Multidimensional data** are usually images, such as infrared thermographs or X-ray images, where data points cannot directly be mapped to a single variable (Jardine et al., 2006). Considering the monitoring frequency of CM data, three types can be distinguished: **Continuous** records implying an equidistant measurement at constant intervals (e.g. every five seconds), **regular** records at a defined inspection interval (e.g. once a day or depending on the current state), and **irregular** records at no particular time (Jardine et al., 2006). Another dimension is the variety of sensors, where it is possible to differentiate between data from **multiple** sensors or data from a **single** sensor (Jardine et al., 2006). For example, in legacy systems the measurements of CM data might be limited to only a few signals like power consumption. Additionally, according to the interviewees, multiple sensor data should be further classified into **heterogeneous** and **homogenous** types in order to distinguish whether the sensors are of different types (e.g. temperature vs. power consumption) or of the same type, since redundancy, for example, may help to avoid measurement errors or deliver better insights from slightly different perspectives. Furthermore, CM data can be grouped into **direct** and **indirect** CM data (Si et al., 2011), where the first type can directly describe the underlying health state of a physical asset (e.g. crack sizes, wear) and the latter can only indirectly or partially indicate the underlying state (e.g. vibration data, temperature) and thus failure event data needs to be derived from additional sources.

In contrast to the CM data, the use of event data is rather rare or of secondary importance within the CBM literature (e.g. for labelling purposes using failure events), since there is the erroneous belief that CM data are sufficient to reduce equipment failures (Elattar et al., 2016; Jardine et al., 2006). However, they play a major role in reliability modelling (e.g. Heng et al., 2009; Jardine et al., 2006) and according to all interviewed data scientists, event data provide an important information base towards the understanding of maintenance-related processes, ranging from simple flags to mark specific events up to comprehensive textual messages delivering more context about faults or maintenance actions carried out. Due to a missing categorization, the interviewees classified event data into the following five groups: machine states, operating steps, machine configurations, malfunctions and maintenance actions. **Machine states** describe the current status of a machine (e.g. running, standby, overloaded), while **operating steps** represent the current activity of a machine (e.g. cleaning, milling) and **machine configurations** can be understood as current settings and machine parameters (e.g. workload, pace, frequencies). **Malfunctions**, on the other hand, describe all kinds of misbehaviour in terms of failures, faults or warnings and **maintenance actions** comprise all activities recorded within the data to maintain a machine and resolve malfunctions occurred (e.g. repair, replacement, oil change).

As the core of MA is the identification and prevention of all kinds of malfunctions, it is also important to classify different types of system behaviour in order to apply the appropriate analytical techniques. In this context, it can basically be distinguished between a **continuous degradation**, which follows some sort of deterioration process until a failure occurs (Hashemian and Bean, 2011), and a rather **sudden, intermittent effect**, which affects or even terminates the asset’s normal behaviour (Elattar et al., 2016; Ahmad and Kamaruddin, 2012; Vogl et al., 2016). In the latter case, it should be further differentiated whether a sudden effect leads to another machine condition in terms of a characteristic **state** (e.g. another health state or failure) or only results in a temporary **incident** (e.g. warning message occurred because a threshold was temporarily exceeded). However, having a wide range of event and CM data, it cannot necessarily be guaranteed that relevant information about the misbehaviour and malfunctions to be prevented is available in order to gain further insights. Thus, a distinction can be made whether the data at hand are **labeled** or **unlabeled**. The latter refers to a situation where information about faults or failures is missing and needs to be derived from other sources and approaches (e.g. consideration of produced quality, consumption of additional resources). One of the main reasons for this situation is that critical assets are usually not allowed to run to failure and thus are replaced or overhauled before they fail. Hence, it can be further distinguished whether records are either **censored** or **uncensored**. Thus, censored data refer to a situation where it is not known how long the asset might have run if it had been left unaffected (Heng et al., 2009).

4.3 Analytical Techniques

From an analytical point of view, there exist a broad range of techniques with many modifications, extensions or specializations (e.g. Bousdekis et al., 2015; Kaisler et al., 2014; Manyika et al., 2011; Lee et al., 2014). For this reason, most of the identified survey papers already provide classification schemes for structuring maintenance-relevant techniques. Some of them even distinguish approaches at a fine grain level and discuss advantages and disadvantages or deliver decision criteria for selecting appropriate techniques according to specific problems (e.g. Heng et al., 2009; Elattar et al., 2016; Bousdekis et al., 2015; Ahmadzadeh and Lundberg, 2014; An et al., 2015). However, to keep the taxonomy at this stage sufficiently generic, some abstraction was required and thus we focused on a high level classification of techniques.

A first distinction can be made with regard to the type of integrated knowledge. While we primarily focus on **data-driven** approaches, where insights are derived from an abundant amount of empirical observations, further knowledge can be derived from either physical models or human expertise (Peng et al., 2010; Elattar et al., 2016; An et al., 2015; Vogl et al., 2016). **Physical models** are usually mathematical representations of physical processes that influence the health state of a technical system. Such models have the advantage of being very accurate, since they are built on natural laws (e.g. specific degradation laws such as Paris' law for fatigue crack growth). However, the development of physical models can be considered as costly and time consuming, as it requires a thorough understanding of the physical mechanisms relevant to the system under consideration. **Expert knowledge**, on the other hand, is based on experiences from human domain experts. Such knowledge can be integrated, for example, in terms specific domain rules, the application of experience-based thresholds for CM variables or towards the establishment of fuzzy models.

At a next level, analytical techniques can be grouped due to their output and thus we distinguish again between descriptive, diagnostic, prognostic and prescriptive approaches. At this point, descriptive and diagnostic approaches are considered together, since there is no strict distinction between the techniques applied. Both approaches aim at providing transparency and explanations about all maintenance-related issues, especially with regard to faults and failures. Hence, there is a broad variety of data-driven techniques to reveal descriptive and diagnostic insights (e.g. Veldman et al., 2011; Jardine et al., 2006; Goyal and Pabla, 2015; Schwabacher, 2005; Accorsi et al., 2017). According to the literature and interview results, the following seven categories could be identified: Summary statistics, hypothesis testing, clustering, classification, anomaly detection, frequent pattern mining and process mining. **Summary statistics** deliver insights in terms of univariate measures (e.g. counts of faults), multivariate measures (e.g. correlation coefficients), aggregated measures (e.g. effectiveness of equipment) and visualizations (e.g. plots, charts, tables). **Hypothesis testing** is based on statistical models and inference and specifies how and why certain empirical phenomena occur (e.g. explanation of causal relationships between malfunctions and specific CM variables). **Clustering** techniques allow the identification of heterogeneous groups whose members are similar to each other (e.g. grouping signals into different fault categories or separation of normal behaviour from abnormal behaviour). **Classification** techniques are built on a given target variable and help to identify to which class a new object belongs (e.g. classification of faults to designated type). **Anomaly detection** deals with the identification of objects which do not conform to an expected behaviour (e.g. detection of malfunctions). **Frequent pattern mining** is concerned with the discovery of regularities in terms of association rules or sequential patterns (e.g. identification of frequent co-occurrence of different faults). **Process mining** can be used to reconstruct sequences of activities based on visual process models, which are extracted from event data (e.g. revision of executed operations and detection of deviations).

Compared to descriptive and diagnostic approaches, the variety of prognostic approaches is much smaller, since the task is limited to the estimation of the system health state or RUL prediction (Jardine et al., 2006). Data-driven prognostics is based on techniques from either **statistics**, using parametric and non-parametric approaches, or artificial intelligence, using **machine learning** (ML) algorithms to learn complex, non-linear interactions between input data (e.g. CM variables) and the target variable (e.g. failure event) (Si et al., 2011; Vogl et al., 2016). ML approaches include, for example, decision trees, support vector machines, artificial neuronal networks (e.g. feed forward, recurrent and polynomial networks) and other regression and classification techniques, which are closely related to statistics (e.g. linear and logistic regression, partial least square, linear discriminant analysis) (Peng et al., 2010; Heng et al., 2009; Elattar et al., 2016; Vogl et al., 2016). Statistical approaches, on the other hand, are primarily based on stochastic processes and state space models, which can be characterized diversely due to a variety of modelling mechanisms (Si et al., 2011). However, the following four categories could be identified for classifying the most commonly used techniques within the literature: **Trend projection models** use trend forecasting techniques for time series data, such as exponential smoothing or autoregressive models (e.g. autoregressive moving average) to provide simple predictions (Heng et al., 2009; Jardine et al., 2006). **Reliability and hazard rate models** are built on time to failure data. While traditional reliability models only focus on event data and use different distribution models (e.g. Weibull, log-normal, Poisson) to estimate the mean time to failure or the probability of reliable operations, more enhanced approaches (e.g. proportional hazard rate models) additionally consider CM data as covariates to provide more accurate predictions (Peng et al., 2010; Heng et al., 2009; Elattar et al., 2016; Jardine et al., 2006; Ahmad and Kamaruddin, 2012). **Stochastic filters**, such as Kalman filter or particle filter, are estimations techniques where Bayesian inference is used to estimate and recursively update unknown variables based on measured data. Such techniques are often applied in combination with physical models, where the model parameters need to be estimated (An et al., 2015; Dragomir et al., 2009). **Graphical models** use graphs as diagrammatic representations to visualize the structure of probabilistic models (e.g. Bayesian networks or hidden Markov models) and thus make it easier to examine model properties and perform complex computations (Peng et al., 2010; Dragomir et al., 2009).

Building on all the aforementioned techniques, the goal of prescriptive approaches is eventually to support the actual phase of decision-making. At this point, it can basically be distinguished between evidence-based techniques, optimization and simulation (Bousdekis et al., 2015; Jardine et al., 2006; Ahmad and Kamaruddin, 2012). **Evidence-based** techniques are the simplest form of decision support as the results of descriptive, diagnostic and prognostic techniques can directly be used for decision guidance (e.g. derived explanations, rules, thresholds, probabilities). However, it is often the case that, under certain conditions, such as costs, resources and security (expressed with the “business” data type), a decision must be made towards the best possible solution among several alternatives (i.e. optimal type and/or time of maintenance action) (Bousdekis et al., 2015). For this purpose, **optimization** techniques can be applied from the field of mathematical programming (e.g. stochastic programming, dynamic programming), where different alternatives, utility functions and constraints are taken into account to find an optimal solution. Furthermore, **simulation** techniques can be used to imitate the behaviour of a system by integrating the results of descriptive, diagnostic and predictive models and thus evaluate different decision alternatives based on the simulated behaviour (Brodsky et al., 2015).

Besides that, it is important to note that analytical techniques often require intensive pre-processing, such as data cleaning, filtering, normalization or noise reduction (Peng et al., 2010; Elattar et al., 2016; Jardine et al., 2006). Even though such aspects are not part of the current research, the interviewees and some survey papers (e.g. Jardine et al., 2006) highlighted some fundamental pre-processing techniques, which constitute important prerequisites for further analysis depending on the data types given. Thus, we organized those techniques within a last dimension and categorized them into the following four groups: **Signal processing** comprises various techniques for dealing with time waveform data to extract useful features, including time-domain analysis (e.g. peak, skewness, kurtosis), frequency-domain analysis (e.g. fast Fourier transform), and time-frequency analysis (e.g. short-time Fourier transform, wavelet transform) (Jardine et al., 2006). **Image processing** deals with multidimensional data for feature extraction in images, including techniques from closely related disciplines like object recognition. **Natural language processing**, on the other hand, comprises techniques to handle textual content in event data, provided either by logging engines or manual data entries of maintenance workers. The fourth group includes techniques for **single value processing**, such as trend analysis or techniques for dimensionality reduction (e.g. feature selection based on information gain or feature extraction using principal component analysis) (Jardine et al., 2006).

5 Discussion

In the following, we discuss our results with regard to the merits and limitations of the taxonomy approach and outline implications for further research and practical applications.

From a methodical point of view, the development of a taxonomy for structuring RDAP of a specific domain can be considered as a suitable approach for several reasons. Having a high number of varying conditions and requirements, the application of BDA is a complex and challenging task. However, by decomposing a complex task into single dimensions, it is possible to bring it into a manageable structure and create well-defined solutions that consist of characteristic properties. Moreover, applying BDA can be considered as a knowledge-intensive process, which requires multidisciplinary expertise like domain understanding, experiences with different data sources and analytical modelling skills. This knowledge is often distributed among various actors involved, including domain experts, IT professionals, analysts, etc. (e.g. Brodsky et al., 2015). To this end, a taxonomy not only organizes knowledge in a structured manner, it also provides a tool for communication and a common understanding. This is particularly helpful to bridge the gap between domain-specific contexts and domain-independent practices, while a convergence is possible from both directions. Thus, domain experts can use the taxonomy as a starting point to understand the analytical tool set referring to standards and best practices, while data analysts can gain insights into the particularities of the domain. In the case of the specially developed taxonomy for MA, it can be seen, for example, that common data analysis problems are characterized by various aspects that are highly specific to the domain, such as the significance of CM and event data to monitor and control machine behaviour, the strong impact of diagnostic and prognostic models to avoid machine failures or the possibility to integrate physical models based on natural laws. Thus, the taxonomy gives a quick and comprehensive overview about contextual aspects relevant for all actors involved to overcome the hurdles of entering a multidisciplinary problem space. Following this argumentation, we claim that this taxonomy is not limited to the field of maintenance, but rather could be used to systematize RDAP in any digitized, data-intensive domain. Possible fields of application could be, for example, areas with strong analogies to industry-related questions such as process performance analysis in manufacturing analytics (e.g. Brodsky et al., 2015), but also areas with completely different contexts and data sources such as social media analytics (e.g. Kleindienst et al., 2015) or sales force analytics (e.g. von Bischhoffshausen et al., 2015).

A critical root element within the overall approach was the choice of the tripartite meta-characteristic to distinguish between outputs, inputs and throughputs of RDAP (i.e. analytical maintenance objectives, data characteristics and analytical techniques). From another perspective, this structure can also be associated with the three aspects *domain understanding*, *data understanding* and *analytical modelling*, which are helpful points to decompose the multidisciplinary nature of BDA tasks as described above. A similar decomposition can be found in common procedure models applied in data mining practices, where data analysis projects are broken down into well-defined steps to provide a road map for performing data analysis in a structured manner. Looking at such models' commonalities (Kurgan and Musilek, 2006), it can be argued that, besides those three parts above, our taxonomy lacks three essential aspects that are not yet fully covered with our tripartite meta-characteristic, but which are necessary to satisfy all important steps provided by common procedure models: *data preparation*, *evaluation* and *deployment*. Of particular interest is a more detailed consideration of the data preparation step, since our taxonomy already classifies major groups of pre-processing techniques, but ignores, for example, the systematization of recurring data quality treatments, such as outliers or missing values, which can occur when dealing with maintenance-related data. But also the classification of evaluation measures and techniques offers potentials to get an even more profound picture on RDAP in MA. At this stage, such additional aspects were not explicitly considered to keep the complexity of the taxonomy manageable. However, they show possible directions for further investigations.

By drawing further analogies to procedure models that help to perform recurring data analysis tasks, some more limitations of the taxonomy approach can be observed. Procedure models usually have an iterative character (Kurgan and Musilek, 2006), since data analysis is often carried out with many iterations, feedback cycles and an experimental design. As such, the attempt to break down a sophisticated data analysis problem and describe it in terms of an oversimplified framework appears contradictory. This issue becomes particularly apparent during the systematization of analytical techniques (see Section 4.1), where a sufficiently generic taxonomy structure required a flat and abstract representation of multifaceted analytical approaches, which, in contrast, should rather be classified hierarchically and multi-perspectively. For example, in our taxonomy, clustering and process mining approaches are organized at the same level, although clustering cannot only be considered as an diagnostic method to detect different machine behaviour but also serves as a pre-processing step to cluster event logs for subsequent process mining analyses. Nevertheless, in agreement with the experts surveyed, a suitable level of abstraction had to be found to preserve one of the most important purposes of the taxonomy, namely the comprehensibility for a quick orientation into the topic. Moreover, it can be argued that a data analysis problem is often that complex that it requires more than just a single technique. It rather demands a sequence of method combinations with algorithmic fine tuning. However, even in this case, it is possible to break down enhanced analysis problems into more fine grain sub-problems or tasks and then recombine them into a coherent whole. As such, we are still confident that our results may be an important step towards the conceptualization of RDAP to establish a reusable knowledge base.

Furthermore, we are confident that, despite the existence of procedure models, the taxonomy approach cannot be regarded as a dispensable tool. Rather, it should be understood as a possible supplement that can be integrated into process models to enrich them with the necessary context. This vision was also endorsed by two of the seven experts surveyed (head of division and division manager), who see potential for further research in such a project. The IT service provider already uses process models in data analysis projects, but they criticize that the existing process models are too generic and domain-independent to be used in a certain application context such as MA. Additionally, the respondents indicated two main scenarios in which they see potential benefits of the taxonomy to support their operational activities: First, as a communication tool in consulting and development meetings, where customer requirements for data-driven solutions are identified together with different stakeholder groups and then translated into solution drafts, and second, for internal and external qualification purposes to structure the field, identify current knowledge gaps and close them with suitable training programs.

6 Conclusion and Outlook

Applying BDA within the industrial practice can be a challenging task due to a variety of data types and multifaceted analytical techniques. At this point, this paper dealt with the establishment of a reusable knowledge base for MA, initiating the identification of dimension and characteristics of RDAP in data-driven maintenance scenarios. As a result, a taxonomy could be derived by applying the development method proposed by Nickerson et al. (2013). By not only using existing MA concepts and survey papers summarizing a large body of knowledge on maintenance research, but also conducting interviews with experts from a medium-sized IT service provider, it was possible to iteratively refine dimensions and characteristics towards a robust, precise and comprehensive taxonomy. Thus, the results can help researchers and practitioners to provide an overview about contextual aspects and deliver a systematization from a BDA perspective, as it structures the field by considering different specifications of analytical maintenance objectives, data characteristics and analytical techniques of RDAP.

In the next step, it is planned to use the results to establish reusable model repositories within an analytical platform for solving recurring analytics tasks in data-driven maintenance environments, such as exemplified by Brodsky et al. (2015). By integrating experiences from practice, the results have been validated in a first cycle of iterations. In further research, it is planned though to apply the results within more iterations to broader maintenance cases from the industrial practice to examine how well the derived taxonomy covers important dimensions and characteristics of real maintenance scenarios. At this stage, the paper does not claim to provide a taxonomy which is fully exhaustive, since the field of analytics is subject to a fast developing environment, where we expect the emergence of more diverse data types and new enhancements of analytical approaches in the near future. However, the results can be considered as a first proposal towards the systematization of RDAP in data-driven maintenance environments and we encourage practitioners and researchers to participate in further developments.

References

Abbasi, A., S. Sarker, and R. H. Chiang (2016). “Big Data Research in Information Systems: Toward an Inclusive Research Agenda.” *Journal of the Association for Information Systems* 17 (2), 1-32.

Accorsi, R., R. Manzini, P. Pascarella, M. Patella, and S. Sassi (2017). “Data Mining and Machine Learning for Condition-based Maintenance.” *Procedia Manufacturing* 11, 1153–1161.

Agarwal, R. and V. Dhar (2014). “Editorial - Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research.” *Information Systems Research* 25 (3), 443-448.

Ahmad, R., and S. Kamaruddin (2012). “An Overview of Time-Based and Condition-Based Maintenance in Industrial Application.” *Computers & Industrial Engineering* 63 (1), 135-149.

Ahmadzadeh, F. and J. Lundberg (2014). “Remaining Useful Life Estimation.” *International Journal of System Assurance Engineering and Management* 5 (4), 461-474.

An, D., N. H. Kim, and J. H. Choi (2015). “Practical Options for Selecting Data-Driven or Physics-Based Prognostics Algorithms with Reviews.” *Reliability Engineering & System Safety* 133, 223-236.

Ao, Y. H. (2011). “A Review on Development and Trend of Intelligent Maintenance System.” In: *Advanced Materials Research* 314, pp. 2365-2369.

Bousdekis, A., B. Magoutas, D. Apostolou, and G. Mentzas (2015). “Review, Analysis and Synthesis of Prognostic-Based Decision Support Methods for Condition Based Maintenance.” *Journal of Intelligent Manufacturing*, 1-14.

Brodsky, A., G. Shao, M. Krishnamoorthy, A. Narayanan, D. Menascé, and R. Ak (2015). “Analysis and Optimization in Smart Manufacturing based on a Reusable Knowledge Base for Process Performance Models.” In: *IEEE International Conference on Big Data (Big Data)*, pp. 1418-1427.

Chen, H., R. H. Chiang, and V. C. Storey (2012). “Business Intelligence and Analytics: From Big Data to Big Impact.” *MIS Quarterly* 36 (4), 1165-1188.

Constantiou, I. D. and J. Kallinikos (2015). “New Games, New Rules: Big Data and the Changing Context of Strategy.” *Journal of Information Technology* 30 (1), 44-57.

Debortoli, S., O. Müller, and J. v. Brocke (2014). “Comparing Business Intelligence and Big Data Skills.” *Business & Information Systems Engineering* 6 (5), 289-300.

Dragomir, O. E., R. Gouriveau, F. Dragomir, E. Minca, and N. Zerhouni (2009). “Review of Prognostic Problem in Condition-Based Maintenance. In: European Control Conference. IEEE Control Systems Society. Budapest, pp. 1587-1592.

Eckert, S. and J. F. Ehmke (2017). “Classification of Data Analysis Tasks for Production Environments.” In: *Business Information Systems Workshops*. Ed. by W. Abramowicz, R. Alt, and B. Franczyk. Lecture Notes in Business Information Processing (263). Springer, Cham, pp. 399-407.

Elattar, H. M., H. K. Elminir, and A. M. Riad (2016). “Prognostics: A Literature Review.” *Complex & Intelligent Systems* 2 (2), 125-154.

Famurewa, S. M., L. Zhang, and M. Asplund (2017). “Maintenance Analytics for Railway Infrastructure Decision Support.” *Journal of Quality in Maintenance Engineering* 23 (3), 310-325.

Flath, C. M. and N. Stein (2018). “Towards a Data Science Toolbox for Industrial Analytics Applications.” *Computers in Industry* 94, 16-25.

Goyal, D. and B. S. Pabla (2015). “Condition Based Maintenance of Machine Tools - A Review.” *CIRP Journal of Manufacturing Science and Technology* 10, 24-35.

Gregor, S. (2006). “The Nature of Theory in Information Systems.” *MIS Quarterly* 30 (3), 611-642.

Hashemian, H. M. and W. C. Bean (2011). “State-of-the-Art Predictive Maintenance Techniques.” *IEEE Transactions on Instrumentation and Measurement* 60 (10), 3480-3492.

Heng, A., S. Zhang, A. C. Tan, and J. Mathew (2009). “Rotating Machinery Prognostics: State of the Art, Challenges and Opportunities.” *Mechanical Systems and Signal Processing* 23 (3), 724-739.

Jardine, A. K., D. Lin, and D. Banjevic (2006). “A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance.” *Mechanical Systems and Signal Processing* 20 (7), 1483-1510.

Jöhnk, J., M. Röglinger, M. Thimmel, and N. Urbach (2017). "How to Implement Agile IT Setups: A Taxonomy of Design Options." In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*. Guimarães: Portugal.

Kaisler, S. H., J. A. Espinosa, F. Armour, and W. H. Money (2014). "Advanced Analytics - Issues and Challenges in a Global Environment." In: *47th Hawaii International Conference on System Sciences*. IEEE Computer Society. Waikoloa, pp. 729-738.

Kans, M. and D. Galar (2017). "The Impact of Maintenance 4.0 and Big Data Analytics within Strategic Asset Management." In: *Proceedings 6th International Conference on Maintenance Performance Measurement and Management*. Ed. by D. Galar and D. Seneviratne. Luleå University of Technology: Sweden, pp. 96-103.

Karim, R., J. Westerberg, D. Galar, and U. Kumar (2016). „Maintenance Analytics - The New Know in Maintenance. *IFAC-PapersOnLine* 49 (28), 214-219.

Kleindienst, D., R. Pfleger, and M. Schoch (2015). „The Business Alignment of Social Media Analytics." In: *Proceedings of the 23rd European Conference on Information Systems (ECIS)*. Münster: Germany.

Kothamasu, R., S. H. Huang, and W. H. VerDuin (2006). "System Health Monitoring and Prognostics - A Review of Current Paradigms and Practices." *The International Journal of Advanced Manufacturing Technology* 28 (9-10), 1012-1024.

Kurgan, L. A. and P. Musilek (2006). "A Survey of Knowledge Discovery and Data Mining Process Models." *The Knowledge Engineering Review* 21 (1), 1-24.

Lee, J., F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel (2014). "Prognostics and Health Management Design for Rotary Machinery Systems - Reviews, Methodology and Applications." *Mechanical Systems and Signal Processing* 42 (1), 314-334.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. Technical Report. McKinsey Global Institute.

Meeker, W. Q. and Y. Hong (2014). "Reliability Meets Big Data: Opportunities and Challenges." *Quality Engineering* 26 (1), 102-116.

Muchiri, P., L. Pintelon, L. Gelders, and H. Martin (2011). "Development of Maintenance Function Performance Measurement Framework and Indicators." *International Journal of Production Economics* 131 (1), 295-302.

Müller, O., I. Junglas, J. v. Brocke, and S. Debortoli (2016). "Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines." *European Journal of Information Systems* 25 (4), 289-302.

Myers, M. D. and M. Newman (2007). "The Qualitative Interview in IS Research: Examining the Craft." *Information and Organization* 17 (1), 2-26.

Nadj M. and C. Schieder (2017). "Towards a Taxonomy of Real-Time Business Intelligence Systems". In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*. Guimarães: Portugal.

Nickerson, R. C., U. Varshney, and J. Muntermann (2013). "A Method for Taxonomy Development and Its Application in Information Systems." *European Journal of Information Systems* 22 (3), 336-359.

Peng, Y., M. Dong, and M. J. Zuo (2010). "Current Status of Machine Prognostics in Condition-Based Maintenance: A Review." *The International Journal of Advanced Manufacturing Technology* 50 (1), 297-313.

Prajapati, A., J. Bechtel, and S. Ganesan (2012). "Condition Based Maintenance: A Survey." *Journal of Quality in Maintenance Engineering* 18 (4), 384-400.

Ramannavar, M. and N. S. Sidnal (2016). "Big Data and Analytics - A Journey Through Basic Concepts to Research Issues." In: *Proceedings of the International Conference on Soft Computing Systems, Advances in Intelligent Systems and Computing. Springer*. New Delhi, pp. 291-306.

Russo B. (2016). "The Need for Data Analysis Patterns (in Software Engineering)." In: *Perspectives on Data Science for Software Engineering*. Ed. by T. Zimmermann, L. Williams, and T. Menzies. Morgan Kaufmann Publishers. Burlington, pp. 19-23.

Püschel, L., M. Röglinger and H. Schlott (2016). “What's in a Smart Thing? Development of a Multi-Layer Taxonomy.” In: *37th International Conference on Information Systems*, pp. 1-19.
Schultze, U. and M. Avital (2011). “Designing Interviews to Generate Rich Data for Information Systems Research.” *Information and Organization* 21 (1), 1-16.
Schumann, C., P. Zschech, and A. Hilbert (2016). „Das aufstrebende Berufsbild des Data Scientist - Vom Kompetenzwirrwarr zu spezifischen Anforderungsprofilen.“ *HMD Praxis der Wirtschaftsinformatik* 53 (4), 453-466.
Schwabacher, M. (2005). “A Survey of Data-driven Prognostics.” In: *Proceedings of the AIAA Infotech@ Aerospace Conference*, pp. 1-5.
Sharma, R., S. Mithas, and A. Kankanhalli (2014). “Transforming Decision-Making Processes: A Research Agenda for Understanding the Impact of Business Analytics on Organisations.” *European Journal of Information Systems* 23 (4), 433-441.
Si, X. S., W. Wang, C. H. Hu, and D. H. Zhou (2011). “Remaining Useful Life Estimation - A Review on the Statistical Data Driven Approaches.” *European Journal of Operational Research* 213 (1), 1-14.
Sipos, R., D. Fradkin, F. Moerchen, and Z. Wang (2014). “Log-based Predictive Maintenance.” In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1867-1876.
Stein, N. and C. Flath (2017). “Applying Data Science for Shop-Floor Performance Prediction.” In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*. Guimarães: Portugal.
Tilly, R., O. Posegga, K. Fischbach, and D. Schoder (2017). “Towards a Conceptualization of Data and Information Quality in Social Information Systems.” *Business & Information Systems Engineering* 59(1), 3-21.
Tsai, C. W., C. F. Lai, M. C. Chiang, and L. T. Yang (2014). “Data Mining for Internet of Things: A Survey.” *IEEE Communications Surveys and Tutorials* 16 (1), 77-97.
Veldman, J., H. Wortmann, and W. Klingenberg (2011). “Typology of Condition Based Maintenance.” *Journal of Quality in Maintenance Engineering* 17 (2), 183-202.
Vogl, G. W., B. A. Weiss, B. A., and M. Helu (2016). “A Review of Diagnostic and Prognostic Capabilities and Best practices for Manufacturing.” *Journal of Intelligent Manufacturing*, 1-17.
vom Brocke, J., A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven (2009). “Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process.” In: *Proceedings of the 17th European Conference on Information Systems (ECIS)*. Verona: Italy.
von Bischhoffshausen, J. K., M. Paatsch, M. Reuter, G. Satzger, and H. Fromm (2015). “An Information System for Sales Team Assignments Utilizing Predictive and Prescriptive Analytics.” In: *IEEE 17th Conference on Business Informatics.* Lisbon: Portugal, pp. 68-76.
Watson, H. (2014). “Tutorial: Big Data Analytics: Concepts, Technologies, and Applications.” *Communications of the Association for Information Systems* 34 (1), 1247-1268.
Webster, J. and R. T. Watson (2002). “Analyzing the Past to Prepare for the Future: Writing a Literature Review.” *MIS Quarterly* 26 (2), xiii–xxiii.
Zschech P., K. Heinrich, M. Pfitzner, and A. Hilbert (2017). “Are You Up for the Challenge? Towards the Development of a Big Data Capability Assessment Model.” In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*. Guimarães: Portugal.
Zhu, L., B. Jiang, Y. Cheng (2016): “Life Prediction Methods Based on Data-Driven: Review and Trend.” In: *IEEE Chinese Guidance, Navigation and Control Conference*, pp. 1682-1686.