Towards an Integrative Big Data Analysis Framework for Data-driven Risk Management in Industry 4.0

Tim Niesen  
IWI at the DFKI  
Saarland University  
tim.niesen@iwi.dfki.de

Constantin Houy  
IWI at the DFKI  
Saarland University  
constantin.houy@iwi.dfki.de

Peter Fettke  
IWI at the DFKI  
Saarland University  
peter.fettke@iwi.dfki.de

Peter Loos  
IWI at the DFKI  
Saarland University  
peter.loos@iwi.dfki.de

Abstract

With the advent of Industry 4.0, industrial manufacturing systems constantly evolve into smart, interconnected production systems. Pervasive integration of information and communication technology into productional components results in massive amounts of various data. To meet the challenges that arise from an increasingly competitive market and more demanding customer requirements, technological drivers have to be leveraged in order to process data effectively. One important aspect in that regard is the efficient management of business processes and process risks. As an integrative concept in these areas is missing, we present a holistic framework for data-driven risk assessment based on real-time data. Besides a conceptual model, we provide a technical concept that combines methods for risk assessment with performance metrics and demonstrate a software implementation in the context of an exemplary use case scenario. Finally, we present the results of expert interviews and a discussion indicating future research directions.

1. Introduction

The manufacturing industry forms the backbone of many nations’ economies. This is especially true for countries like Germany where its share in the overall economic performance is much higher than in several other national economies that reach a comparable stage of development [1]. With a contribution of 25.9% to the GDP, it constitutes an important foundation for a prospering economy and – against the background of an advancing globalization – a core competency regarding international competition [1].

With the rise of new technologies and ways of working, traditional manufacturing faces tremendous changes that impose massive challenges on established business models as well as the way that manufacturing processes are organized today. Among the main drivers for these changes is the ongoing integration of information and communication technology (ICT) in industrial components. At the same time, numerous opportunities arise from these developments, leading to new business models that, for instance, focus on the individualization of manufacturing as well as cost and efficiency gains due to shorter innovation cycles, reduced time to market or self-adapting production [2]. Because of its “revolutionary” nature, current research in the field of interconnected ICT-enhanced production systems and associated developments is referred to as Industry 4.0 [3]. In an international context, similar concepts are also subsumed as Advanced Manufacturing or Manufacturing 2.0 in academia and practice [4, 5].

Industry 4.0 focuses on developing concepts and methods to make production processes more flexible and transforming currently fixed production lines into automated, autonomously organized assembly lines. The collection and analysis of extensive amounts of real-time data from various production resources provides the basis for achieving this goal. Processing this data by means of adapted methods from the fields of business intelligence and business analytics is key to establishing an integrated management information system, which supports appropriate strategic and operational decision-making. In particular, an integrated analytics database opens up new perspectives with respect to company-wide risk management.

However, integrating ICT into established manufacturing plants creates new danger potential that has until now only been of concern to information technology (IT) components and is not at all reflected in current concepts for industrial risk assessments. Figuratively speaking, ICT provides the “technological breeding grounds” for IT-related malware and spyware as well as IT risks like cyber attacks or system intrusion [6]. Prevalent concepts of risk management and the assessment of security threats are often based on the experience or intuition of decision makers and are, thus, reactionary and highly subjective. Reliable data mostly originates from technical specifications like manufacturer information concerning the operational life span or expected error behavior of a production machine. Based on the lack of valuable real-time data,
risk management is, therefore, often conducted irregularly and in a reactionary manner, which is not suitable for detecting operational risks during production time. With the progressing interconnection of production components through ICT and integrated analytics systems, live monitoring of relevant variables becomes feasible and enables real-time risk assessment.

In this paper, we argue that assessing risks and security threats should be based on objective and transparent metrics that build on integrated real-time data in Industry 4.0 scenarios. Our goal is, thus, to develop an integrated big data analytics framework for data-driven risk management. As a first step, we perform a requirements analysis, which yields a catalogue of risks that are of major importance to the manufacturing industry and especially Industry 4.0. Next, we develop a conceptual framework to establish methodical foundations for our concept by integrating approaches from the fields of business process management, process performance management, and risk management. Based on that framework, we finally propose a concept for real-time risk assessment that integrates operational data from a variety of heterogeneous sources into a unified database and provides a calculation scheme for the probability of occurrence for specific types of risks.

Furthermore, we present a software implementation of the concept that shows the idea’s applicability in an exemplary use case. Based on that concept, we further argue that many relevant types of risk in the context of the manufacturing industry can be assessed in an automated way, which in turn helps to leverage the potential of big data analytics for industrial risk management. This aspect was also confirmed in the evaluation of our approach based on expert interviews.

In particular, the following two research questions will be addressed throughout the paper:

(RQ1) Which kinds of risks can be assessed in an automated way using an integrated database of real-time data from manufacturing systems?

(RQ2) Which conceptual approaches are suitable for developing a comprehensive framework that provides an objective and transparent method for risk and security assessment?

The remainder of this paper is structured as follows: after this introduction, section 2 describes conceptual foundations to provide the relevant context of the work presented in this paper. Section 3 then explicates our research approach before section 4 describes our three-step approach towards a data-driven risk management framework. In section 5, a prototypical software implementation is demonstrated. Afterwards, section 6 describes evaluation results based on expert interviews. Section 7 discusses the potential as well as challenges of our concept before section 8 provides conclusions and an outlook on further research.

2. Conceptual Foundations

2.1. Industry 4.0

We refer to the ongoing technological advancement of manufacturing components towards intelligent and interconnected systems as Industry 4.0 [7]. After an initial mechanization through hydro and steam power, mass production and electrification as well as automation of production, this fourth industrial revolution focuses on the establishment of intelligent manufacturing components, smart objects and new production processes [8]. To provide for consistency and interoperability, various international organizations like the Industrial Internet Consortium (IIC) and the Object Management Group (OMG) are concerned with the standardization of interaction and management of the aforementioned concepts.1

Industry 4.0 is characterized by a progressing integration of ICT into manufacturing systems. Based on that, so-called cyber-physical systems (CPS) emerge at the intersection of IT components for information processing as well as data exchange and mechanical or electrical machine components [9]. Within the industrial context of interconnected manufacturing plants, these systems are also referred to as cyber-physical production systems (CPPS) [10]. This broad pervasion of ICT aligns with the vision of an Internet of Things and Services and supports a close integration along established structures for value creation [11]. Integration can be differentiated into vertical and horizontal integration. Vertical integration denotes an increasing information exchange and collaboration among different levels of hierarchy (management, corporate planning, production scheduling) within an enterprise. On the other hand, horizontal integration describes a close collaboration between multiple enterprises within the same value creation network [12]. A main driver for both forms of integration is the broad availability of efficient and affordable sensor networks (e. g., radio frequency identification, RFID). Hereby, intelligent or smart objects and devices are created that allow for real-time communication between machines, working resources and application systems. Taken together, these technological developments provide the basis for implementing new manufacturing processes and business models in so-called smart factories [13]. In a smart manufacturing environment, self-adaptive, flexible assembly lines allow for individual products to autonomously lead their way through the production life cycle [13]. Thus, future potential for mass customization and efficient reconfigurable manufacturing systems can be realized [14].

2.2. Big Data in the Context of Industry 4.0

With the ever-increasing number of data sources and a slump in prices regarding IT storage capacity and computational resources in recent years, collecting and analyzing large amounts of data became feasible. Under the designation big data a lot of research has been conducted in various disciplines including among others sales prediction, production planning or user relationship mining and clustering [15]. However, in this paper we focus on the area of smart manufacturing environments and their characteristics. The requirements in this area involve the analysis of unstructured machine-generated data from various sources, like machine-to-machine (M2M) communication data, operational sensor data (temperature, pressure, engine speed) as well as data from manufacturing execution systems (MES) for production scheduling [16]. Data mining approaches can be applied to “historical” databases to identify error patterns and correlations, which can in turn be used to perform predictive analyses. Thus, influential variables that affect the correct behavior of a technical component can be proactively monitored to avoid sudden failures (predictive maintenance [17]).

2.3. Risk Management

Risks – in the sense of potential deviations from a planned situation – are inherent components of every entrepreneurial activity and, thus, an important factor in business-related decision-making processes [18]. Awareness of different types of risks as well as metrics for risk assessment and techniques for risk control and governance are critical to an enterprise’s success and typically conceptualized as risk management (RM). ISO 31000 represents a family of standards that seeks to provide unified and generic guidelines by means of an industry-independent risk management approach.2 Throughout this paper, we use the term risk to refer to both IT-related and non-IT-related security threats.


A business process is typically defined as a set of interrelated activities, which defines an end-to-end work and creates an economic value for a customer [19]. In business process management (BPM), concepts for designing, modeling, implementing, executing and monitoring business processes are provided and supported by appropriate methods and tools [20]. To ensure a continuous management of business processes, related activities are organized in the form of BPM life cycles [21]. In order to evaluate the efficiency of a process, i.e. performance, process performance management (PPM) aims at quantifying their outputs by multidimensional approaches. Based on specific key performance indicators (KPI) process characteristics are measured and compared to desired values to obtain a performance score [22]. With industrial manufacturing shifting from strictly ordered production lines towards flexible, self-organizing smart machines, a clear understanding and proper governing of core business processes becomes indispensable. The broad dissemination of ICT in manufacturing environments introduces considerable complexity in terms of dynamical changes to process executions and, thus, affects the way that processes have to be managed [23]. In order to cope with the decentralization and modularization of manufacturing components, which result from an increasing demand for flexibility and highly dynamic value creation networks, traditional BPM techniques have to be extended. E.g., interoperability of business processes is essential in such scenarios because individual manufacturing steps are highly interwined and can impact modules involved in associated processes. Therefore, resilient BPM and continuous monitoring of process-related risks is key to managing smart manufacturing systems in an Industry 4.0 environment.

3. Research Approach

In our research, we use a design-oriented research approach [24]. We seek to develop an innovative artifact in the form of a methodical framework as well as a technical concept that can be applied in the domain of industrial risk management for Industry 4.0. This research was conducted in the context of a joint research project with partners from both academia and industry. As a first step in developing our approach to data-driven risk management, important risks and threats to production processes were collected using systematically retrieved literature from international scientific databases (e.g. Scopus, ScienceDirect) as well as existing standards for industrial risk and security management (e.g. ISO 27001). Thus, a comprehensive overview of a variety of risks from different business areas, ranging from operational to strategic threats was developed (deductive approach). After that, we identified factors affecting the achievement of process goals for an exemplary process in our research project and discussed them with all project partners to extend the collection by risks particularly relevant from a practical perspective (inductive approach). This resulted in a comprehensive catalogue of risks, which provides a structure for threats in the context of Industry 4.0.

4. Concept Development

4.1. Step 1 – Requirements analysis

As a first step towards developing a concept for data-driven risk management, we conducted a requirements analysis. The outcome of this analysis constitutes a systematic overview of relevant risks and security threats in the context of Industry 4.0. It focuses on security aspects that are not inherent to an industrial setting and predominantly arise due to a strong dissemination of ICT in that area. Therefore, general risks concerning financial, personal or strategic aspects are not explicitly mentioned. Furthermore, it emphasizes process-related threats, i.e. process risks. As argued before, a clear understanding of core processes and influencing factors is key to building flexible manufacturing systems. To provide a structured classification, collected risks were then consolidated in the form of a catalogue of risks with five sections as presented in Table 1. This catalogue aims to provide an overview of the aggregate of identified risks and to ease a situation-dependent concentration on specific risk categories. It is not intended to represent an all-encompassing collection of industry-related endangerment. Instead, it focuses on risks that are not very typical in a “traditional” industrial setting but are gaining importance in the advent of Industry 4.0.

Table 1: Catalogue of risks (overview)

<table>
<thead>
<tr>
<th>Section 1: Industrial espionage</th>
<th>Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT security-related threats that involve directed or undirected attacks on enterprise systems in order to steal business-critical, internal information.</td>
<td></td>
</tr>
<tr>
<td>data theft, exploitation of maintenance access, network vulnerability scans, phishing/pharming, security breaches</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2: Organisation and planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threats concerning entrepreneurial decision making and organizational aspects.</td>
</tr>
<tr>
<td>deficient event logging, deficient maintenance, system overload</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 3: Process-related risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threats related to efficient and compliant execution of core business processes.</td>
</tr>
<tr>
<td>costs, efficiency, flexibility, quality, time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 4: Sabotage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threats arising from deliberately manipulating, overloading or destroying hardware or software components of production systems.</td>
</tr>
<tr>
<td>Bring Your Own Device (BYOD), Denial-of-Service (DoS), malware, manipulation of communication, manipulation of data, manipulation of hardware/software</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 5: Technical misconduct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical threats that describe unspecified behavior of communication, hardware or software components due to defects or aging.</td>
</tr>
<tr>
<td>Hardware: environmental influences, IT infrastructure, production system, supportive infrastructure, wear and tear</td>
</tr>
<tr>
<td>Software: access control management, authentication, data consistency/integrity</td>
</tr>
</tbody>
</table>

Our approach to data-driven risk management presented in section 4.3 focuses on process-related and resource-related risks (see Table 2) since their influence on the performance of production processes is paramount.

4.2. Step 2 – Meta Model of Risk Assessment

Having established a structured catalogue of risks with a focus on Industry 4.0, we define a conceptual framework to ground our approach for risk assessment based on real-time data. It draws on methodical principles from the fields of BPM and PPM and combines them with elements from RM into a novel concept. As we argue that RM in smart manufacturing environments must incorporate concepts from both BPM and PPM, we proceed on the following assumptions:

1. Governance of business processes and examining process risks are essential for RM based on real-time operational data in Industry 4.0,
2. To investigate performance, risk and goal attainment of processes, approaches from BPM, PPM and RM have to be integrated and combined,

3. Risks have to be assessed by means of clearly defined data structures and indicators in a designated calculation scheme building upon these structures.

With regard to the previous assumptions, substantial constructs and relations of the framework have been captured in the form of a meta model. Figure 1 depicts a graphical representation of this model. It is inspired by preliminary work regarding process-related risk taxonomies and multi-perspective modeling methods for IT risk assessment [25]. We extend these works by incorporating aspects from PPM, yielding an abstract representation of structural and semantic relations between process-, performance- and risk-related elements. Thus, it constitutes a reference framework that can serve as a basis for our further work.

In this model, we distinguish three perspectives, i.e. process, risk and performance. Relations linking different perspectives illustrate the integration of methodical principles from different fields.

Starting from the process perspective, a process constitutes the central entity in the model. Processes are organized in hierarchical structures and can be composed of sub-processes. They consist of an arbitrary number of activities, which represent specific tasks in a process sequence, and they are assigned to one or more supportive resources. Among applications or information systems, technical components or other means of production can be relevant resources. They serve as an interface between the process and risk perspective and supply various kinds of data, which are in turn input for KPI calculations. Like processes, KPIs can be arranged in a hierarchical structure to reflect various abstraction levels. For one, indicator values are used to assess process performance in terms of the dimensions cost, quality and time based on predefined thresholds. Furthermore, these values are employed to perform risk assessment by using risk-specific evaluation schemes building on selected KPIs (see section 4.3). Schemes, in addition, conceptualize the relations between performance and risk perspectives and, thus, brace our data-driven approach to RM. Risks affect entities within the process perspective in two different ways: firstly, in the form of process risks and, secondly, in the form of resource risks. As process risks we subsume impacts on process characteristics like cycle time, correct execution or compliance to reference processes. On the other hand, resource risks threaten process-supporting IT systems and technical components. This two-step approach allows for an explicit integration of risks into the field of industrial production as well as into an overall process model structure. Each risk comprises a hierarchical order similar to processes or KPIs and is linked to a measure for risk response. The occurrence of a risk almost always compromises the attainment of certain goals, which again are facilitated by business processes. Establishing a structure of sub-goals enables us to separate strategic from operational goals.

4.3. Step 3 – KPI-based Calculation Scheme

The calculation scheme presented below draws on the conceptual relations in the meta model and provides a means for data-driven risk assessment. In order to perform risk assessment automatically, i.e. to estimate the probability of occurrence per risk based on current data from the production system, the following assumptions have to hold:
1. In principle, a risk can be reliably detected from specifics in data records. That is, if appropriate data is available, a causal relationship or statistical correlation between data records and the probability of risk occurrence may be assumed.

2. Computing the probability of occurrence can be automated using a consistent method providing repeatable results for the same characteristics of data records, i.e., it is reliable and objective. Building on these assumptions, Figure 2 outlines the context in which BPM, PPM and RM are integrated throughout our concept.

**Figure 2. Integration of BPM, PPM and RM**

The figure is divided into the three layers process, resource and data. Within the process layer, the exemplary business process P – consisting of events E₁ to E₃ together with activities A₁ and A₂ – is supported by resources 1 to 4, for instance technical devices or software systems. During runtime, resources produce numerous operational data ranging from sensor values (e.g., temperature) to aggregated values (e.g., total process cycle time). Referring to the notion of risks stated in the meta model, risks are modeled in the process perspective (PR₁ to PRₙ) as well as in resource perspective (RR₁ to RRₙ). With a focus on the overall business process, this separation is important: while resource risks are local to individual components in manufacturing systems, process risks arise from the occurrence of resource-related risks. They, hence, subsume endangerments to the entire process on a higher level. Specifically, our concept for automated risk assessment focuses on the following risks.

**Table 2. Process- and resource-related risks**

<table>
<thead>
<tr>
<th>Process-related risks:</th>
<th>delay of subsequent processes, increased cycle time, inferior efficiency of production, inferior flexibility of production, increased production costs, inferior throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource-related risks:</td>
<td>inferior production quality, increased use of resources, technical faults</td>
</tr>
</tbody>
</table>

In order to finally establish a calculation scheme for an automated evaluation of the risks depicted in Table 2, we compiled a total number of 33 KPI from a variety of sources. For instance, we analyzed several catalogues with German industry standards to derive indicators. In addition, we paid particular attention to the characteristics of the application scenario in our research project when deducing indicators. After compilation, the indicators were arranged into a hierarchical structure with respect to the three performance dimension cost, quality and time. An overview of the hierarchy is visualized in Figure 3. Entries in light shading denote process-related indicators while entries in dark shading denote technical indicators. Striped entries eventually denote a coupling of process-related and technical (sub-)indicators.

**Figure 3. KPI hierarchy (Overview)**

Different levels within the hierarchy break down the above process performance dimensions into sub-indicators that – on the lowest level – are based on operational data. In Figure 3, lines and bulleted lists denote hierarchical relations. At each level, a score is calculated that provides a measure for the performance of the indicator. The scoring builds on a four-stage scale ranging from 10 points (high performance) to 5 points, 1 point to 0 points (no considerable performance). On the lowest level (basic indicators), operational data is compared with predefined classes of thresholds to obtain a scoring. For instance, transport time denotes the total time that is spent for transport within a production process. If the measured value for that indicator deviates from a desired value by amount x, the performance-based scoring is incrementally decreased towards 0 points. On higher levels in the hierarchy (compound indicators), the scoring is based on the weighted average of the scoring of subordinate indicators. For example, to obtain a score for cycle time, the arithmetic mean of the scores of processing time, set-up time, transport time and waiting time is calculated. In order to systematically describe KPIs and to link them to process- and resource-related risks (see Table 2), we introduce the concept of KPI profiles and risk profiles. Table 3 shows an exemplary profile for the indicator Level of allocation.
A KPI profile is divided into four sections. (1.) General contains the name and abbreviation of an indicator as well as its original source and a description. Furthermore, the target value defines the maximum value, i.e. best performance, of the indicator while the scoring scheme provides a mapping between indicator values and corresponding scoring points. Indicator values correspond to the result of the formula contained in the Data section. (2.) Risks presents which risk evaluations are affected by the particular indicator. This is a cross-link to the respective risk profiles discussed below. (3.) Data describes the sources from which input data is retrieved along with the measured values and their unit. To obtain an indicator value for the mapping scheme, the calculation formula defines how measured values are processed. For compound indicators, the weighted average of subordinate indicators $I_1$ to $I_p$ is calculated as $a_1 \cdot \text{score}(I_1) + ... + a_p \cdot \text{score}(I_p)$ with $a_1 + ... + a_p = 1$ where $\text{score}(I)$ is defined according to indicators’ profiles. Then, hierarchical relations in terms of superior and subordinate KPIs are given in (4.) Hierarchy.

To draw inferences about the probabilities of occurrence for a particular risk from the indicator scorings, risk profiles have been specified analogously to the KPI profiles. An example is depicted in Table 4. (1.) General provides the name, abbreviation and description of a risk, as well as its type. The probability of occurrence is determined by a mapping between summed KPI scores and an estimation of probability on a five-stage scale from 1 to 5 (scorings originating from calculations in section Data and calculation, see below). An expected extend of damage must be specified on the same five-stage scale, then allowing for final risk assessment according to the evaluation scheme formula. (2.) Data and calculation lists KPIs that are considered significant for the assessment of a risk and, hence, should be taken into account during evaluation.

Because of its modular composition, the hierarchy of KPIs may be extended and organized as desired. Thus, it can be adapted to particular use cases, with respect to the data available and depending on the risks that are considered to be most important. By linking various indicators – from potentially different levels of the hierarchy – to a risk laying weight on specific risk factors becomes feasible.

## 5. Implementation

In order to provide an exemplary implementation of the developed concept, we first designed a modular software architecture that comprises different function blocks (see Figure 4). On the server side, the back-end components are implemented in JavaScript (JS) using the Node.js platform. Data is stored using a MongoDB database and transferred using JSON (JavaScript Object Notation) throughout the entire platform. On the client side front-end component, we also use JavaScript to display results of the analysis and to manipulate options or parameters in the risk assessment scheme. The module DataIntegration contains interfaces to integrate unstructured and heterogeneous data from various data sources; right now we provide bindings to OPC UA, which is the de-facto standard for industrial M2M communication, as well as a simulation-based test component. We further plan on implementing a REST interface to manufacturing execution systems (SAP ME) and an import module to process log files from application systems. In the module DataManager, integrated data is persistently stored in a non-relational database (NoSQL).
Because the data is highly heterogeneous in terms of structure, this kind of document-oriented storage can best be leveraged for further processing. Please note that we do not aim at structuring data in the sense of classical data warehousing. We rather intend to efficiently and temporarily save relations between data for further analysis; subsequent to the analysis, only aggregated information is stored for historical trend analysis. Methods to periodically retrieve selected data from the storage are provided in the module DataAnalytics. Serving as input for the calculation scheme introduced in section 4.3, real-time risk assessment is conducted based on this data. There exist individual JS classes per KPI and risk profiles that implement their specified characteristics and interrelations. The frequency of data retrieval is defined as a parameter and can be adjusted by the user. Aiming at a real-time risk assessment, the frequency should be high enough to reflect this demand.

However, since updating risk evaluations in intervals of several seconds or even shorter might be too volatile because of temporary extremes in the database, the frequency is initially set to 5 minutes. This means that real-time data is available for risk assessment at any time but the user can define the update interval according to their needs. On the front-end, the module RiskManagement handles user interactions with the analytics back-end component. Figure 5 represents a screenshot of the client-side user interface providing a management cockpit perspective. It is vertically divided into two areas. On the left side, KPI values in the simulation component can be adjusted by sliders to investigate influences on risks and relationships between multiple indicators. The right side provides different views on the results from the back-end analyses. In Figure 5, a visualization of risks is given as a risk heat map, which is a standard form of visualization in RM. The x-axis shows the probability of risk occurrence while the y-axis shows the expected extend of damage. Each risk is placed in one of the 25 fields, which originate from the five steps on the defined scales for risk assessment. Rendering of the map is done in real-time and can smoothly be adapted to a desired update interval. In addition, other views showing a real-time KPI value hierarchy, a real-time or a historical risk assessment are provided in corresponding tabs above the visualization pane.
6. Evaluations

To further evaluate our concept and its implementation, we conducted guided interviews with two industry experts in May 2015. The first expert draws on several years of experience in RM as well as in GRC (governance, risk, compliance). The second expert is specialized in production planning and scheduling (PPS) and deals with risk-related issues in the context of intelligent order processing.

Interview questions were organized in three sections: 1.) general information about the interviewee, 2.) questions regarding the transformation of industry toward Industry 4.0 and 3.) questions about the concept presented in this paper. Asked about new challenges for RM in the era of Industry 4.0, the experts state data explosion and information overload due to the advancing integration of ICT as an important aspect. In this context, interpreting data in a reasonable way is considered the key to efficiently assessing risks. Both experts see IT-related threats like cyber attacks or data theft among the most critical challenges for Industry 4.0. At the same time, the potential of ICT-enhanced manufacturing systems is considered to be very high. In particular, big data and efficient analytics will gain importance for industrial production, both experts agree: because real-time data becomes widely available, decisions in the RM process can be taken much faster based on reliable data. In addition, new service concepts concerning maintenance and usage statistics provide novel insights into the impact of risks and their relationships. At the time of our interviews, both experts were familiar with the presented concept as it was provided to them for internal use. We asked them whether they recognize benefits with regard to adequately assessing risk situations based on operational data. The expert in the field of RM emphasized the consistency of the concept as well as the possibility to focus on specific levels in the hierarchy. Thus, he says, it is possible to customize the results of the analysis for individual user groups. According to the PPS expert, the approach’s simplicity and clear structure provide transparent instructions for the calculation of probabilities of risk occurrences. At this point, one interviewee remarks that creating and maintaining transparency across the entire RM process also entails great effort and must be reviewed regularly. Furthermore, the consideration of the process dimensions cost, quality and time is highlighted as a very useful characteristic. Asked about comparable tools, the experts describe the modular and clear structure of the concept as an advantage over existing solutions. Since the hierarchy can be extended, individual relations between indicators and risks can be modeled making it highly customizable. Further, existing and partly commercial solutions known to the interviewees mostly rely on historical data, which means that they do not implement data analytics on real-time data. According to one of the experts, the concept’s transparency can also be seen as a disadvantage since it does not involve experience-based knowledge of expert risk managers very much (“the gut feeling could be overruled by the calculated figures”).

7. Discussion of Potential and Challenges

The presented approach particularly supports the process of RM in terms of identifying relevant risks, linking them to performance indicators and defining evaluation schemes. Leveraging available real-time data from manufacturing systems, thus, provides adequate means for RM in Industry 4.0. For future work, we concentrate on the following three areas: 1.) Stronger consideration of IT-related threats by analyzing appropriate data e.g., application log files or reports from intrusion detection systems (IDS). Building on existing security reference models like the Common Vulnerabilities and Exposure system (CVE) [26] may help to define relevant indicators for such IT threats. 2.) We seek to adapt concepts from the area of predictive maintenance to proactively monitor machine statuses and conditions. In that regard, data mining and machine learning approaches could be used to learn patterns automatically. 3.) We will investigate how the process of RM could be further automated. In addition to an ongoing risk assessment, appropriate measures for risk handling could be triggered automatically when certain risks occur. Incorporating IT-related threats into an automated concept for risk assessment also issues new challenges. Because security risks are highly diverse, they can often not be detected from data characteristics. For instance, breaking into an information system can be performed in different ways. Furthermore, if so-called zero day exploits are used to take advantage of yet unknown security flaws, detecting an attack is nearly impossible. However, drawing on data from IDS, unusual patterns might be detected, indicating potential security risks.

8. Conclusions and Future Work

In this paper, we focused on extending traditional approaches to industrial RM by a data-driven analysis framework. Leveraging big data in the context of Industry 4.0, we provided an approach integrating heterogeneous data sources to enable automated risk assessment. To evaluate our concept, we presented a software implementation and the results of expert
interviews that were conducted with RM experts from organizational practice. In the future, our prototype will be used in practical use case scenarios to further examine its performance and to investigate potential for further development.

Acknowledgement: The research described in this paper was partly supported by a grant from the German Federal Ministry of Education and Research (BMBF), project name: LinAprom, support code 01IS12050. Moreover, the authors would like to thank the anonymous reviewers and the mini-track co-chairs for their constructive comments which contributed to the improvement of this article.

9. References


