Electronic Markets
https://doi.org/10.1007/s12525-019-00383-6

RESEARCH PAPER

Value of data meets IT security – assessing IT security risks in data-driven value chains

Laura Bitomsky[1] · Olga Bürger[2,4] · Björn Häckel[3,4] · Jannick Töppel[3,4]

Received: 10 August 2018 / Accepted: 17 October 2019

Abstract
Digitalization forces manufacturing companies to shift towards customer-oriented, highly data-driven forms of value creation. This results in a changing IT security risk landscape as data becomes an attractive target for adversaries leading to an increasing number of attacks. In order to successfully protect data, it is essential that it is assessed in an integrated manner. Although IT security and data-based value creation have been studied by large research bodies, the existing literature fails to provide guidance on IT security risk analysis in data-based value chains. To contribute to the closure of this research gap, we propose a modeling approach which allocates different data types to value activities and analyses the data types in relation to the properties of relevant IT security risks. The evaluation, conducted with industry experts, reveals that it is not only a company's primary assets that are of concern but also less important data types subject to significant levels of exposure that bear considerable IT security risks.

Keywords Data-driven value creation · Value of data · IT security · Data security · Risk assessment

JEL classification O3

Introduction

Digitalization continues to significantly transform the way companies conduct business across all sectors of the economy, often enhanced by technological enablers such as big data analysis, cloud computing, mobile technologies, and integrated sensor networks (Müller et al. 2016). Major trends such as servitization and the Internet of Things (IoT) further amplify these changes, promoting the shift from product-centric to customer-centric, highly data-driven value creation.

Responsible Editor: Volker Bach

✉ Laura Bitomsky
laura.bitomsky@fim-rc.de

[1] FIM Research Center, University of Augsburg, Universitätsstrasse 12, 86159 Augsburg, Germany

[2] FIM Research Center, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

[3] FIM Research Center, University of Applied Sciences Augsburg, Friedberger Straße 2a, 86161 Augsburg, Germany

[4] Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Universitätsstrasse 12, 86159 Augsburg, Germany

However, interlinking data with a company's value creation also entails a shift in the company's risk landscape. Firstly, the increasing importance of data as a new value driver makes it an attractive target for adversaries, leading to an increasing number of attempts to steal, manipulate or deny the use of data. Further, data-driven value chains involve integrating data into products and services and sharing them with external partners and in-house, which leads to new vulnerabilities and increases the attack surface. Moreover, the increasing dependency of products and services, as well as the value chains themselves, on data can lead to considerable damage when data breaches do occur. The increasing professionalism of hacker attacks and exponential increases in the quantity of malicious software (BSI 2016) should also be taken into account.

To protect their data in an appropriate manner, companies need to assess the IT security risks that arise as a result of the shift towards data-driven value creation and derive adequate security measures. Various frameworks, such as the NIST cybersecurity framework, and standards, such as the ISO/IEC 27000 series and – on a national level – the German IT-Grundschutz, have emerged, providing guidance to organizations working to identify and manage cybersecurity risks. An integral part of these efforts is risk assessment and

Published online: 18 February 2020

management in the case of a security breach. Therefore, data must be assessed from both a value and a security perspective in order to enable an economically well-founded evaluation and the prioritization of mitigation measures.

To date, however, it has been extremely challenging to quantitatively assess the risk of data breach. This is largely due to the multitude of parameters that need to be taken into consideration. Moreover, a large research gap exists in the literature between data and value creation, despite the fact that scholars and practitioners have paid significant interest to data-based value creation, particularly in the context of big data (Ekbia et al. 2015; Lim et al. 2017; Ostromet et al. 2015; ur Rehman et al. 2016; Yaqoob et al. 2016). While existing studies identify the use of data as a key success factor in customer satisfaction and have discussed the positive effects of integrating data into services and products, these studies fail to offer guidance on quantifying data in terms of value contribution or affiliated IT security risks. Hence, although existing security standards and frameworks require the value of data to be determined with respect to their importance for key business processes, these standards and frameworks provide no further guidance on how to perform this assessment. The result is that a majority of companies still face the challenge of identifying which of their current and future data contribute to value creation (the so-called 'crown jewels') and are critical from a security perspective, and thus need to be protected.

In order to bridge this research gap, we develop a model to 1) assist the identification of data types involved in value creation and their allocation to individual value activities, and 2) assess them from an IT security risk perspective. This allows users to assess the integration of data types into the value creation process in terms of both their individual value contribution and the associated IT security risks. Our model also helps users to identify a company's 'crown jewels' – that is, their most valuable data – and allows them to simulate and assess potential future value creation in different business models, e.g. with a shift from a product-centric value creation towards the increased integration of information-intensive services. Therefore, the objective of our approach is to lay the groundwork for data assessment in relation to value creation, guiding companies in the identification of appropriate IT security investment strategies and bridging the existing research gap by connecting data and value creation from an IT security perspective.

The remainder of this paper is organized as follows: Firstly, we provide a brief overview of existing literature related to data-driven value creation and data security. We then introduce the underlying methodology and develop the model for value chain analysis and the calculation of the *Probability Weighted Risk Indicator* for risk analysis in data-driven value creation. Next, we illustrate the applicability of the model with two real-world use-cases before conducting a more general evaluation. We conclude by discussing the results and implications for companies' IT security investment strategies and providing an outlook for further research.

Literature review

Data-driven value creation

A traditional and widely acknowledged concept for modeling value creation is the *value chain* developed by Porter (1985), which states that, within the course of value creation, a company performs *value creation activities* to "design, produce, market, deliver and support its product" (Porter 1985, p.36). The *primary activities* (i.e., inbound logistics, operations, outbound logistics, marketing, as well as sales and services) are "involved in the physical creation of the product and its sale and transfer to the buyer as well as after-sale assistance" (Porter 1985, p. 38). While the primary activities each represent one step in the process of value creation, *supporting activities* (i.e., procurement, technology development, human resource management, and firm infrastructure) provide support for primary activities and for each other. Therefore, they can be associated with a specific value activity and with the entire value chain.

However, digitalization forces traditional manufacturers to deliver smart, connected products and services, which disrupts the traditional value chains and leads to changes to existing value activities (e.g., marketing, or sales and services) and a need for new activities (e.g., product data analytics and security) (Porter and Heppelmann 2014). Moreover, in the past decades, there has been a shift from hierarchical, integrated, sequential supply chains toward strategic partnerships with external entities, which has resulted in fragmented networks (Bitran et al. 2007) and ecosystems. Digitally-enabled networks, in particular, have increased in popularity as new digital technologies fundamentally reshape traditional business models into modular processes that are globally distributed and cross-functional (Sambamurthy et al. 2003; Straub and Watson 2001; Wheeler 2002).

Various approaches can be used to describe these new business models, networks, and ecosystems. For example, the E3-value business model is a well-known approach used to describe e-business models (Liu and Jia 2010; Shoukry et al. 2019). It considers *actors* (e.g., a company or a person), *market segments* (i.e., segments with common properties), *value activities* (i.e., processes which add value), *value objects* (e.g., services or money), *value ports* (i.e., port used to provide or request value objects), *value interfaces* (i.e., a service offered to or requested from a value activity), and *value exchanges* (i.e., the potential trading of value objects between two connected value ports) (Liu and Jia 2010). Liu and Jia (2010) show how this model can be used for an IoT-based drug

supply chain by defining nine value activities for eight actors, whereby the actor "drug manufacturer" has only one generic value activity: producing and selling drugs. Papert and Pflaum (2017) develop an ecosystem model for the realization of IoT services in supply chain management. They define 19 roles and their value contribution, including the role of "product manufacturer" which has only one generic value contribution: the provision of physical objects for smart products.

In sum, we can state that recent research mainly focuses on modeling value creation in ecosystems wherein a manufacturer is only one of several actors. Although this reflects the increasing fragmentation of value creation in digitized ecosystems and helps to describe the interrelationships between value creation partners, it means that the value creation activities of individual actors are only described at a higher level of granularity. This limits the applicability of these approaches in relation to our concrete research goal, as we aim to allocate value creation-relevant data types to individual value activities. Therefore, we require a more nuanced view of value activities from a company perspective, as offered by more traditional approaches such as that developed by Porter (1985).

With the shift to digital, data-driven business models, a profound understanding of data-driven value creation has become essential in the modern, data-rich economy. However, despite the growing body of research on data-driven value creation – particularly in the context of big data – there still exists a notable shortage of research that takes a holistic approach to examining the links between data and value creation (Ekbia et al. 2015; Lim et al. 2017; Ostromet et al. 2015; ur Rehman et al. 2016; Yaqoob et al. 2016). In general, existing research agrees that big data and big data analytics (BDA) can create business value (for a more detailed literature review see, e.g., Akter and Wamba 2016; Günther et al. 2017) by providing transactional, informational, and strategic benefits (Akter and Wamba 2016; Wixom et al. 2013). Moreover, empirical studies confirm that using BDA can be beneficial for companies. For example, Chen et al. (2015) provide empirical evidence of the impact BDA has on business growth. Swanson (2001) finds a significant positive relationship between the use of data-based proactive maintenance strategies, such as predictive maintenance, and overall company performance rates. Yoo et al. (2014) show how hospitals enable medical practitioners and hospital administrators to enhance the quality of their services through the collection and analysis of operational data. These studies identify the use of data as a key success factor when it comes to customer satisfaction and discuss the positive effects of integrating data into services and products.

However, there exists a research gap regarding the mechanisms behind these benefits, in other words, how different activities and resources (i.e., data) need to work together in order to create value (Lim et al. 2018). Akter and Wamba (2016) specify four types of big data that can be used to create value in e-commerce: transaction or business activity data; click-stream data; video data, and voice data. Porter and Heppelmann (2014) show product data to be fundamental for value creation. For example, analyzing product usage and performance data improves the after-sale service (e.g., through predictive maintenance) and marketing (e.g., customized offerings) (Porter and Heppelmann 2014). Based on the idea that value is created with the use of information and by applying information within a process, Lim et al. (2018) designed the "Data-Value Chain" in the context of information-intensive services (IIS). Their intention is to provide a comprehensive framework to analyze the overall spectrum of data-based value creation. While the study does important groundwork for understanding the transformation of data into value, the authors fail to provide answers as to how one might adequately measure the value of data or the contribution such data makes to the overall value created.

IT security risks

One widely studied concept relating to information security threats is the distinction between unauthorized information release (confidentiality), unauthorized information modification (integrity), and unauthorized denial of use (availability). The three are collectively known as the 'CIA triad' (e.g. Anderson 1972; BSI 2016; Saltzer and Schroeder 1975) and are considered to be the basic protection goals of information security. In response to the constant and dynamic development of both information technology and information security threats, the CIA triad has been refined and extended throughout the years. However, our literature review shows that, until now, there has been no agreed-upon set of goals exceeding the CIA triad.

As data-driven products and services continue to change existing business models, information security gains new significance. This rising significance is evident in the exponential growth in the number of security threats (BSI 2016) resulting in increased costs in the event of a successful breach. For example, Grobauer et al. (2011) outline that, in cloud computing, the threat not only comes from the emergence of new vulnerabilities but also from the enhancement of well-established vulnerabilities. Based on estimations by Information Week and PricewaterhouseCoopers LLP in 2000, computer viruses and hacking took a "$1.6 trillion toll on the worldwide economy and $266 billion in the U.S alone" that year (Denning 2000). Thus, further research on the value of data is needed with a particular focus on the criticality of data and the impact of IT security breaches.

Quantifying the exact impact of an IT security breach is highly challenging due to the multitude of parameters that must be considered. The cost of an IT breach does not only involve the short-term cost incurring during the period of the breach, such as lost business, decreased productivity due to

the unavailability of necessary resources, and so forth (D'Amico 2000). In addition to these obvious costs, long-term costs are incurred, such as costs related to customers who lost faith in the company and switched to competitors, legal liabilities, etc.. However, these long-term costs are often difficult to estimate (Cavusoglu et al. 2004a). Consequently, most existing studies fail to adequately quantify the economic impact that IT security breaches have on companies, as these studies are based on self-reported company data, undermining the credibility of estimations due to a widespread tendency among companies to under-report the actual financial impact or not to report it at all (Garg et al. 2003).

One approach commonly used in such research is an event-study methodology based on Fama et al. (1969), which is used to analyze abnormal returns during a pre-determined period of time around the event in question. This methodology is highly popular in the accounting and finance literature (e.g. Friedman and Singh 1989; Koh and Venkatraman 1991). However, it has produced divergent results when applied to IT security breaches. Goel and Shawky (2008) analyzed the impact of 168 security breach incidents on the market value of publicly traded companies, and present evidence that such announcements led to a 1% reduction in market value during the days prior to and after the security breach. Cavusoglu et al. (2004a) reported an average reduction of 2.1%, translating into an average loss of $1.65 billion in market capitalization per incident. Garg et al. (2003) conducted a similar study which focused on only 22 events, but which analyzed the security breaches, classifying each as one of four major incident types (web site defacement, DoS, theft of customer information, theft of credit card information). The study indicates that theft of credit card information has the most severe impact on market value, with an average fall in share price of 9.3% on the day of the announcement and a significant fall of 15% over a three-day period. Overall, their research demonstrates that such events have a much higher impact than is indicated by other studies.

A different approach was taken by Longstaff et al. (2002), who developed a Hierarchical Holographic Model (HHM) to assess security risks of IT, based on the idea of integrating both exogenous and endogenous events into the risk analysis. By doing so, they aimed to achieve a more holistic approach for modeling complex systems which, as their approach highlights, are both interdependent and interconnected. Complementing these impact studies, research into the optimal investment strategies for IT security also emerged (e.g. Gordon and Loeb 2002). Cavusoglu et al. (2004b) developed a conceptual framework focusing on the optimal level of information security investments, which takes account of the criticality of information and the associated loss of such criticality. They conclude that organizations should concentrate on the protection of information with midrange vulnerabilities, as the benefits of protecting highly vulnerable information might not justify the inordinate associated expenses.

However, the approach taken in the majority of the existing literature is too generic to identify the vulnerability of specific data or the loss associated with this vulnerability, as most studies take a holistic approach rather than mapping IT security risks to the data itself. The result is that these approaches offer companies no guidance on how to identify their 'crown jewels' – that is, those pieces of data particularly worthy of protection – or on how to evaluate and reduce their risk exposure.

With data-based services revolutionizing the ways in which companies conduct business and create value, it is important to consider changes in both value creation and the risk landscape in order to protect emerging data-related crown jewels. However, recent research lacks approaches which can be used to measure the value contribution of data or analyze the associated IT security risks in data-driven value chains. Nevertheless, it is essential to find quantification approaches which can be used to manage the risk of data breaches as – in the context of risk management – risk identification, risk analysis, and risk evaluation are integral parts of risk assessment (Purdy 2010). Our objective, therefore, is to address this research gap by providing a modeling approach that links individual business data to value creation activities and enables the user to assess each data type in terms of its criticality and potential loss in case of a successful security breach. By distinguishing individual data types and their contribution to a company's value creation, data types are made comparable. When carrying out simulations of potential changes in future value creation, these data types are also made intertemporally comparable by the early identification of a shift in the company's crown jewels. This lays the groundwork for a thorough analysis of a company's current and future IT risk landscape and indicates possibilities for the adjustment of IT security investment strategy.

Model

Methodology

Our approach is largely based on the Noy and McGuinness' (2001) method of 'ontology development' and on 'normative analytical modeling' as outlined by, for example, Meredith et al. (1989). We used ontology development to structure the development process of our model and to determine the model's key parameters. Following the normative analytical modeling approach, we developed a key measure for quantifying IT security risks.

Within the scope of this paper, we understand ontologies as "explicit specifications of conceptualizations" (Gruber, 1993, p. 199), meaning a formal and declarative representation of an abstract, simplified view of a real-world problem or situation. There have been many empirical studies concerning the

proper development of ontologies. The method applied within the scope of this paper follows the approach presented by Noy and McGuinness (2001), who developed a seven-step guide for the development of ontologies. The advantage of this approach is that it stringently structures the development process of an artifact and thus offers valuable guidance. Table 1 shows how our approach follows their guidelines.

Normative analytical modeling captures the essentials of a decision problem, using mathematical representations to produce a prescriptive result. Such analyses provide support when structuring decision problems and resolving trade-offs among different criteria against a given target function. They also enable the user to make informed choices about the available options (Keeney and Raiffa 1993). Based on the classes and their properties derived through ontology development, we primarily follow Meredith et al. (1989) in developing a model for IT security risk analysis in data-driven value chains. We host two evaluative workshops with industry experts in order to validate the model's applicability.

Model development

In our model, we provide a two-phase approach. In the first phase, we derive an instrument for a value chain analysis that enables the identification of strategically important value activities within a company and the most important data types affiliated with these activities. In the second phase, we provide a procedure for IT security risk analysis based on the value chain developed in phase 1 to measure the value contribution of value activities and the associated data, and their security-related criticality.

Phase 1: Value chain analysis

In order to properly assess the value contribution and criticality of data types, a company must first identify its strategically important value activities and their affiliated data types.

Assumption 1: Value activities Despite the development of more sophisticated value creation networks in recent years, we base our model on the nine value activities of Porter (1985). This enables us to measure value creation and so keep complexity manageable within a first modeling approach. Thus, we define a *Value Activity* VA_i with $i = 1, \ldots, n$ as an activity that is either directly involved in the value creation of the company or supports it.

In the context of data-driven value creation, Arcondara et al. (2017) also use the value chain by Porter (1985) to illustrate a big-data-enabled value chain. They state that primary activities generate real-time data, which is screened and

Table 1 Reference overview of ontology development

1) Determine the domain and scope of the ontology	The scope and intention of this model have already been declared in the Introduction.
2) Consider reusing existing ontologies	Existing ontologies could not be found.
3) Enumerate all important terms in the ontology	As this paper focuses on the identification of value-adding data types and their respective risk-relevant properties, we did not conduct this step to the required extent when developing a domain ontology but rather focused on the steps 4) to 7). Therefore, our model is based solely on the results of our literature review.
4) Define the classes and the class hierarchy	*Value Activities,* with two subclasses – 'Primary' and 'Support' Activities – and further subclasses within these. The subclasses of Primary Activities are: Inbound Logistics, Operations, Outbound Logistics, Marketing & Sales, and Service. The subclasses of Supporting Activities are: Firm Infrastructure, Human Resource Management, Procurement, and Technical Development.
	Data, with the subclasses (= *data types*) Logistic Data, R&D Data, Production Data, Distribution Data, Customer Data, IT Data, Financial Data, Personnel Data, and Strategic Planning Data.
5) Define the properties (slots) of classes	Six properties have been identified overall. For the superclass *Value Activities*, these include the intrinsic property *Value Contribution* VC_i as well as the two types of inverse inter-class relations "*generates*" and "*uses*". The intrinsic and extrinsic properties *value contribution* vc_{ji}, *criticality* k_{ji}, *partners* p_{ji}, and *server interfaces* s_{ji} were allocated to the superclass *Data*.
6) Define the facets of the slots	All intrinsic and extrinsic properties defined are single cardinality slots, meaning they can only have one value at a time.
7) Create instances	Instances are created in the Section "Model Evaluation".

processed according to the company's strategy so as to facilitate supportive activities. Via this process, companies create knowledge that they utilize to support their daily operations and management. Nonetheless, we are aware that Porter's (1985) value chain can probably not depict all of the idiosyncrasies of value creation in highly digitalized and connected companies nor within digitalized business models. However, we can use the value chain as a starting point for describing the value activities of classic manufacturing companies that reflect the potential shift from a product-centric to a service-centric world. The value chain also helps to bridge the research gap on connecting data types to a company's value activities in order to evaluate the impact that IT security breaches have on overall value creation. It is important to stress, however, that the value chain should not be understood as a strict sequential flow. Taking trends such as vertical and horizontal integration into account, the value activities should instead be seen as individual modules which can be used to map a company's individual value creation processes.

Moreover, companies can use other value activities for IoT-based networks or supply chains, where the focus on value creation extends beyond the boundaries of a manufacturer and includes other actors. For example, Liu and Jia (2010) define nine value activities for eight actors in an IoT-based drug supply chain, whereby the actor "drug manufacturer" has one generic value activity: producing and selling drugs. Papert and Pflaum (2017) define 19 roles and their value contribution, whereby the role "product manufacturer" also has only one generic value contribution: providing physical objects for smart products. When modeling more generic value activities, companies must be aware that they may lose information at such aggregation levels. Furthermore, while Porter (1985) understood *Services* to be of a physical nature, e.g. maintenance or on-site installation, in the context of this paper we understand *Services* to also include data-driven services.

Assumption 2: Data types In order to link data with the value creation of a company, we need to identify the data types that contribute to value creation. Within the scope of this work, 'data' is defined as a set of qualitative and quantitative variables which exist in different forms and carry specific information that can be collected and analyzed. As addressed in the Literature Review, the prior research lacks concepts which define data types relevant to value creation. According to Peffers et al. (2007), this means we must follow a combined research approach.

In a first step, we draw on the insights gained from a large industry consulting project, which provide us with an initial indication of potential data types. The aim of the project was to develop a framework for the financial evaluation of IT security risks and a well-founded, future-oriented portfolio of mitigation measure, and included, among other things, risk assessments of different types of data. In a second step, we evaluate the selected data types using the literature on value activities along the value chain, e.g., the literature on data-driven inbound logistics or manufacturing. Finally, we identify the following nine key data types by grouping the data types derived from both practice and research: *Logistic Data, R&D Data, Production Data, Distribution Data, Customer Data, IT Data, Financial Data, Personnel Data,* and *Strategic Planning Data*.

To the best of our knowledge, consistent definitions of these data types have not been established in the existing literature. Therefore, within the scope of this paper, these nine data types are defined as summarized in Table 2. Thus, we define a *Data Type* D_j with $j = 1, \ldots, m$ as a key data type that contributes to the value creation of a company. We are aware that these data types are generic and contain multiple sub-categories which may differ in their characteristics. Therefore, further individual specifications are needed for each company, which will also prevent the overlapping of data types within sub-categories. However, we abstain from a more detailed mapping in our first modeling approach in order to limit complexity.

Assumption 3: Allocation of Data Types to Value Activities Each *Value Activity* VA_i "uses and creates information, such as buyer data (order entry), performance parameters (testing), and product failure statistics" (Porter 1985, p.38). As *Data* has been defined as variables in various forms carrying information, the two inter-class relations "*generates*" $rg_{ij} \in \{0, 1\}$ and "*uses*" $ru_{ij} \in \{0, 1\}$ shall be assigned to each individual VA_i. Variable rg_{ij} is a binary integer describing whether the *Data* D_j is created by *Value Activity* VA_i, thereby taking the value 1 if *Value Activity* VA_i creates *Data* D_j, and 0 otherwise. An example could be *Customer Data,* which is generated during the *Value Activity Marketing and Sales*, e.g., within the scope of market research, as well as during *Customer Services,* e.g., in the form of user data. Variable ru_{ij} is also a binary integer, describing whether the *Data* D_j is used by *Value Activity* VA_i, thereby taking the value 1 if *Value Activity* VA_i uses *Data* D_j, and 0 otherwise. An example could be *Strategic Planning Data,* which is used during the *Value Activity HR* in order to develop capabilities necessary in the future. *Both* are inversely related as they depend on the value of another slot (Noy and McGuinness 2001).

The distinction between "*generate*" and "*use*" has to be considered in the subsequent risk analysis (phase 2) as, according to the CIA principle, a data type which is not available after an IT security breach would primarily affect activities that "*use*" this data. A confidentiality incident, however, would affect both using and generating activities as hackers can access data in both cases. Regarding the intended use of this model, only combinations of $i = 1, \ldots, n$ and $j = 1, \ldots, m$ are considered, if their $ru_{ij} = 1$, ergo the *Data* D_j is *used* in

Table 2 Overview of defined data types

Name	Definition	Sources
Logistic Data	Data associated with receiving, storing, and issuing production-relevant inputs (e.g., supplier information, delivery status information, inventory information, route planning, vehicle fleet information, and so forth).	Desrochers et al. 1992
R&D Data	Data involved in and generated during an organization's efforts to either optimize a product and/or process or get an innovation ready for the market.	Griliches 2007
Production Data	Data associated with or generated during the transformation of inputs into the final product (e.g., process information, information about the product, the machine park, equipment maintenance, and quality testing).	Lee et al. 2014
Distribution Data	Data involved in and generated during the collection, storage, and distribution of final goods to customers (e.g., warehousing and inventory information of finished goods, retailer information, distribution channel characteristics, delivery route planning, vehicle fleet information, and order processing information).	Gaynor et al. 2004
Customer Data	Data related to or associated with the final customer and end-user of the product (e.g., personally identifiable information; information generated by sources such as customer service requests, mobile applications, social media networks, purchasing preferences and history; online browsing data).	Linoff and Berry 2011
IT Data	Data related to the technical infrastructure of a company, comprising of hardware, software, and networks; IT development and any kind of coding generated or used within a company's operations.	Jeffery and Leliveld 2004
Financial Data	Data related to financial transactions, financial property, and financial analysis (e.g., payment information; accounting details such as balance sheets, profit and loss statements, cash flow analysis, and stock information).	Merton 1976
Personnel Data	Data associated with activities related to or involved in the recruitment, hiring, training, development, compensation, and dismissal of staff (e.g., training material, professional development strategies, and compensation schemes).	Harter et al. 2002
Strategic Planning Data	Data related to or generated during a company's process of determining the company vision or identifying associated goals and objectives (e.g., a company's expansion and investment plans, vision statements, and business plan; actual state analysis, market and trend analysis).	Schwenk 1995

Value Activity VA_i. Figure 1 illustrates a template assisting the proposed value chain analysis.

Phase 2: Risk analysis

In this phase, we provide the key figures for quantifying IT security risks for the value activities and data types derived in phase 1. Therefore, we first define four properties of *Data* which focus on attributes relevant from an IT security risk perspective, in order to consider the criticality of individual data. In order to focus on the most salient properties, extensive simplification is needed. Therefore, the following three questions should be answered:

1. What is the Data's value and its contribution to the company's success?
2. How critical is the Data? What would be the consequences if the Data was leaked, compromised, or made (temporarily) unavailable?
3. What does the company's risk landscape look like and how many potential points of attack exist?

These questions are in line with the recommendations of the ISO/IEC 27002: 2005, which state that companies should classify their information by sensitivity, criticality, and its significance for the company's value contribution. Based on these questions, we then determine four properties – of which two cover the potential points of attack – to be incorporated into developing an indicator for measuring the risks for *Data*.

Assumption 3.1: Value contribution The concept of Value Activities was developed by Porter (1985) in order to systematically examine and analyze the activities a company performs in its attempts to gain competitive advantage. This implies the need to identify important activities and their contribution to the overall value creation. On this basis, we define *Value Contribution* $VC_i \in \mathbb{R}_0^+$ as the value contributed by the *Value Activity* VA_i to the total value created throughout a company's operation. It is important to stress that we neglect the value contributed by physical activities and solely focus on the value added by the use of data in activities. Thereby, value contribution VC_i is a cardinal value expressed in monetary units with

$$\sum_{i=1}^{n} \boldsymbol{VC_i} = \boldsymbol{Total\ value}\ (\boldsymbol{TV}). \quad (1)$$

Analogously, each *Data* D_j makes a *value contribution* vc_{ji} within the *Value Activity* VA_i. The value contribution of the *Value Activity* shall further be the sum of the value contributed

	Primary Activities										Supporting Activities							
	Inbound Logistics		Operations		Outbound Logistics		Marketing and Sales		Customer Services		Firm Infrastructure		Procurement		HR		Technical Development	
Data Type	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use
Logistic Data																		
R&D Data																		
Production Data																		
Distribution Data																		
Customer Data																		
IT Data																		
Personnel Data																		
Financial Data																		
Strategic Data																		

Fig. 1 Inter-class relations overview

by the individual data types used in this activity. Thereby, *value contribution* $vc_{ji} \in \mathbb{R}_0^+$ is a cardinal value expressed in monetary units with

$$\sum_{j=1}^{m} \boldsymbol{vc_{ji}} = \boldsymbol{VC_i}, \text{for all } i = 1, 2, \ldots, n. \tag{2}$$

This only holds up on the assumption that the overall value contribution solely reflects the data-driven added value and neglects the value added by physical components.

Assumption 3.2: Criticality The second property is *criticality* $k_{ji} \in [0, 1]$, measuring the criticality of *Data* D_j in *Value Activity* VA_i. The most common approach when characterizing the criticality of critical infrastructures is "to assess the impact level in the presence of security-related threats" (Theoharidou et al. 2009, p. 36). As discussed above, the CIA triad is a popular conceptual model for information security threats. Therefore, in order to determine the criticality value, each data type should be analyzed with regard to this concept. Thereby, a higher impact from a security threat leads to a higher criticality value. Hence, three parameters should be considered when allotting the criticality value: the impact of a *confidentiality breach* c_{ji}, an *integrity breach* i_{ji}, and an *availability breach* a_{ji} of *Data* D_j in *Value Activity* VA_i. According to the CIA principle, all three indicators hold cardinal values between 0 (minimal risk) and 1 (maximal risk). Under the assumption of equal weighting among these three parameters, the final value for *criticality* is per definition within this paper the maximum of the three parameters as seen in (3) resulting in

$$\boldsymbol{k_{ji}} = \boldsymbol{max}\left\{ \boldsymbol{c_{ji}}, \boldsymbol{i_{ji}}, \boldsymbol{a_{ji}} \right\}. \tag{3}$$

Assumption 3.3: Potential points of attack In relation to the potential points of attack, factors both internal and external to the company are considered due to the constantly increasing use of cloud services and both horizontal and vertical integration in times of digitalization. Furthermore, the use of malware to attack both software and hardware has increased over the past years (BSI 2016, p. 18–21). To address this trend, within this paper the properties *partners* $p_{ji} \in \mathbb{N}_0 \{ p_{ji} | \, p_{ji} \in \mathbb{N}, p_{ji} \geq 0 \}$ and internal *server interfaces* $s_{ji} \in \mathbb{N}_0 \{ s_{ji} | \, s_{ji} \in \mathbb{N}, s_{ji} \geq 0 \}$ are defined as potential points of attack. As sharing information with collaboration partners leads to a simultaneous expansion of a company's potential attack surface, the property *partners* p_{ji} describes the number of partners the *Data* D_j is shared with within *Value Activity* VA_i, representing the company-external view of attack points. *Server interfaces* s_{ji} follows the same logic from a company's internal view. By storing the same information on multiple server interfaces, a company distributes its IT security risk for each data type, as a security breach on one server might not result in a complete loss or unauthorized modification of data. On the other hand, this strategy increases the number of attack points as the data is then accessible from not only one but multiple servers in the case of a successful security breach.

We are aware that defining four data properties relevant to risk analysis cannot cover the wide range of possible properties. However, the selected properties can help to characterize the risk contribution of the data without losing validity in real-life application. Thus, the properties for every data type in every value activity can be expressed in vectors like

$$D_{ji} = \begin{pmatrix} vc_{ji} \\ k_{ji} \\ p_{ji} \\ s_{ji} \end{pmatrix} \tag{4}$$

Assumption 4: Probability weighted risk Indicator Based on the identified value activities, data types, and their respective IT security-relevant properties, in the next step, we provide a key figure for measuring the IT security risks of data types.

For risk measurement, expected loss (EL) is a common key figure (Sonnenreich et al. 2006). It is usually defined as *probability of default* (PD) times *impact of default* (I), in other words

$$\boldsymbol{EL} = \boldsymbol{PD} \times \boldsymbol{I}. \tag{5}$$

However, until now, determining expected loss resulting from data security breaches has been extremely challenging due to the multitude of parameters which must be considered, as we discussed in the literature review. Therefore, we drew on the idea of the EL to develop an 'impact indicator' which, based on the value contribution and criticality of a particular data type, can be used to estimate potential damage. This also allows the user to make different data types comparable in order to allocate adequate IT security measures. For this comparison, we adjust the EL and introduce a *Probability Weighted Risk Indicator (PWRI)* in order to perform the risk analysis.

Firstly, we calculate an *impact indicator* X_{ji} measuring the impact of a successful security breach regarding *Data* D_{ji} ($rg_{ij} = 1$ or $ru_{ij} = 1$).

$$\boldsymbol{X_{ji}} = \left(\boldsymbol{\alpha} \times \frac{\boldsymbol{k_{ji}}}{\sum_{j=1}^{m} \boldsymbol{k_{ji}}} + (\mathbf{1}-\boldsymbol{\alpha}) \times \frac{\boldsymbol{vc_{ji}}}{\sum_{j=1}^{m} \boldsymbol{vc_{ji}}} \right) \times \boldsymbol{VC_i} \qquad (6)$$

with α being an internal weighting factor determined by the company signifying the importance of either value contribution or criticality. Thus, with $\alpha < 0.5$ ($\alpha > 0.5$) the value contribution of the data type is more (less) important than the data type's IT security-related criticality. This formula implies that the risk indicator of a security breach regarding the *Data* D_j used in *Value Activity* VA_i is the product of the overall *Value Contribution* VC_i of that *Value Activity* VA_i and a weighted average of the data's value contribution and criticality regarding that specific activity. Thus, we expand the concept of EL by introducing parameters reflecting both the value and criticality of the respective data type. In the model developed here, value contribution and criticality are modeled to be independent. We are aware that, in practice, this might not always be the case, as a higher value contribution may correlate with the data type's criticality. As this does not always necessarily hold (e.g., IT Data might have a low value contribution, but a high criticality due to its widespread use to support the value creation process), we abstain from modeling a correlation.

The next step in calculating the *PWRI* is to define the *threat probability* (*TP*) of the *impact indicator* X_{ji}. As elaborated in the model development, the attack surface of a data type is addressed by the internal and external points of attack, the quantity of which are represented by p_{ji} and s_{ji}. Further, within the scope of this paper, π_p is defined as the probability of one of the external points of attack being successfully compromised per year, and π_s is defined as the probability of one of the internal points of attack being successfully compromised per year. Using the best-practice security approach of IT segmentation (Binz et al. 2012), differentiated values for π_s and π_p might be assigned depending on the respective security level of the interface used, but we have chosen not to do so in this study for reasons of simplification. Finally, the probability that one or more external points of attack are being compromised can be expressed as the counter probability that no external point of attack has been successfully breached:

$$\boldsymbol{TP_{p_{ji}}} = \mathbf{1}-\left(\mathbf{1}-\boldsymbol{\pi_p}\right)^{\boldsymbol{p_{ji}}}. \qquad (7)$$

The probability that one or more internal points of attack are being compromised within a given time period $\boldsymbol{TP_{s_{ji}}}$ can be determined analogously. Therefore, the overall threat probability TP_{ij} can be defined as

$$\boldsymbol{TP_{ij}} = \mathbf{1}-\left(\mathbf{1}-\boldsymbol{\pi_p}\right)^{\boldsymbol{p_{ji}}} + \mathbf{1}-\left(\mathbf{1}-\boldsymbol{\pi_s}\right)^{\boldsymbol{s_{ji}}}. \qquad (8)$$

Now that the impact indicator X_{ji} and the threat probability TP_{ij} of that impact have been defined, we can follow the mathematical logic of the EL calculation and determine the $PWRI_{ji}$ as follows:

$$\boldsymbol{PWRI_{ji}} = \boldsymbol{X_{ji}} \times \left(\mathbf{1}-\left(\mathbf{1}-\boldsymbol{\pi_p}\right)^{\boldsymbol{p_{ji}}}\right) + \boldsymbol{X_{ji}} \times \left(\mathbf{1}-\left(\mathbf{1}-\boldsymbol{\pi_s}\right)^{\boldsymbol{s_{ji}}}\right). \qquad (9)$$

The $PWRI_{ji}$ gives an indication of the expected potential loss in the case of a successful security breach of data type *Dj* being used in the *Value Activity* VA_i.

In order to compare the different data types and determine adequate security measures, the overall *PWRI* of *Data* D_j must be considered. To maintain the simplicity of the model, the overall *PWRI* of *Data* D_j can be calculated as follows:

$$\boldsymbol{PWRI_j} = \sum_{i=1}^{n} \boldsymbol{PWRI_{ji}}, \text{ for } j = 1, 2, \ldots, \text{m}. \qquad (10)$$

Companies can use these calculations from the model in order to compare different data types within their company and so determine the crown jewels and their exposure and can then allocate adequate measures to reduce risk.

Furthermore, this model and its implications can be used to forecast IT security risks that reflect changes in the IT security landscape, e.g., due to the emergence of new digital business models. To do so, the time component $t \in \mathbb{N}_0$ must be introduced. A company must identify its current situation $t = 0$, fill the values of the property slots accordingly, and calculate the *PWRI*.

$$\boldsymbol{PWRI_{j_{t=0}}} = \sum_{i=1}^{n} \boldsymbol{PWRI_{ji_{t=o}}}, \text{ for } j = 1, 2, \ldots, \text{m}. \qquad (11)$$

In the next step, the company must determine a prospective business model that reflects its strategic vision, repeat the value chain analysis and calculate the new *PWRI* for $t + 1$

$$\boldsymbol{PWRI_{j_{t+1}}} = \sum_{i=1}^{n} \boldsymbol{PWRI_{ji_{t+1}}}, \text{ for } j = 1, 2, \ldots, \text{m}. \qquad (12)$$

By comparing $PWRI_{j_{t=0}}$ and $PWRI_{j_{t+1}}$, adequate measures can be deduced, depending on the delta $\Delta PWRI_j$

$$\boldsymbol{\Delta PWRI_j} = \boldsymbol{PWRI_{j_{t+1}}} - \boldsymbol{PWRI_{j_{t=0}}}, \text{for } j = 1, 2, \ldots, \text{m}. \qquad (13)$$

A positive $\Delta PWRI_j$ indicates an increase in the damage expected to result from a security breach regarding the *Data* D_j, and hence indicates the increasing importance of effectively protecting this data type. Analogously, a negative $\Delta PWRI_j$ indicates decrease in the damage predicted in the case of a security breach regarding the *Data* D_j. The early identification of future needs will enable the company to invest in the necessary security measures and ensure tailored data protection. Of course, this is a simplified approach to determining the impact of IT security breaches and thus should merely be used as an indicator. However, our approach is an important first step towards quantifying data and its contribution to both value creation and risk.

Evaluation

Model application

In order to test the practical intelligibility and applicability of our model, we evaluated it in interviews with experts from two manufacturing companies. To gain different perspectives, we selected two companies that differ in their organizational setup, industry and their level of digitalization maturity. We conducted qualitative, semi-structured group interviews (Myers and Newman 2007) with experts involved in the business IT solutions of each company. These experts have a deep understanding of their company's value creation processes and the associated activities, and, thus, have a well-balanced expertise of IT and business know-how.

The first company (C1) is a corporation that operates internationally with approximately 15,000 employees around the world and annual sales of around 2 billion Euros. The company produces specialty glass and glass-ceramics for a variety of industries and considers itself an innovative, international leading technology group with a sole focus on B2B-interactions. We interviewed C1's director of business services and solutions (experience >10 years), the head of process technology (experience >20 years), and an IT Infrastructure & Security manager (experience >5 years). Interviewing executives from both the operational, value creation perspective and the IT perspective ensured credible results. To date, C1 has relied on integrated IT solutions to monitor and control production processes, yet the product itself does not feature further applications that make it a 'smart product', nor is it likely to feature such applications in the future. Instead, C1 is increasingly searching for additional information-intensive services to complement their products and generate additional value for the customer.

The second company (C2) is a multinational corporation with approximately 27,000 employees worldwide and annual sales of approximately 4.4 billion Euros. The company develops and manufactures products, systems, software, and services for the construction and energy industries, catering mainly to professional end-users (B2C). We interviewed C2's head of security and risk management IT (experience >10 years), head of IT enterprise risk management (experience >20 years), and an IT Infrastructure & Security manager (experience >10 years).

After a short presentation on the developed model, the underlying assumptions and its intended use, the interviews were structured around the two phases of the model development, consisting of an interactive value chain analysis followed by a risk analysis. Moderated by us, each workshop took place at the experts' site and lasted approximately 3 h. For simplification and to aid facilitation, we defined three categories (low (1), medium (2), and high (3)) representing the underlying parameters of the risk analysis. This provided us with a uniform scale for facilitating communication and parameterization in the workshops. The distinctive characteristics of these categories can be seen in Table 3. For further, more in-depth analysis, companies can use precise values in place of the scale provided in Table 3.

Results phase 1: Value chain analysis

The objective of the value chain analysis is to first identify the main value activities involved in a company's value creation and allocate the data types associated with these activities accordingly. As each activity uses a multitude of data, we focus on the most salient and important data types, limiting the allocation to a maximum of three data types per relationship, where possible. For C1, the experts identified the most heavily-used data types to be Logistic Data, Production Data, Financial Data, and Customer Data, each being among the most salient data types used within the four value activities. This is in line with C1's strong focus on manufacturing. As production and input supply are closely linked and mutually dependent, production planning must be coordinated with the availability of necessary inputs for optimal operational activities. Furthermore, Customer Data is required during production for customer-specific features and the customer-specific issue of a quality certificate. Figure 2 depicts the full data allocation of phase 1 conducted by the experts at C1, complemented by the value contribution of the value activities identified in phase 2.

In comparison, due to C2's orientation towards B2C-interactions, the experts identified Production Data, Distribution Data, and Financial Data as the data types most heavily-used to better cater to individual customer needs. During the value chain analysis, a need for more differentiated data types in order to achieve more sound results was expressed by experts from both companies. However, both companies' experts validated the real-world fidelity of the identified data types and value activities, agreeing that the model covers all relevant

Table 3 Distinctive characteristics of underlying parameterization categories

	Low (1)	Medium (2)	High (3)
Value contribution	Mere support process	Standard process (generates value, no core competency)	Core competency, main value driver
Criticality			
Confidentiality	Data partially publicly accessible	For internal use only, widely accessible to all employees	Strictly confidential internal data
Integrity	Manipulated data is identified and output revised quickly	Manipulated data is identified or output revised quickly	Manipulated data cannot be identified quickly and output cannot be revised
Availability	Irregular data accessing	Near-time data usage	Real-time data usage
Point of attacks			
External partners	Only internal data storage	Access to a few selected business partners	Shared access with a multitude of external partners
Internal server interfaces	Marginal data usage by services/applications. Access via company intranet only	Occasional data usage by services/applications. Access via intranet only	Regular, widespread data usage through diverse services/applications

constellations that typically occur in their companies. They also confirmed that the specifications of the model are intelligible for industry experts.

Results phase 2: Risk analysis

The risk analysis was conducted in a three-step process. Firstly, the industry experts used the categories low (1), medium (2), and high (3) – as displayed in Table 3 – to weight the identified main value activities according to each activity's contribution to the overall value. Based on these weighting factors, the relative contribution of each value activity was calculated to enable a better comparison. For a comprehensive overview of the weighted value contribution per value activity, refer to Fig. 2. For C1, the interviewees identified Operations and Technical Development as the value activities with the highest value contribution, accounting for almost 40% of the company's value creation. This is in line with C1's business model, which focuses on manufacturing and innovation. Secondly, the experts conducted the parametrization of the identified IT security-relevant risk properties, resulting in a risk property vector for each data type per value activity.

In a third step, we used this parametrization to evaluate each data type by means of the previously introduced *PWRI*. For simplification and facilitation purposes, the underlying parameters are again categorized into low (1), medium (2), and high (3), as displayed in Table 3. Furthermore, α, the weighting factor for the impact of a data type's value contribution and criticality, is pre-set to 0.5, and the probabilities of a successful IT security breach per year for internal and external attack points both pre-set to 5%. For simplification, we decided to only base our evaluation on data types with the label "*use*" within all activities and to exclude the "*generate*" column in Fig. 2. These intrinsic values are determined only as an example, and we are aware that they are otherwise company-specific and need to be adapted for individual use

	Primary Activities										Supporting Activities							
	Inbound Logistics		Operations		Outbound Logistics		Marketing and Sales		Customer Services		Firm Infrastructure		Procurement		HR		Technical Development	
Data Type	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use	generate	use
Logistic Data	x	x		x								x	x	x				
R&D Data								x				x					x	x
Production Data		x	x	x		x			x	x								
Distribution Data			x		x	x			x	x								
Customer Data				x		x	x	x	x	x								
IT Data											x	x					x	x
Personnel Data															x	x		
Financial Data	x	x	x		x			x			x		x	x	x	x		
Strategic Data											x	x		x		x		x
Value Contribution	2		3		2		1		1		2		1		1		3	
Weighted Value Contribution VCi	0.1250		0.1875		0.1250		0.0625		0.0625		0.1250		0.0625		0.0625		0.1875	

Fig. 2 Value chain analysis results at C1

Table 4 Underlying model parameters

	Low (1)	Medium (2)	High (3)
Value contribution	0.1	0.5	1
Criticality	0.1	0.5	1
External partners	0	1	5
Internal server interfaces	1	2	10

cases. For the values applied within the analysis, please refer to Table 4.

Applying these pre-set categories to the risk property vectors of each data type per value activity resulted in 27 vectors at C1 and 26 vectors at C2. To take potential future changes in their respective business model into account, we also canvassed the industry expert's expectations on future developments, resulting in an additional 27 (26) vectors at C1 (C2). In order to quantify and compare data types, we inserted the information collected from the experts at C1 and C2 into our model and calculated the impact of a successful security breach as a function of the data type's value contribution and criticality, as well as the threat probability for each data type dependent on the data type's dispersion. We then inserted the information of value contribution and criticality into formula (6) to calculate the impact indicator X_{ji} and the information collected from internal and external partners into formula (8) in order to calculate the threat probability TP_{ji}. Plugging these results into formula (10), we determined the *PWRI* per both *Data* D_j and the *Value Activities* VA_i. An exemplary risk property vector, including results of Operations at C1, is presented in Table 5.

Result analysis

In this section, we show how results yielded from our model can be analyzed and interpreted. In a first step, a company can identify its data-related crown jewels via the calculated impact indicator X_{ji}. As the impact consists of both the data type's weighted proportional value contribution and criticality, a higher X_{ji} implies a greater significance in terms of the company's value creation. Using this information, companies can determine measures to secure their data crown jewels. According to our model, for C1, Financial Data, Production Data, and Customer Data have the highest impact indicator, together accounting for 50% of the overall impact indicated. For all three data types, this can be explained by the high criticality associated with these data types. Thereby, confidentiality is the key criticality-driving factor for Financial Data and Customer Data, as these are subject to strict privacy policies and large fines are imposed in cases of unauthorized disclosure. The key criticality-driving factor of Production Data is integrity, as the C1's products are subject to strict quality specifications which determine the stability and safety of the product when in use. In terms of crown jewels, the analysis yielded the same results for C2.

Companies can also rank data types in an integrated manner by comparing the calculated $PWRI_j$. According to our model, at C1, the data types with the highest $PWRI_j$ are IT Data (PWRI = 0.041), Financial Data (PWRI = 0.027), Production Data (PWRI = 0.026), and Logistic Data (PWRI = 0.025), which together make up 70% of the overall *PWRI*, with IT Data alone holding a surprisingly large share of 25% (see Fig. 3, grey bars). Financial Data, Production Data, and Logistic Data yielded very similar results with a delta smaller than 0.2% of the overall share. The high $PWRI_j$ for IT Data is mainly driven by its application within Firm Infrastructure, as IT Data does not only contribute significant value and is highly critical but, more importantly, is widely distributed both internally and externally, resulting in an exceptionally high threat probability. The same holds for Logistic Data, which is widely shared with external partners in both Inbound Logistics and Procurement, resulting in a relatively high overall threat probability for this data type. In contrast, Customer Data, despite having a high impact value, is generally kept in-house and shared with a minimal number of parties, resulting in a significantly lower threat probability and therefore a lower overall *PWRI*.

For C2, however, Customer Data (PWRI = 0.036) yielded by far the highest *PWRI*, being nearly twice as high as the

Table 5 Operations risk property vector at C1 (Only "Use", model input values in brackets)

	Logistic Data		Production Data		Customer Data	
	Current values	Exp. future values	Current values	Exp. future values	Current values	Exp. future values
Value contribution	1 (0.1)	1 (0.1)	3 (1)	3 (1)	3 (1)	3 (1)
Criticality	2 (0.5)	2 (0.5)	3 (1)	3 (1)	2 (0.5)	2 (0.5)
External partners	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	2 (1)
Internal server interfaces	1 (1)	1 (1)	2 (2)	3 (10)	1 (1)	2 (2)
Impact indicator Xji	0.019	0.019	0.061	0.061	0.045	0.045
TPij	0.05	0.05	0.098	0.401	0.05	0.148
PWRIji	0.001	0.001	0.006	0.024	0.002	0.007

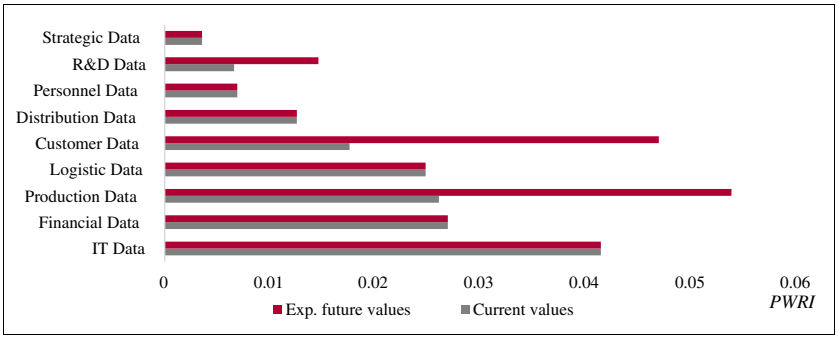

Fig. 3 Shift in threat landscape due to business model changes at C_1

second most critical data type, IT Data (PWRI = 0.018). In addition to a high impact, Customer Data is widely distributed and shared with a large number of internal and external parties, resulting in an exceptionally high threat probability and therefore a high overall *PWRI*. The same applies to IT Data and Distribution Data (PWRI = 0.012), both of which yield a high threat probability due to their widespread distribution among internal and external partners. The widespread distribution of these three data types can be explained by C2's business model, which focuses on B2C-interactions and offers a wide range of IT-enabled services complementing the physical products they offer their customers. For full results of the PWRI ranking refer to Table 6.

Another benefit of our model is that it offers an integrated view of a company's value activities, which the user can access by looking at the activity's cumulative *PWRI*. Intuitively, it might seem that the activities which make the highest contribution to the company's value creation will be at most at risk from IT security-related attacks. However, according to our model, this is not the case at C1. While Firm Infrastructure, Inbound Logistics, and Services represent the top three with regard to their respective PWRI, the experts identified them as contributors of a medium or even low level of value. This can again be traced back to the data distribution in these activities, as they all feature data types widely shared with external and internal parties, offering a wide range of potential attack targets. This insight helps to raise companies' awareness of their potential weak links in their value creation process as viewed from an IT security perspective. At C2, the results did not deviate as strongly from the sole value contribution perspective.

Finally, our model can be used to analyze changes in the IT security risk landscape that may result from predicted future shifts in business models (i.e., stronger integration of information-intensive services, increased integration of smart products and smart solutions, etc.). According to the experts at C1, these shifts are primarily expected to affect manufacturing processes and distribution activities, with changes predicted to include the increased integration of smart solutions for better data analysis, individualized production, and transparency in relation to the customer, resulting in a significant rise in the threat probability for Production Data and Customer Data. It also predicts that R&D Data will be affected by a significant increase in the level of threat, which can be traced back to the increasing need to establish new collaborations in order to enhance IT innovations. This implies an increasing need for integrated security measures which take complex collaboration-based ecosystems into account. Figure 3 (red bars) illustrates the shift for C1. In contrast, having already reached a high digitalization maturity level, C2 does not expect a noticeable change to its business model in the coming years. Therefore, C2 aims to achieve an overall increase in the value contribution made by each data type, rather than a shift in the nature of the value contribution or criticality of these data types.

Table 6 PRWI ranking at C1 and C2

	Data Type	PWRI ranking at C1		Data Type	PWRI ranking at C2
1	IT Data	0.041	1	Customer Data	0.036
2	Financial Data	0.027	2	IT Data	0.018
3	Production Data	0.026	3	Distribution Data	0.012
4	Logistic Data	0.025	4	Financial Data	0.010
5	Customer Data	0.018	5	Logistic Data	0.009
6	Distribution Data	0.013	6	Personnel Data	0.007
7	Personnel Data	0.007	7	Production Data	0.005
8	R&D Data	0.007	8	R&D Data	0.003
9	Strategic Data	0.004	9	Strategic Data	0.001

As our analysis of the results shows, our model supports the assessments of the value contribution and criticality of each data type. Thereby, our aim is not to provide a model for calculating the exact loss that would be incurred in the case of a security breach for each data type, but rather to determine a key figure for comparing and ranking data types in an integrated manner. By distinguishing individual data types and their contribution to a company's value creation, data types become both intra-temporally and inter-temporally comparable, which enables the simulation and assessment of potential future developments. Furthermore, critical value activities and a company's crown jewels can be identified. Finally, companies can use such analysis as a first step in adjusting their IT security investment strategies in response to future changes in the threat landscape.

Model evaluation

We agree with Nickerson et al. (2013) that a useful model must be concise, robust, comprehensive, extendible, and explanatory. Within the model evaluation, academic focus-group members (one distinguished and one associate professor, and six research assistants, excluding the authors) and the interviewed industry experts confirmed the utility of the model in terms of the named conditions.

We can confirm that our model is concise (i.e., it involves a limited number of terms and characteristics for reasons of simplicity) (Nickerson et al. 2013) as it contains nine key value creation activities and nine key data types, which can be linked through two types of inter-class relations in the value creation analysis. Additionally, our model limits the number of key figures for risk analysis to the Probability Weighted Risk Indicator, which draws on four IT security risk characteristics describing data types. While providing a simplified version of reality, the model should help to minimize the cognitive demands placed on decision-makers and help to overcome the difficulties of application (Nickerson et al. 2013), which is what our model does. The experts confirmed that both Porter's approach and the data types identified represent an adequate initial structuring approach. At the same time, they noted that a fine-granular view of the data types is necessary on a case-by-case basis. Moreover, they pointed out that the broad categories of the identified data types may lead to the potential overlapping of data within subcategories.

Secondly, we can confirm that our model is robust (i.e., it includes enough terms and characteristics to clearly differentiate the objects of interest) and comprehensive (i.e., it is complete in that it includes of all relevant terms and characteristics) (Nickerson et al. 2013), as it enables the analysis and comparison of different data types in terms of their value contributions and risk properties. It also enables the user to consider various use cases through different model parametrization. The interviewed experts confirmed that the model provides transparency about value creation activities, data types and the dependencies between them, and that it makes the data types comparable in terms of their value contributions and risk properties. The experts also acknowledged that the structure of our model covers the essential terms and characteristics and that they do not miss further elements.

Additionally, we can confirm that the model is extendible (i.e., there are no restrictions on its future extension) (Nickerson et al. 2013), as it allows the inclusion of further value creation activities, data types, and/or IT security risk characteristics. Finally, our model is explanatory (i.e., it enables the instantiation of real-world use cases) (Nickerson et al. 2013), as illustrated in the subsection Model Application.

Conclusion

Digitalization forces companies to reevaluate their business models and shift to data-driven alternatives. For manufacturing companies, in particular, this leads to extensive changes in value creation processes and in IT security risk landscapes. If they are to successfully protect their newly emerged, data-based crown jewels, companies must be able to identify the data that contribute to their value creation, both at present and in the future, and assess the associated IT security risks. Despite the extensive body of research on IT security and data-driven value creation, approaches that link these disciplines and measure both the value contribution and associated IT security risks of data in an integrated manner are still missing.

Our approach aims to contribute to the closure of this research gap and support companies in their analysis of the IT security risks in their data-driven value chains. With this goal in mind, we provide a two-step approach. The first step, comprising a value chain analysis, enables the user to identify strategically important value activities and data types generated or used for these activities today and in the future. In the second step, comprised of an integrated risk analysis, we determine the Probability Weighted Risk Indicator; a key figure for assessing the value contribution and associated IT security risks of different data types. Among other things, our model provides companies with guidance on the identification and exposure of their crown jewels and makes different data types comparable. We invited research and industry experts to evaluate our model in order to confirm its real-world fidelity, applicability, and usability.

Our approach contributes to both research and practice. From an academic perspective, we lay important groundwork at the interface of IT security and data-driven value creation by combining these two research streams to develop an integrated modeling approach for IT security risk analysis. In particular, our study complements existing IT security frameworks such as the NIST cybersecurity framework and standards such as

ISO/IEC 27000 series or, on a national level, the German IT-Grundschutz. These frameworks, however, only address the need for a quantitative risk assessment of the involved assets – in this case, the respective data types – and provide no concrete guidance for implementation. Our work, on the other hand, provides an approach for analyzing different data types from both a value creation and security risk perspective and enables a quantitative assessment of the underlying data type.

In practice, our approach can be used to help companies identify valuable assets in order to allocate scarce IT security budgets in an economically sounds manner. In particular, it can be used for various analyses, e.g., to analyze the current state of value creation by identifying the most important value activities and data types – that is, the crown jewels of the company. Practitioners can also use the model to carry out an integrated assessment of potential strategic business model developments and the associated shifts in value creation. The application of our model is thus a first step toward identifying potential IT security risks associated with these shifts, in that it enables the analysis of different data types in current and future value creation. This allows the user to identify the most critical data types and initiate discussions on mitigation measures. The model has the same advantages for project owners, who can use our approach to evaluate new project solutions involving innovation or digitalization in order to illustrate the impact of their project solutions on the current value contribution and associated IT security risks of the data involved. Practitioners could further adjust our approach to assess a company's idiosyncrasies, such as stronger IT security guidelines, by changing or expanding the model parameters.

Despite providing novel insights into IT security risk analysis of data-driven value chains, our approach has some limitations which could be used as a starting point for further investigations. As we could not find any approaches that analyze both the value contribution and IT security properties of data types, we could not compare our results with the results of other models. And, although we evaluated our model with experts from two companies to illustrate its applicability in practice, further empirical evaluation of the model in a given organizational context might help to strengthen our findings (Meredith et al. 1989; Wacker 1998). We chose to draw on Porter's (1985) approach to value creation as a first step in the development of value activities. Researchers might use this as a starting point when further developing the approach presented here, or as a means to evaluate alternatives and identify value activities within more complex interdependent value creation networks.

We also derived key data types used in the value creation process from literature and evaluated them with industry experts. Future research might focus on other methods for identifying key data types, such as surveys or empirical investigations, in order to ensure the generalizability of key data types. As the data types identified here fall into broad categories, the potential overlapping of data within subcategories is not addressed, nor are the different degrees of criticality of that data. One additional research path could, therefore, focus on splitting the identified key data types into more detailed sub-data types in order to enable a more fine-grained analysis. This would allow practitioners to consider mitigation measures in more detail, and to identify, for example, which specific systems they should restrict access to, and which employees may need a higher degree of vetting or more advanced training.

Moreover, for the risk analysis, we define a risk indicator based on the concept of excepted loss and consider four key properties of *Data* used to calculate the risk indicator. Investigations focusing on other appropriate risk parameters and measurement approaches would provide further valuable insights. For simplification and facilitation purposes, in the risk analysis with practitioners we classified the underlying parameters as low (1), medium (2), or high (3). However, assessments based on such simplified categories could be considered subjective, and additional research could help to better estimate these model parameters.

Another point for researchers to consider is that our model has, so far, been used as a one-period model only, hence all decisions and outcomes have occurred simultaneously. Thus, dynamic aspects, such as spillover effects from a successful breach in one value activity to another have not been considered and might yet be incorporated into the model. Furthermore, our model does not consider interdependencies and spread effects within the value chain and risk analysis. Further investigations on how these aspects can be incorporated into the approach could be helpful. Despite these limitations, our approach serves as an important first step toward IT risk analysis in data-driven value chains, and as a starting point for further investigations in this area.

References

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: A systematic review and agenda for future research. *Electronic Markets, 26*(2), 173–194.

Anderson, J. P. (1972). Information security in a multi-user computer environment. *Advances in Computers, 12*, 1–36.

Arcondara, J., Himmi, K., Guan, P., & Zhou, W. (2017). Value oriented big data strategy: Analysis & case study. *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Binz, T., Leymann, F., Nowak, A., & Schumm, D. (2012). Improving the manageability of enterprise topologies through segmentation, graph transformation, and analysis strategies. In enterprise distributed object computing conference, 2012 IEEE 16th international (pp. 61-70). IEEE.

Bitran, G. R., Gurumurthi, S., & Sam, S. L. (2007). The need for third-party coordination in supply chain governance. *MIT Sloan Management Review, 48*(3), 30.

Bundesamt für Sicherheit in der Informationstechnik (BSI) (2016). The State of IT Security in Germany 2016.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004a). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce, 9*(1), 70–104.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004b). A model for evaluating IT security investments. *Communications of the ACM, 47*(7), 87–92.

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems, 32*(4), 4–39.

D'Amico, A. D. (2000). What does a computer security breach really cost. Secure Decisions, Applied Visions Inc.

Desrochers, M., Desrosiers, J., & Solomon, M. (1992). A new optimization algorithm for the vehicle routing problem with time windows. *Operations Research, 40*(2), 342–354.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology, 66*(8), 1523–1545.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review, 10*(1), 1–21.

Friedman, S. D., & Singh, H. (1989). CEO succession and stockholder reaction: The influence of organizational context and event content. *Academy of Management Journal, 32*(4), 718–744.

Garg, A., Curtis, J., & Halper, H. (2003). Quantifying the financial impact of IT security breaches. *Information Management & Computer Security, 11*(2), 74–83.

Gaynor, M., Moulton, S. L., Welsh, M., LaCombe, E., Rowan, A., & Wynne, J. (2004). Integrating wireless sensor networks with the grid. *IEEE Internet Computing, 8*(4), 32–39.

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC), 5*(4), 438–457.

Griliches, Z. (Ed.). (2007). R&D, patents and productivity. University of Chicago Press.

Grobauer, B., Walloschek, T., & Stocker, E. (2011). Understanding cloud computing vulnerabilities. *IEEE Security & Privacy, 9*(2), 50–57.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191–209.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis.

ISO 27002. (2005). *Information technology, security techniques, code of practice for information security management*. Geneve: International Organization for Standardization ISO.

Jeffery, M., & Leliveld, I. (2004). Best practices in IT portfolio management. *MIT Sloan Management Review, 45*(3), 41.

Keeney, R., and Raiffa, H. 1993. *Decisions with multiple objectives*, Cambridge [England]: Cambridge University press.

Koh, J., & Venkatraman, N. (1991). Joint venture formations and stock market reactions: An assessment in the information technology sector. *Academy of Management Journal, 34*(4), 869–892.

Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia Cirp, 16*, 3–8.

Lim, C., Kim, K. H., Kim, M. J., Heo, J. Y., Kim, K. J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management, 39*, 121–135.

Lim, C., Kim, M. J., Kim, K. H., Kim, K. J., & Maglio, P. P. (2017). Using data to advance service: Managerial issues and theoretical implications from action research. Journal of Service Theory and Practice.

Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: For marketing, sales, and customer relationship management. John Wiley & Sons.

Liu, L., & Jia, W. (2010). Business model for drug supply chain based on the internet of things. In 2010 2nd IEEE international conference on network infrastructure and digital content (pp. 982-986). IEEE.

Meredith, J. R., Raturi, A., Amoako-Gyampah, K., & Kaplan, B. (1989). Alternative research paradigms in operations. *Journal of Operations Management, 8*(4), 297–326.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics, 3*(1–2), 125–144.

Müller, S. C., Böhm, M., Schröer, M., Bahkirev, A., Baiasu, B. C., Krcmar, H., & Welpe, I. M. (2016). Geschäftsmodelle in der digitalen Wirtschaft. Vollstudie (No. 13-2016). Studien zum deutschen Innovationssystem.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization, 17*(1), 2–26.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336–359.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Papert, M., & Pflaum, A. (2017). Development of an ecosystem model for the realization of internet of things (IoT) services in supply chain management. *Electronic Markets, 27*(2), 175–189.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77.

Porter, M. E. (1985). Competitive advantage: Creating and sustaining superior performance. Simon and Schuster.

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review, 92*(11), 64–88.

Purdy, G. (2010). ISO 31000: 2009—Setting a new standard for risk management. *Risk Analysis, 30*(6), 881–886.

Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE, 63*(9), 1278–1308.

Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS Quarterly*, 237–263.

Schwenk, C. R. (1995). Strategic decision making. *Journal of Management, 21*(3), 471–493.

Shoukry, A., Khader, J., & Gani, S. (2019). Improving business process and functionality using IoT based E3-value business model. *Electronic Markets*, 1-10.

Sonnenreich, W., Albanese, J., & Stout, B. (2006). Return on security investment (ROSI)-a practical quantitative model. *Journal of Research and Practice in Information Technology, 38*(1), 45.

Straub, D. W., & Watson, R. T. (2001). Research commentary: Transformational issues in researching IS and net-enabled organizations. *Information Systems Research, 12*(4), 337–345.

Swanson, L. (2001). Linking maintenance strategies to performance. *International Journal of Production Economics, 70*(3), 237–244.

Theoharidou, M., Kotzanikolaou, P., & Gritzalis, D. (2009). Risk-based criticality analysis. Critical infrastructure protection III, 35-49.

Ur Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management, 36*(6), 917–928.

Wacker, J. G. (1998). A definition of theory: Research guidelines for different theory-building research methods in operations management. *Journal of Operations Management, 16*(4), 361–385.

Wheeler, B. C. (2002). NEBIC: A dynamic capabilities theory for assessing net-enablement. *Information Systems Research, 13*(2), 125–146.

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive, 12*(2).

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management, 36*(6), 1231–1247.

Yoo, S., Kim, S., Lee, K. H., Jeong, C. W., Youn, S. W., Park, K. U., & Hwang, H. (2014). Electronically implemented clinical indicators based on a data warehouse in a tertiary hospital: Its clinical benefit and effectiveness. *International Journal of Medical Informatics, 83*(7), 507–516.