Fruhwirth, Michael; Pammer-Schindler, Viktoria; Thalmann, Stefan

Article

Knowledge leaks in data-driven business models? Exploring different types of knowledge risks and protection measures

Schmalenbach Journal of Business Research (SBUR)

Provided in Cooperation with:
Schmalenbach-Gesellschaft für Betriebswirtschaft e.V.

Suggested Citation: Fruhwirth, Michael; Pammer-Schindler, Viktoria; Thalmann, Stefan (2024) : Knowledge leaks in data-driven business models? Exploring different types of knowledge risks and protection measures, Schmalenbach Journal of Business Research (SBUR), ISSN 2366-6153, Springer, Heidelberg, Vol. 76, Iss. 3, pp. 357-396, https://doi.org/10.1007/s41471-024-00189-z

This Version is available at:
https://hdl.handle.net/10419/312594

Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (2024) 76:357–396
https://doi.org/10.1007/s41471-024-00189-z

ORIGINAL ARTICLE

Knowledge Leaks in Data-Driven Business Models? Exploring Different Types of Knowledge Risks and Protection Measures

Michael Fruhwirth · Viktoria Pammer-Schindler · Stefan Thalmann

Received: 16 February 2022 / Accepted: 17 June 2024 / Published online: 30 July 2024

Abstract Data-driven business models imply the inter-organisational exchange of data or similar value objects. Data science methods enable organisations to discover patterns and eventually knowledge from data. Further, by training machine learning models, knowledge is materialised in those models. Thus, organisations might risk the exposure of competitive knowledge by sharing data-related value objects, such as data, models or predictions. Although knowledge risks have been studied in traditional business models, little research has been conducted in the direction of data-driven business models. In this explorative qualitative study, we conducted 28 expert interviews in three rounds (two exploratory and one evaluatory) and identified five types of risks along the three basic types of value objects: data, models and predictions. These risks depend on the context, i.e., when competitive knowledge could be discovered from shared value objects. We found that those risks can be mitigated by technology, contractual regulations, trusted relationships, and adjusting the business model design. In this study, we show that the risk of knowledge leakage is a relevant risk factor in data-driven business models. Overall, knowledge risks should be

Michael Fruhwirth · Viktoria Pammer-Schindler
Know-Center GmbH, Sandgasse 34/II, 8010 Graz, Austria

Institute for Interactive Systems and Data Science, Faculty of Computer Science and Biomedical Engineering, Graz University of Technology, Sandgasse 36/III, 8010 Graz, Austria
E-Mail: michael.fruhwirth@student.tugraz.at; viktoria.pammer-schindler@tugraz.at

Present Address:
✉ Michael Fruhwirth
Silicon Austria Labs GmbH, Sandgasse 34/IV, 8010 Graz, Austria
E-Mail: michael.fruhwirth@silicon-austria.com

Stefan Thalmann
Business Analytics and Data Science-Center (BANDAS-Center), School of Business, Economics and Social Sciences, University of Graz, Universitätsstraße 15 Building F/III, 8010 Graz, Austria
E-Mail: stefan.thalmann@uni-graz.at

considered already during business model design, and their management requires an interdisciplinary approach via a balanced assessment. The level of knowledge protection from a technology perspective highly depends on computer science innovations and thus is a moving target. As an outlook, we suggest that knowledge risk will become even more relevant with the extensive usage of machine learning and artificial intelligence in data-driven business models.

Keywords Business model innovation · Data analytics · Data-driven business models · Knowledge risks · Risk management · Value objects

1 Introduction

Developments in big data technologies and artificial intelligence (AI), as well as the availability of large data sets, hold the opportunity for the development of new products, services, and business models (Günther et al. 2017; Woerner and Wixom 2015), so-called data-driven business models (DDBMs) (Hartmann et al. 2016; Wiener et al. 2020). Such business models often imply the exchange of data and similar data-related value objects. Further, in such business models, sensitive information and competitive knowledge are materialised in data or models. At the same time, data science methods allow extracting information or knowledge from fine-granular, heterogeneous data, leading to potential risks when data is shared. Whereas before, knowledge needed to be represented in a much more explicit manner. Thus, it is challenging for organisations to evaluate what knowledge could be discovered from shared data sets (Zeiringer and Thalmann 2020). For instance, simply "looking at the data" (i.e., at the headers of a database or descriptive statistics over a single dataset) is not enough to assess which knowledge could be drawn from the data. Sharing data implies the risk—which we refer to as knowledge risks—that competitive knowledge could leak and spill over to other organisations.

For example, we found such risks in a case study with an industrial company (Fruhwirth et al. 2019). In this case, novel knowledge of a real-world physical phenomenon (i.e., predicting the residual lifetime of a physical component) was generated from data and materialised in a model. Building new DDBMs around this model (i.e., offering the model) could imply the risk of leaking core knowledge, as one workshop with managers of this company showed. Further, the willingness to share data is often a prerequisite for a DDBM, but potential knowledge leakages negatively influence this willingness. Thus, DDBMs require balancing between sharing and protecting knowledge. Further, IP might be shared or could be re-engineered when offering machine learning (ML) models through an API (Application Programming Interface) (Hanzlik et al. 2021).

Knowledge risks have been studied in strategic alliances (Hernandez et al. 2015; Jiang et al. 2016; Kale et al. 2000) and traditional business models (Al-Aali and Teece 2013). However, as shown above, DDBMs imply new types of risks, particularly that knowledge may spill over to competitors via sharing data and similar value objects. Although such risks exist, little has been written about how different types of offerings of DDBMs, or exchanged value objects in particular, relate to

knowledge risks. Therefore, we address the following research question in this paper: *What knowledge risks are associated with sharing different types of data-related value objects in data-driven business models, and what are protection measures?*

To answer this research question, we interviewed 28 experts from industry and academia to explore cases of knowledge risks. We structured different types of risks, contextual factors and protection measures based on the three basic types of value objects: data, models and predictions. Based on our findings, we suggest three fields of action to mitigate knowledge risks in DDBMs: using technology, adjusting the business model design and establishing trustful relationships and contractual regulations. Managing knowledge risks in DDBMs requires a balanced view and interdisciplinary approach already during the design of a DDBM.

2 Background

2.1 Data-Driven Business Models

Data-driven business models (DDBMs) have a conceptual focus on value creation from data (Guggenberger et al. 2020). A business model is a conceptual tool that allows a simplified description of how organisations create, deliver and capture value (Osterwalder and Pigneur 2010; Osterwalder et al. 2005; Teece 2010).

Firms with a DDBM utilise data as a key resource for new business (Hartmann et al. 2016). They generate customer value through data analytics and machine learning (Schüritz et al. 2017b). Data analytics and machine learning techniques are used to discover insights from data (Kühne and Böhmann 2019). These insights are delivered as data analytics-based features, products, or services and support customers in their decision-making process (Schüritz et al. 2019) and enable the generation of new revenue streams (Schüritz et al. 2017a). Thus, data intermediation is the central value proposition (Dorfer 2016). Developing a DDBM requires business and technological capabilities (Stahl et al. 2023).

Literature started to analyse and categorise DDBMs from different perspectives. Two common approaches are to differentiate based on the type of data sources used (e.g., internal existing or self-generated data vs externally acquired, customer-provided or free available data; see, e.g., Hartmann et al. 2016) or the type of analytics used (e.g., descriptive, diagnostic, predictive, vs prescriptive; see, e.g., Hunke et al. 2019). As data intermediation is the central value proposition (Dorfer 2016), it is also worthwhile to distinguish DDBMs based on the type of value proposition and offerings. For instance, Schüritz et al. (2019) differentiate between data, insights, and actions as offerings. Dehnert et al. (2021) further differentiate between data, information/knowledge, actions and non-data products and services in DDBMs. Hirt and Kühl (2018) describe Model-as-a-Service and Prediction-as-a-Service as two other types of offerings.

These offerings can be differentiated by the type of exchange of value objects (Leski et al. 2021). A value object, as described in the e-3 value ontology, *"is of value for one or more actors. Actors may value an object differently and subjectively, according to their own valuation preferences"* (Gordijn and Akkermans 2003,

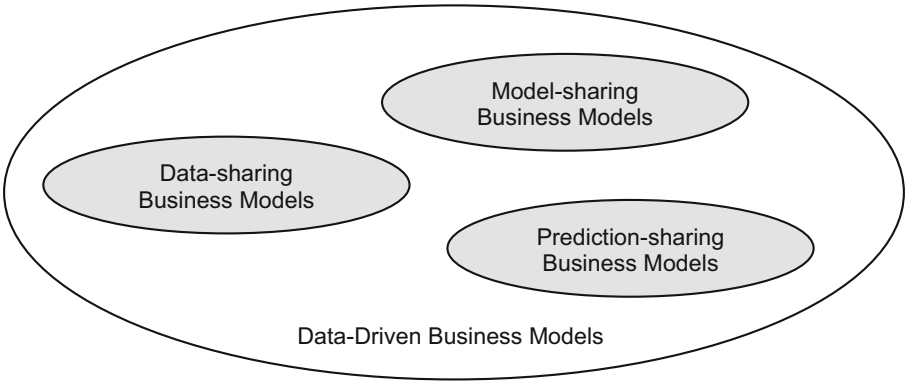

Fig. 1 Subtypes of DDBMs based on exchanged value objects

p. 120). Concerning DDBMs, such a value object can be *data* (e.g., Dehnert et al. 2021), *models* (e.g., Hirt and Kühl 2018) or *predictions* or insights in general (e.g., Schüritz et al. 2019).

By data, we understand a tradeable collection of *"codified observation[s] fixed in a tangible medium"* (Thomas et al. 2023, p. 256). Shared data can be in the form of specific data points, whole data sets (or data streams) or aggregated data (e.g., via descriptive statistics). By *model*, we understand a program or function that can identify patterns or provide predictions based on previously unseen input data. A model is a result of applying a machine learning algorithm to a set of (training) data. A model consists of its code and configuration. Hirt and Kühl (2018) differentiate between base models specific to one particular problem and transfer models that can be applied or transferred to a set of similar problems. The type of *prediction* encompasses identifying patterns, predicting events or attributes, or recommending actions based on incoming data applied to a learning model (Hirt and Kühl 2018). Predictions also represent target-specific insights that are shared to solve a specific (decision) problem of the customer, create customer benefit, and, in return, generate revenue.

As Fig. 1 illustrates, data-, model, and prediction-sharing business models can be understood as three subtypes of DDBMs. Differentiating DDBMs based on exchanged value objects is still under-represented in the DDBM literature, but a reasonable differentiation when it comes to knowledge risks: We assume that sharing different types of value objects leads to different types of risks.

Examples of DDBMs that provide *data* as an exchanged value object are API-based data-sharing business models in logistics (e.g., Möller et al. 2020). In such data-sharing business models (Schweihoff et al. 2023) or "data-as-a-service" business models (Chen et al. 2011), the business model owner grants other parties access to his own data set in exchange for compensation (Schweihoff et al. 2023; Vesselkov et al. 2019). One major obstacle to data sharing in organisations is the concern about exposing sensitive data and giving competitors a competitive advantage (Gelhaar and Otto 2020; Schweihoff et al. 2023). Thus, security aspects, such as usage restrictions

or cryptography, need to be implemented in such business models (Schweihoff et al. 2023).

Examples of DDBMs that provide *models* as an exchanged value object are Language-Model-as-a-Service (Sun et al. 2022). In such a model-as-a-service business model, the user provides or uploads data to the service provider who builds (trains) a model based on this training data and his own human and/or machine intelligence (Hirt and Kühl 2018).

Examples of DDBMs that provide *predictions* as an exchanged value object are prediction APIs (Santhosh et al. 2019). In a "prediction-as-a-service" or more general "analytics-as-a-service" business model, the provider applies a (machine learning) model to the input data provided by the customer to generate a prediction of events, recommendations or to identify patterns and finally to support decisions or automate actions for the customer (Hirt and Kühl 2018; Schüritz et al. 2019). We subsume these different terms under the term prediction for the context of this paper.

2.2 Knowledge Risks Emerging from Data Sharing

DDBMs involve new types of risks. Large-scale data sharing can cause leakage of competitive knowledge and intellectual property (Zeiringer and Thalmann 2020; Zeng et al. 2012). This risk is called knowledge risk and comprises potential knowledge attrition, loss, leakage or spill-over of knowledge that could adversely affect the organization's strategic advantage (Durst and Zieba 2017; Perrott 2007).

Competitive knowledge of a firm can be discovered from shared data sets using advanced analytics methods (Ilvonen et al. 2018). Further, it is difficult for firms to evaluate which knowledge could be discovered by external actors from shared data (Zeiringer and Thalmann 2020). Known approaches for external acquisition of competitive knowledge that endanger a firm's intellectual property are information leakage in supply chains (Zhang et al. 2012), industrial/data espionage (Thiel and Thiel 2015), or data breaches (Khan et al. 2021). An adversarial actor could also obtain valuable knowledge by reverse-engineering the firmware of a physical product to reconstruct an embedded algorithm (Thiel and Thiel 2015). For instance, it is technically possible to reverse-engineer black-box neural networks (e.g., Oh et al. 2019), or to steal machine learning models via API access (e.g., Tramèr et al. 2016).

The described attacks can lead to unintended leakage or spill-over of knowledge, denoted as knowledge risk (Ilvonen et al. 2018; Zeiringer and Thalmann 2020). A knowledge risk is the *"measure of the probability and severity of adverse effects of any activities engaging or related somehow to knowledge that can affect the functioning of an organisation on any level"* (Durst and Zieba 2018, p. 2). Knowledge risks can be analysed by the factors that cause them and the preventive measures organisations can take (Durst and Zieba 2017). Managing knowledge risks in terms of knowledge protection is one core strategy of knowledge management (Loebbecke et al. 2016). It is crucial for organisations as knowledge is essential for competitive advantage (Jennex and Zyngier 2007). Therefore, knowledge protection prevents unwanted knowledge leakage to non-authorized people and organisations (Manhart and Thalmann 2015). Existing knowledge protection literature focuses on formal and explicit knowledge. It does not consider tacit knowledge in organisations (Manhart

and Thalmann 2015) and the knowledge that can be discovered from data streams (Ilvonen et al. 2018). While explicit knowledge (e.g., materialised in data-related value objects) could quickly leave a company, tacit knowledge is more difficult to transfer and informal knowledge protection practices are needed (Thalmann et al. 2024).

Finally, developing business models can be understood as a set *"of concrete choices and the consequences of these choices"* (Casadesus-Masanell and Ricart 2010, p. 198). Managers must balance expected risks and estimated returns when deciding between different business model design options (Casadesus-Masanell and Ricart 2010; Tesch et al. 2017). Such risks can threaten the profitability of the business model or even the firm's value (Brillinger 2018), making it necessary to manage the risks. Risk management generally involves identifying, assessing and monitoring risks (Brillinger et al. 2020; Hallikas et al. 2004). Risks are usually evaluated by assessing the probability of a risk event and its impact on the business model (Hallikas et al. 2004; Brillinger et al. 2020). The problem with assessing non-financial risks, such as cyber security risks, is that little quantitative information is available, especially no reliable probability distributions (Franke 2020). Despite this, identifying and deciding how to deal with risks already in the business model design is crucial (Girotra and Netessine 2011). After identifying and being aware of risks, managers can adapt the business model design as a risk management measure (Brillinger et al. 2020).

Our Conclusion from the Literature DDBMs can be differentiated based on the offering or, in particular, exchanged values. Based on the literature, we have stated that offerings in DDBM can be distinguished by three types of value objects: data, models and predictions. Further, knowledge protection literature recognises data sharing as a knowledge risk in general and that extracting knowledge from shared data is possible via data science methods. We already have the first evidence from previous research that exchanging data-related value objects can lead to knowledge risks (Fruhwirth et al. 2019). Nevertheless, the relationship between knowledge risks and exchanged data-related value objects in DDBMs has not been studied, and this connection has not been made by previous literature.

3 Research Method

Our study aims to explore knowledge risks specific to DDBMs due to the specific nature of value objects. Given the novelty of the problem and lack of understanding of how and if knowledge risks occur in DDBMs, we applied an exploratory, qualitative research design that is appropriate for investigating why a certain phenomenon occurs (Yin 2009). The research design is qualitative, as we analysed interview data (see data collection section below), and exploratory, as we used a bottom-up data analysis method as informed inductive coding (see data analysis section below).

Table 1 Overview of our data collection process

	Interview Round 1	Interview Round 2	Interview Round 3
Interview participants	16 Interviews 7 Researchers (R1–R7), 9 Industry Experts (I1–I9) active in data-driven services, business model innovation and knowledge risks	7 Interviews 3 Researchers (R8–R10), 4 Industry Experts I10–I13 active in data-driven services	5 Interviews 5 Industry Experts (I14–I18) active in data-driven services
Duration	35–75 min	38–59 min	40–59 min
Goal, main questions and content	Focus on knowledge risks in DDBMs in general	Focus on knowledge risks from sharing data-related value objects (data, models and predictions)	Evaluation of results Presentation of 5 types of knowledge risks
Main outcomes	Knowledge risks differ if data, models or predictions are shared	Identified five types of risks based on the three types of value objects	Subtypes of risks for each type of shared value objects & contextual factors

3.1 Data Collection

Due to the tacit and sensitive nature of the topic for organisations, we decided on expert interviews in three rounds as our primary data source (see Table 1), as interviews allow comprehensive discussions (Yin 2009). As interview partners, we selected 28 experts, 18 from industry (I1–I18) and 10 from research institutions (R1–R10) (see Table 3 in Appendix A).

We followed a purposive sampling strategy (Etikan 2016) and, in particular, an expert sampling strategy that is useful "when investigating new areas of research" and in particular when "there is currently a lack of observational evidence" (Etikan 2016, p. 3). As it was challenging to identify suitable cases (i.e., organisations) where knowledge risks have or could occur, as such information is not publicly available, we also selected consultants and researchers as informants who reported such cases. Academic experts reported on their experience and cases of knowledge risks in DDBMs based on their collaboration with industry (e.g., as part of research or consulting projects). We selected experts based on their knowledge and experience in developing DDBMs or supporting organisations in that process. For academic experts, we considered their recent publications on DDBM as an additional selection criterion. The selection of experts in the initial interview round was broader: we also selected experts in business model innovation and knowledge risks in general to explore the topic. We searched for experts in our immediate network and through an extended network on the *LinkedIn* platform (2nd order contacts).

We conducted the interviews as face-to-face meetings or via digital communication software and audio-recorded them. Appendix A provides a detailed description of the experts who were interviewed.

The scope of the *first interview round* was very broad, serving as a starting point to explore knowledge risks in DDBMs. After initial data analysis, we found that differentiating and analysing knowledge risks in DDBMs based on exchanged value objects is interesting and reasonable. Therefore, we conducted seven additional and

more focused interviews with additional experts. In this *second interview round*, we presented and discussed the three data-related value objects (data, model, and predictions) and asked about cases and their relation to knowledge risks. In the first round, not all value objects were covered in each interview as the insights emerged over time. Further, we investigated motivations and practices in the design phase of a DDBM in detail, as we could now ask more focused questions in the second round.

At the beginning of our semi-structured interview guideline, we presented working definitions of central concepts and an abstract problem definition, illustrated with a case example. The interview was divided into two parts: The first two-thirds of the interview focused on exploring the problem of knowledge risks in DDBM. The last one-third (only in interview round 1) focused on discussing requirements for ICT tools identifying and describing knowledge risks in DDBMs (Fruhwirth et al. 2021). We asked the interview partners for real examples from their context to concretise and ground the discussion as much as possible within their experience. The guideline was tested with a PhD student from the same subject (with practical experience) and methodological knowledge (training) regarding the guideline's comprehensibility, question flow and structure. We adjusted our interview guideline for the second set of interviews through detailed questions (e.g., regarding protection measures) and a short presentation of our interim results. We presented each type of value object shortly and asked the experts how they perceived the knowledge risk related to each value object.

To validate our results, we conducted a *third interview round* with five additional industry experts in data-driven services and data analytics. The interviews lasted between 40 and 59 min. We again presented our problem definition, the concepts from the data analysis step after the two previous rounds and the five types of risks identified. Further, we provided one slide per type of risk with a short description and one example from the initial expert interviews. The three guiding questions for the evaluation interviews were: 1) Do you perceive these risks as relevant for your business? 2) Are there any other types of risks missing in that context? 3) Is the description of each risk reasonable for you?. Table 1 summarizes our data collection process.

3.2 Data Analysis

Interviews were fully transcribed and cleaned. Quotes used in this publication from interviews conducted in German were translated into English (marked with a "*") and reviewed by a second researcher. We analysed this data following a qualitative content analysis approach via informed inductive coding (Mayring 2015) using MAXQDA V.11.

For analysing the *first round of interviews*, the dimensions of analysis were themes that corresponded to the leading interview questions and developed a provisional coding scheme to structure the data. The major themes from the interview guideline have been "causes for knowledge risks", "consequences of knowledge risks", and examples. For the theme of the causes, we generated "influencing factors" and "mechanism" as our major categories. We distinguished between "type of knowledge

risk" and "second-order consequence" for the consequence theme. We have informed our inductive categories based on the literature presented in the background section, as Mayring (2015) suggested.

The intermediate category system was iteratively discussed among three researchers to focus the exploratory research until we arrived at three different types of data-related value objects (data, models, and predictions) that were suitable to analyse and distinguish knowledge risks in DDBMs, as they imply different types of knowledge risks. The value object types are based on the literature, e.g., Gordijn and Akkermans (2001) for the basic concept of value object; and, e.g. Hirt and Kühl (2018) for specific value object types data, models and predictions.

For analysing the data from the *second round of interviews*, we tightened the codes and category scheme and dropped all unrelated to knowledge risks associated with exchanged value objects (e.g., dropping codes like knowledge loss arising from leaving data scientists). If we identified new themes, we created new codes and matched them to the existing category scheme. We extended our categorisation scheme to describe the relation between value objects and knowledge risks. We constructed further categories to analyse and describe the risks, such as "protection measures", "knowledge retrieval mechanism", "type of knowledge", the "reason for sharing/exchanging", and "influencing factors" with different subcategories and codes informed by the literature. Appendix D shows exemplary text segments and corresponding codes mapped via the type of value object to the "Knowledge Protection Measure" category. In the last iteration of data analysis after the second interview round, we derived for each type of value object one or two types of risks (see results section).

For analysing the *third round of interviews*, we focused on identifying subtypes of risks to have a more differentiated view of the risks. Therefore, we also recoded the data from the first two rounds of interviews. Further, we aimed to identify contextual factors that influence the risks. Therefore, we simplified our category system to types of knowledge risks (integrating "reason for sharing/exchanging" and "knowledge retrieval mechanism" from round 2), contextual factors (integrating "influencing factors" and "type of knowledge" from round 2) and "knowledge protection measures". We found that the risk is higher when competitive knowledge is involved and that balancing expected benefits and risks was perceived as important. Further, we found from our additional interviewees that trusted relations and contractual regulations are two important protection measures.

4 Results

Sharing different types of data-related value objects lead to different types of risks: the risk of leaking competitive knowledge from shared data, the risk of leaking competitive knowledge by using a data service, the risk of leaking competitive knowledge from shared model, the risk of inference on the original training data from a shared model and the risk of reconstructing a model from shared predictions. For each type of value object, we present different types of risks, contextual factors influencing the risk and knowledge protection measures. Table 2 gives an overview of

Table 2 Overview of different risks for each type of shared value object and corresponding contextual factors and protection measures

Shared Value Object	Risks (and sub-risks)	Contextual Factors	Knowledge Protection Measures
Sharing Data	The risk of leaking competitive knowledge from shared data. The risk may arise – when sharing data sets in open data initiatives, – when sharing data with partners for joined data service development, or – when offering data-as-a-service, the customer could resell the data to third parties The risk of exposing competitive knowledge by using a data (prediction) service	The risk of knowledge leakage depends on the context, i.e., the value of the knowledge that could be discovered from shared data. The risk was perceived as critical if shared data was related to competitive knowledge. – Example: Competitive knowledge about production processes or product configurations or their development could be discovered from shared data	Classify data sources Involve a data platform Use secure technologies (e.g., encryption) Share only synthetic data Use data anonymization Build trusted relationships Set up appropriate contracts (e.g., NDAs) Share models instead of data Do not share data as an over-cautious measure Run the service on-premise or apply federated learning mechanisms
Sharing Model	The risk of leaking competitive knowledge from a shared model. The risk may arise – if the user is able to reconstruct the parameters or configuration from a shared black-box model, – if the provider needs to explain how the model comes to certain decisions or predictions, or – if the model is leaked to a third party, e.g., while collaborating with a startup to build the model The risk of inference on the original training data from a shared model	The risk of knowledge leakage depends on the context, i.e., if competitive knowledge can be derived from a shared model. – Competitive knowledge can be knowledge of the model generation process or domain knowledge that is materialized in the model. – Example: In consulting and engineering, preserved domain knowledge from experts is leaked when white-box models are shared. The knowledge risk depends on whether the model can be transferred to other application scenarios or if it is very specific to the application. In particular the risk depends on the volatility of the model or if appropriate data is available and can be applied	Implement legal protection mechanisms (e.g., regulate IP regarding model creation in contracts) Define and identify what information should be revealed by the model Adjust the business model (e.g., pricing model) Take technical measures (e.g., share only black-box models or use synthetic data for training a model) Offer the model as a service via a platform or an API (i.e., only sharing predictions)

Table 2 (Continued)

Shared Value Object	Risks (and sub-risks)	Contextual Factors	Knowledge Protection Measures
Sharing Prediction	The risk of reconstructing a model from shared predictions. Nevertheless, it is not so easy to draw clear conclusions—the inference is subject to probabilities	The risk of knowledge leakage depends on the context, i.e., – if prior information on the model is available, – the variance of input data, and – if the required expertise is available and the required effort is less than the expected gain	Control the access to the service (limit the number of requests per time unit or limit the allowed value range) Build the business model around dynamic data as a key resource Implement a pay-per-use revenue model

our results by stating the type of value object, then the (sub-) type of risk, contextual factors and knowledge protection measures. Subsequently, Sect. 4.1, 4.2 and 4.3 describe the risks associated with the value objects data, models, and predictions, respectively.

4.1 Sharing Data

4.1.1 The Risk of Leaking Competitive Knowledge from Shared Data

Organisations share data in a DDBM in return for economic or other non-financial benefits. Our interviewees are aware that knowledge can be discovered from data sets if domain knowledge, complementary data sets, or the necessary data analytics capabilities are available. The knowledge is represented implicitly in the data, and with data analytics methods, this knowledge can be discovered from the data, as one expert explained:

> "With the help of methods, you try to make the implicit information in this data explicit. For me, the data are the shell of the information. If you share it with someone who knows how it works, then he can generate an incredible amount of insights from it, which definitely have a business-critical factor." (I8*, CEO Data Science Company A)

The interviewee mentioned "an incredible amount of insights", which shows that he is aware of the risks but that it is very difficult to say which knowledge can be discovered exactly. This means that the data provider cannot specify the risk explicitly. Rather the risk is vague, and everyone has to be prepared for the unknown, as another expert mentioned:

> "Indeed, it seems to me that the risk is a very high one. Because it's so undefined because you're extracting something from this data that wasn't expected." (I5*, Director Digital Business)

This vagueness is a big challenge and makes the systematic assessment of knowledge risks in shared data sets challenging, especially if the business model owner is not aware of this implicit knowledge. We found that sharing data can lead to different knowledge risks depending on the business model.

Sharing Data Sets in Open Data Initiatives Could Lead to a Spill-Over and Thus Imply Knowledge Risks One motive for sharing data in the reported cases was to foster innovation and to create promising future business opportunities. The most extreme case for this direction was sharing data sets in open data initiatives so that others can build new services and the provider benefits from indirect revenues or reputation. However, this means that competitors could also access that data, which could lead to an unintended spill-over of knowledge and thus imply knowledge risks, as one interviewee mentioned:

> "All my competitors can also access this open data. And then, of course, I don't want them to gain too many insights into my operations so that they can dis-

> cover my competitive advantage and re-engineer and exploit it. As an open data provider, I try to find this balance between knowledge protection and knowledge sharing." (R8*, Researcher Data-Driven Service)

The interviewee mentioned a very important aspect: the balance between knowledge sharing and protection. Thus, while sharing data to foster innovation, organisations contrast the potential benefits with the potential negative impact of losing competitive knowledge. This is also the case if organisations share knowledge in a defined group, e.g., among project partners, to jointly develop a DDBM. Such stakeholders could also be in a competitive relationship.

Risk when Exposing Knowledge for Joined Data Service Development When developing a new service or platform, the organisation must share data with its partners and thus implicitly also competitive knowledge. Thus, there is the risk of leaking competitive knowledge when jointly developing a DDBM and, therefore, sharing data (e.g., for training a machine learning model). One expert mentioned here a potential case from the automotive domain:

> "If manufacturers A, B and C [...] are now jointly considering to build a data platform in order to generate telematics services that work everywhere, then this is fundamentally a knowledge risk." (I8*, CEO Data Science Company A)

The interviewee finally points to the knowledge risk resulting from sharing data sets.

Risk of Reselling Data by the Customer to Third Parties when Offering Data-as-a-Service One interviewee also mentioned the risk of reselling data by their customers to third parties when offering data-as-a-service. At the same time, he mentioned that they handle this via contracts:

> "That is, of course, standard in our contracts, for the data they have only a pure right of use but no right of exploitation. The right of use, the separation is difficult again, because if I sell the data to some consultant, he interprets the data for himself in some way, he has used it, and with the knowledge generated he now advises someone else." (I18*, Managing Director Data Service Company)

However, the expert also acknowledges the challenge of enforcing such contracts.

4.1.2 The Risk of Exposing Competitive Knowledge by Using a Data Service

One particular type of risk in data sharing evolves when an organisation or their employees use another organisation's data or prediction service. In this case, the service user often has to share his data with the service provider and, by that, risk that competitive knowledge might spill over. This risk will become even more important when using AI services and data science pipelines in the cloud. One expert mentioned the case where knowledge might be leaked through an AI service:

> "If you have employees who use ChatGPT, you also have the risk that information leakage happens—that information from your company goes somewhere else without anyone wanting it to. If you write a technical problem as a prompt, then OpenAI will also get your company secrets." (I15, Consultant Data Protection & AI)

4.1.3 Contextual Factors

The severity of the risk emerging from sharing data depends on the context, particularly the *value of the knowledge*, i.e., if it is competitive knowledge that might be leaked. Competitive knowledge about production processes or product configurations or their development could be discovered from shared product-related data with the help of data science methods. Such knowledge is especially critical in complex engineering products, such as vehicles, that require special engineering knowledge and huge development efforts. Shared data sets often allow the retrieval of such knowledge for unintended reasons in addition to the purpose for which it was collected and shared. One interviewee mentioned an incident where a car manufacturer shared data with a production equipment provider for predictive maintenance and where the provider could discover competitive knowledge on the production process from the data:

> "who could use this data to determine precisely when the customer was retooling his production line, how many units of a particular vehicle type were produced. Because he could derive exactly this data through various analyses." (I6*, Manager Data Analytics Consulting)

The interview partner described a concrete incident from his practice and linked it to the challenge of complex analytics. Complex analytics comprise multiple data analysis methods applied to the shared data set and combining it with other (publicly available) data. One experienced manager further mentioned one imaginable example from the automotive domain where a car manufacturer would share data of his vehicles on a platform:

> "You can't upload all the data from the CAN bus, from the ECU. Otherwise, someone with malicious intent could extract a lot of information from it about the development of the vehicles, about the performance of the vehicles, about the quality. All of this could be extracted from such data" (I8*, CEO Data Science Company A)

The interviewee is aware of the potential knowledge leakage and takes this into account while sharing. The consequence he described, in this case, is "you cannot share everything". You rather have to select and share based on the expectation that others can retrieve. In our interviews, we found that a differentiated consideration of knowledge risks is necessary, as one manager from the semiconductor industry pointed out:

> "On the other hand, when I talk about data that directly relates to the product, with which it is possible to draw conclusions about the architecture and technological specifics. Here, of course, the situation is different and the sensitivity of the information is higher." (I17*, Manager Data Analytics Semiconductor Company)

On the other hand, the manager also mentioned that sharing operational data from their production machines for maintenance or optimisation was perceived as less critical, as no conclusions on competitive knowledge are possible.

The risk of knowledge leakage through data sharing depends on the context. If data is shared that relates to competitive knowledge, i.e., about their products or core processes, that allows an external party to make conclusions on the architecture or technology used, then it is perceived as critical. If the data relates to a more common context, such as the maintenance of machines, sharing data was perceived as less critical. Thus, what is competitive knowledge is very specific to the company and depends on its business model.

One interviewee, therefore, pointed to the direction that *internal balancing is necessary*, i.e., at what stage is the retrieval of knowledge not acceptable for the company anymore? They need to take measures:

> "The internal discussions have to be held about when we have reached a level where drawing conclusions about the data or, for example, the vehicle's configuration, the production, the development, is no longer acceptable for us, and we therefore have to do something else." (I14*, Consultant Data-Driven Services)

4.1.4 Knowledge Protection Measures

As we have seen above, knowledge risks in DDBMs are very contextual, i.e., if the shared data relates to competitive knowledge. One protection measure that our interviewees mentioned was to *classify the data sources* and to decide if this data can be shared or not, as one manager from the semiconductor industry mentioned:

> "And you have to have business processes in place. That's what we have at our company in place, where you evaluate the data according to categories, from public to strictly confidential, for example." (I17*, Manager Data Science Semiconductor Company)

Another mechanism to tackle knowledge risks and enable data sharing is to *involve a data platform*. It mediates the data exchange between actors with technical measures implemented in the platform while preserving the provider's knowledge. The automotive manager further mentioned here:

> "That's why there are all these data-sharing platform initiatives, [enabling] data exchange under the premise of knowledge preservation. So, I can retain my knowledge but still share data. However this may work, it's a task that probably needs to be solved so that it really takes off." (I4*, Manager Data Analytics)

The interviewee highlighted that knowledge protection concerns seem to be one of the main motivations for the rise of data platforms. However, he also acknowledges that protection concerns must be addressed properly before implementing a DDBM. There are also technical measures regarding *secure technologies*, like encrypting or decentralising data when performing data analytics and thus applying methods such as multi-party computation or homomorphic encryption. Another approach mentioned was to *share only synthetic data*, i.e., data generated by generative AI with similar properties necessary for sharing.

Our interviewees frequently also mentioned using *contracts* such as NDAs (Non-Disclosure Agreements) to tackle this risk. Nevertheless, they cannot prevent knowledge leakage when the contract is breached. Further, our interviewees frequently mentioned trusted relationships as a measure to mitigate knowledge risks. One practical approach mentioned was to begin sharing smaller and less critical data sets and to intensify the relationship over time.

Firms and customers might be *over-cautious* and over-protective and, therefore, *unwilling to share* their data for fear of knowledge risks. This would imply that the DDBM is not implemented. This is especially the case as there is currently much awareness of data-related risks. Our interviewees reported the fear that others could benefit more from sharing and, therefore, as a consequence, decided not to share the data. This is perceived as a barrier for DDBMs, as one data science manager in the automotive industry mentioned:

> "Because all the companies in the [supply] chain are so afraid of losing know-how, they don't share the data. [...] This leads to the fact that it is sometimes difficult in the data environment for me to do business." (I4*, Manager Data Analytics)

Not realising a DDBM is the most extreme knowledge protection measure which is chosen if the perceived (vague) risks outweigh the perceived benefits of the DDBM. Therefore, our interviewees suggested balancing the expected benefits and possible risks:

> "And then there is also the question of the benefit: How much information can I gain when I give out data for further processing, versus the risk, what am I giving away?" (I18*, Managing Director Data Service Company)

Thus, the risk can be reduced by *running* a data service or prediction model *on-premise*, i.e., locally at the customer's premise, so that the data does not have to be shared. Another approach would be to use *federated learning architectures*, where the data stays local and only (transfer) models are shared or the weights of a neural network.

A further knowledge protection measure is *sharing models instead of data*. Models are exchanged to protect the underlying data and allow a bidirectional flow of information without exposing competitive knowledge, as one data science professor explained:

> "To build a model in order not to share the data. The model is already a risk mitigation method. With the goal, though, that you then have a flow of information in both directions." (R4*, Professor for Data Science)

The important aspect mentioned here is that exchanging models is a risk mitigation strategy, which is part of DDBMs.

4.2 Sharing Models

4.2.1 The Risk of Leaking Competitive Knowledge from Shared Model

Competitive knowledge might be leaked by sharing models, as knowledge from experts (e.g., engineers) is introduced to the model in the process of creating or training (e.g., engineering knowledge about the ageing behaviour of a certain technical component). Models could also reveal information they have learned but not intended to be shared. If the model is *shared in a white-box-like manner (i.e., sharing* the code with parameters and configuration), competitive knowledge is likely shared, leading to a knowledge risk. For instance, models are delivered as part of a consulting or engineering project to support the customer in developing a DDBM, as one interviewee reported:

> "We are a service provider for model development and algorithms, and we sell those directly to our customers, then we always sell a bunch of knowledge too." (I4*, Manager Data Analytics)

The interviewee highlights that, with the model, a huge amount of knowledge is transferred to the customer. Thus, our interviewees acknowledge that competitive knowledge could spill over to other actors if models are shared. The interviewed manager is already aware of this problem and mentioned later that there are hardly any organisational guidelines to ensure that shared models are not misused regarding knowledge leakage.

Reconstruct the Parameters or Configuration from a Black-Box Model Even if models are shared as black boxes, i.e., the configuration and parameters of the model are hidden, there is also the risk that knowledge can be retrieved through re-engineering of the model through specific data science methods from a theoretical point. Overall this risk was perceived as low compared to white-box models. One data analytics consultant reported here on one case:

> "In general, you can re-engineer nearly every model if you know the input and the output. Then there are also algorithmic methods to decompose analytics models. There are methods from explainable AI to understand them. [...] We see this more, and more frequently, our customers try to better understand how our models work." (I6*, Manager Data Analytics Consulting)

This example shows that business customers are already trying to understand and re-engineer models and that providers are aware of this fact. However, similar to other security mechanisms, it is a question of effort.

Needing to Explain how the Model Comes to Certain Decisions or Predictions The requirement of fair, accountable, and transparent AI (FAT AI) creates a demand to explain how models come to a certain decision or recommendation. One professor in Business Analytics sees this as a challenging trend from the perspective of knowledge protection and reported on one case from an industry project:

> "And there is a pressure here from the customer to the provider. Because you have to explain how a chatbot comes up with that conclusion. So, in that way, you are kind of exposing the algorithm behind this. [...] the openness of the algorithm means that you also expose knowledge." (R3, Professor for Business Analytics)

This example shows that providers could be forced to expose their underlying models and algorithms, and thereof knowledge could be retrieved from the exposed model. Thus, FAT AI-compliant models or explainable AI approaches could reduce the protective effect of models in DDBM.

Leaking the Model to a Third Party, e.g., when Collaborating with a Startup to Build the Model A knowledge risk from sharing models could also arise when a model is jointly developed with a partner (e.g., an AI start-up) and the model is leaked there to a third party. One manager, for instance, mentioned one potential scenario:

> "Let's say I have a transformer model that knows exactly how I make a chip at our company. If I lose something like that out of my hands, for example by cooperating with a startup or a partner company, whether it's small or large. Then I lose all know-how at the push of a button." (I17*, Manager Data Science Semiconductor Company)

4.2.2 The Risk of Inference of the Underlying Training Data

Further, data science methods, such as model inversion attacks, allow someone to *infer the original data* used to train the model. Competitive knowledge might spill over when the model user can reconstruct the original training data from a shared model, in particular, to infer the structure of the data (e.g., particular data fields) or the structure in the data (e.g., properties of the sample and the bias in the data). So-called model inference is technically possible in particular cases, according to data science literature (e.g., Fredrikson et al. 2015). Our experts mentioned that this can happen if a model is overfitting. This is particularly important for generative models, where not the original training data is generated, but only similar data. One of our interviewees mentioned here one hypothetical example where this model inference could happen:

> "[...] Then there is the risk that you are revealing information about your own data with the models. [...] Let's assume we take two insurance companies. They want to improve fraud detection. They exchange meta-information or train models together to do that. From that, you can get the structure of the data

used for training. And that underlying structure can already give one insurance company, which of course is a competitor, a lot of information about the other." (I9*, CEO/Co-Founder Data Science Company B)

4.2.3 Contextual Factors

The risk of knowledge leakage from sharing models depends on the context, i.e., if competitive knowledge can be derived from a shared model. Especially in consulting and engineering, preserved domain knowledge from experts is leaked when white-box models are shared, as one interviewee reported:

"If I take these models and give them away, then I've taken the knowledge that I've discovered from people, from their actions, from their labelling, from their input, preserved it in the model, and sold it to the outside world. That's a tremendous risk." (I4*, Manager Data Analytics)

This case shows that expert knowledge from employees is materialised into models. As part of an engineering business, models are shared with their customers. Moreover, through sharing the model, the materialised knowledge of their experts could spill over to their customers.

One expert from the semiconductor industry (I17) also mentioned a future example in terms of generative AI and transformer models that could explain how to build a technical system (e.g., a microchip). This could be a huge risk if such a model was trained with company-specific data and leaked (e.g., through a collaboration with a start-up).

The risk depends on how easily the model can be applied and *transferred to other application scenarios*, as one manager mentioned:

"If it [the model] is very specific to a problem, I'm not afraid. [...] If the model is very generic and easily transferable to different types of problems, to a different data set, to a different context, [...], then we have to be careful." (I4*, Manager Data Analytics)

The interviewer mentions, "I am not afraid" and "we have to be careful". Both phrases clearly indicate that this is a well-evaluated decision. Beyond abstract transferability, another organisation also needs the *capability* and knowledge to apply the model. Further, the availability of appropriate data sets where the model can be applied influences the risk, as one consultant mentioned:

"Without the raw data, the algorithm is less useful for me. [...] has the other party also the same raw data or other data with similar formats? If yes, then that is a big risk. [...] And the highest risks are in cases in which when the algorithm is leaked, and the raw data is available or reproducible." (I6*, Manager Data Analytics Consulting)

The interviewee points to the strategy of keeping the training data in the back and just sharing the model. This is especially important, as many successful DDBMs rely on unique dynamic data sets generated through using the service (e.g., location

data of traffic participants to predict traffic jams). Thus, the model only has value if it is used in combination with this *available data*. Another influencing factor is the *volatility* of the model: The risk increases if the model is valid for a longer period. Whereas the risk is lower if a dynamic model is constantly adjusted and updated.

Models implicitly contain the knowledge represented by the data used to train the model. Building a model also comprises knowledge of how to create value-added information from raw data, as one consulting manager explained:

> "[You need] a combination of knowledge of the data scientist who just looks at the raw data, at the graph, very simply speaking, and the engineer who knows exactly how the machine works, who knows exactly what it means when there's a pressure drop in the hydraulic arm of the welding robot." (I6*, Manager Data Analytics Consulting)

This statement shows that, on the one hand, domain-related knowledge, e.g., from engineering, is needed to train a model. On the other hand, knowledge from the data science discipline is also needed. Domain (expert) knowledge about a real-world phenomenon can add value to the model, such as specific casualties or relationships that cannot be discovered from the data itself but need additional contextual knowledge on the domain. Data science knowledge involves the labelling, preparing, and aggregating of the data and subsequent analytics and algorithmics and their combination.

4.2.4 Knowledge Protection Measures

Our interviewees mentioned that traditional *legal protection mechanisms* for IPR (e.g., patents) do not work for models. As the knowledge is only implicitly contained in the model, a lawsuit to convict the guilty seems very challenging. Therefore, the owner of the know-how and IP should be defined in contracts, e.g., the IP regarding the model creation remains at the provider.

Further, our interviewees mentioned *defining and identifying what information should be revealed by the model* and which not to build the model accordingly and ensure that the model is only used as intended. Models should be designed so that they only disclose the intended minimum amount of information (e.g., only the transfer function without revealing the influencing parameters (e.g., I17)). This, again, requires alignment and balance between sharing and protecting knowledge.

The risk also depends on the balance between generated returns and the estimated risk. For instance, one expert mentioned that the monetary value of selling a model would be significantly higher than only sharing predictions, in particular, if the code of the model can be accessed. Thus, the risk can also be mitigated by *adjusting the business model*, or more precisely, the *pricing model*.

Thus, protecting knowledge in DDBMs is currently mainly performed via *technical measures*. One simple knowledge protection mechanism is to share *only black-box models:*

> "For example, if I share the source code, where I can see every parameter of the [decision] tree, then it's clear that I'm selling critical knowledge. But in contrast, if I make predictions black box-like, then I would find it difficult to reconstruct the parameters." (I2*, Data Scientist Automotive Company A)

Our interviewees suggest applying data science methods to prevent model inversion attacks, such as randomisation in training, differential privacy, or other anonymization methods. For example, our interviewees mentioned using different loss functions or synthetic data for model training.

Another protection mechanism is to keep the model within the organisation's knowledge boundary and *offer the model as a service via a platform or an API*. However, the user of the model has to share his data now with the service provider, which could create a knowledge risk for the user. The provider shares only the results.

4.3 Sharing Predictions

4.3.1 The Risk of Reconstructing a Model from Shared Predictions

Competitive knowledge might spill over when plenty of predictions are shared, and the receiver can reconstruct the model or parts of the model based on these predictions. According to computer science literature, reconstructing models based on predictions is technically possible in particular cases (e.g., Tramèr et al. 2016). However, such attacks can be mitigated easily by restricting the number of predictions or the value range. Thus, this risk was perceived as low.

One way to discover knowledge is to reconstruct the underlying model by provoking lots of predictions. Moreover, the model allows inferences about the materialised knowledge. One academic expert in knowledge protection pointed to the problem:

> "If you sell many outcomes, yeah, then it would be even then possible to re-engineer the algorithm itself. If you are looking at what kind of results are created by what kind of data." (R5, Senior Researcher Knowledge Management)

However, the interviewee refers to "what kind of data", and another interviewee, a data scientist, specifies this in more detail:

> "If you take a look at the predictions now, you'll probably see a few features and check for which group it's working better or worse. You'll be able to reconstruct something there." (I2*, Data Scientist Automotive Company A)

As he says, "to reconstruct something there", he acknowledges the big challenge of discovering competitive knowledge from a prediction-based value proposition. However, our interviewees perceived the risk of knowledge leakage through sharing predictions as low as, for instance, one interviewee said:

> "For example, the customer only gets the results back. In that case, I think the risk is very low that any knowledge will drain from the provider because the customer doesn't have access to that knowledge." (R8*, Researcher Data-Driven Services)

This statement shows that the knowledge is hidden and that the customer has no direct access to the model and the materialised knowledge in the model. Further, our experts (e.g., I17) noted that it is not so easy to derive clear conclusions—the inference is subject to probabilities.

4.3.2 Contextual Factors

Reconstructing the model from predictions is possible from a theoretical point of view. However, in reality, this is not trivial and requires some *prior information on the model available*. How easy it is to reconstruct the model also depends on its complexity and the input data variability, as one expert in the field of DDBMs explained:

> "The heart of a good model is the variance of the input factors. And if I just offer an API, where I only provide a result to certain input values, but the input data that have led to that model has more variety than I'm allowing through the API, I can actually [prevent that well]." (I5*, Director Digital Business)

As this quote shows, re-engineering a model based on lots of "results" that we call predictions *depends on the variance of the input data* if it covers the whole input space. The knowledge is hidden and is materialised in the prediction model itself. The single prediction thus offers only a small and scattered glimpse of the model. Many predictions need to be collected or even provoked in a systematic attack to re-engineer knowledge:

> "If you send enough different queries, you can already [reconstruct] what knowledge is materialised in the model. Depending on the complexity of the problem, this might be a task at the moment, which do not allow model re-engineering due to the complexity." (I1*, Data Analytics Consultant)

This quote shows that reconstructing knowledge is possible but *requires significant effort and expertise*. If insights about the model are successfully collected, knowledge could be discovered. One mentioned example of knowledge that could be reconstructed is the bias that the model has learned. Further, one must balance the effort if it is worth it for the attacker.

4.3.3 Knowledge Protection Measures

When predictions are shared through access to a prediction model, one simple protection measure is to *control the access* in terms of the number of allowed requests, the minimum time span between two requests, and the range of input values. Limiting the number of requests prevents brute force attacks for reconstruction and also denial of service attacks. Potential attacks could be recognised through atypi-

cal requests, e.g., uniformly distributed across the input space, as training and re-engineering a model requires a broad range of input data. Our interviewed expert continued:

> "First of all, when someone penetrates me and asks me questions over the entire vector space, then I notice that this is atypical. That would be a uniform distribution in the query, which is totally atypical for such a thing, there you rather have a normal distribution in the queries." (I5*, Director Digital Business)

One protection measure is to build a prediction service that relies on *dynamic data*, such as real-time vehicle location data, generated through service usage and not shared with other actors. Even if the prediction model could be reconstructed based on many predictions, the knowledge cannot be applied as one malicious actor cannot access the necessary data. Another protection measure lies in the design of the business model: in prediction-as-a-service business models, *pay-per-use revenue models* are often used, which means that requesting lots of predictions gets expensive, and by that, even if something could be reconstructed from the model, it was compensated monetarily.

5 Discussion of Results

5.1 Discussion of Problem and Risk Relevance

In our interviews, we found that knowledge risks are a relevant topic in data-driven business models. For the three types of value objects data, model and prediction, we identified five types of risks that arise when they are exchanged in a DDBM: The risk of leaking competitive knowledge from shared data, the risk of exposing competitive knowledge by using a data service; the risk of leaking competitive knowledge from a shared model; the risk of inference of the underlying training data; and the risk of reconstructing a model from shared predictions.

The validation interviews confirmed the five types of risks, i.e., no additional types were suggested or emerged, and the description of the existing ones was sufficient. The risk of exposing competitive knowledge by using a data service was perceived as the most relevant risk in the validation interviews, as one expert brought it to the point: *"I think that is the biggest, but also very hard to grasp, threat or fear that the management in the industry has now"* (Industry Expert 14*). One problem is that the risk is very difficult to grasp. Therefore, there is sometimes a lot of fear, and as a consequence, companies are very cautious, and DDBMs may not be realised.

Knowledge Risks in DDBMs Depend on Contextual Factors of the DDBM Itself
The risk depends on the area of the company from which data-related value objects are shared. For instance, if data is shared to optimise an ancillary process (e.g., maintenance of production machines), the risk is perceived as less critical. Whereas, if data from their core process allows inference on their core processes, e.g., the design and configuration of products, the knowledge risk was perceived as critical. Thus, it must always be assessed if the (potential) leaked knowledge is competitive

and business-critical. Our data also suggest that knowledge risks are particularly relevant in knowledge-intensive businesses that want to innovate towards DDBMs in addition to their existing business model. In such business models, domain expert knowledge (e.g., engineering) is materialised in models that are shared with customers and partners as part of a DDBM. Thus, competitive knowledge might be put at risk. Further, in business models with complex systems and high competition (e.g., the automotive or semiconductor industry), organisations are very restrictive about data sharing, as corporate secrets might be shared with the data.

Knowledge Risks in DDBM Differ from Knowledge Risks Associated with Traditional Business Models As more areas of an organisation are digitised, there is a risk that more competitive knowledge is materialised in (AI) models. These, however, are easy to transfer compared to traditional business models, where engineers from the competition need to be headhunted or a product needs to be reverse engineered. In DDBMs, leaking a model could be sufficient for knowledge leakage. With the spread of generative AI and transformer models, we assume this aspect will become even more important in the upcoming years (cp. Tredinnick and Laybats 2023). Thus, the question of how to protect knowledge and IP in DDBMs will become more important.

5.2 Discussion of Protection Measures—How to Deal with the Risk?

We found that knowledge risk in DDBMs can be mitigated by technology (which might be fast changing), by business model design options, and by ensuring transparency, building trust and contractual regulations. As a synthesis of these three areas of action, one major strategic implication of our work is that knowledge risk mitigation in DDBMs needs a differentiated and balanced assessment of whether the perceived risk has a negative economic impact or is acceptable compared to the expected return in the DDBM.

5.2.1 Technology to Mitigate Knowledge Risks

A knowledge risk can often be reduced upfront by technology. Computer science literature discusses several technical attacks to retrieve something from data and models (see, e.g., Kaissis et al. 2020). Such attacks encompass training data leakage, model stealing, reverse engineering or membership inference (Hanzlik et al. 2021). Preventing such attacks or exacerbating the knowledge discovery process can be done by technical measures that relate to contemporary computer science research (see, e.g., Kaissis et al. 2020). Privacy-preserving technologies tailored to the context of big data analytics ensure the confidentiality of the data (e.g., Yakoubov et al. 2014). Examples of such privacy-preserving technologies are multi-party computation (e.g., Archer et al. 2018), data anonymization (Zeiringer et al. 2024), homomorphic encryption (e.g., Alabdulatif et al. 2020), watermarking (e.g., Regazzoni et al. 2021) or meta- and transfer machine learning (e.g., Hirt and Kühl 2018), which were also mentioned by our experts as technical protection measures. Such technology, like multi-party computation, has already been found to be a pro-

tection measure to mitigate knowledge risks in data-centric collaborations (Zeiringer 2021).

Summing up, the level of knowledge protection from a technology perspective highly depends on data science innovations and thus is a moving target. Thus, it is important to continuously monitor advances in computer science, both in terms of potential attacks and retrieval mechanisms and technical protection measures. This means technical expertise is needed in the strategic discussions for designing DDBMs.

5.2.2 Measures in Business Model Design to Mitigate Knowledge Risks

Proper design of a DDBM, particularly a proper choice of value object itself, is also a knowledge protection measure. A model can be shared when sharing data is considered too risky (i.e., competitive knowledge could be discovered from the data). Sharing models instead of data as a protection measure has been shown in the case of an R&D collaboration in the semiconductor industry (Kaiser et al. 2021). Also, instead of providing data to use a (prediction) service, federated machine learning can be applied (Hirt and Kühl 2018). In federated machine learning, the model is distributed to where the data is instead of gathering the data where the model is (Kaissis et al. 2020). When sharing a model is considered as too risky (i.e., competitive knowledge could be discovered from the model), predictions can be shared. Instead of giving out the prediction model, it can be accessed via an API enabling pay-per-use business models (Hanzlik et al. 2021). Also, detailed adjustments of the offering, such as limiting the number of access queries or the allowed data range of input values, could reduce the risk.

Nevertheless, there is a trade-off between sharing model (where the service provider risks a knowledge leakage) and sharing data (where the service user risks a knowledge leakage). Running a machine learning model on a client's computing systems can raise the fear of leaking details of the model, giving away the service provider's competitive knowledge (or IP) (Hanzlik et al. 2021). A protection against direct model access is an offline deployment of machine learning as a service (i.e., client site execution where model and computation remain secret) (Hanzlik et al. 2021).

Summing up, we assume from our explorative study that addressing knowledge risk concerns already during the design phase of a DDBM via suitable business model design (and in particular, a proper choice of the value object as part of the value proposition) is a key success factor for DDBM. This also depends on available technology and thus aligns with the field's technical developments.

5.2.3 Transparency, Trust and Contracts to Mitigate Knowledge Risks

Beyond addressing knowledge risk by technology and adjustments in the business model design, we found transparency, establishing trustful relationships and proper contracts as a third opportunity to mitigate knowledge risks in DDBMs. We assume that doing business with data in a B2B context will be only sustainable and profitable in the long term when data transparency and trust are part of the value proposition.

Trustful relationships through openness and security standards have also been noted as one measure to address knowledge risks in data-centric collaborations (Zeiringer 2021). Further, it is important to have proper contract regulations regarding the allowed usage of shared data-related value objects and instruments/sanctions for breaches. Nevertheless, these aspects have not been the focus of this study but have been mentioned frequently by the interviewees, e.g., that they have contractual regulations, such as NDAs, in place.

5.2.4 Strategic Implication: Multi-Perspective Assessment and Balancing of Knowledge Risks

To manage knowledge risks in DDBMs, it is important to assess the risk differentiated and balance sharing and protecting knowledge in a DDBM, as we perceived that there is partly very much fear, insecurity or overcautiousness regarding sharing data-related value objects. Knowledge protection literature suggests assessment and preventive measures (i.e., a clear risk assessment) and awareness for managing knowledge risks in data sharing via digital supply chains (Zeiringer and Thalmann 2022). Also, our interviewees suggested a differentiated view on the risks: When is competitive knowledge shared, or can it be discovered from a shared data-related value object? Is it company-critical knowledge? Is there an imminent business risk if something goes wrong? Second, there is also the question of to whom the knowledge goes. Is it a competitor, where it leads to a competition problem or to other stakeholders, where it is less problematic? Third, the effort and outcome of attacks also need to be evaluated. What is the effort to reconstruct something compared to the expected gain? Can reliable statements be discovered or only probabilities? Further, it must be evaluated up to which point it is acceptable that conclusions on the competitive knowledge can be drawn and at what point the risk is so high that measures must be taken. Thus, it is important to balance knowledge sharing and protection (Manhart and Thalmann 2015) and balancing estimated returns and acceptable risks in a DDBM (Casadesus-Masanell and Ricart 2010).

This balancing and differentiated view are, in particular, important, as some of our interview partners mentioned that ideas for DDBMs are often not realised because actors are afraid of sharing data and thus implicitly risk unwanted knowledge spillovers. Data exchange represents an obstacle due to confidentiality and privacy concerns (Miorandi et al. 2012). Thus, knowledge risks can be a barrier to innovation and influence the adoption of DDBMs. Considering them already during business model design and understanding the choice of value objects as a possible knowledge protection measure can help overcome this barrier.

5.3 Embedding the Discussion into Current Literature Streams

Our findings also relate to current literature streams in the context of DDBMs: data privacy and security in DDBMs, enhancing inter-organisational data sharing via data intermediates and trust-enhancing technologies, or the advancement of DDBMs towards AI-based business models.

With our study we connect to current literature on privacy and security in DDBMs. Privacy can be a threat or an opportunity (competitive advantage) in a DDBM. Therefore, privacy and data-driven business must go hand in hand (Schäfer et al. 2023a). Cybersecurity and privacy have been found as important capabilities for a DDBM to ensure confidentiality (Stahl et al. 2023). Ensuring data security via secure processes, legal frameworks and usage policies has been also found as a design principle for DDBMs (Azkan et al. 2022). Thus, security is an important factor in implementing DDBMs (Rashed et al. 2022). Security can be implemented via technological measures (e.g., encryption) and organisational measures (e.g., contractual agreements) to increase trust and transparency in data sharing (Azkan et al. 2022; Stahl et al. 2023). Overall, the strategic integration of IT security is seen as a key challenge in digitalization projects (Guggenmos et al. 2022) and DDBMs in particular. And therefore, risk management activities need to be aligned with the process of developing DDBMs (Schäfer et al. 2023a). In recent studies, the fear or risk of leaking sensitive information and competitive knowledge has been listed as one of many barriers in data sharing and DDBMs (e.g., Fassnacht et al. 2023; Azkan et al. 2022). In this study, we provided an in-depth study of this specific risk. Therefore, with our work, we extend existing literature on privacy and security in DDBMs, that often has a focus on personal data, with the additional perspective of knowledge risks, particularly that competitive knowledge can be leaked when exchanging data-related value objects in a DDBM. Further, we identified particular measures to manage knowledge risks as part of ensuring security in DDBMs.

Our results also connect to the current discussion in the literature on secure data exchange across the value chain with the help of data intermediaries (e.g., Stachon et al. 2023), such as data spaces (e.g., Gieß et al. 2023), and trust-enhancing technologies (Schäfer et al. 2023b), such as Multi-Party Computation (e.g., Agahari et al. 2022). These solutions address the risk that companies could lose competitive advantage when they participate in data sharing (e.g. Agahari et al. 2022) or the fear that shared data could be misused against them (e.g., Opriel et al. 2021). Data spaces also aim to solve the issue of data sovereignty when sharing data (e.g., Gieß et al. 2023). In this paper, we point to specific protection measures via data intermediates (like Data Marketplaces) and secure technologies (like Multi-Party Computation) and provide an application scenario in the context of DDBMs to prevent knowledge risks.

Our results also connect to the current debate in the literature on the advancement of DDBMs towards business models built around machine learning and AI (e.g., Vetter et al. 2022; Weber et al. 2022), where the issues of organisational data sharing will become even more important (Kanbach et al. 2023). (Generative) AI is data-driven and requires large amounts of data and, therefore, will affect organisational data sharing (Strobel et al. 2024). In such business models, data is used to train AI models instead of generating insights; these AI models are then embedded in services and products (Weber et al. 2022). AI-based business models induce, in particular, the automation of knowledge work through AI (Coombs et al. 2020). AI can complement or substitute humans at work (Murray et al. 2021). This delegation of tasks is related to agentic Information Systems (Baird and Maruping 2021). Such AI systems generate models that "provide descriptions and explanations for orga-

nizational knowing processes", contain prediction and decision functions and can perform real actions in the environment (Shollo et al. 2022, p. 9). Thus, competitive knowledge can be materialized in AI models. When these models are used in a service or is part of an offering, competitive knowledge could be leaked—a risk that we denoted as knowledge risks in this study. In our results we have already pointed in that direction. Thus, we contribute the perspective of knowledge risks to the topical literature on understanding and realizing AI-based business models.

6 Conclusion

In this interview study, we explored different types of knowledge risks in DDBMs with experts from research and industry. We explored the risks along three basic types of value objects: data, models and predictions and identified contextual factors and protection measures.

6.1 Implications

This study offers four implications that contribute to a deeper understanding of knowledge risks in DDBMs and, thus, to the ongoing debate on data sharing by adding the perspective of knowledge risks to DDBMs.

First, we contribute that the risk of knowledge leakage is a relevant risk factor in DDBMs. Knowledge risks in DDBMs differ from knowledge risks associated with traditional business models, as competitive knowledge is materialized in data or (AI) models, which makes knowledge more explicit to transfer. Thus, with our findings, we contribute a new risk that could occur in a DDBM and by that extending the existing debate on data privacy and security in DDBMs (e.g., Schäfer et al. 2023a; Azkan et al. 2022). We add the perspective of competitive knowledge that needs to be protected.

Second, we contribute that knowledge risks should be considered already in the design phase of a DDBM, and their management requires an interdisciplinary approach via a differentiated and balanced assessment. By studying knowledge risks and protection measures, we contribute to existing research on the design and realization of DDBMs (Rashed et al. 2022) and, in particular, by addressing a DDBM-specific risk (Fruhwirth et al. 2020). Further, we contribute to existing work on risk management in business model innovation in general (e.g., Brillinger 2018; Brillinger et al. 2020).

Third, we contribute that the level of knowledge protection from a technology perspective highly depends on computer science innovations and thus is a moving target. With our findings, we contribute to the literature stream on inter-organisational data sharing supported by data intermediates and trust-enhancing technologies (e.g., Agahari et al. 2022; Gieß et al. 2023; Schäfer et al. 2023b; Stachon et al. 2023) by providing application scenarios in the context of DDBMs where such technology could be needed. Simultaneously, we contribute to the field of knowledge risks in data-centric collaborations (e.g., Ilvonen et al. 2018; Kaiser et al. 2021; Zeiringer

and Thalmann 2022) by adding the case of DDBMs as one form of data-centric collaborations.

Fourth, we contribute that knowledge risk in DDBMs will become even more relevant with the extensive usage of machine learning and AI in DDBMs, particularly in knowledge work and knowledge-intensive businesses as competitive knowledge can be materialized in (AI) models. Thus, we contribute to the literature stream on AI-based business models (e.g., Farayola et al. 2023; Kanbach et al. 2023; Weber et al. 2022) by highlighting knowledge risks that could be associated with such business models.

6.2 Limitations and Future Research

Our research certainly comes with some limitations and opportunities for future research. First, due to the novelty of this topic, the availability of research data was limited, because of cases where a knowledge risk in a DDBM was identified are challenging to identify and experts that encountered such a risk in that context are hard to find. We have conducted exploratory research as a starting point to bring this problem to the discussion. Thus, further research is needed to establish a theoretical framework for knowledge risks in DDBMs and to investigate subtypes of risks in more detail.

Second, regarding the data collection process, we relied on expert opinions and their perception of knowledge risks in DDBMs. Collecting data from real-world cases via company representatives was difficult, as such information is usually not publicly available and shared. Further, sometimes the interviewees mentioned no real-world cases but described knowledge risks that they assumed to be relevant.

Third, in this exploratory research, we did not focus on quantifying the risks, i.e., estimating the probability and economic impact, as these depend highly on the individual context. Further research could develop and evaluate a set of criteria to assess and quantify the risk of knowledge leakage through shared value objects.

Fourth, based on our cases, we can see that the type of business model also influences the risk, e.g., sharing data in open data initiatives (with an open circle of stakeholders) implies a different risk than sharing data to develop a data-driven service with dedicated partners jointly. Therefore, we see one avenue for future research to investigate contextual factors of knowledge risks in DDBM in more detail. We assume that knowledge risks are, in particular, critical for knowledge-intensive businesses and business models with complex systems and high competition (e.g., the automotive or semiconductor industry). Such organisations are very restrictive with data sharing as corporate secrets might be shared with the data.

Fifth, we see important areas for future research in the context of AI. In this paper, we striped this topic at the edge. We assume that knowledge risk will become more important through the widespread deployment of AI in organisations: First, through the intensive usage of AI tools in the cloud, like large langue model-based tools (e.g., *ChatGPT* or *deepl* for translators) by employees of an organisation, sensitive information and therefore competitive knowledge might be leaked. Second, such AI models might expose information they have learned but were not intended to, e.g., in large language models. Third, through the increasing importance of explainable

and trustworthy AI, an organisation might have to open their models and expose competitive knowledge. Finally, with developments in generative AI, models will become more powerful, especially in engineering and knowledge-intensive companies. If AI can replace knowledge work, then leaking such a model would imply a huge risk.

7 Appendix

7.1 Appendix A

Table 3 List of Interviewed Experts

Round	ID	Type of Position	Industry	Duration (min)	Language
1	R1	Professor for Digital Business	Research	36	EN
1	R2	Professor for Business Model Innovation	Research	61	DE
1	R3	Professor Business Analytics	Research	35	EN
1	R4	Professor for Data Science	Research	56	DE
1	R5	Senior Researcher for Knowledge Management	Research	67	EN
1	R6	Professor for Knowledge Management	Research	60	DE
1	R7	Professor for Knowledge Management	Research	59	EN
1	I1	Consultant Data Analytics	Consulting	39	DE
1	I2	Data Scientist	Automotive	52	DE
1	I3	CEO/Co-Founder	Cyber Security	63	DE
1	I4	Manager Data Analytics	Automotive	62	DE
1	I5	Director Digital Business	Information Technology	76	DE
1	I6	Manager Data Analytics	Consulting	67	DE
1	I7	Manager Digital Business	Automotive	48	DE
1	I8	CEO/Co-Founder	Data Science	70	DE
1	I9	CEO/Co-Founder	Data Science	45	DE
2	R8	Researcher Data-Driven Services	Research	50	DE
2	R9	Research Group Leader Data Analytics	Research	53	DE
2	R10	Senior Researcher Data-Driven Services	Research	38	DE
2	I10	Manager Data Analytics	Consulting	55	DE
2	I11	Manager Data Science	Data-Driven Service	45	DE
2	I12	Managing Director	Data-Driven Service	48	DE
2	I13	Consultant Business Model Innovation	Consulting	59	DE
3	I14	Consultant Data-Driven Services	Information Technology	59	DE
3	I15	Consultant Data Protection, Artificial Intelligence	Consulting	43	DE
3	I16	Founder and Managing Director	Consulting	50	DE
3	I17	Manager Data Science	Semiconductor	42	DE
3	I18	Managing Director	Data-Driven Service	40	DE

7.2 Appendix B: Interview Guideline

7.2.1 Guiding Questions Interview Round 1

- Problem description
- Do you see this as a relevant problem? And do you know any similar examples?
- What other causes of risk could you imagine in this context in data-driven business models?
- What consequencies do you see based on these risks?
- Presentation of exemplary consequencies (knowledge loss, knowledge leakage, knowledge spill-over)
- How do you assess each of these consequences as a possible/relevant problem in data-driven business models? For each, do you know any example?
- What other consequences could arise from such knowledge risks?
- What examples from the practice of companies do you known to you where the topic of knowledge risks in data-driven business models is, was or could be relevant?
- 2nd part (not the scope of this paper): Presentation of a tool and evaluation questions

7.2.2 Guiding Questions Interview Round 2

- Problem description
- How do you assess this problem of knowledge risks just described as a relevant problem in your business model/in general?
- Can you tell an examples from practice you aware of where the issue of knowledge risks in data-driven business models is, was or could be relevant?
- What potential mechanisms can you think of to reconstruct or access knowledge in the three types as customer and attacker at the same time?
- What would be the potential consequences of such knowledge risks for your/an organization?
- What factors influence the risk of knowledge leakage through the exchange of data, models or predictions?
- What protection measures have you implemented to avoid or prevent such knowledge risks?

7.2.3 Guiding Questions Interview Round 3

- Presentation of problem knowledge risks in DDBMs
- Presentation of of interim results (main concepts, 5 types of risks, for each a short description and exemplary quotes from the interviews)
- Do you perceive these risks as relevant for your business?
- Are there any other types of risks missing in that context?
- Is the description of each risk reasonable for you?

7.3 Appendix C

Table 4 Coding scheme with main categories after interview round 2

Category	Description
Motives for sharing	This category describes motives why a type of value object is shared with other stakeholders
Type of knowledge	This category describes different types of knowledge that can be discovered from data-related value objects
Knowledge retrieval mechanism	This category describes mechanisms of how the knowledge can be discovered from the data-related value object by another party leading to a knowledge leakage
Influencing factors	This category describes the circumstances that make knowledge retrieval and, thus a, knowledge leakage possible. These factors influence the probability of the risk
Knowledge protection measures	This category describes measures of how technical or business model design measures could prevent such knowledge leakage

7.4 Appendix D

Table 5 Snapshot of the coding scheme and exemplary text segments

Text segment	Code	Type of value object	Category
"Ich mach das Ganze dann als Software-as-a-Service. Das wäre so die beste Mitigation." (I10) "Wenn man das Modell nur als API zur Verfügung stellt, dann kann jemand zwar Anfragen stellen, da kann jemand das Modell aber noch nicht rekonstruieren." (I9)	Offer Model-as-a-Service as a protection measure	Model	Knowledge protection measure
"... dass man verschlüsselte Daten für so eine Dienstleistung verwendet." (I1) "Daten sollten auf jeden Fall verschlüsselt übertragen werden." (I11)	Using encrypted data	Data	

Acknowledgements The research based on this paper has received funding from the Austrian COMET Program—Competence Centers for Excellent Technologies—under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry for Digital and Economic Affairs and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

Conflict of interest M. Fruhwirth, V. Pammer-Schindler and S. Thalmann declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agahari, Wirawan, Hosea Ofe, and Mark de Reuver. 2022. It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing. *Electronic Markets* 32(3):1577–1602. https://doi.org/10.1007/s12525-022-00572-w.

Al-Aali, Abdulrahman Y., and David J. Teece. 2013. Towards the (strategic) management of intellectual property: retrospective and prospective. *California Management Review* 55(4):15–30. https://doi.org/10.1525/cmr.2013.55.4.15.

Alabdulatif, Abdulatif, Ibrahim Khalil, and Xun Yi. 2020. Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption. *Journal of Parallel and Distributed Computing* 137:192–204. https://doi.org/10.1016/j.jpdc.2019.10.008.

Archer, David W., Dan Bogdanov, Yehuda Lindell, Liina Kamm, Kurt Nielsen, Jakob I. Pagter, Nigel P. Smart, and Rebecca N. Wright. 2018. From keys to databases—real-world applications of secure multi-party computation. *The Computer Journal* 61(12):1749–1771. https://doi.org/10.1093/comjnl/bxy090.

Azkan, Can, Frederik Moller, Lennart Iggena, and Boris Otto. 2022. Design principles for industrial data-driven services. *IEEE Transactions on Engineering Management* 71:2379-2402 https://doi.org/10.1109/TEM.2022.3167737.

Baird, Aaron, and Likoebe M. Maruping. 2021. The next generation of research on IS use: a theoretical framework of delegation to and from agentic IS artifacts. *Management Information Systems Quarterly* 45(1):315–341. https://doi.org/10.25300/MISQ/2021/15882.

Brillinger, Anne-Sophie. 2018. Mapping business model risk factors. *International Journal of Innovation Management* 22(05):1840005. https://doi.org/10.1142/S1363919618400054.

Brillinger, Anne-Sophie, Christian Els, Björn Schäfer, and Beate Bender. 2020. Business model risk and uncertainty factors: toward building and maintaining profitable and sustainable business models. *Business Horizons* 63(1):121–130. https://doi.org/10.1016/j.bushor.2019.09.009.

Casadesus-Masanell, Ramon, and Joan E. Ricart. 2010. From strategy to business models and onto tactics. *Long Range Planning* 43(2–3):195–215. https://doi.org/10.1016/j.lrp.2010.01.004.

Chen, Ying, Jeffrey Kreulen, Murray Campbell, and Carl Abrams. 2011. *Analytics ecosystem transformation: a force for business model innovation*. 2011 Annual SRII Global Conference., 11–20. https://doi.org/10.1109/SRII.2011.12.

Coombs, Crispin, Donald Hislop, Stanimira K. Taneva, and Sarah Barnard. 2020. The strategic impacts of intelligent automation for knowledge and service work: an interdisciplinary review. *The Journal of Strategic Information Systems* 29(4):101600. https://doi.org/10.1016/j.jsis.2020.101600.

Dehnert, Maik, Alexander Gleiss, and Reiss Frederik. 2021. *What makes a data-driven business model? A consolidated taxonomy*. Proceedings of the Twenty-Ninth European Conference on Information Systems (ECIS 2021).

Dorfer, Laura. 2016. Datenzentrische Geschäftsmodelle als neuer Geschäftsmodelltypus in der Electronic-Business-Forschung: Konzeptionelle Bezugspunkte, Klassifikation und Geschäftsmodellarchitektur.

Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 68(3):307–369. https://doi.org/10.1007/s41471-016-0014-9.

Durst, Susanne, and Malgorzata Zieba. 2017. Knowledge risks—towards a taxonomy. *International Journal of Business Environment* 9(1):51–63. https://doi.org/10.1504/IJBE.2017.084705.

Durst, Susanne, and Malgorzata Zieba. 2018. Mapping knowledge risks: towards a better understanding of knowledge management. *Knowledge Management Research & Practice* 17(1):1–13. https://doi.org/10.1080/14778238.2018.1538603.

Etikan, Ilker. 2016. Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics* 5(1):1–4. https://doi.org/10.11648/j.ajtas.20160501.11.

Farayola, Oluwatoyin A., A. Abdul Adekunle, Blessing O. Irabor, and Evelyn C. Okeleke. 2023. Innovative business models driven by AI technologies: a review. *Computer Science & IT Research Journal* 4(2):85–110. https://doi.org/10.51594/csitrj.v4i2.608.

Fassnacht, Marcel, Carina Benz, Daniel Heinz, Jasmin Leimstoll, and Gerhard Satzger. 2023. Barriers to data sharing among private sector organizations. In Proceedings of the 56th Annual Hawaii International Conference on System Sciences, January 3–6, 2023., ed. Tung X. Bui, 3695–3704. Honolulu: Department of IT Management Shidler College of Business University of Hawaii.

Franke, Günter. 2020. Management nicht-finanzieller Risiken: eine Forschungsagenda. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 72:279-320. https://doi.org/10.1007/s41471-020-00096-z.

Fredrikson, Matt, Somesh Jha, and Thomas Ristenpart. 2015. *Model inversion attacks that exploit confidence information and basic countermeasures*. CCS '15: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security., 1322–1333. https://doi.org/10.1145/2810103.2813677.

Fruhwirth, Michael, Viktoria Pammer-Schindler, and Stefan Thalmann. 2019. To Sell or Not to Sell: Knowledge Risks in Data-Driven Business Models. In *Proceedings of the 2019 Pre-ICIS SIGDSA Symposium*. 11.

Fruhwirth, Michael, Christiana Ropposch, and Viktoria Pammer-Schindler. 2020. Supporting data-driven business model innovations: a structured literature review on tools and methods. *Journal of Business Models* 8(1):7–25.

Fruhwirth, Michael, Viktoria Pammer-Schindler, and Stefan Thalmann. 2021. A Network-based Tool for Identifying Knowledge Risks in Data-Driven Business Models. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, ed. Tung X. Bui, 5218-5227.

Gelhaar, Joshua, and Boris Otto. 2020. *Challenges in the emergence of data ecosystems*. Twenty-Third Pacific Asia Conference on Information Systems. UAE, 2020, Dubai.

Gieß, Anna, Frederik Möller, Thorsten Schoormann, and Boris Otto. 2023. *Design options for data spaces*. Thirty-first European Conference on Information Systems (ECIS 2023).

Girotra, Karan, and Serguei Netessine. 2011. How to build risk into your business model. *Harvard Business Review* 89(5):100–105.

Gordijn, Jaap, and Hans Akkermans. 2001. Designing and evaluating e-business models. *IEEE Intelligent Systems* 16(4):11–17. https://doi.org/10.1109/5254.941353.

Gordijn, Jaap, and J.M. Akkermans. 2003. Value-based requirements engineering: exploring innovative e-commerce ideas. *Requirements Engineering* 8(2):114–134. https://doi.org/10.1007/s00766-003-0169-x.

Guggenberger, Moritz T., Frederik Möller, Karim Boualouch, and Boris Otto. 2020. *Towards a unifying understanding of digital business models*. Twenty-Third Pacific Asia Conference on Information Systems, UAE, 2020, Dubai.

Guggenmos, Florian, Björn Häckel, Philipp Ollig, and Bastian Stahl. 2022. Security first, security by design, or security pragmatism—strategic roles of IT security in digitalization projects. *Computers & Security* 118:102747. https://doi.org/10.1016/j.cose.2022.102747.

Günther, Wendy A., Mohammad H. Rezazade Mehrizi, Marleen Huysman, and Frans Feldberg. 2017. Debating big data: a literature review on realizing value from big data. *The Journal of Strategic Information Systems* 26(3):191–209. https://doi.org/10.1016/j.jsis.2017.07.003.

Hallikas, Jukka, Iris Karvonen, Urho Pulkkinen, Veli-Matti Virolainen, and Markku Tuominen. 2004. Risk management processes in supplier networks. *International Journal of Production Economics* 90(1):47–58. https://doi.org/10.1016/j.ijpe.2004.02.007.

Hanzlik, Lucjan, Yang Zhang, Kathrin Grosse, Ahmed Salem, Maxmilian Augustin, Michael Backes, and Mario Fritz. 2021. *MLCapsule: guarded Offline deployment of machine learning as a service*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 3300–3309.

Hartmann, Philipp M., Mohamed Zaki, Niels Feldmann, and Andy Neely. 2016. Capturing value from big data—a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management* 36(10):1382–1406. https://doi.org/10.1108/IJOPM-02-2014-0098.

Hernandez, Exequie, G. Sanders, and Anja Tuschke. 2015. Network defense: pruning, grafting, and closing to prevent leakage of strategic knowledge to rivals. *Academy of Management Journal* 58(4):1233–1260. https://doi.org/10.5465/amj.2012.0773.

Hirt, Robin, and Niklas Kühl. 2018. Cognition in the era of smart service systems: inter-organizational analytics through meta and transfer learning. In *Proceedings of the 39th International Conference on Information Systems—Bridging the Internet of People, Data, and Things*. Francisco., ed. Jan Pries-Heje, Sudha Ram, and Michael Rosemann

Hunke, Fabian, Christian Engel, Ronny Schüritz, and Philipp Ebel. 2019. Understanding the anatomy of analytics-based services: a taxonomy to conceptualize the use of data and Analytics in service. In *Proceedings of the 27th European Conference on Information Systems—Information Systems for a Sharing Society*. Stockholm Uppsala., ed. Jan Vom Brocke, Shirley Gregor, and Oliver Müller

Ilvonen, Ilona, Stefan Thalmann, Markus Manhart, and Christian Sillaber. 2018. Reconciling digital transformation and knowledge protection: a research agenda. *Knowledge Management Research & Practice* 16(2):235–244. https://doi.org/10.1080/14778238.2018.1445427.

Jennex, Murray E., and Suzanne Zyngier. 2007. Security as a contributor to knowledge management success. *Information Systems Frontiers* 9(5):493–504. https://doi.org/10.1007/s10796-007-9053-4.

Jiang, Xu, Bao Yongchuan, Yan Xie, and Shanxing Gao. 2016. Partner trustworthiness, knowledge flow in strategic alliances, and firm competitiveness: a contingency perspective. *Journal of Business Research* 69(2):804–814. https://doi.org/10.1016/j.jbusres.2015.07.009.

Kaiser, Rene, Stefan Thalmann, and Viktoria Pammer-Schindler. 2021. An investigation of knowledge protection practices in inter-organisational collaboration: protecting specialised engineering knowledge with a practice based on grey-box modelling. *VINE Journal of Information and Knowledge Management Systems* 51(5):713–731. https://doi.org/10.1108/VJIKMS-11-2019-0180.

Kaissis, Georgios A., Marcus R. Makowski, Daniel Rückert, and Rickmer F. Braren. 2020. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence* 2(6):305–311. https://doi.org/10.1038/s42256-020-0186-1.

Kale, Prashant, Harbir Singh, and Howard Perlmutter. 2000. Learning and protection of proprietary assets in strategic alliances: building relational capital. *Strategic Management Journal* 21(3):217–237.

Kanbach, Dominik K., Louisa Heiduk, Georg Blueher, Maximilian Schreiter, and Alexander Lahmann. 2023. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science* https://doi.org/10.1007/s11846-023-00696-z.

Khan, Freeha, Jung H. Kim, Lars Mathiassen, and Robin Moore. 2021. Data breach management: an integrated risk model. *Information & Management* 58(1):103392. https://doi.org/10.1016/j.im.2020.103392.

Kühne, Babett, and Tilo Böhmann. 2019. Data-driven business models: building the bridge between data and value. In *Proceedings of the 27th European Conference on Information Systems—Information Systems for a Sharing Society*. Stockholm Uppsala., ed. Jan Vom Brocke, Shirley Gregor, and Oliver Müller

Leski, Florian, Michael Fruhwirth, and Viktoria Pammer-Schindler. 2021. Who Else do you need for a data-driven business model? Exploring roles and exchanged values. In *34th bled econference digital support from crisis to progressive change*. June 27–30, 2021., ed. Andreja Pucihar, Mirjana Kljajić Borštnar, Roger Bons, Helen Cripps, Anand Sheombar, and Doroteja Vidmar, 365–378.

Loebbecke, Claudia, Paul C. van Fenema, and Philip Powell. 2016. Managing inter-organizational knowledge sharing. *The Journal of Strategic Information Systems* 25(1):4–14. https://doi.org/10.1016/j.jsis.2015.12.002.

Manhart, Markus, and Stefan Thalmann. 2015. Protecting organizational knowledge: a structured literature review. *Journal of Knowledge Management* 19(2):190–211. https://doi.org/10.1108/JKM-05-2014-0198.

Mayring, Philipp. 2015. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*, 12th edn., Weinheim: Beltz.

Miorandi, Daniele, Sabrina Sicari, Francesco de Pellegrini, and Imrich Chlamtac. 2012. Internet of things: Vision, applications and research challenges. *Ad Hoc Networks* 10(7):1497–1516. https://doi.org/10.1016/j.adhoc.2012.02.016.

Möller, Frederik, Maleen Stachon, Christina Hoffmann, Henrik Bauhaus, and Boris Otto. 2020. Data-driven business models in logistics: a taxonomy of optimization and visibility services. In *Proceed-*

ings of the 53rd Annual Hawaii International Conference on System Sciences (HICSS2020), ed. Tung Bui, 5379–5388.

Murray, Alex, Jen Rhymer, and David G. Sirmon. 2021. Humans and technology: forms of conjoined agency in organizations. *Academy of Management Review* 46(3):552–571. https://doi.org/10.5465/amr.2019.0186.

Oh, Seong J., Max Augustin, Bernt Schiele, and Mario Fritz. 2019. Towards reverse-engineering black-box neural networks. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, ed. Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller, 121–144. Cham: Springer.

Opriel, Sebastian, Frederik Möller, Ute Burkhardt, and Boris Otto. 2021. Requirements for usage control based exchange of sensitive data in automotive supply chains. In *Proceedings of the 54th Annual Hawaii International Conference on System Sciences*, ed. Tung Bui, 431–440.

Osterwalder, Alexander, and Yves Pigneur. 2010. *Business model generation: a handbook for visionaries, game changers, and challengers*, 1st edn., Hoboken: Wiley.

Osterwalder, Alexander, Yves Pigneur, and Christopher L. Tucci. 2005. Clarifying business models: origins, present, and future of the concept. *Communications of the Association for Information Systems* https://doi.org/10.17705/1CAIS.01601.

Perrott, Bruce E. 2007. A strategic risk approach to knowledge management. *Business Horizons* 50(6):523–533. https://doi.org/10.1016/j.bushor.2007.08.002.

Rashed, Faisal, Paul Drews, and Mohamed Zaki. 2022. *A reference model for data-driven business model innovation initiatives in incumbent firms*. Proceedings of the Thirtieth European Conference on Information Systems (ECIS 2022), Timișoara.

Regazzoni, Francesco, Paolo Palmieri, Fethulah Smailbegovic, Rosario Cammarota, and Ilia Polian. 2021. Protecting artificial intelligence IPs: a survey of watermarking and fingerprinting for machine learning. *CAAI Transactions on Intelligence Technology* 6(2):180–191. https://doi.org/10.1049/cit2.12029.

Santhosh, Gautham, Fabrizio de Vita, Dario Bruneo, Francesco Longo, and Antonio Puliafito. 2019. *Towards trustless prediction-as-a-service*. 2019 IEEE International Conference on Smart Computing (SMARTCOMP)., 317–322. https://doi.org/10.1109/SMARTCOMP.2019.00068.

Schäfer, Fabian, Heiko Gebauer, Christoph Gröger, Oliver Gassmann, and Felix Wortmann. 2023a. Data-driven business and data privacy: Challenges and measures for product-based companies. *Business Horizons* 66(4):493–504. https://doi.org/10.1016/j.bushor.2022.10.002.

Schäfer, Fabian, Jeremy Rosen, Christian Zimmermann, and Felix Wortmann. 2023b. *Unleashing the potential of data ecosystems: establishing digital trust through trust-enhancing technologies*. Thirty-first European Conference on Information Systems (ECIS 2023).

Schüritz, Ronny, Stefan Seebacher, and Rebecca Dorner. 2017a. Capturing value from data: revenue models for data-driven services. In *Proceedings of the 50th Hawaii International Conference on System Sciences*. Waikoloa Village., ed. Tung Bui, 5348–5357.

Schüritz, Ronny, Stefan Seebacher, Gerhard Satzger, and Lukas Schwarz. 2017b. *Datatization as the next frontier of Servitization: understanding the challenges for transforming organizations*. Proceedings of the Thirty-Eighth International Conference on Information Systems (ICIS), Seoul.

Schüritz, Ronny, Killian Farrell, Barbara H. Wixom, and Gerhard Satzger. 2019. *Value co-creation in data-driven services: towards a deeper understanding of the joint sphere*. Proceedings of the Fortieth International Conference on Information Systems (ICIS), Munich.

Schweihoff, Julia, Ilka Jussen, Valentin Dahms, Frederik Möller, and Boris Otto. 2023. *How to share data Online (fast)—A taxonomy of data sharing business models*. Proceedings of the 56th Hawaii International Conference on Systems Sciences (HICSS).

Shollo, Arisa, Konstantin Hopf, Tiemo Thiess, and Oliver Müller. 2022. Shifting ML value creation mechanisms: a process model of ML value creation. *The Journal of Strategic Information Systems* 31(3):101734. https://doi.org/10.1016/j.jsis.2022.101734.

Stachon, Maleen, Frederik Möller, Moritz T. Guggenberger, and Martin Tomczyk. 2023. *Understanding data trusts*. Proceedings of the Thirty-first European Conference on Information Systems (ECIS 2023), Kristiansand.

Stahl, Bastian, Björn Häckel, Daniel Leuthe, and Christian Ritter. 2023. Data or business first?—manufacturers' transformation toward data-driven business models. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 75:303–343. https://doi.org/10.1007/s41471-023-00154-2.

Strobel, Gero, Frederik Möller, Thorsten Schoormann, and Boris Otto. 2024. Introduction to the 2nd Mini-Track on Designing Data Ecosystems: Values, Impacts, and Fundamentals. In *Proceedings of the 57th Annual Hawaii Interna8onal Conference on System Sciences*, ed. Tung X. Bui, 4236–4237.

Sun, Tianxiang, Shao Yunfan, Huang Xuanjing, and Xipeng Qiu. 2022. Black-Box Tuning for Language-Model-as-a-Service. *Proceedings of the 39th International Conference on Machine Learning*, 20841–20855.

Teece, David J. 2010. Business models, business strategy and innovation. *Long Range Planning* 43(2–3):172–194. https://doi.org/10.1016/j.lrp.2009.07.003.

Tesch, Jan, Anne-Sophie Brillinger, and Dominik Bilgeri. 2017. Internet of things business model innovation and the stage-gate process: an exploratory analysis. *International Journal of Innovation Management* 21(5):1740002-1–1740002-17. https://doi.org/10.1142/S1363919617400023.

Thalmann, Stefan, Ronald Maier, Ulrich Remus, and Markus Manhart. 2024. Connect with care: informal knowledge protection practices to enhance knowledge sharing in networks of organizations. *VINE Journal of Information and Knowledge Management Systems,* Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/VJIKMS-02-2022-0051

Thiel, Christian, and Christoph Thiel. 2015. Hare and tortoise: can industry 4.0 win the race against counterfeiting and piracy? *Datenschutz und Datensicherheit* 1(0):663–667.

Thomas, Llewellyn D.W., Aija Leiponen, and Pantelis Koutroumpos. 2023. Profiting from data products. In *Research handbook on digital strategy*, ed. Carmelo Cennamo, Giovanni Battista Feng Zhu Dagnino, 255–272. Cheltenham: Edward Elgar Publishing.

Tramèr, Florian, Fan Zhang, Ari Juels, Michael K. Reiter, and Thomas Ristenpart. 2016. *Stealing machine learning models via prediction APis*. Proceedings of the 25th USENIX Security Symposium, 601–618.

Tredinnick, Luke, and Claire Laybats. 2023. The dangers of generative artificial intelligence. *Business Information Review* 40(2):46–48. https://doi.org/10.1177/0266382123118375.

Vesselkov, Alexandr, Heikki Hämmäinen, and Juuso Töyli. 2019. Design and governance of mhealth data sharing. *Communications of the Association for Information Systems* 45(1):299–321. https://doi.org/10.17705/1CAIS.04518.

Vetter, Oliver A., Felix S. Hoffmann, Luisa Pumplun, and Peter Buxmann. 2022. *What constitutes a machine-learning-driven business model? A taxonomy of B2B start-ups with machine learning at their core*. Proceedings of the Thirtieth European Conference on Information Systems (ECIS 2022), Timișoara.

Weber, Michael, Moritz Beutter, Jörg Weking, Markus Böhm, and Helmut Krcmar. 2022. AI startup business models. *Business & Information Systems Engineering* 64(1):91–109. https://doi.org/10.1007/s12599-021-00732-w.

Wiener, Martin, Carol Saunders, and Marco Marabelli. 2020. Big-data business models: a critical literature review and multiperspective research framework: a critical literature review and multi-perspective research framework. *Journal of Information Technology* 35(1):66–91. https://doi.org/10.1177/0268396219896811.

Woerner, Stephanie L., and Barbara H. Wixom. 2015. Big data: extending the business strategy toolbox. *Journal of Information Technology* 30(1):60–62. https://doi.org/10.1057/jit.2014.31.

Yakoubov, Sophia, Vijay Gadepally, Nabil Schear, Emily Shen, and Arkady Yerukhimovich. 2014. *A survey of cryptographic approaches to securing big-data analytics in the cloud*. 2014 IEEE High Performance Extreme Computing Conference (HPEC), 1–6. https://doi.org/10.1109/HPEC.2014.7040943.

Yin, Robert K. 2009. *Case study research: design and methods*, 4th edn., Los Angeles: SAGE.

Zeiringer, Johannes P. 2021. *Tackling knowledge risks in data-centric collaborations: a tackling knowledge risks in data-centric collaborations: a literature review literature review*. PACIS 2021 Proceedings.

Zeiringer, Johannes P., and Stefan Thalmann. 2020. Knowledge risks in digital supply chains: a literature review. In *WI2020 Zentrale Tracks: 15th International Conference on Wirtschaftsinformatik*. Potsdam, March 9–11, 2020., ed. Norbert Gronau, Moreen Heine, K. Poustcchi, and H. Krasnova, 370–385. GITO Verlag. 15. Internationale Tagung Wirtschaftsinformatik.

Zeiringer, Johannes P., and Stefan Thalmann. 2022. Knowledge sharing and protection in data-centric collaborations: an exploratory study. *Knowledge Management Research & Practice* 20(3):436–448. https://doi.org/10.1080/14778238.2021.1978886.

Zeiringer, Johannes P., Stefan Thalmann, and Jürgen Fleiss. 2024. Data anonymization as instrument to manage knowledge risks in supply chains. In *Proceedings of the 57th Annual Hawaii International Conference on System Sciences*, ed. Tung X. Bui, 5503–5512.

Zeng, Yong, Wang Lingyu, Deng Xiaoguang, Cao Xinlin, and Nafisa Khundker. 2012. Secure collaboration in global design and supply chain environment: Problem analysis and literature review. *Computers in Industry* 63(6):545–556. https://doi.org/10.1016/j.compind.2012.05.001.

Zhang, Da Yong, Cao Xinlin, Wang Lingyu, and Yong Zeng. 2012. Mitigating the risk of information leakage in a two-level supply chain through optimal supplier selection. *Journal of Intelligent Manufacturing* 23(4):1351–1364. https://doi.org/10.1007/s10845-011-0527-3.