Design Principles for Quality Scoring–Coping with Information Asymmetry of Data Products

Tobias Moritz Guggenberger
Fraunhofer ISST
TU Dortmund University
tobias.guggenberger@isst.fraunhofer.de

Marcel Altendeitering
Fraunhofer ISST
TU Dortmund University
marcel.altendeitering@isst.fraunhofer.de

Chris Langdon Schlueter
Peter Drucker School of Management
Claremont Graduate University
chris.langdon@cgu.edu

Abstract

Data products are a hot topic within companies since more and more organizations are implementing data meshes and data product management. Product management is closely related to product quality. The concept of data quality is long established, and there are several notions to measure it. However, these approaches are less practical when data is shared between different domains and organizations with undefined contexts. Such as physical products, data products need a clear definition of what is inside and of what quality they are. With this article, we summarize existing approaches to data quality regarding data products, discuss how a lack of information provision restrains efficient data markets, and provide prescriptive knowledge. To do so, we identify meta-requirements from literature and derive design principles for data scoring systems that cope with information asymmetry for data products to enable efficient data markets.

Keywords: Data Products, data quality, quality scoring, information asymmetry, data market

1. Introduction

Organizations generate a vast amount of data with their operations (IDC, 2020). Like any other business operation, data generation is intended to provide a starting point for value creation. Value creation from data is manifold. From an internal perspective, data is used for business intelligence (Guggenberger et al., 2020; Hasan & Legner, 2023; Hummel et al., 2021). Externally, data can, in principle, be traded like any other good (Jussen et al., 2023). This is also intended by the European Commission, who propose a single European market for data (European Commission, 2021). However, the utilization of data lags expectations and only 0.5% is used (Burn-Murdoch, 2012).

Identifying the right data is like finding a needle in a haystack, nearly impossible. And once found, its quality is not obvious. Data has no well-accepted criteria to measure its fit for specific purposes. While data quality management is not new, it's performed mostly manually, leading to high efforts in every data science project (Langdon & Sikora, 2020). We found practitioners struggle with the definition of quality and even more with agreeing on specific measures. However, 82% of our sample of 65 managers would invest their next dollar in quality (Schlueter Langdon, 2019a, 2019b).

One important aspect of e-business models is to decrease information asymmetry between buyers and sellers to allow buyers to make better-informed decisions. Lower information asymmetry increases the efficiency for creating value (Amit & Zott, 2001). For data, this is a hard challenge. They must be described independently from their context, data models must be shared, and semantics must be made interoperable. Compared to other sectors, data lags product management and established quality and usage measures (Altendeitering & Guggenberger, 2021; Altendeitering et al., 2022; Crosby & Schlueter Langdon, 2019; Hasan & Legner, 2023).

In this article, we emphasize a particular point of data products–its quality. When dealing with data products, this approach is important, as the data owner and consumer may provide incorrect quality information. This might cause the propagation of low-quality data. Thus, measuring quality scores should support organizations in deciding which data product is the right for a specific use case. Concretely, we provide design knowledge for data quality scoring

URI: https://hdl.handle.net/10125/106928
978-0-9981331-7-1

systems (DQS). To do so, we use examples from industry to explain how information asymmetry works and how to cope with it. Nutrition scoring, for example, helps customers choose the right product for their needs. Consumers can decide which market option is more worthwhile by using verifiable information. If a customer has a recipe for cooking, they can choose the product that will help them solve their problem. Like data-engineers should be able to choose data for their algorithms. Implementing DQS helps to reduce information asymmetry and enables efficient markets for data to emerge. Our approach, data quality scoring systems, is one solution for the data economy. Following this, the question we address in this article is *how to design data quality scoring systems that support efficient data markets.*

The rest of this article is structured as follows. First, we review the literature regarding value creation with data products, information asymmetry and its effect on transactions and markets, and data quality and its scoring. Second, we outline the research approach we take and describe how we generate design principles from the literature. Third, we describe the identified meta-requirements and the design principles we derived from them. Finally, we summarize our article, discuss limitations, and give an outlook on fruitful research avenues.

2. Background

The following section summarizes the conceptual background of this article. Thus, we start by discussing value creation with data products. We then continue explaining an important issue of e-business: information asymmetry. Coping with this is paramount to enable efficient business models. One solution is to provide detailed quality scoring, which is the last part of this section.

2.1. Value Creation with Data Products

In the European Union, billions are invested in data infrastructures, such as data spaces. This is good and needed, but these infrastructures are usually used to digitize existing supply chains. Catena-X is an example of this struggle. The data space is intended to connect the whole automotive industry (Catena-X, 2022), but what is beyond? The proposed single market for data requires integration across industry boundaries (European Commission, 2021). If data are goods that are tradable across industrial sectors and create value, we need to discuss data products and the market efficiency.

Data products are a relatively new concept that stems from the data mesh paradigm. Data meshes is an approach for decentralized data architectures (Dehghani, 2022; Hasan & Legner, 2023). However, the idea is not new as the concept of information products arose at the end of the 1990s with the seminal article of Wang (1998) (Hasan & Legner, 2023; Legner & Schemm, 2008). "A data product is a managed artifact that satisfies recurring information needs and creates value through transforming and packaging relevant data elements into consumable form" (Hasan & Legner, 2023). Using data in organizations requires a large effort that is usually conducted by high-skilled workers. Data is mostly stored in an unrefined form, in puddles, not lakes. Making products from this puddle, analyse, and label them with quality scores will reduce uncertainty. Like bottled water compared to the raw product (Langdon & Sikora, 2020; Schlueter Langdon, 2020). Applying product management to data enables organizations to realize business use-cases 90% faster and reduce the manual effort, which is up to 80% of every data science project (Desai et al., 2022a, 2022b; Press, 2016). With data products, data can be organized efficiently and external data products, like analytical ones, can be launched faster (Hasan & Legner, 2023). In conclusion, data must be managed as products to maximize its value (Crosby & Schlueter Langdon, 2019; Desai et al., 2022a, 2022b).

Trading external data products relies on the efficiency of the data market. In the past, several data marketplaces have been found. They have one thing in common: they all failed or did not scale significantly yet and remain in niches (Azcoitia & Laoutaris, 2022). But why? Of course, there might be a plethora of antecedents, such as missing or inefficient pricing models or shortcomings in market analysis. However, in this article, we focus on the economic perspective and discuss two important aspects we defined in the hypotheses:

First, finding the right data products is difficult. It depends on sound descriptions and the communication of the information content of the data. Second, the user needs to be informed about the quality of the data product (Langdon & Sikora, 2020; Schlueter Langdon, 2020). However, the provider of a data product typically has much more information than the consumer does. This conflict is called information asymmetry and will be discussed in the following section.

2.2. Information Asymmetry

Information asymmetry is a problem in every market because sellers usually have more information about the product than buyers (Ackerlof, 1970; Ho & Michaely, 1988; Levin, 2001). This problem is central to the principal-agent theory, which studies the behavior in the relationship between two entities under information asymmetry. Both entities want to maximize the outcome of a relationship, e.g., based on a buying contract. Agents (or sellers) have more information than principals (or buyers), which want to get a fair price (Eisenhardt, 1989).

This imbalance of information results in multiple problems, which are an important research stream of information economics and the value of information. Examples of such problems are adverse selection and moral hazard (Eisenhardt, 1989). Adverse selection causes buyers to go for cheaper products because they cannot assess the quality or other hidden characteristics of high-priced products. So the average quality of the market is decreasing (Ackerlof, 1970; Eisenhardt, 1989).

To cope with information asymmetry, at least two strategies can be identified. First, from the principal's perspective, signaling of information can reduce information asymmetry and help principals to evaluate characteristics, which are usually hidden (Spence, 1973). Examples of signaling are shown in Table 1. Second, screening refers to identifying qualities of goods, individuals, and other entities (Stiglitz, 1975). Especially in digital markets, signaling is crucial and easy because of the technology (Amit & Zott, 2001), but screening is usually not possible. E-businesses for physical products rely on traditional measures and means. However, these do not exist for data products.

2.3. Data Quality Scoring

The concept of data quality is long established and considered an essential building block in data management (Legner et al., 2020). It is usually defined as the 'fitness for use' by data consumers (Wang & Strong, 1996). The fitness for use principle implies two challenges that complicate quality measurements and make quality scores difficult to comprehend. First, data quality is a multi-dimensional concept, including completeness, accuracy, or timeliness, that must be combined into a single score (Altendeitering et al., 2022). Second, the interpretation and importance of quality dimensions vary and depend on the data consumer (Geisler et al., 2021).

Many measurement frameworks tried to address the complexity of data quality assessment (e.g., Pipino et al. (2002) or Wang and Strong (1996)), but the problems prevail. Technological and organizational advancements, including data ecosystems and locally managed data products, emerged in light of digital transformation and data becoming a strategic asset (Hasan & Legner, 2023; Legner et al., 2020). These developments and the establishment of decentralized data architectures exacerbate the complexity of data quality management and the communication of metrics. For instance, data products are often repurposed and consumed long after their creation, leading to outdated data quality information and a potential offering of low-quality data products (Altendeitering et al., 2022; Zhang et al., 2019). As a result, it is imperative to establish a data quality scoring system that is easily comprehensible for data consumers (Holland et al., 2020). Users from different backgrounds should be able to assess the quality of a data set quickly using a single score. This single score combines the quality information from several dimensions and offers a clearly defined and standardized quality score to avoid ambiguity and potential misinterpretations (Altendeitering et al., 2022; Geisler et al., 2021).

Quality scoring systems are well-established services in various industries. For example, in the automotive industry, "J.D. Power's quality scores from the vendor's Initial Quality Study (IQS, problems after three months) and Vehicle Dependability Study (VDS, problems after three years)" (Langdon & Sikora, 2020). Table 1 shows examples of quality scoring systems to cope with information asymmetry in different markets. However, the quality scoring for data is more or less a white spot in research and practice. Against this background, we call for a data quality nutrition label that offers an accessible form to the complexity of data quality and can reduce the information asymmetry in data markets.

3. Methodology

This section outlines the research methodology. The main purpose of this article is the generation of design principles for data scoring systems to enable efficient data markets. Thus, we describe first the background of design principles and then the information gathering based on a structured literature review.

3.1. Design Principle Generation

Design Principles are meta-artifacts that represent a general solution for a class of problems (Baskerville et al., 2018; Iivari, 2015) and, therewith, a starting point for developing design theories (Gregor & Hevner, 2013). They offer codified design knowledge as prescriptive guidelines. Within defined constraints, DPs guide and constrain actions (Baskerville & Pries-Heje, 2010; Chandra et al., 2015; Gregor et al., 2020). Thus, DPs help developers to make the development processes

Table 1. Scoring Concepts and Impact of Signalling on Information Asymmetry (Ackerlof, 1970; Serrano-Cinca & Gutiérrez-Nieto, 2016; Teisl & Levy, 1997).

Scoring Concept	Description and Example	Impact on Information Asymmetry
Nutrition scoring	Given the dietary recommendations, nutrition labels offer consumers an aggregated score based on a product's nutrition information. This way, the label offers a simple way for consumers to make an informed decision about the healthiness of a product.	Nutrition labels help the principal to evaluate the quality of food regarding its healthiness. Several studies confirmed the positive effect on user perception and 'nudging' consumers towards healthier foods (Teisl & Levy, 1997).
Credit Scoring	Credit Scoring Systems are implemented to evaluate borrowers' credibility. The credit score is based on credit scorecards and is used to define the risk or profitability of a population segment. Therewith, lenders can minimize default rates or maximize profitability.	The credit market is facing significant information asymmetry. Lenders are typically not having a complete overview of the financial situation and behavior of borrowers. With a credit scoring system, lenders send quantitative and proofed signals to the borrower.
Quality Scoring for used cars	In his seminal article Ackerlof (1970) emphasize the information asymmetry in the market for used cars. There are two groups: low- and high-quality cars. Sellers have better information about cars and their hidden characteristics. Thus, buyers would make adverse selections because they cannot evaluate the quality and reason for higher prices.	Without quality scoring, the average quality of used cars tends to be lower because the market provides incentives for sellers of "lemons". Quality scoring helps reduce information acquisition costs, incentive higher quality, and finally leads to a higher average market price.
Educational Scoring	The talent and value of a worker can be inherently high, but employers cannot evaluate this for every applicant. Thus, they need to measure them via grades and school reputations. This reduces the information asymmetry and thus the employer's effort for market screening.	Without proper certifications from well-recognized schools, chances for good jobs are low. Certifications are representing excellent signals from applicants to employers.

more efficient (Chandra Kruse et al., 2016; McAdams, 2003). They are recognized as an excellent medium to communicate design knowledge with managers and other experts (Hevner et al., 2004).

Developing DPs is a research strand of its own. For our purpose, the DPs are a necessary part of a larger process. It represents the ultimate starting point of a design science research project for data quality scoring systems. However, this article reflects the information gathering and design preparation stage. We use the process view of DP development suggested by Möller et al. (2023).

Thus, the research approach we are taking here is supportive, as we are using the DPs to prepare the development of a specific artifact using Design Science Research. We use literature and theories as sources for deriving meta-requirements. The DPs are derived in a single iteration based on the knowledge we gathered in advance. Finally, we evaluate the DPs argumentatively.

3.2. Structured Literature Review

We apply a structured literature review to identify meta-requirements for data quality scoring systems. Therefore, we follow the well-established guidelines of vom Brocke et al. (2009) and Webster and Watson (2002). Our research on data quality scoring systems expands upon our previous literature review, specifically focusing on data quality measures and tools (see Altendeitering & Tomczyk, 2022). However, regarding the specific focus of this study, we complemented these findings with a literature review regarding data quality in the context of information asymmetry.

Thus, we choose "Scopus" and "AIS e-Library" to search for relevant articles, because they comprise all relevant journals and conferences in our scope: information systems research, management, and economics. We have used the following search term in Scopus and, comparably, in AISeL:

ALL("Data" OR "Information" AND "Quality") AND ALL ("Information Asymmetry") AND ALL("Scoring")

We limited the scope of relevant topics to, e.g., business, economics, and engineering. With these strings, we identified 110 articles in AISeL and 506 in Scopus. The next step was removing all duplicates and screening the individual articles, first, by their title and, second, by their abstract. Finally, we read every relevant article completely. To exclude articles, we used the following criteria at every stage: focus on DQ measures that are already part of the previous literature reviews, providing only marginal value to the quality scoring systems, too generic discussions, and results that are not complementing the previous reviews. The final step was conducting a forward and backward search (Webster & Watson, 2002). The final set of 21 articles was then analyzed and coded. By aggregating the codes, we could identify four meta-requirements.

4. Meta-Requirements for Data Quality Scoring Systems

In the next section, we present a generic concept for data quality scoring systems. Therefore, we build on the conceptual background we discussed above. To reduce information asymmetry in data markets, we need to consider two things: the quality of data and the information it contains. Table 2 summarizes the meta-requirements that are discussed in the following. We identified four requirements to be considered while designing the artifact of data quality scoring systems.

Meta-Requirement #1: *Quality signals from agents to principals must be increased*

The first meta-requirement (MR) tackles the main reason for implementing such data quality scoring: the provision of a system to minimize the information asymmetry between agents and principals. According to signaling theory, agents need to signal their quality or the quality of their data products to the principal. This is important because of multiple reasons: first, high information asymmetry is associated with lower average market quality and, thus, lower average prices. This results from the fact that customers that are not aware of quality differences and will usually buy cheaper products (Ackerlof, 1970; Rose, 1993).

Second, high information asymmetry is associated with low transaction efficiency. However, decreasing the information asymmetry comes with agency costs. As transaction cost theory suggests, higher efficiency is associated with lower transaction costs (Amit & Zott, 2001; Williamson, 1981). Customers have to pay for signalling, which is an effort to create transparency about the quality of the products in the market. They also have to screen the market and compare different products (Stiglitz, 1975). However, even if these costs cannot be eliminated, they can be reduced through the DQS.

Meta-Requirement #2: *Operational efforts for scoring must be minimized*

The second MR addresses agency costs associated with reducing information asymmetry, i.e., signaling and screening (Jensen & Meckling, 1976; Spence, 1973; Stiglitz, 1975). Data quality scoring relies on the description of information content and the quality of data. This could be achieved through manual work. However, this would not scale. Especially when considering continuously changing data products, e.g., real-time data. In practice, it is important to have quality scoring for every data product and, thus, in every data factory (Langdon & Sikora, 2020).

Thus, it is important to reduce agency costs using a scalable system. This is important for both the agent and the principal. If the solution is trustworthy, principals can rely on the scoring. They can avoid further agency costs for comparing and evaluating data products. The data products' quality and information content is communicated in advance and bad buys can be avoided. Data products are hard to evaluate without knowing the data collection and processing environments. This is also important for enabling principals to create high-qualitative products on these products.

Meta-Requirement #3: *Integratability into existing solutions must be enabled*

The second MR aims to promote the interoperability of data products and the DQS into different data markets and ecosystems. Organizations will probably participate in several data ecosystems to share data with multiple organizations (Jussen et al., 2023). Consequently, the DQS must be integrable into existing data factories and ecosystem technologies. The interoperability of a DQS supports the avoidance of different metadata models and reduces potential information asymmetries between data consumers and providers (Geisler et al., 2021). Achieving this goal requires the DQS to adapt to general and domain-specific standards and incorporate real-time updates to the DQS. This way, data products can be used across data domains, and the costs for adapting the DQS to other contexts and data factories can be reduced.

Meta-Requirement #4: *Context-independent measures must be implemented*

The information asymmetry can vary across different data-sharing scenarios. For instance, within a single data science project, the principal and agent might work within the same organizational unit and have a

Table 2. Overview over Meta-Requirements and Descriptions.

MR	Meta-Requirement	Description
1	Quality signals from agents to principals must be increased	The core requirement of data quality scoring systems is the minimization of information asymmetry to enable efficiency in data markets. Thus, quality signals need to be increased.
2	Operational efforts for scoring must be minimized	Enabling efficient data product provision requires the reduction of human effort because it is the major obstacle for scaling data science projects.
3	Integratability into existing solutions must be enabled	To ensure maximum flexibility in any scenario, data quality scoring systems must be integrable into existing solutions.
4	Context-independent measures must be implemented	To enable use-case independent scoring measures are needed that are not only valid within a specific use-case but across cases and organizational borders.

common understanding of the requirements. However, the information asymmetry in inter-organizational data-sharing scenarios is very high. Often, only data domain experts have the knowledge to evaluate the usefulness of a data set. In contrast, external consumers often lack knowledge about the data context, content, or quality and thus face difficulties evaluating data (Altendeitering et al., 2022).

Independent quality measures are needed by DQS to overcome the problem and decrease information asymmetry. Measures that are commonly agreed upon and easy to comprehend facilitate the creation of high-quality data products.

5. Design Principles for Data Scoring Systems

Building on the meta-requirements for data scoring systems, we present a set of design principles. They enable system designers to develop instantiations of such systems. Figure 1 shows the relationships between the meta-requirements and design principles.

DP1: *Provide the DQS with a component to measure the information content of the data product, so that the service can provide indicator(s) that enable consumers to evaluate the usefulness for their use case.*

The first design principle relates to providing one of the core components of the DQS: the scalable measurement of information content. By generating content signals for agents, this significantly reduces information asymmetry. The principal of a data product transaction can evaluate these signals regarding its needs. Without such an indicator, they must either use a sample of the product or make a blind purchase. Ultimately, this also reduces the effort for both transaction participants. DQS eliminates the need for suppliers to update samples frequently and customers to screen the market constantly. Context independence is essential for the design of a concrete measurement method.

DP2: *Provide the DQS with a component to measure the data quality, so that the service can provide indicator(s) that enable consumers to evaluate the matching for their use case and the effort to use the data product.*

The second design principle relates to providing the second core component of the DQS: the scalable measurement of data quality. This generates the quality signal of the agents. The principal of a data product transaction can evaluate the signals before buying it. Together with DP2, the important signalling in the principal-agent relationship is conducted. Thus, information asymmetry is reduced. Essential for the instantiation of the design principle is the achievement of context independence.

DP3: *Provide the DQS with an interface to data catalogs, so that the service can publish data products including the indicator(s) to enable consumers to easily find data products meeting their requirements.*

The third design principle relates to providing metadata about the data products in catalogs. These catalogs can, for example, be located at data marketplaces (Azcoitia & Laoutaris, 2022) or established as (federated) catalogs in Data Spaces (Otto, 2022). The purpose of the provision is to enable transaction by increasing the discoverability of data products. Making product measures and indicators available is crucial for principals to recognize quality and content signals. Indicators include, e.g., further information about the product, such as pricing or usage conditions.

DP4: *Provide the DQS with automation of quality and information content measurement, so that the service can be integrated into data factories easily to*

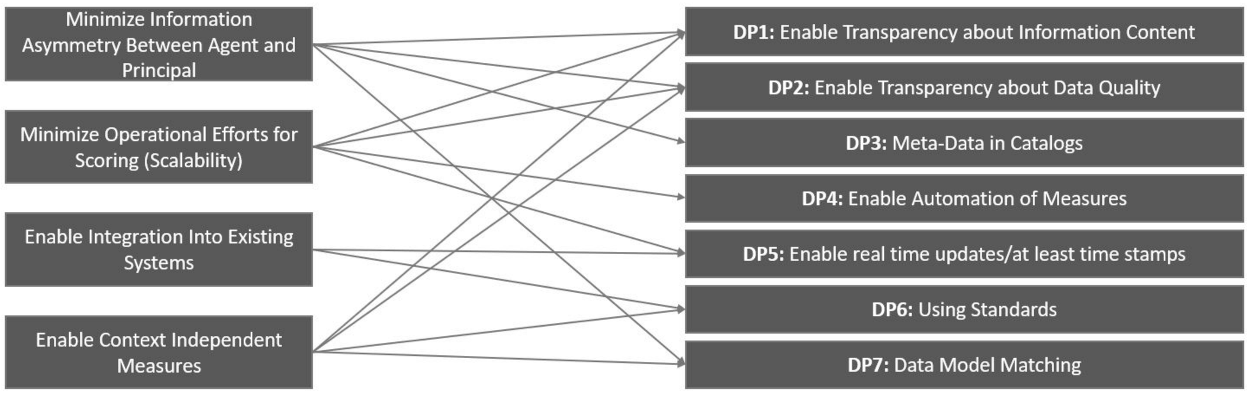

Figure 1. Relations Between Meta-Requirements and Design Principles.

avoid manual effort.

The fourth design principle relates to automated quality measurements to enhance scalability in light of ever-growing and more diversified data sets. A lack of automation would lead to cumbersome manual processes, often accompanied by low data quality coverage (Altendeitering & Guggenberger, 2021). Data quality tools should feature two functionalities to promote automation. First, the use of AI and ML-based algorithms to assess large amounts of data automatically based on established assessment frameworks (e.g., Pipino et al. (2002) or Wang and Strong (1996)). Second, the automated derivation of quality rules or integrity constraints supporting continuous data validation (see also DP5).

DP5: *Provide the DQS with capabilities for real-time measurement of the quality and information content, so that consumers can trust the measures even when the underlying data sources are changing or developing.*

The fifth design principle relates to timely updates of the DQS to create trust and avoid outdated information. Frequently changing data sets (e.g., social media) and a proliferation of digital devices and sensors, producing continuous data streams, raise challenges for scalability and system integration (Gröger, 2021). Both developments call for DQS that can be frequently updated. An up-to-date quality score provides data consumers with timely information and informs them about potential quality issues early on (Swami et al., 2020). This increases the trust in the DQS and can aid in establishing inter-organizational data sharing. At the same time, real-time updates require automated quality measurements to avoid delays and provide timely updates (DP4).

DP6: *Provide the DQS with standardized measures and interfaces, so that consumers can trust the measures even when the underlying data sources are changing or developing.*

The sixth design principle relates to following standardized quality measurements and interfaces to improve context-independent comprehensibility and trust in the DQS. Signals in principal-agent-relationships can be ambiguous because of different backgrounds and varying interpretations of quality scores. The ambiguity makes it challenging to follow up on low DQS and can damage trust in the data. An efficient DQS requires clearly defined quality measures acceptable to all data consumers (Geisler et al., 2021). Moreover, the DQS report should offer explanations for the quality scores and include possibilities to follow up on potential data quality issues (Swami et al., 2020). Establishing data quality standards, such as the Data Quality Vocabulary, can benefit the DQS and simplify system integration.

DP7: *Provide the DQS with a data model matching capability to provide support for consumers to evaluate the usefulness of data products, so that consumers can evaluate the integrability in their use case.*

The seventh design principle relates to the capability of integrating data products into different technological and organizational contexts. Data consumers are from various organizations and industries. This diversity comes with multiple data models and different interpretations of data products. To decrease this complexity, DQS should offer standardized models (DP6). Besides, the DQS should offer functionalities for data model matching and connecting synonymous data attributes (Geisler et al., 2021). This will increase the usefulness of data products for consumers and lower information asymmetry.

6. Summary, Limitations, and Outlook

Data quality scoring is an important aspect of every data transaction inside and across organizations.

Without case-independent knowledge sharing about what is inside the data and of which quality it is, data markets will never be successful. There are no general solutions yet, neither in practice nor in research. However, based on our practical experience in multiple data-sharing projects, we recognized that as an important issue for many use cases. Especially when data is shared, that is not associated with a particular industry or real-world object, information content and its value for a data science project are important measures.

This article provides **contributions to management**. With the design principles, we codified the knowledge regarding data quality scoring that can increase the business value of almost every data science project. Data products are a hot topic within organizations and so is the problem of quality scoring. Thus, managers can use this information to increase operational efficiency in their data science projects and their data organization when implementing DQS.

Scientific contributions of this article are the following: first, the design principles represent codified knowledge regarding data quality scoring systems. With the design principles, the first step towards a design theory is done. Second, building on that, we contribute to the research strand of data quality measures by providing a starting point for innovative thinking of context- and use-case independent measuring and communication of quality and information content of data products.

This research has **limitations**. First, we used empirical data from our own experience as the kick-start of the project. However, the design principles rely on theoretical knowledge collected from research papers. Second, as the design principles are developed a priori, we cannot guarantee that their implementation would lead to successful artifacts. At this stage, they are at a high-level of abstraction. Third, reducing information asymmetry does not come without agency costs. Furthermore, following Levin (2001), reducing information asymmetry has ambiguous effects. It is important to find the right degree of information asymmetry to have a balance between efficiency and agency costs.

Outlook can be seen from two perspectives: coping with limitations and further developing the research objectives. Instantiating the design principles in a real-world solution is the most obvious outlook for further research. To do so, we plan to, first, conduct a second iteration of DP development and enrich the knowledge base with empirical data, e.g., with interviews. Second, implement and validate the design principles empirically. This article provides a starting point for researchers to focus on different aspects of information content and quality scoring. Especially modeling use-case independent quality scores is a white spot in research. Additionally, defining indicators of quality and content and the efficiency of providing respective signals should be a subject of further investigations. We aim to provide a research agenda on these issues.

In **summary**, this article presents a novel class of artifacts coping with information asymmetry for value creation from data products. Data quality scoring systems are needed when data products are shared across organizational borders and should be also used within. Until now, there are no practical examples of such systems nor research articles dealing with this integrated view. Thus, this research can help practitioners in designing such services for their data organizations to make their business models more valuable.

References

Ackerlof, G. A. (1970). The market for ”lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Altendeitering, M., & Guggenberger, T. (2021). Designing data quality tools: Findings from an action design research project at boehringer ingelheim. *ECIS 2021 Research Papers*.

Altendeitering, M., Dübler, S., & Guggenberger, T. (2022). Data quality in data ecosystems: Towards a design theory. *AMCIS 2022 Research Papers*.

Altendeitering, M., & Tomczyk, M. (2022). A functional taxonomy of data quality tools: Insights from science and practice. *Wirtschaftsinformatik 2022 Proceedings*.

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, *22*(6-7), 493–520. https://doi.org/10.1002/smj.187

Azcoitia, S. A., & Laoutaris, N. (2022). A survey of data marketplaces and their business models. *ACM SIGMOD Record*, *51*(3), 18–29. https://doi.org/10.1145/3572751.3572755

Baskerville, R., Baiyere, A., Gergor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Baskerville, R., & Pries-Heje, J. (2010). Explanatory design theory. *Business & Information Systems Engineering*, *2*(5), 271–282. https://doi.org/10.1007/s12599-010-0118-4

Burn-Murdoch, J. (2012). Study: Less than 1% of the world's data is analysed, over 80% is unprotected. https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume

Catena-X. (2022). Catena-x operating model whitepaper: Release v2 - 21.11.2022.

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive knowledge in is research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. *2015 48th Hawaii International Conference on System Sciences*, 4039–4048. https://doi.org/10.1109/HICSS.2015.485

Chandra Kruse, L., Seidel, S., & Purao, S. (2016). Making use of design principles. In J. Parsons, T. Tuunanen, J. Venable, B. Donnellan, M. Helfert, & J. Kenneally (Eds.), *Tackling society's grand challenges with design science* (pp. 37–51). Springer International Publishing. https://doi.org/10.1007/978-3-319-39294-3_3

Crosby, L., & Schlueter Langdon, C. (2019). Data is a product. *Marketing News, American Marketing Association*.

Dehghani, Z. (2022). *Data mesh : Delivering data-driven value at scale*. O'Reilly Media, Incorporated.

Desai, V., Fountaine, T., & Rowshankish, K. (2022a). A better way to put your data to work. *Harvard Business Review*, (July-August).

Desai, V., Fountaine, T., & Rowshankish, K. (2022b). Manage data like a product to unlock full value. *McKinsey*.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57. https://doi.org/10.2307/258191

European Commission. (2021). Commission welcomes political agreement to boost data sharing and support european data spaces. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_6428

Geisler, S., Vidal, M.-E., Cappiello, C., Lóscio, B. F., Gal, A., Jarke, M., Lenzerini, M., Missier, P., Otto, B., Paja, E., et al. (2021). Knowledge-driven data ecosystems toward data transparency. *ACM Journal of Data and Information Quality (JDIQ)*, *14*(1), 1–12. https://doi.org/10.1145/3467022

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Gregor, S., Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information systems*, *21*, 1622–1652. https://doi.org/10.17705/1jais.00649

Gröger, C. (2021). There is no ai without data. *Communications of the ACM*, *64*(11), 98–108. https://doi.org/10.1145/3448247

Guggenberger, T. M., Möller, F., Boualouch, K., & Otto, B. (2020). Towards a unifying understanding of digital business models. *Proceedings of the Twenty-Third Pacific Asia Conference on Information Systems*.

Hasan, M., & Legner, C. (2023). Understanding data products: Motivations, definition, and categories. *Thirty-first European Conference on Information Systems (ECIS 2023)*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Ho, T. S. Y., & Michaely, R. (1988). Information quality and market efficiency. *The Journal of Financial and Quantitative Analysis*, *23*(1), 53. https://doi.org/10.2307/2331024

Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2020). The dataset nutrition label. *Data Protection and Privacy*, *12*(12), 1.

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1), 205395172098201. https://doi.org/10.1177/2053951720982012

IDC. (2020). Worldwide global datasphere forecast, 2020–2024: The covid-19 data bump and the future of data growth.

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, *24*(1), 107–115. https://doi.org/10.1057/ejis.2013.35

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Jussen, I., Schweihoff, J., & Möller, F. (2023). Data sharing funda-mentals: Characteristics and definition. *Proceedings of the 56th Hawaii International Conference on System Sciences*.

Langdon, C. S., & Sikora, R. (2020). Creating a data factory for data products. In *Lecture notes in business information processing* (pp. 43–55). Springer International Publishing. https://doi.org/10.1007/978-3-030-67781-7_5

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating design knowledge with reference models: Insights from 12 years' research into data management. *Journal of the Association for Information Systems*, *21*(3), 2. https://doi.org/10.17705/1jais.00618

Legner, C., & Schemm, J. (2008). Toward the inter-organizational product information supply chain – evidence from the retail and consumer goods industries. *Journal of the Association for Information Systems*, *9*(4), 119–150. https://doi.org/10.17705/1jais.00156

Levin, J. (2001). Information and the market for lemons. *The RAND Journal of Economics*, *32*(4), 657. https://doi.org/10.2307/2696386

McAdams, D. (2003). Identification and codification of principles for functional tolerance design. *Journal of Engineering Design*, *14*(3), 355–375. https://doi.org/10.1080/0954482031000091095

Möller, F., Guggenberger, T., & Otto, B. (2023). Towards a method for design principle development in information systems. *Proceedings of the 15th International Conference on Design Science Research in Information Systems and Technology*.

Otto, B. (2022). A federated infrastructure for european data spaces. *Communications of the ACM*, *65*(4), 44–45. https://doi.org/10.1145/3512341

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Commun. ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010

Press, G. (2016). Cleaning big data: Most time-consuming, least enjoyable data science task, survey says.

Rose, C. (1993). Equilibrium and adverse selection. *The RAND Journal of Economics*, *24*(4), 559. https://doi.org/10.2307/2555745

Schlueter Langdon, C. (2019a). Data: How to measure it? https://dih.telekom.com/en/dataquantity/

Schlueter Langdon, C. (2019b). Data: Quantity or quality? https://dih.telekom.com/en/quantity-vs-quality/

Schlueter Langdon, C. (2020). Data factories for data products: Dih. https://dih.telekom.com/en/datafactory/

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending. *Decision Support Systems*, *89*, 113–122. https://doi.org/10.1016/j.dss.2016.06.014

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355. https://doi.org/10.2307/1882010

Stiglitz, J. E. (1975). The theory of "screening," education, and the distribution of income. *The American Economic Review*, *65*(3), 355.

Swami, A., Vasudevan, S., & Huyn, J. (2020). Data sentinel: A declarative production-scale data validation platform. *36th International Conference on Data Engineering (ICDE)*, 1579–1590. https://doi.org/10.1109/ICDE48307.2020.00140

Teisl, M. F., & Levy, A. S. (1997). Does nutrition labeling lead to healthier eating? *Journal of Food Distribution Research*, *28*(856-2016-56991), 18–27.

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., R., P., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *Proceedings of the 17th European Conference on Information Systems*.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65. https://doi.org/10.1145/269012.269022

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*, xiii–xxiii.

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American Journal of Sociology*, *87*(3), 548–577. https://doi.org/10.1086/227496

Zhang, R., Indulska, M., & Sadiq, S. (2019). Discovering data quality problems: The case of repurposed data. *Business & Information Systems Engineering*, *61*, 575–593. https://doi.org/10.1007/s12599-019-00608-0