

Don't Buy A Pig In A Poke A Framework for Checking Consumer Requirements In A Data Marketplace

Sebastian Lawrenz
Clausthal University of Technology
sebastian.lawrenz@tu-clausthal.de

Andreas Rausch
Clausthal University of Technology
andreas.rausch@tu-clausthal.de

Abstract

The 21st century is the century of data. Data is the new oil and powers many new applications and business models. The increasing value of data leads to the fact requires new possibilities and methods to trade with "data" as a trading good. One of these possibilities is a data marketplace, a digital e-commerce platform like eBay or Amazon, but specifically targeting data, information, and datasets. Since, other than physical trading goods, data is not so easy to verify, a framework is required to characterize data and to validate data against user requirements. This leads to a conflict between security and privacy requirements of data providers and the necessity of checking the data consumer requirements. In this paper, we discuss the specific challenges of data trading and introduce a framework to help users check their specific requirements (Data Quality) as a reference point for a purchase decision.

1. Introduction

How can we trade something we cannot touch or see physically? Data and statistics are key components of innovation in our data-centric world [1]. Data builds the foundation for AI-based algorithms and provides the basis for informed decisions. Furthermore, due to the Internet, smart devices, and the Internet of Things, it has never been easier to collect data.

1.1. Problem Statement and Objective

Still, there is a gap between *data providers*, who produce and collect data, and *data consumers*. Data consumers need specific data, which they cannot produce or collect themselves, such as a training dataset for an AI-based system. Data providers own data, which they want to share or sell. A bridge between these two stakeholders is a *data marketplace*. A data marketplace is a platform on which data from a variety of sources is

collected, aggregated, processed, enriched, bought, and sold [2].

There are already professional data marketplaces online, such as the German "mobilitätsdatenmarktplatz" *mdm*, *streamr*, and *advaneo*. Yet, the biggest difference between data and other trading goods, like cars or books, is that one cannot test a dataset without buying it. The data itself constitutes the actual value and by showing it before making a deal, this value might decrease. The value of data depends on its novelty, which means an unknown dataset has a high value, while the value of a known dataset tends towards zero. This can be illustrated by a simple example: Assume, we have two datasets: one with the lottery numbers from yesterday, and another one with the lottery numbers for next week. Of course, nobody would spend money for the old data, while the future data would be of incredible value (as long as it is true).

Besides the question of whether the data is true or not (the integrity of the dataset), a data consumer always has to deal with one key question: *Does a particular dataset really fit the requirements?* Back to the lottery example above: Are the numbers for the German lottery, or for another country? If this question cannot be answered, the data consumer would buy a pig in a poke, or just play again in a lottery. The saying "to buy a pig on poke" means that something is sold or bought without the buyer knows its true value, especially when buying without inspecting the item before.

The objective of this paper is to develop a framework that helps check data quality requirements considering the requirements of the stakeholders in the context of a data marketplace.

1.2. Research Methodology

The research presented in this paper has been conducted using the *Design Science* methodology. Instead of formulating hypotheses, Design Science concentrates on developing an artifact in a context [3]. Artifacts are, for example, methods, techniques, notations, and algorithms used in software systems [3].

Specifically, we use the three-cycle view as proposed by Hevner [4]. Hevner introduces three closely related cycles of activities. First, the *Relevance Cycle* analyses the environment and the specific requirements of the stakeholders. Second, the *Rigor Cycle* provides domain knowledge and, finally, the *Design Cycle* supports designing the new artifact [4]. In Section 1.4, we provide a scenario, which is aligned with the Design Science methodology and that we use to present our framework.

1.3. Contribution

In this paper, we propose a framework for checking consumer requirements in a data marketplace. The Design Science approach is used to understand the data marketplace application domain and the stakeholders involved. Based on the state-of-the-art, related domains (e.g., books or movies), and the requirements, we design the framework and evaluate it using the scenario described in Section 1.4.

1.4. Scenario

To provide a better understanding of the challenges of data marketplaces, we introduce a small scenario. In this scenario, we have one data consumer with a specific need for a dataset. The dataset should contain customers that are receptive to advertising. Our data consumer has two requirements:

- 1.) The data consumer expects at least 4,000 potential customers over 18 years.
- 2.) Every customer over 18 years has an e-mail address.

Further, we have a data provider that owns a dataset with three columns (*Name, Age, E-Mail Address*) and 5,000 entries. Our data provider is willing to sell this dataset for a price x . An important precondition in our scenario is the fact, that there is no previous relationship between the data consumer and the data provider.

1.5. Outline

The remainder of the paper is organized as follows: Section 2 presents the context, relevant background, and refines the initial problem statement. In Section 3, we analyze related domains with similar challenges. Section 4 introduces our framework for which Section 5 provides a proof of the concept. Section 6 discusses the limitations of our solution, before we conclude the paper in Section 7.

2. Context and Background

The *Design Science* approach focusses on iteratively developing an artifact according to the requirements. This section analyzes the minimum requirements of different stakeholders based on the scenario introduced in Section 1.4, and focusses on *Relevance Cycle*. Specifically, we discuss the relationship between a data provider and data consumer, derive the minimum set of requirements based on our previous work and the current state-of-the-art, and we introduce the relevant background for the rest of the paper. This section is concluded with the refined problem statement that drives our research.

2.1. Trust Relationship between a Data Provider and a Data Consumer

As stated in Section 1.4, we require no previous relationship between the data provider and the data consumer present, which means not knowing each other implies a lack of *trust*. Defining trust is a difficult task. McKnight and Chervany propose two kinds of trust typologies, a classification system for types of trust and definitions of six related trust types [5]: Structural, dispositional, attitude, feeling, expectancy, belief, and intention. Regardless of an exact definition of the word “trust”, this term describes a relationship between two parties. And it is quite obvious, that we unlikely trade with someone if there is not at least a minimum trust. In Economics, trust means *expectation upon a risky action under uncertainty and ignorance based on the calculated incentives for the action* [6]. This definition supports our assumption that trust is a precondition for a trading scenario, which is also supported by [7], [8], and [9].

2.2. Data Provider

A data provider is a stakeholder that owns a dataset and is willing to sell it for a specific price x . In a market structure, the data provider is the *seller*. The data provider requires access to a marketplace as a trading infrastructure. Until the dataset is sold, the dataset needs protection, since the data is the value-creating element. The value is characterized by the level of surprise, or the information content itself [10], [11] (cf. Section 2.5). Based on this information, we conclude that the provider is interested to hide the content of his dataset until the dataset is sold.

Based on our previous work, our scenario, and the state-of-the-art, the data provider’s requirements can be summarized as follows [12], [13], [14], [15], [16], [17]:

- Data security guarantee: The data provider wants to protect the information content of the dataset until a trade is completed.
- System safety guarantee: The data provider expects a reliable marketplace environment.
- Process reliability: The data provider wants to receive the money safely and wants to ensure that the sold record is only sent to the buyer.

Furthermore, another challenge (depending on the knowledge and experience of the data provider) is the question, how to characterize and describe the dataset in the offer [13] (we consider this challenge, but it is not the focus of this paper).

2.3. Data Consumer

Based on our scenario, the data consumer is a stakeholder with a need for a specific dataset. In a market structure, the data consumer is the *buyer*. The data consumer is supposed to have no previous relation with the data provider. Consumers trust, for example, the credibility, compatibility, and expertise [18] of providers, but all this is not present in our scenario.

If a personal relation justifying a certain level of trust is not present, the data consumer can trust the platform (the marketplace). However, in most cases, the platform is just a broker providing trust in the trading process. Trust in the dataset (the trading good), is still missing.

Therefore, the data consumer needs a mechanism to check the quality of the dataset, such that the risks coming along with an unknown dataset can be minimized and that the novelty of the data package content can be assessed. The resulting requirements can be summarized as follows, based on [12], [13], [14]:

- System Safety guarantee: The data consumer expects a reliable marketplace.
- Process reliability: The data consumer wants to receive a bought dataset safely and wants to be sure that the money is transferred to the consumer
- Data quality: The data consumer expects to buy a high-quality dataset, without incorrect or missing values and without fake data (*data integrity*) (see Section 2.6)
- Purchase decision support: The data consumer needs a method to estimate the investment risk. *Is it worth to invest in this dataset?*

The latter causes the data consumer to avoid buying the proverbial “*Pig in a Poke*”. Therefore, we recall the data consumers’ requirements (Section 1.4): a minimum of 4,000 customers over 18 years and every customer

over 18 years has an e-mail address. Furthermore, we have a new requirement: The data consumer expects a security guarantee for meta requests, motivated by the fact that a second person could conclude the data consumer’s (business) idea from marketplace queries.

2.4. Data Marketplace as a Bridge

If trust is missing between the data provider and the data consumer, other options for establishing a relation of trust are necessary. For this, norms, regulations, laws, and contracts are often used to establish such a relation between the parties [19], [20]. Pavlou and Gefen [21] showed that an online marketplace can provide a base for trust between buyers and sellers. Instead of trusting each other, the stakeholders can trust the platform. For this, in previous work, we propose a data marketplace as a secure platform for data trading. A data marketplace provides a legal framework for data trading, as well as accompanying regulations, laws, and contracts.

2.5. The Value of Datasets and the Consumers Purchase Decision Process

Data trading is challenging. Data is a kind of digital goods, which are distinguished from other goods by six characteristics: Digital goods are *non-rival*, *infinitely expandible*, *discrete*, *aspatial*, and *recombinant* [22]. A key challenge in data trading is to define the value of datasets. Shannon [10] quantified the news content of information by the *level of surprise* for a particular outcome. If the data provider can estimate this value for a dataset, this estimate can be used as an indicator for pricing, which underlines the provider’s requirement to protect the content of a dataset.

There are various models for the purchase-decision process of consumers available, such as the work of Horward and Sheth [23], Ehrenberg, and Nicosia [24], and Engel et al. [25]. These models are pretty detailed, however, we use a simplified, well-established model by Kotler [26] that describes the five stages of a consumer buying process:

- 1.) Problem recognition
- 2.) Information search
- 3.) Evaluation of alternatives
- 4.) Purchase decision
- 5.) Postpurchase behavior

Kotler’s model begins with the problem recognition. In our scenario, the basic demand is a need for data for advertisement purposes. The second step is the search for information and the evaluation of the alternatives (if applicable). Based on the evaluation results, the data consumer makes a purchase decision, followed by the

postpurchase behavior, which describes the satisfaction with the product.

In the *Decision Theory*, the value of information (similar use for data) is calculated by the difference between the expected outcome of a decision with and without information [27]. Here, an information paradox occurs. *How does a data consumer predict the outcome of a decision with information, without knowing the exact information content?* This question shows the conflict between the data providers and the data consumers. The data provider wants to share as little information about the dataset as possible to not decrease the value of the dataset, and the data consumer wants to know as much as possible about the dataset before buying it to make the right purchase decision.

2.6. Data Quality

An important indicator for the evaluation of alternatives and the purchase decision is the quality of a product, notably, data and information quality. *But what does quality mean?* In general, the word “quality” must not include a rating (good, average, bad), yet, people usually associate quality with a rating, e.g., “good” or “bad” quality. Garvin identified five basic approaches to define quality [28]:

- Transcendent: Quality can be understood in mind but cannot be explained verbally. Accordingly, quality is like love or beauty.
- Product-based: Quality is defined as a set of precise and measurable variables, which reflect the attributes of a product.
- User-based: Quality is considered personal and depends on the users. Quality is defined by the satisfaction of user preferences.
- Manufacturing-based: Quality is concerned with specific manufacturer requirements.
- Value-based: Quality is defined in terms of cost and prices. High quality is expressed by an optimal price-service-ratio.

Since it is difficult to define quality by one definition, Garvin described quality in eight dimensions [28]: Performance, features, reliability, conformance, durability, serviceability, aesthetics, and perceived quality. For data quality, Wong et al. identified further dimensions, e.g., actuality, completeness, and integrity [29], [30], [31].

Although not all these dimensions are in the scope of this paper, we want to briefly discuss the dimension of *data integrity*. Data integrity refers to the source of a dataset and describes the credibility of the dataset. For example, in our scenario, a high level of data integrity indicates all the customer data being real and not fake.

There exist metrics and labels for high data quality standards, which are good approaches for “product”- and “manufactures”-based quality views. Nevertheless, quality is requirement-driven, i.e., data quality criteria depending on the specific stakeholder requirements. In our scenario, this specific stakeholder is the *data consumer*. For the marketplace and the data provider, the motivation of a data consumer to buy a dataset is unknown and, therefore, if at all, hard to generalize. Motivated by a multitude of reasons for buying a dataset and the fact that we cannot generalize the data consumer requirements, we define data quality as follows:

Data quality describes the value of a dataset from the data consumer’s point of view with regards to the requirements. High data quality means a good match of the requirements and the characteristics of the dataset.

2.7. Problem Statement

So far, we analyzed the data marketplace domain, and we derived the minimum set of requirements for our framework. Table 1 summarizes these requirements.

Table 1: Overview of the requirements for a data marketplace

Requirement	Data Provider	Data Consumer
Data Privacy	Privacy of his dataset	Privacy of his specific requirements
System safety	Reliable marketplace	Reliable marketplace
Process reliability	Receives the payment	Receives the dataset
Assistance	Assistance with describing the dataset	Decision support method

Both, the data provider and the data consumer, expect a reliable and secure system. Furthermore, the privacy of the dataset and the specific requirements have a high priority. For supporting the matching data consumers and providers, information transparency is required, which conflicts with privacy requirements. In an ideal (data) marketplace, market transparency would be 100%. That is, a data consumer would see the whole dataset and the data provider would know the specific requirements of the consumer, which would also lead to a fair pricing model.

Vision: In an ideal data marketplace, complete information transparency is present, which enables data consumers to reduce the risk of their purchase decision.

Problem: A complete product transparency (dataset) is not possible, since the disclosure of the dataset will

reveal the information content. The data provider wants to protect the product and the data consumer does not want to disclose the requirements.

Method: To address this above issue, a framework is required that allows for comparing the data consumer requirements with the dataset. This includes not only the minimum set of requirements, e.g., completeness and uniformity, but also research-based requirements, like content features. According to the privacy requirements of the stakeholders, the framework must ensure that the requirements and the dataset are kept inside the framework.

3. Evaluation of Related Domains

Before designing the framework, the *Rigor Cycle* provides domain knowledge. Transparency as well as objective quality criteria are general challenges in the area of trading digital goods. Therefore, in this section, we analyze related products and how they handle these challenges.

Related products are for example *books*, *movies*, and *shows*. From an abstract point of view, such products are also characterized by the information content. To advertise *movies* and *shows*, film scenes are cut together for a trailer, which is supposed to make the consumer “want more” [32]. These critics that have seen trailers before the release, recommendations of friends’ that already saw the movie, and the reputation of directors, actors, and so on, are also reasons for watching or buying a movie [33].

Books follow a similar model. They also live from the reputation of the author, from critics and samples. Furthermore, in a bookstore, the clients can have a look into the book and read it for a while before buying it.

In contrast to data and information, these examples are products for a consumer goods market. This allows the provider to set the cost lower per product thus reducing the investment risk for the consumer. *It is easier to buy a movie for 10 €, which I maybe like, instead of a dataset for 1,000 €, which is maybe useful.*

Data and information packages are usually products for a business-to-business (B2B) market. There are less consumers and less producers, which leads to higher prices and to the fact that contracts are only concluded with high collateral. Consumers as well as the providers are not willing to take a risk here.

Purchase indicators, such as recommendations do not have the same effect as in the consumer goods market. Neither can data consumers have a look into the dataset before the purchase, as this will reduce the value of the dataset itself. One working indicator could be the reputation of the data provider, if the provider has already achieved a good reputation.

Some professional data exchange platforms and marketplaces are already established. One of the biggest marketplaces is *kaggle.com* – a free dataset exchange platform supported by Google. Its main purpose is the organization of data science competitions as well as research in this area. The datasets are for free and *kaggle* does not have any quality criteria, but a standard model for the data description and metadata, such as usage information (license and visibility), provenance (source and collection methodology), maintainers (Dataset owner), updates and data coverage.

Few blockchain-based approaches are available or in development, e.g., the *iota data marketplace* or *streamr*. The purpose of these platforms is subscribing to real-time data like sensor data. Another platform is *datarade.ai*, which, however, only supports matching, i.e., to bring together data providers and data consumers. It is not possible to buy datasets.

All these platforms do not offer a decision-support model for consumers. Furthermore, we can conclude that the indicators from other areas (books, movies) cannot be transferred easily to data trading. Since the value of a dataset is described by the information content itself, we cannot offer a consumer the possibility to have a deeper look into the dataset before buying it. Therefore, for establishing a data marketplace, new solutions are required.

4. Overall Approach

The main goal of our research is to develop a framework that supports data consumers to reduce the risk of investing in a dataset. Our *Design Cycle* is driven by the question:

How can we enable a fair data-trading process considering the requirements of the data provider and the data consumer?

Figure 1 shows the abstracted architecture of our data marketplace as the central platform for data trading and the interaction with the data provider who owns a Dataset DS and the data consumer. Our framework meets the requirements of Table 1. Furthermore, we use the scenario from Section 1.4. to explain the structure and the behavior.

4.1. Structure

Our data marketplace consists of the following components. The *Offer* class contains all offered datasets, including price and description. The *Search* component helps the data consumer find offers in the marketplace. Every offer consists of two main parts:

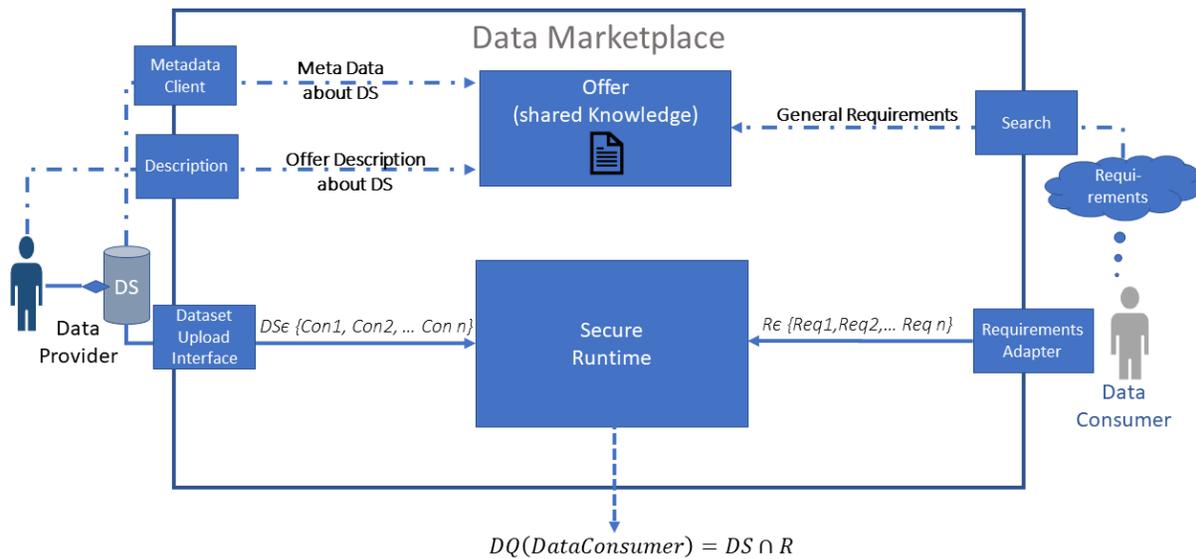


Figure 1: Logical View of the data marketplace

- **Metadata:** The metadata is automatically generated by a client that the data provider can download and run on his device. The *Metadata Client* generates meta-information such as size, the names of the columns, and the number of values per column. This information gives an objective overview of the content of the dataset and makes that data discoverable by others.
- **Description:** The data provider has the chance to write a description of the dataset as a free text. He can give further information about the dataset, such as the main content, the source, and all potentially relevant information.

The *Secure Runtime* represents a closed area in the marketplace itself, where neither data providers nor data consumers have access. The runtime is used to compare the data consumer requirements with the dataset. It defines an interface to the data provider (*Dataset Upload Interface*) and to the data consumer (*Requirements Adapter*).

4.2. Behavior

Every Stakeholder has access to the *Offer (shared knowledge)*. This is the first orientation guide in our marketplace. The data consumer searches over all the available offers and chooses the ones, which could fit the *general requirements*. If the shared knowledge does not provide enough information to make the purchase decision, the data consumer can use the *Requirements Adapter* (see above). Based on predicate logic, the *Requirements Adapter* translates the data consumer

requirements into formal requirements, which are sent to the *Secure Runtime*. The data provider can upload a dataset as well, by using the *Dataset Upload Interface*. The secure runtime computes the matching of the specific requirements with the dataset as follows:

- $DS \in \{Con1, Con2, \dots, Con n\}$
DS: Dataset; Con=Condition
- $Re \in \{Req1, Req2, \dots, Req n\}$
R: Set of Requirements; Req: Requirement
- $DQ(DataConsumer) = DS \cap R$
DQ: Data Quality

Based on the outcome *DQ*, the data consumer can make a final purchase decision. This process is also shown in Figure 2 and will be explained in more detail in Section 5.

4.3. Summary

Our framework supports the data provider as well as the data consumer and contains three main elements. The *Metadata Client* (1) analyzes datasets and create a meta-information objectified. The *Requirements Adapter* (2) enables the data consumer to formulate the specific requirements for a dataset and the *Secure Runtime* (3) provides a neutral zone in which algorithms can check the specific requirements for a dataset.

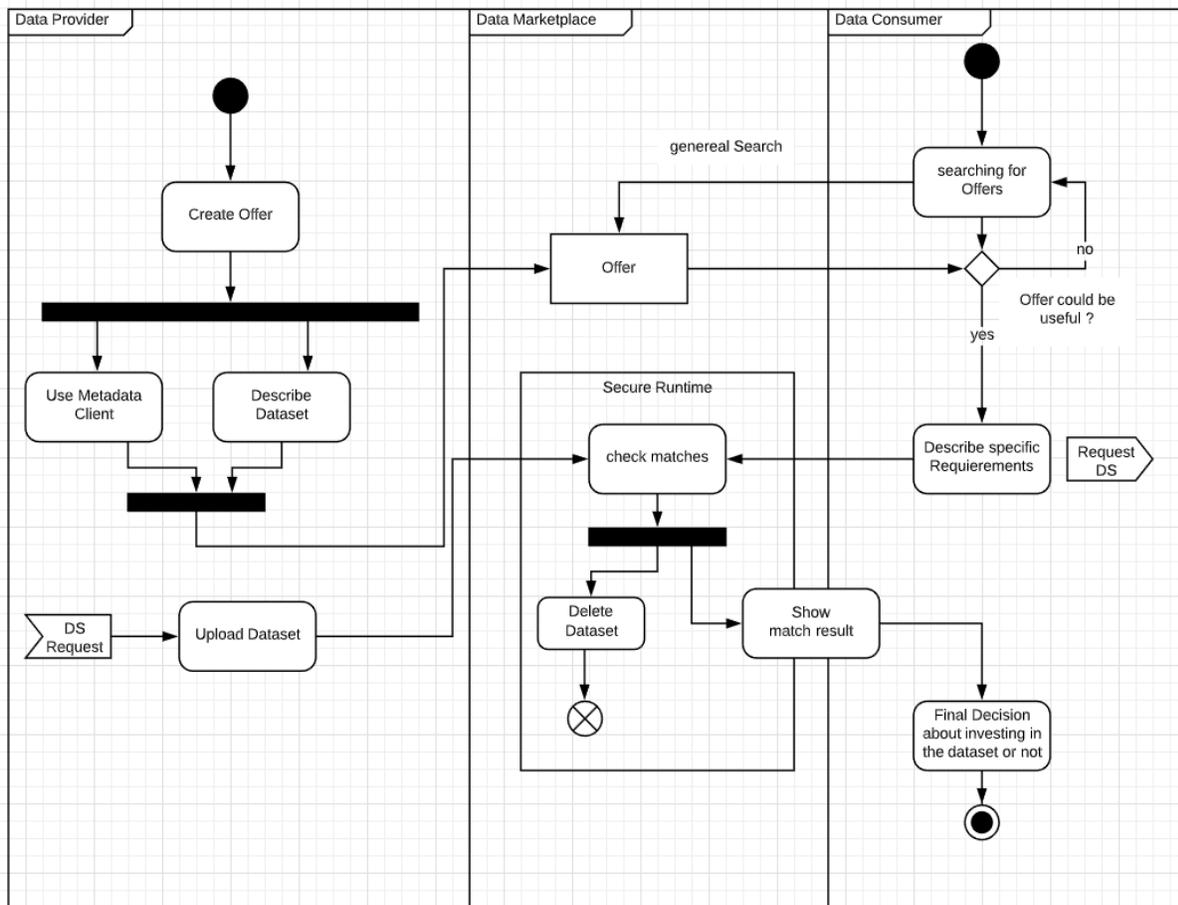


Figure 2: Process view of the decision framework

5. Proof of Concept and Implementation

This section shows the current status of the implementation of our framework and evaluates the current artifact. Even though the framework is still work in progress, a technical proof of concept is already implemented. Therefore, we use our scenario from Section 1.4 (which is not complying with the general data protection regulation).

5.1. Implementation

For testing our scenario we created a dataset with 5,000 customers. The dataset consists of three columns: *Name*, *Age*, and *E-Mail Adress* with randomly created values (obviously our dataset won't fulfill the data integrity criteria). After creating the instance of our offer, it was added to our data marketplace (Figure 3).

Unfortunately, the data provider did not write any information in the offer (shared knowledge) regarding the data distribution of the variable *Age*. Furthermore, the metadata client already found approximately 450

missing values in the *E-Mail Adress* column. The data consumer has now to decide if he is willing to buy the dataset, or if he would buy *the pig in the poke*.

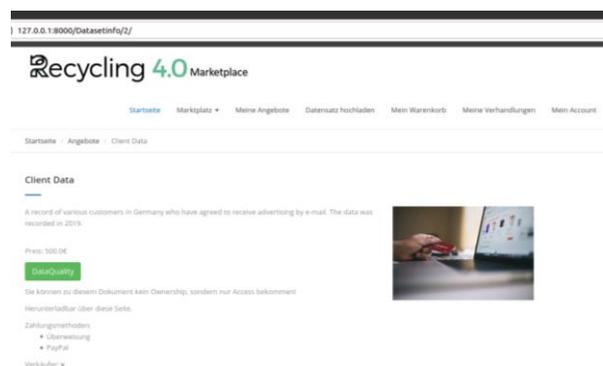


Figure 3: Screenshot from the example offer

The sequence for the complete process is shown in Figure 2 starting with the creation of the offer. The data provider uses the *Metadata Client* and describes the offer. The client's output is shown in Figure 4, and identified the three columns *name*, *mail*, *age*, 5,001 rows, and three columns. The client detected 9% of

```
{ "sha256" : "da7ed6882efc89e1ca2504796d1a38add2fba87576c0f51ecf79f68d2f5d0675", "name" :
"advertisingClient.csv", "size" : "140289", "noRows" : "5001", "noCols" : "3", "cols" :
"[Name,Mail,Age]", "perNullVal" : "0.09", "percStringFields" : "66.57", "percNumberFields"
: "33.31", "noCharacters" : "140286" }
```

Figure 4: Metadata Client Output

missing values (figure 4 perNullVal), i.e., the 450 missing values.

The data consumer starts now searching for offers and finds our offer with the customer data. From the description, the offer could fulfill the need. However, to avoid buying the pig in the poke, the data consumer uses the requirements adapter and describes requirement 1 (4,000 customers over 18 years) and requirement 2 (Every customer over 18 years has an e-mail address). Therefore, the data consumer has to describe the requirements as follows:

<Subject> <Predicate>.

The description is in the form of predicate logic. The subject is one of the columns (*name, mail, or age*) and the predicate is a function $P: X \rightarrow \{true, false\}$ [34].

5.2. Evaluation

In our example for the first requirement the data consumer formulates the predicates as follows:

R1:= <age> <4000 rows greater 17>
R2:=<mail> <!= null>

Furthermore, the data consumer can connect both logic queries with logical *AND*: $R1 \wedge R2$. These queries are implemented in Prolog. The response, to these queries, will then be *true* in case the requirements are fulfilled, or *false* otherwise. The execution happens in the secure runtime, and cannot be accessed by the data provider or the data consumer. The dataset and the consumer's requirements will be deleted after the execution. Based on the result (with our example dataset *true*), the data consumer can make the final purchase decision.

6. Discussion and Limitations

In the previous section, we demonstrated the use case of our scenario. With the current state of the implementation, we can check every specific consumer requirement, as long as the consumer can formulate it in logic. However, there are still some limitations for the data consumer as well as for the data provider.

First: The response from the secure runtime to our queries: *DQ* can only contain two values {0 (*false*) or 1 (*true*)}. It limits the data consumer to the opportunity to

check if the dataset fits the requirements for 100%. *But what is about datasets that may fit some 90%?* The current state of the implementation does not yet support these cases. If the data provider has several offers (that do not fulfill the 100%-requirement), the data consumer cannot select the best-matching offer. This issue will be solved in future implementations by offering the data consumers the option for a weighting/ranking. For example, in our scenario, this would mean that R1 has a weight of 0,7 (70%) and R2 0,3 (30%). If R1 is fulfilled, but R2 is not, then *DQ* would be still 0,7.

Second: The current scenario is strongly focused on the data consumer and it is assumed that the data consumer is not abusing the framework. Right now, the data consumer can write as many queries as possible. It opens, for example, the possibility to guess the content of the dataset without buying. With logical queries such as checking is the age in a raw is exactly 18 or 19 or 20 and so on. The data consumer can get knowledge about the dataset and the information value and accordingly the value for the data provider decreases. For avoiding this issue, in the future, we are planning to introduce a pricing model for the queries. The pricing model has to be based on the information content of the dataset and the maximum of information a data consumer could obtain per request. Furthermore, the pricing model should not be too expensive.

7. Conclusion

The objective of this paper was to design a framework for checking consumer requirements in a data marketplace. Based on the design science approach, we first analyzed the domain of data marketplaces, deducted the requirements of the involved stakeholders, and have narrowed down the problem, that a framework for checking consumer requirements is still missing.

In summary, we identified the data provider who is willing to sell a dataset but wants to protect the information content of this dataset, until it is sold and the data consumer who wants avoiding to buy a pig in a poke. Therefore, the data consumer needs assistance with the purchase decision. We evaluated related domains, such as books, or movies, and had a closer look at other data marketplaces. Based on our findings we designed a framework that allows a data consumer to check his requirements against the dataset, even without seeing it. Our Framework contains right now

three main elements to support the data provider and the data consumer.

We demonstrated and evaluated our implementation using our scenario. Nevertheless, there are a couple of limitations in the status quo, such as the limited value range for $DQ \{0, 1\}$, and the possibility to abuse the framework against the provider. Therefore, we already sketched up possible solutions to overcome these issues. The implementation of these solutions and the evaluation against a larger test group are the next steps of our project.

8. Acknowledgment

This paper evolved from the research project Recycling 4.0 (Digitalization as the Key to the Advanced Circular Economy using the Example of Innovative Vehicle Systems) which is funded by the European Regional Development Fund (EFRE | ZW 6-85017703, ZW 6-85017297) and managed by the Project Management Agency NBank, Germany.

Special thanks go to Marco Kuhrmann and Arthur Strasser for their helpful feedback! Further, the authors would like to thank Sona Otaryan, Vera Stein, and Robert Werner for their help with the implementation of our working prototype.

9. References

- [1] S. G. Pantula, "Statistics: A key to innovation in a data-centric world!," *J. Am. Stat. Assoc.*, vol. 106, no. 493, pp. 1–5, 2011.
- [2] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," *Vietnam J. Comput. Sci.*, vol. 3, no. 3, pp. 137–143, 2016.
- [3] R. J. Wieringa, *Design science methodology: For information systems and software engineering*. 2014.
- [4] R. Hevner Alan, "A Three Cycle View of Design Science Research," *Scand. J. Inf. Syst.*, vol. 19, no. 2, pp. 87–92, 2007.
- [5] Mcknight D. Harrison and N. L. Chervany, "The Meanings of Trust," no. January 1996, 1996.
- [6] H. S. James, "The trust paradox: A survey of economic inquiries into the nature of trust and trustworthiness," *J. Econ. Behav. Organ.*, vol. 47, no. 3, pp. 291–307, 2002.
- [7] F. A. G. Den Butter and R. H. J. Mosch, "Trade, trust and transaction costs," 2003.
- [8] R. Bunduchi, "Business relationships in internet-based electronic markets: The role of goodwill trust and transaction costs," *Inf. Syst. J.*, vol. 15, no. 4, pp. 321–341, 2005.
- [9] P. M. Doney and J. P. Cannon, "An examination of the nature of trust in buyer-seller relationships," *J. Mark.*, vol. 61, no. 2, pp. 35–51, 1997.
- [10] C. E. Shannon and W. Weaver, "THE

MATHEMATICAL THEORY OF COMMUNICATION," 1949.

- [11] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Res.*, vol. 49, no. 10, pp. 1295–1306, 2009.
- [12] S. Lawrenz, P. Sharma, and A. Rausch, "Blockchain Technology As an Approach for Data Marketplaces," in *Proceedings of the 2019 International Conference on Blockchain Technology*, 2019, pp. 55–59.
- [13] S. Lawrenz, P. Sharma, and A. Rausch, "The Significant Role of Metadata for Data Marketplaces," *Int. Conf. Dublin Core Metadata Appl.*, vol. 0, pp. 95–101, 2019.
- [14] S. Blömeke *et al.*, "Recycling 4.0: An Integrated Approach Towards an Advanced Circular Economy," in *Proceedings of the 7th International Conference on ICT for Sustainability*, 2020, pp. 66–76.
- [15] L. Jiang, L. D. Xu, H. Cai, Z. Jiang, F. Bu, and B. Xu, "An IoT-Oriented Data Storage Framework in Cloud Computing Platform," *IEEE Trans. Ind. Informatics*, vol. 10, no. 2, pp. 1443–1451, 2014.
- [16] K. Mišura and M. Žagar, "Data marketplace for Internet of Things," in *2016 International Conference on Smart Systems and Technologies (SST)*, 2016, pp. 255–260.
- [17] J. D. Bokefode, A. S. Bhise, P. A. Satarkar, and D. G. Modani, "Developing A Secure Cloud Storage System for Storing IoT Data by Applying Role Based Encryption," *Procedia Comput. Sci.*, vol. 89, pp. 43–50, 2016.
- [18] J. A. Wood, J. s. Boles, W. Johnston, and D. Bellenger, "Buyers' Trust of the Salesperson: An Item-Level Meta-Analysis," *J. Pers. Sell. Sales Manag.*, vol. 28, no. 3, pp. 263–283, 2008.
- [19] L. G. Zucker, "Institutional theories of organization," *Annu. Rev. Sociol.*, vol. 13, no. 1, pp. 443–464, 1987.
- [20] C. Castelfranchi, "Commitments: from individual intentions to groups and organizations," *Proc. First Int. Conf. Multiagent Syst. ICMAS 1995*, pp. 41–48, 1995.
- [21] P. A. Pavlou and D. Gefen, "Building effective online marketplaces with institution-based trust," *Inf. Syst. Res.*, vol. 15, no. 1, 2004.
- [22] D. Quah, "Digital Goods and the New Economy," 2003.
- [23] J. A. Howard and J. N. Sheth, "A Theory of Buyer Behavior," *Perspect. Consum. Behav.*, 1968.
- [24] A. S. C. Ehrenberg and F. M. Nicosia, "Consumer Decision Processes: Marketing and Advertising Implications," *J. Mark. Res.*, 1968.
- [25] J. F. Engel, D. T. Kollat, and R. D. Blackwell, "Consumer behavior," *Holt, Rinehart and*, 1973.
- [26] P. Kotler, *Marketing management*. 2009.
- [27] A. Borek, A. K. Parlikad, J. Webb, and P. Woodall, *Total Information Risk Management: Maximizing the Value of Data and Information Assets*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013.
- [28] D. A. Garvin, *Managing quality: The strategic and competitive edge*. Simon and Schuster, 1988.

- [29] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manag. Inf. Syst.*, vol. 12, no. 4, pp. 5–34, 1996.
- [30] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data quality in context," *Commun. ACM*, vol. 40, no. 5, pp. 103–110, 1997.
- [31] R. Y. Wang, "A Product Perspective on Total Data Quality Management," *Commun. ACM*, vol. 41, no. 2, pp. 58–65, 1998.
- [32] S. Karray and L. Debernitz, "The effectiveness of movie trailer advertising," *Int. J. Advert.*, 2017.
- [33] R. J. Faber and T. C. O'guinn, "Effect of Media Advertising and other Sources on Movie Selection," *Journal. Mass Commun. Q.*, vol. 61, no. 2, pp. 371–377, 1984.
- [34] D. W. Cunningham, "Predicate Logic," in *A Logical Introduction to Proof*, New York, NY: Springer New York, 2013, pp. 29–60.