

Data Contracts: Structuring Promises and Expectations in Data Exchange

Laura Schuiki

University of Stuttgart

Stuttgart, Germany

laura-sophie.schuiki@ipvs.uni-stuttgart.de

Arif Wider

HTW - University of Applied Sciences

Berlin, Germany

arif.wider@htw-berlin.de

Simon Harrer

Entropy Data GmbH

Monheim am Rhein, Germany

simon.harrer@entropy-data.com

Abstract—Modular service-oriented approaches to managing enterprise data, i.e., so-called *data products*, have gained in popularity in industry. In fast-innovating organizations, new data products are created every day, and access to them is requested daily. Ensuring robustness of resulting data pipelines and compliance with regulations has become a major data governance challenge. *Data contracts* are a new approach to managing data exchange in such settings. However, the interpretation of data contracts in industry is inconsistent, and tool support and standardization are nascent. While most interpretations view data contracts as provider-driven artifacts, there is a growing trend in scientific literature to include consumers in their creation. We propose the incorporation of the consumer’s perspective into data contracts and suggest the adaptation of data governance tooling to enable higher levels of governance automation in data sharing.

Index Terms—data governance, data contracts, data products, data mesh, service-oriented architecture

I. INTRODUCTION

The definition and documentation of interfaces for robust inter-service communication have always been core topics of service-oriented architectures. With data-driven services and AI applications becoming more prevalent, data sharing is on the rise, and the need for robust data flows within and between organizations is increasing [5]. Managing the exchange of analytical data, however, brings different challenges than the exchange of operational data via APIs in a typical microservices architecture. Therefore, practices and tools from the services community cannot be applied as is.

A *data contract* is a relatively new concept that aims to make data exchange robust, compliant, and verifiable [10], [11]. Data contracts are often introduced with efforts to modularize an organization’s data landscape, similar to the transition from a monolithic to a microservices architecture [3].

Thereby, the concept of a *data product* is applied, which is a self-contained unit of analytical data that is self-describing, ready for consumption, and designed for composition. Similar to microservices, it is an independently deployable unit to create higher-order services. Data products interact through output ports, which are clearly defined interfaces that specify the schema of the shared data. Any terms associated with the output ports, such as service-level agreements (SLAs) or constraints, are outlined in corresponding data contracts. As illustrated in Fig. 1, higher-order data products consume data from one or more source data products by connecting to their

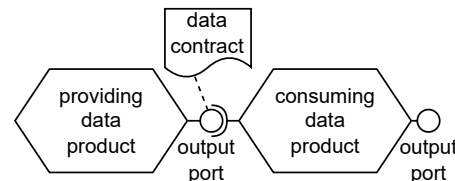


Fig. 1. **Data products and their composition.** The sharing of data is defined through data contracts that specify the usage terms of the output ports.

output ports, typically transforming the data internally, and re-sharing results through their own output ports.

One of the main goals of modularizing an organization’s data landscape and introducing data contracts is to facilitate data governance automation, reducing the burden on data providers [7], [9], [12]. However, the usage of data contracts in industry is nascent, and both the interpretation of the concept and its application are inconsistent. In particular, while most view data contracts as provider-driven artifacts, there is a growing trend towards including consumers’ perspectives as well [8], [10]. In order to automate governance around data products, it is critical to include consumers’ perspectives, e.g., to automatically match a consumer with a data product fulfilling their needs. Therefore, this work aims to establish the relevance of consumer-driven data contracts in practice by presenting a general concept of data contracts (Section III).

II. RELATED WORK

A substantial amount of literature exists on data products; data contracts are primarily discussed in industry spaces.

Wasser et al. [11] compiled an extensive systematic gray literature review demonstrating the relevance of data contracts for practitioners. Oppold et al. [6] reinforce this with their publication, detailing the use of data contracts at ZEISS. Dolhopolov et al. [4] present a formal governance model for data products in data mesh. It includes provider-driven contracts but does not consider the consumer point of view.

Wider et al. [14] describe data product input and output ports via data contracts, incorporating both provider and consumer perspectives. Subsequently, Wider et al. [13] provide more detail on various factors that influence data contracts and argue that agreements between providers and consumers require expectations and promises from both sides. In their most recent work [12], they propose a tool using large language

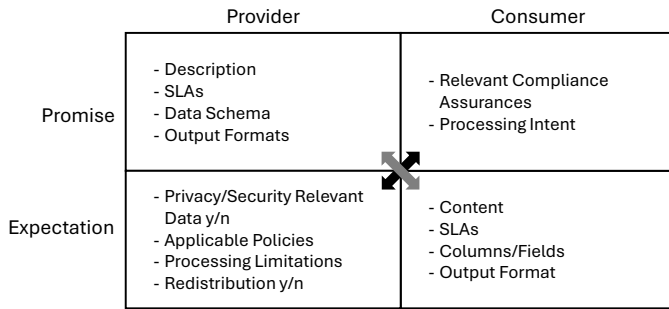


Fig. 2. Four aspects of data contracts and their relation to each other.

models (LLMs) to generate data governance policies directly from legal documents that also influence data contracts.

While data contracts are a popular topic in practice, it is evident that there is not yet a large body of scientific literature on the subject. Existing literature still mostly focuses on provider-driven contracts, with some theoretical approaches also acknowledging the need to consider the consumer’s perspective. However, a structured representation of all the aspects needed to create a complete data contract and automate data governance tasks does not currently exist.

III. A GENERAL CONCEPT OF DATA CONTRACTS

Data contracts are often viewed as one-sided promises made by data providers to their consumers, e.g., by stating SLAs. However, in enterprise data landscapes, relations between data providers and consumers are more complex: First, consumers should also state their expectations of providers, which is already a common practice with consumer-driven contracts in microservices architectures. Second, providers have expectations as well, e.g., that consumers adhere to policies when consuming sensitive data. Third, consumers should provide promises regarding how they will handle data, so that providers can be confident that their data will be treated properly.

Unfortunately, these three aspects of data contracts are often not documented explicitly or overlooked entirely, which increases the risk of data incidents. Explicit data contracts are mostly used to define a data product provider’s promises to the consumers of their data product. The expectations and promises of data consumers are, at best, only covered by implicit conventions or understandings. Because these aspects are often overlooked, current data governance tools lack support to define, manage, and verify such complex data contracts. We suggest explicitly implementing provider expectations and consumer-driven contracts, thereby defining data contracts fully from both points of view and allowing consumers to explicitly influence agreements.

This results in our concept, in which both provider and consumer make promises and have expectations towards the data exchange, creating four distinct aspects that are described in detail in the following subsections, depicted in Fig. 2, with example attributes provided in Table I.

A. Provider Promises

As the name suggests, provider promises are used to convey promises of a data product provider to potential consumers,

TABLE I
EXAMPLE ATTRIBUTES FOR THE DIFFERENT DATA CONTRACT ASPECTS.

Aspect	Example attributes
Provider Promises	Schema of sales data, 98% availability, weekly updates, ...
Consumer Expectations	Sales data and consumer group, daily updates, new data until 2027, ...
Provider Expectations	No further sharing, basic security class, compulsory privacy metric, ...
Consumer Promises	Agree to no sharing, assure compliance with privacy metric, ...

i.e., to advertise their data product. The roles involved are, on the one hand, the provider who creates this aspect of a data contract, and on the other hand, the consumer who checks the promises. If the provider’s promises match the consumer’s expectations, an agreement might be formed. Therefore, it is very important to accurately present the provider’s promises and record them explicitly to maximize the chance of consumers using the data product.

At this time, provider promises are the most widely implemented aspect of data contracts; typically, if a document is labeled as a data contract, it will contain provider promises.

For example, a data provider wants to share their data product containing data on product sales starting in 2023. Part of the deployment process is a description of the benefits their data product will deliver to consumers, e.g., a description of the data being provided, SLAs, and a database schema.

B. Consumer Expectations

Consumer expectations are used by consumers to convey their needs to providers of existing or not yet existing data products. The roles involved are the consumer who creates this aspect of a data contract and the provider who checks the expectations. For an agreement to be formed, a consumer’s expectations have to match the provider’s promises. Therefore, it is critical to enable consumers to adequately voice their expectations by documenting them. In turn, this enables providers to inquire about the expectations of consumers to discover missing data products or whether their data products need to be changed. Additionally, providers can verify if their data products fulfill expectations that were not known to them, which can then be included as additional promises. Lastly, formalization of provider promises and consumer expectations enables automatic matching of data products to consumers.

Although consumer expectations are beneficial and might be implicitly considered, the majority of practitioners do not expect explicit documentation. Therefore, consumer expectations are not supported by most available platform tools.

The attributes that make up consumer expectations largely mirror those of provider promises, but state a demand rather than give a promise.

For example, a consumer finds several suitable-looking products, but they are not certain that they will be a good fit. They therefore start documenting their expectations to send to providers, hoping to find a good match.

C. Provider Expectations

Provider expectations are used by providers to convey their expectations to potential consumers, i.e., to ensure consumers are informed about obligations and restrictions on their usage rights. The roles involved are the provider who creates this aspect of a data contract and the consumer who considers the defined restrictions and decides whether they agree. In light of laws and regulations that have come into effect in recent years, it is paramount to ensure lawful sharing and usage of data. Therefore, it is necessary to document restrictions and obligations regarding data consumption, as well as to obtain guarantees of compliance from consumers. Additionally, consumers can verify early on whether a data product allows them to consume data in a manner fit for their use case or if restrictions are too limiting.

Although this aspect of data contracts is beneficial and provider expectations might be implicitly considered by providers, e.g., through company-wide privacy policies, the majority of users do not expect them to be explicitly documented. As a consequence, provider expectations are not supported by most available platform tools.

For example, the data provider of the sales data product also defined expectations. Consumers need to agree not to share data further, acknowledge that there is privacy-relevant data, and describe what kind of data processing they plan to do.

D. Consumer Promises

Consumer promises are used by consumers to convey to a provider their promises regarding agreement to the terms of use and how they plan to use a data product. The roles involved are the consumer who creates this aspect of a data contract and the provider who evaluates whether or not this fits their expectations, e.g., adherence to privacy policies. This aspect of the data contract is critical to reaching a sharing agreement. A potential consumer assesses the provider's expectations through the data contract and gives a counteroffer. The provider evaluates this counteroffer and decides if they are comfortable with giving access on these terms.

Currently, this is mostly done implicitly or through company-wide policies and processes. Thereby, consumers might implicitly agree with provider expectations when requesting access. This can lead to problems because consumers do not understand the implications of agreeing to the terms of use, which might cause conflict and revoked access later, or even legal problems if regulations are not adhered to.

For example, a consumer discovering the sales data product must agree not to share the data with other consumers, acknowledge that the data product contains privacy-relevant data, and describe the type of processing they plan to do.

IV. VALIDATION AND FUTURE WORK

In a preliminary analysis, we examined how the prevalent Open Data Contract Standard [1] (ODCS) and ODPS [2] support our *general concept of data contracts*. ODCS is well suited for provider promises and consumer expectations, but relies mostly on free text fields for provider expectations

and consumer promises, which lack formalization. ODPS integrates ODCS by referencing data contracts in data product ports, indicating that the existing standards can in principle cover the general concept, even if two of its aspects are not yet formalized. A detailed evaluation is left to future work.

We validated our general concept of data contracts through interviews with experts in various roles across three companies. All interviewees were aware of the four aspects specified in the general concept, and their answers could be categorized based on the four aspects. This supports the validity of the general concept and shows that it can be useful in helping enterprises adopt data contracts. All three companies started their journey towards data contracts with interface descriptions. Depending on their current maturity of adoption, their next step is either to create the first data contracts, propagate the data contract concept, or enhance existing data contracts with consumer expectations and promises.

While the maturity of adoption varies significantly between companies, they all show a similar progression in their adoption of data contracts, which aligns with the general concept. We identified a fixed sequence of maturity levels that each of our interviewees followed. From this, we plan to derive a maturity model for data contract adoption in industry, which we aim to validate through additional interviews.

REFERENCES

- [1] Bitol: Open Data Contract Standard (ODCS). LF AI & Data Foundation (2025), <https://bitol-io.github.io/open-data-contract-standard>
- [2] Bitol: Open Data Product Standard (ODPS). LF AI & Data Foundation (2025), <https://bitol-io.github.io/open-data-product-standard>
- [3] Deghani, Z.: Data Mesh: Delivering data-driven value at Scale. O'Reilly (2022), <https://www.oreilly.com/library/view/data-mesh/9781492092384/>
- [4] Dolhopolov, A., Castelltort, A., Laurent, A.: Implementing federated governance in data mesh architecture. *Future Internet* **16**(4), 115 (2024)
- [5] Goedegebuure, A., , Kumara, I., Driessen, S., van den Heuvel, W.J., Monsieur, G., Tamburri, D.A., Di Nucci, D.: Data mesh: A systematic gray literature review. *ACM Computing Surveys* **57**(1), 1–36 (2024)
- [6] Oppold, S., Fritz, M., Woltmann, L.: Data contracts to leverage (de-)centralized data management in manufacturing industries: An experience report. In: *BTW 2025*, pp. 731–743. Bamberg (2025)
- [7] PricewaterhouseCoopers: Data mesh - the next-generation enterprise data platform? (2022), <https://www.pwc.de/en/digitale-transformation/data-mesh-the-next-generation-enterprise-data-platform.html>
- [8] Sanderson, C.: The consumer-defined data contract (2023), <https://dataproducts.substack.com/p/the-consumer-defined-data-contract>
- [9] Vestues, K., Hanssen, G.K., Mikalsen, M., Buan, T.A., Conboy, K.: Agile data management in NAV: A case study. In: *XP 2022. LNBP*, vol. 445, pp. 220–235. Springer, Cham (2022)
- [10] Warlin, D.: Deep-dive and FAQ: What are data contracts and what are they not? (2023), <https://blog.mesh-ai.com/deep-dive-and-faq-what-are-data-contracts-and-what-are-they-not-bdb6b73225>
- [11] Wasser, J., , Kumara, I., Monsieur, G., van den Heuvel, W.J., Tamburri, D.A.: Data contracts in data mesh: A systematic gray literature review. In: *BMSD*. pp. 21–38. LNBP, Springer, Cham (2026)
- [12] Wider, A., Harrer, S., Dietz, L.W.: Ai-assisted data governance with data mesh manager. In: *ICWS 2025*. pp. 963–965. IEEE, Piscataway, NJ (2025)
- [13] Wider, A., Jarmul, K., Akhtar, A.: Towards automating federated data governance. In: *ICWS 2024*. pp. 10–19. IEEE, Piscataway, NJ (2024)
- [14] Wider, A., Verma, S., Akhtar, A.: Decentralized data governance as part of a data mesh platform: Concepts and approaches. In: *ICWS 2023*. pp. 746–754. IEEE, Chicago, IL, USA (2023)