

A Layered Architecture for Large-Scale Data Space Federation using a Common Carrier Layer

Sebastian Pretzsch¹, Harrie Bastiaansen², Bernadette Farias Lóscio² and Gert Kruithof²

¹Fraunhofer IVI, Dresden Germany

²TNO, 9747AA Groningen, The Netherlands

sebastian.pretzsch@ivi.fraunhofer.de, harrie.bastiaansen@tno.nl,
bernadette.fariaslosocio@tno.nl, gert.kruithof@tno.nl

Abstract. Data sharing between organizations is fundamental to promote innovation and to drive value co-creation. In the last years, not only the interest on sharing data has increased but also the scale on how data has been shared. This is mainly due to i) the dissemination of data spaces; ii) the increasing interest on data sharing across sectors or geographical borders; and iii) the emergence of various data sharing applications such as Digital Product Passports, Digital Twins and Privacy Enhancing Technologies. In this paper, we propose a layered system architecture with a Common Carrier Layer (CCL) as an enabler for large-scale data space federations. The CCL enables the establishment of dynamic trust relationships and addresses interoperability issues of heterogeneous parties.

Keywords: Data Space, Common Carrier Layer, Discoverability, Data Sovereignty, Trust, Interlinking Layer, Federation, Interoperability.

1 Introduction

The concept of data spaces is central to Europe's Data Strategy¹, which expresses the ambition to create "Common European Data Spaces"². Data spaces aim to provide a scalable mechanism for the sovereign, trusted, and interoperable sharing of data among numerous participants and stakeholders. Presently, the data space landscape includes both the emerging sectoral data spaces as part of the broader Common European Data Spaces initiative, and a plethora of individual data sharing initiatives. In the mobility and logistics domain, for example, there are already several regional data spaces in operation, such as the German MDS³, Eona-X⁴, FSDS⁵ and the Dutch BDI⁶.

¹ European Commission (2020). 'A European Strategy for Data'. <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

² European Commission. 'Common European Data Spaces'. <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

³ <https://mobility-dataspace.eu/>

⁴ <https://eona-x.eu/>

⁵ <https://www.imec-int.com/en/flemish-smart-data-space>

⁶ <https://bdinetwork.org/en>

However, it is to be realized that these data spaces are organizationally and technically not interoperable with each other. Furthermore, the data spaces use cases, and the needs of their participants generally extend beyond the boundaries of a single data-sharing community or data space. Mobility and logistics data sharing use cases, for example, inherently cross regional or national borders and will involve data stemming from other sectors such as energy, tourism, and urban infrastructure. In the context of the common European mobility data space (EMDS), the EC already reacted by proposing an *Interlinking Layer* as a core part of the EMDS to enable cross-data space data discovery by collecting metadata in a federative catalogue.

The need for cross-community data sharing even becomes more pronounced with the rise of new data sharing application areas such as Digital Product Passports (DPPs) [3], collaborative *Digital Twins (DTs)* [4], and *Privacy Enhancing Technologies (PETs)* [5]. These emerging application areas integrate data from various sectors and therefore operate across their associated data sharing communities.

Developing interoperable and federated data spaces [1][2] for cross-community use cases requires significant investment and time, especially due to the need for interoperability across governance, legal, semantic, and technical levels. Therefore, bilateral federations between data spaces are not scalable for cross-cutting use cases. In this context, innovative solutions to enable large-scale data space federation are needed.

In this paper we propose an (alternative) business, legal and technical data space system architecture for large-scale interoperability and federation of data spaces [7]. The proposed architecture leverages new data space standards, such as the Data Space Protocol (DSP), the Decentralized Claims Protocol (DCP) and the OpenID for VC (OI4VC). Utilizing a Common Carrier Layer (CCL), it enables dynamic, trusted data sharing across community boundaries. The CCL onboarding model offers a single point-of-entry for participants to access data services across different data spaces.

This paper has the following structure. Section 2 introduces the *participation-driven* and *community-driven* models for data-sharing federation, as well as a combined approach. Section 3 elaborates the new data space business roles considering the alternative layered system architecture based on a CCL. Subsequently, Section 4 elaborates the CCL, highlighting the pivotal role of the data space connector and the data space protocols as key enablers for dynamically establishing data sharing relationships. Finally, Section 5 presents the overarching conclusions.

2 Data-Sharing Federation Models

Interoperability within data spaces extends beyond just the technical components. The European Interoperability Framework⁷ (EIF), developed by the European Commission offers a structured approach to tackle interoperability challenges. The EIF identifies four levels of interoperability—technical, semantic, organizational, and legal—under a

1. ⁷ https://ec.europa.eu/isa2/sites/isa2/files/eif_brochure_final.pdf.

comprehensive, integrated governance model. For federation of data spaces, interoperability on each of the four levels must be established.

For use cases within well-demarcated (“closed”) communities of participants and strong commonalities in data sharing needs, a static ‘top-down’ *community-driven* data sharing model may be well suited, with community-specific agreements on each of the EIF’s interoperability levels, tailor-made to the community and use case specific needs.

However, this *community-driven* model appears to be less suited for use cases that cross the community / data space borders. Interoperability and federation require bilateral agreements between data spaces on each of the four levels, with major complexities in alignment. These especially occur on the governance, legal and trust relationships.

As the number of data spaces grows, bilateral interoperability and federation agreements will not scale effectively for cross-cutting use cases. Hence, an alternative solution is proposed that covers the federation scalability aspects by design. It is referred to as a ‘bottom-up’ *participation-driven* data sharing model, see Fig. 1.

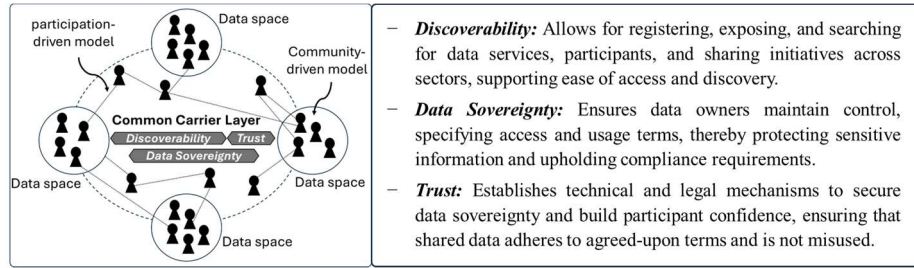


Fig. 1. Participation-driven and community-driven models combined.

Fig.1 shows various data spaces composed of participants willing to share data. Within the borders of a data space, the *community-driven* model is employed. It focuses on the community specific (e.g. sector or geographical area) data sharing needs and encompasses a set of community-specific agreements and services to support the community specific use cases. As Fig.1 shows, there can also be participants that do not belong to a community and use cases that cross the borders of a community, for which the *participation-driven* model may be used. In these cases, the CCL's discoverability, data sovereignty, and trust capabilities may be used to dynamically establish trusted data sharing relationships across community borders. The CCL provides data sharing participants with a single point of entry to access data services from data providers across the boundaries of data space communities.

3 Layered system architecture and business role models

The *participation-driven* model can complement and coexist with the *community-driven* one. Moreover, a combined *community-driven* and *participation-driven* model enables a layered business, legal and technical data space system architecture, as depicted in Fig. 2. The layered architecture supports large-scale interoperability and federation through a dual onboarding process, creating new business opportunities for

participants and data sharing communities, while also benefiting Connector-as-a-Service (CaaS) providers operationally. These aspects will be described in the following.

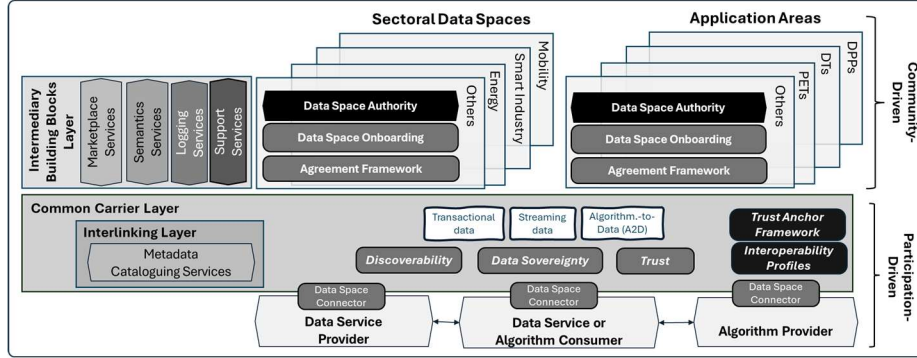


Fig. 2. Layered business, legal and technical data space system architecture.

3.1 Layers of the Data Space System Architecture

The layered business, legal and technical data space system architecture (Fig. 2) combines the benefits of coexisting *community-driven* and *participation-driven* models. It presents the CCL as extension to the Interlinking Layer, and its relation to Data Spaces and Application Area Layer and Intermediary Building Blocks Layer.

The *Interlinking Layer (IL)*, as proposed by the EC for the common European Mobility Data Space, enables the architecture by a federated discoverability mechanism across participation-driven and community-driven participants: a harvesting mechanism collects meta-data from data spaces and participants and stores it in the Interlinking Layer to provide a single access point for search and discovery.

The *CCL* extends the IL by providing a scalable, flexible, and extensible infrastructure layer capable of supporting a multitude of cross-community data sharing use cases. Through the dynamic establishment of data sharing relationships, it facilitates interoperability on the key capabilities for discoverability, data sovereignty, and trust between participants, even if they belong to different data sharing communities. Moreover, it offers Interoperability Profiles to simplify its operations, consisting of templates for commonly used ODRL⁸-based usage policy sets and optional components of legally binding data sharing contracts.

The *Data Spaces and Application Area Layer* focuses on the community specific (e.g. sector or geographical area) or application area specific data sharing needs. It relies on a framework of community-specific agreements for both generic and use case specific data sharing capabilities and services to support the community specific use cases.

The *Intermediary Building Blocks Layer* consists of support and added-value services that are generically re-usable or cross-usable for (multiple) data sharing communities or data spaces, further promoting interoperability and scalability. As shown in Fig. 2, they may be categorized into (1) *Marketplace Services*, offering opportunities

⁸ <https://www.w3.org/TR/odrl-model/>

to promote data, products and services, contract and policy templates, and app stores, (2) *Semantics Services*, providing tools for vocabulary management, semantic transformation and semantic configuration, (3) *Logging Services*, including registries for metadata query, contract and transaction logging, and (4) *Support Services*, encompassing legal, security, and marketplace services for monetization. This layer gives data spaces also the opportunity to streamline their own service offerings by outsource non-core services to external operators.

The layered data space system architecture with the dual onboarding process allows for differentiation in data spaces business models. Developing community-specific agreements and capabilities for discoverability, data sovereignty, and trust may no longer be reasonable when these generic features can be reused from the CCL. In this manner, *community-driven* data spaces can focus on developing value-added services for their participants, i.e. separation of concerns. It enables them to be more efficient and prepares them for large-scale interoperability and federation across data spaces.

3.2 Large-Scale Federation and Interoperability: Dual Onboarding

The *participation-driven* model is enabled by onboarding to the CCL, which provides participants with a single point-of-entry to access data services as provided by other participants, irrespective of the community boundaries of specific data spaces. Membership and onboarding in the CCL adhere to a well-defined process, utilizing a trust anchor framework (Fig. 2) that includes trust mechanisms like those from Gaia-X⁹ or iSHARE¹⁰. It operates independently from specific data spaces by certified trust ensuring organizations. The trust ensuring organizations issue certificates on the minimum set of trust capabilities for participant membership of the CCL, as will be addressed in the following. Moreover, the framework for becoming trust ensuring organizations for the CCL must be broadly accepted and under EC and national governance.

There are several options for aligning the onboarding process of the *participation-driven* model (with the CCL) and *community-driven* model (within data spaces). Third-party service providers can offer onboarding to the CCL as a service, which can be used for onboarding to individual data spaces, simplifying the process for individual data spaces. Conversely, onboarding to individual data spaces can be reused for the CCL. A combined approach supports a two-level onboarding process, providing dual membership in both the participation-driven CCL and community-driven data spaces.

3.3 CCL Operations Perspective: Benefits for CaaS-Providers

Although still requiring a rigorous onboarding process based on well-defined criteria, it is to be noted that there are major benefits in CCL onboarding over onboarding to specific data spaces. One of those benefits is the simplicity. For CCL onboarding, only minimal trust capabilities are needed to establish dynamic, trusted data sharing relationships. This includes trustworthy identification and authentication of legal

⁹ <https://gaia-x.eu/>

¹⁰ <https://ishare.eu/>

participants, which can be achieved through a Qualified eSignature, supporting legally binding, bilaterally negotiated data sharing contracts.

Economies of scale and scope exist in providing these trust capabilities at the CCL. Moreover, interoperability in providing dynamic, trusted, data sharing relationships strongly depends on the interoperability of its data space connectors, as will be further elaborated in the next section. Hence, clear business opportunities exist in operating the CCL by a limited number of major, trusted, third-party, service providers.

As such, the emerging Connector-as-a-Service (CaaS) providers are in an excellent position to provide the CCL services, ensuring interoperability-by-design across data spaces. By alignment and agreement on connector and protocol standards to be used and operated within the CCL, they are well-positioned to reap its benefits of economies of scale and scope.

4 The Common Carrier Layer

The CCL enables the dynamic establishment of trusted data sharing relationships. It provides the basic organizational and infrastructural capabilities for discoverability, data sovereignty, and trust, enabling the dynamic establishment of data sharing relationships. This offers a single point-of-entry to access data services as provided by other participants, irrespective of the community boundaries of specific data sharing communities. It is enabled by the Dataspace Connector architecture, the Dataspace Protocol standards and the underlying trust mechanisms and the yielding data sharing contracts, as will be addressed in the following.

4.1 The Dataspace Connector and the Dataspace Protocol Standards

The recent development of the dataspace connector architecture, the Decentralized Claims Protocol (DCP)¹¹, the OpenID for VC (OID4VC)¹² and the Dataspace Protocol (DSP)¹³ standards lay the groundwork for the CCL. Moreover, they align with the concepts, reference architectures and standards as being defined in the context of the EU Data Strategy, enabling broad acceptance and adoption. Their inclusion in the DSSC Blueprint v2.0 [6] marks a significant milestone for European data spaces.

The data space connector and protocols adhere to the architecture design principle with separation between the Control Plane and the Data Plane for federated data sharing. The key advantage of this separation is flexibility. The control plane protocols define how identity, trust information and metadata are exchanged between autonomous entities (participants and data space connectors). As Fig. 3 depicts, it includes discoverability, usage policies, a contract negotiation identity and trust protocols.

Jointly, the data space connector with the protocols provides the control plane functions to dynamically establish trusted data sharing relationships through peer-to-peer

¹¹ <https://projects.eclipse.org/projects/technology.dataspace-dcp/>

¹² <https://openid.net/sg/openid4vc/>

¹³ <https://internationaldataspaces.org/offers/dataspace-protocol/>

interactions. The trust framework of the CCL defines the required claims that participants need to provide as Verifiable Credentials/Presentations and how these claims are provided and signed by the trusted network of CCL providers. These credentials are exchanged over the OID4VC and DCP between connectors and the trust framework.

The discovery of data and negotiation of contracts is realized via the DSP. The discovery can be extended by connecting also to the Interlinking Layer to search the data offers of further data spaces. In turn, participation-driven participants might want to directly publish the metadata of their offerings to the Interlinking Layer. This minimizes the need for centralized services or building blocks to authorize, control and operate legally binding data sharing transactions, which has previously been mainly formalized and operated by data spaces.

Moreover, a data plane facilitates the actual transfer of (potential sensitive) primary data between the connectors. Various transfer protocols, such as HTTP, MQTT and Kafka, can be supported by a data plane. It is expected that only a limited set of transfer protocols will suffice to handle the various types of data sharing. The Eclipse Data Space Components (EDC) framework and the FIWARE Data Space Connector are currently attracting attention as (open source) implementations of this architecture

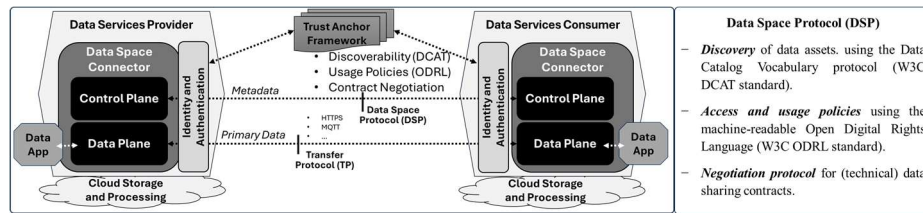


Fig. 3. The DSP defining the control interfaces (metadata) between *data space connectors*.

4.2 Legal Embedding: Signing Bilaterally Negotiated Data Sharing Contracts

According to EU law, electronically negotiated data sharing contracts are legally binding if they meet the following four conditions: i) the contract must be accessible to all parties, ii) the authenticity of the contract must be assured, iii) the creation of the contract must be established with certainty, and iv) the identities of the parties must be verifiable. For electronically, bilaterally, negotiated contracts in dynamically established data sharing agreements to be legally binding, mutual signing by both the data service provider and consumer of the data sharing contract is needed. It requires a qualified electronic signature, e.g. as mandated by the recent EU eIDAS regulation¹⁴. This forms a main trust capability for the CCL service portfolio.

In addition to these mandatory conditions for being legally binding, data-transaction agreements can include optional components such as intellectual property rights, liability terms, and dispute resolution mechanisms. These optional components may similarly be electronically negotiated by means of the DSP in the CCL or be defined within specific data spaces as overarching agreement.

¹⁴ <https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation>

5 Conclusions

Historically, discussions around the development of "Common European Data Spaces" have operated under the assumption that data spaces are *community-driven*, with sector-specific use cases forming the foundation. However, the need for a scalable approach to federation and interoperability between data spaces is becoming increasingly critical. This demand is driven by the growth of use cases that extend beyond the borders of specific data sharing communities or individual data spaces. In this context, the *participation-driven* model has the potential to support such cross-community data sharing use cases, through a scalable federation approach of the key capabilities for discoverability, data sovereignty and trust. A key aspect of this model is the introduction of a new membership and onboarding model, which allows participants to access data services through a single point of entry. This can be done by means of the dynamic establishment of legally binding and trusted data sharing relationships, irrespective of their community affiliations. By adhering to the emerging EU reference architectures and standards, it can foster a trusted, legally-binding and technically interoperable ecosystem of data sharing participants. Through collaboration and ongoing development, the proposed *participation-driven* model may serve as a cornerstone for the future of large-scale, federated data sharing networks.

References

1. Schleimer, A.M., Jahnke, N., Otto, B. Architecture Design Options for Federated Data Spaces (2023) Proceedings of the Annual Hawaii International Conference on System Sciences, 2023-January, pp. 3643-3652.
2. Dennis Heimbigner and Dennis McLeod. 1985. A federated architecture for information management. ACM Trans. Inf. Syst. 3, 3 (July 1985), 253–278. <https://doi.org/10.1145/4229.4233>.
3. European Commission (2022). 'Ecodesign for Sustainable Products Regulation'. <https://eur-lex.europa.eu/eli/reg/2024/1781/oj>.
4. TNO. 'Reliable and Real-Time Digital Twin Systems'. <https://www.tno.nl/en/digital/digital-innovations/data-sharing/reliable-real-time-digital-twin-systems>.
5. OECD (2023), 'Emerging Privacy Enhancing Technologies'. <https://www.oecd.org/publications/emerging-privacy-enhancing-technologies-bf121be4-en.htm>.
6. EU Data Spaces Support Centre. 'The DSSC Blueprint - How a Data Plane and Control Plane Work Together'. <https://dssc.eu/space/BVE2/1071254937/>
7. Centre of Excellence for Data Sharing and Cloud (CoE DSC). 'Towards Scalable Data Space Interoperability and Federation'. Discussion Paper. <https://coe-dsc.nl/towards-a-common-carrier-layer-for-large-scale-data-space-interoperability-and-federation>.