Data Spaces | DOI:10.1145/3512341

A Federated Infrastructure for European Data Spaces

BY BORIS OTTO

THE EUROPEAN STRATEGY FOR DATA CALLS for "common data spaces" as a foundation for the data economy in the European Union.[1] President Ursula von der Leyen expressed her view that Gaia-X should play a key role in response to these calls by providing a federated data infrastructure in the cloud.[a]

One data space example is the Mobility Data Space[b] that was launched in 2020 as a multistakeholder project in Germany in close alignment with the German Federal Government. It aims at data-driven services in the mobility sector and data sovereignty of the data holders and trust among all participants. Smart service examples are intermodal end-to-end services for individual travelers, traffic management services for smart cities, and services to increase road safety for individual drivers.

Figure 1 shows the basic functionality of the Mobility Data Space following the provisions of the International Data Spaces (IDS) Reference Architecture Model (RAM).[4] It does not pool data in a central data store but rather connects data providers and data users by the respective use of the IDS Connector component. A catalog allows data providers to present and describe their data resources, together with conditions under which the data can be used. Data users search the catalog for data they need for their smart service. If data demand and supply match, the exchange of the data itself is carried out just between the participants, with no involvement of the Mobility Data Space operator. Only metadata on data exchange transactions are monitored and logged to ensure conditions for the use of the data are correctly exchanged and followed.

The Mobility Data Space balances the interest of the individual data provider regarding data sovereignty and the interest of the community to increase the data availability and its use. Data has a value when used[7] and, thus, should be used as widely as possible to seize its innovation opportunities. However, data use must always take into consideration the interests of the data holder. Providing data, particularly high-quality data, comes at a cost.[6] In this context, data sovereignty is defined as the capability of a data holder to be self-determined regarding the use of their data.[5]

Gaia-X supports the goals of interoperability of services, data portability, data sovereignty as articulated in the European data strategy, and addresses a requirement gap in existing infrastructure models.

Gaia-X as a Federated Data Infrastructure

Data spaces represent a data integration concept that follows Linked Data design principles.[2] First, data spaces do not require physical data integration but leave the data at the source and only make it accessible when needed. Second, data is not forced into one common schema but linked, that is, integrated on a semantic level. Third, the distributed architecture of data spaces allows data redundancies, that is, multiple data objects may exist in a data space describing the same real-world object. Finally, data spaces can be overlapping and nested so that data holders and data users, respectively, can be participants in multiple data spaces.

To implement the European Strategy for Data and create common data spaces, interoperability of data and collaboration of participants is required across the boundaries of individual data spaces. Gaia-X aims to achieve this by providing so-called federation services that function within and across different data spaces.

The Gaia-X initiative is organized as a not-for-profit association headquartered in Brussels.[c] With more than 300 members, the association aims at a federated data architecture that comprises not only data and smart services, but also cloud infrastructure services (see Figure 2). Gaia-X specifies four so-called federation services, namely "identity and trust," "sovereign data exchange," "federated catalog," and "compliance."[3] They form a blueprint for data spaces and allow for "federations of ecosystems," hence, support interoperability and coordination across data spaces. The latter requires federation of identities of data space participants, of catalog entries, and of data transaction logs.

a See https://ec.europa.eu/commission/presscorner/detail/ov/SPEECH_20_1655

b See https://mobility-dataspace.eu/

c See https://www.gaia-x.eu/

Data Infrastructure Patterns

Gaia-X supports the goals of interoperability of services, data portability, data sovereignty as articulated in the European data strategy and addresses a requirement gap in existing infrastructure models. The latter were—and still are—dominated by hyperscaling platform providers on the one hand side and state-controlled infrastructures on the other hand. Both do not meet the European requirements mentioned here. Consequently, Gaia-X represents an alternative design pattern for data infrastructures.

In contrast to alternative patterns, Gaia-X follows a cooperative approach (see approximate juxtaposition in the accompanying table). The federated infrastructure is open to be used, owned by the community itself, and thus avoids concentration of control over data and services. C

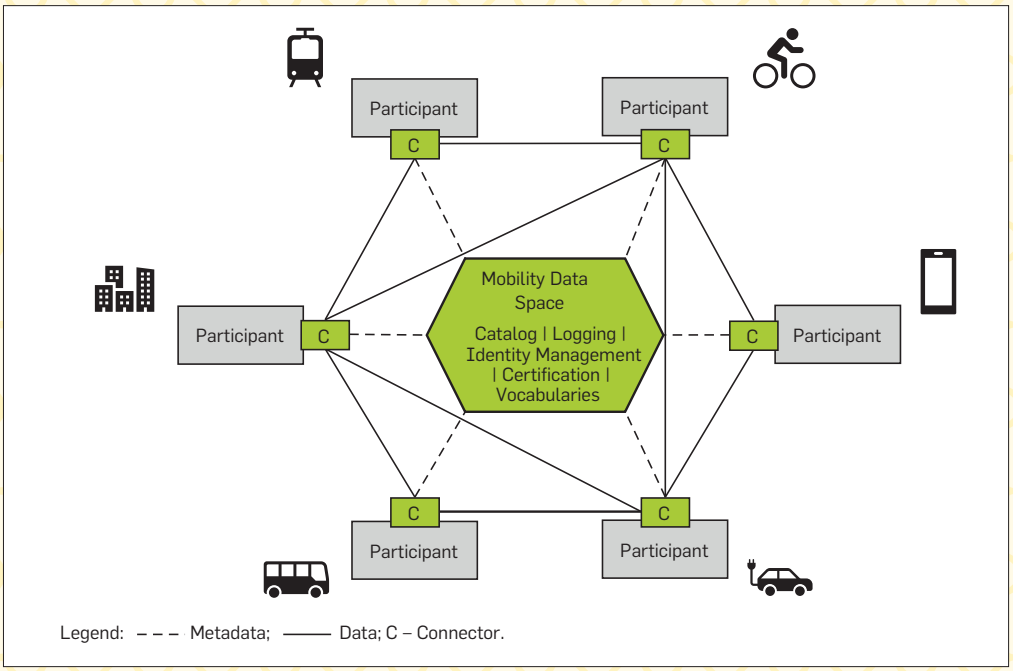

Figure 1. Mobility data space overview.

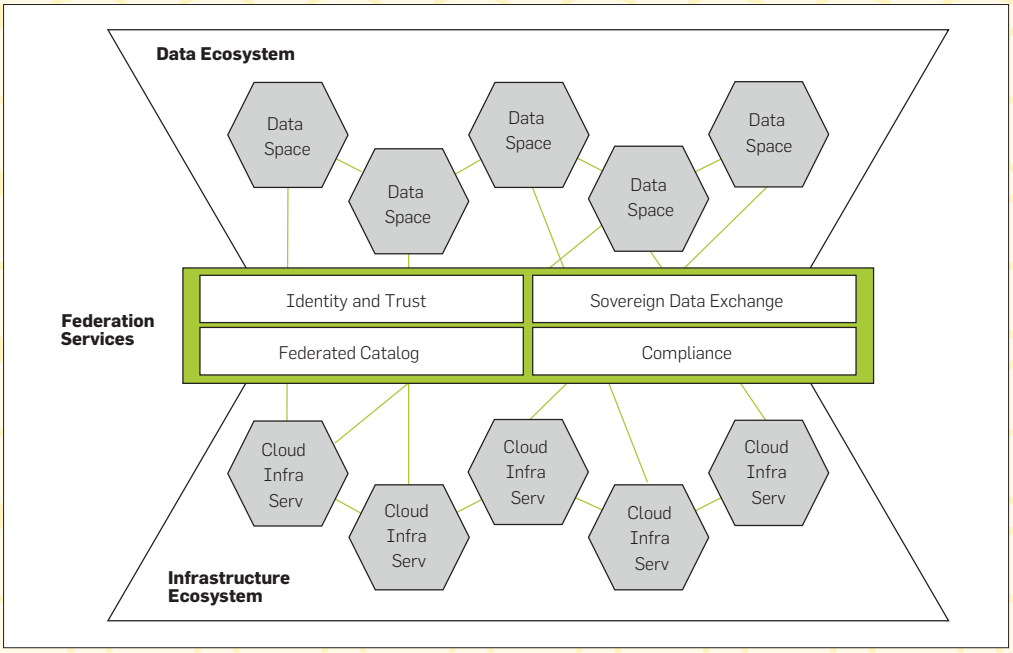

Figure 2. Gaia-X ecosystem and federation services.

	«Hyperscaler»	State-Controlled	Gaia-X
Economic Effect	"Winner takes all"	Digiriste	Cooperative
Platform Owner	1	1	Many
Platform Architecture Design	Central	Central	Federated
Platform Development	Closed/Hybrid	Closed	Open
Code Basis	Proprietary and Open	Proprietary and Open	Open
Data Sovereignty	Basic	Rudimentary	Core Value Proposition
Data Management and Data Exchange	Central	Central	Federated and Bilateral

Data infrastructure patterns.

References

1. European Commission. A European strategy for data. Brussels, Belgium, 2020.
2. Franklin, M., Halevy, A., Maier, D. From databases to dataspaces. *ACM SIGMOD Record 34*, 4 (2005), 27–33; https://doi.org/10.1145/1107499.1107502.
3. Gaia-X. Gaia-X Architecture Document (21.09 Release). Gaia-X European Association for Data and Cloud AISBL, 2021, Brussels, Belgium; https://bit.ly/3sJkUsd
4. IDS Association. Reference Architecture Model (Ver. 3.0). International Data Spaces Association. Berlin, Germany, 2019; https://internationaldataspaces.org/download/16630/.
5. Jarke, M., Otto, B., Ram, S. Data sovereignty and data space ecosystems. *Business & Information Systems Engineering 61*, 5 (2019), 549–550; https://doi.org/10.1007/s12599-019-00614-2.
6. Redman, T.C. The impact of poor data quality on the typical enterprise. *Commun. ACM 41*, 2 (Feb. 1998), 79–82; https://doi.org/10.1145/269012.269025.
7. Tayi, G.K. Ballou, D.P. Examining data quality. *Commun. ACM 41*, 2 (Feb. 1998), 54–57; https://doi.org/10.1145/269012.269021.

Boris Otto is Chair for Industrial Information Management at TU Dortmund University and Director of the Fraunhofer Institute for Software and Systems Engineering in Dortmund, Germany.

Copyright held by author/owner. Publication rights licensed to ACM.