Creating Data Spaces based on GAIA-X and IDS

Boris Otto

Executive Director, Fraunhofer ISST;
Professor for Industrial Information Management, TU Dortmund;
Member of the Board, IDS Association

Boris Otto's research focus lies on industrial data management, data ecosystems and data sovereignty. He looks back on more than 20 years of experience both in the scientific and in the practitioners' community. He is involved in several data space research and implementation projects.

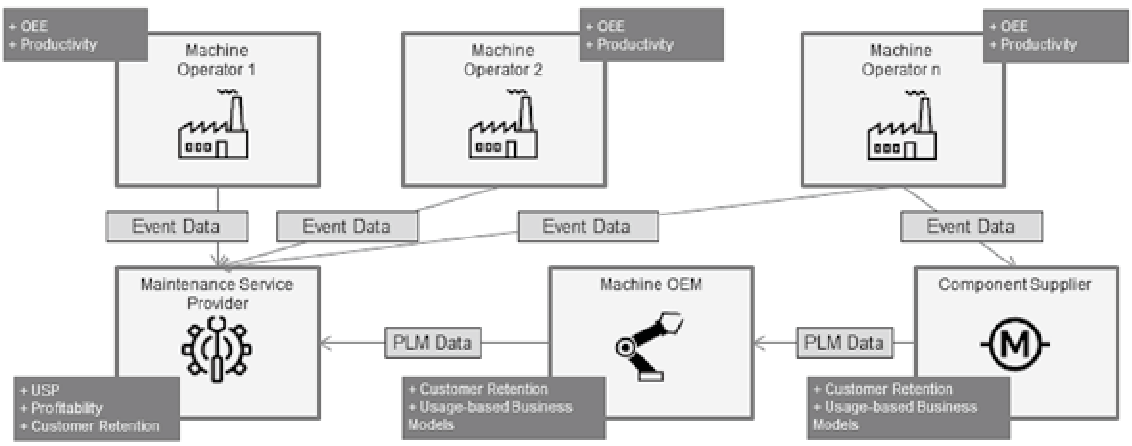
Data has become a strategic resource for competitive advantage of business in all sectors and for societal prosperity. Because of that, demands for data sovereignty are growing. Data sovereignty can be seen as the capability of data providers to act in a self-determined way when it comes to managing, using, and sharing their data. To live up to these demands, calls for trusted data infrastructures and data spaces have been articulated both inside and outside the European Union. This paper motivates data sharing as a key capability in the data-driven economy, discusses the notion of data spaces and introduces the contribution of GAIA-X and the International Data Spaces (IDS) initiative to the creation of European data spaces.

1. The Data-Driven Economy

1.1 An Industrie 4.0 Example

Condition monitoring of production assets has long been a prerequisite for predictive maintenance. The use of data from the manufacturing processes such as vibration parameters, power consumption, temperature parameters etc. of a welding robot, for example, is a source of knowledge to optimize asset maintenance in automotive production. The analysis of these data may lead to better configuration of the robot, more efficient scheduling of maintenance intervals, i.e. increased overall equipment effectiveness (OEE). This effect increases with the number of data sources and the availability of data to be used for analytics. Thus, operators are interested in sharing their data because they can expect economic benefits and every single of them has an interest that other companies operating the same asset share their data as well. So, everyone profits from the fact that others share their data as well. Because of that, though, there is a strong interest of all operators that data is not misused. Data must solely be used for the purpose determined by the data provider. In addition to that, other operators must only be allowed to use the results of the analytical processes but must not be provided access to the raw data from the individual source.

Figure 1 outlines this collaborative predictive maintenance scenario in which all participants, i.e. machine operators,



Legend: PLM – Product Lifecycle Management. USP - Unique Selling Point

Source: Fraunhofer ISST, 2020.

Figure 1 Collaborative Predictive Maintenance Scenario maintenance service providers and machinery manufacturers benefit from data sharing. In fact, the individual participants benefit if the ecosystem as a whole is working properly.

Prerequisites for this are interoperability, a sustainable business model for all ecosystem members, data sovereignty, and trust[1].

1.2 Data Sharing in Ecosystems

The Industrie 4.0 example points to the fundamental characteristics of data-driven services as they can be found in many sectors.

First, value creation and value capture are based on smart services that address a business problem instead of providing a tangible product. In fact, in many industrial use cases value creation increasingly stems from the combination of digital services and "traditional" services such as maintenance and the delivery of tangible products.

Second, in order to provide such a smart service, different types of data must be combined from different sources. Typically, no single company possesses all the data required. Instead, in order to create a data ecosystem, an orchestration of various data items and sources is needed.

Third, data sovereignty of the data provider and trust of all involved members of the ecosystem must be ensured. Trust, for example, refers to the fact that data providers and data consumers can rely on the identity of other members and on the fact that data is only used under specified conditions.

Fourth, as in the given example, the offering of smart services in many cases requires the existence of a cloud infrastructure. Traditional application systems do not provide the required analytical and data sharing functionality because they are designed to support mainly internal business processes and are run behind a company's firewall. For data-driven services to be used in ecosystems, though, Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) are required. Cloud sovereignty in this context refers to the capability of entities to act in a self-determined way when it comes to where their data are stored and on which cloud services the software runs that manage and use their data.

1.3 Data Strategies in Europe

In 2019, the German Federal Government published principles of their data strategy[2]. The document acknowledges the fact the data shall be more accessible and more used in the future while at the same time the interests of individual data providers, i.e. data sovereignty, shall be ensured. Among identified fields of action are the improvement of data access and data provisioning, responsible use of data and leveraging of innovation potential, increasing "data competence" and establishing a "data culture" to make the state a "front-runner" of the data economy.

A similar position was taken in the Netherlands. The "Dutch Vision on Data Sharing between Businesses" identifies three principles[3]. First, voluntary data sharing is preferable. Second, data sharing will be made compulsory if necessary. Third, people and businesses retain control over their data.

In February 2020, the European Commission published a European Strategy for Data[4] which acknowledges the importance of the role of data for the European Single Market and points to the innovation potential that data brings about. Furthermore, it calls for common rules and mechanisms that ensure that "data can flow within the EU and across sectors", that "European rules and values […] are fully respected", and that "the rules for access to and use of data are fair, practical and clear", with "clear and trustworthy data governance mechanisms in place". Nine common data spaces are to be built, e.g. in manufacturing, healthcare and mobility. This action plan was recently confirmed when both the European Commission and the Member States on 15 October 2020 signed a joint declaration aimed at "building the next generation cloud for businesses and the public sector in the EU"[5].

2. Designing Data Spaces

2.1 Data Spaces Defined

With the explicit call for nine European data spaces in the European Strategy for Data, demand increases for knowledge what data spaces actually are and how to create them.

In general, a data space is a data integration concept that, for the first time, was discussed in the scientific information systems and computer science communities around the year 2005. Four fundamental design principles for data space can be identified[6,7].

First, data spaces do not require physical data integration, but support distributed data stores. In fact, data spaces do not require central data store architectures as discussed, for example, as "data lakes" over recent years.

Second, data spaces do not require a common schema Instead, integration is achieved foremost on semantic level, mainly through the so-called vocabularies that make use of the Resource Description Framework (RDF).

Third, data spaces envisage data networking, data visiting and data co-existence, i.e. a certain redundancy of data.

Finally, data spaces allow for nesting and overlapping, i.e. data spaces themselves can be part of other, larger data spaces. In fact, data spaces grow through the evolution of the number of their participants, i.e. data sources and data sinks.

The International Data Spaces (IDS) initiative added three further design principles that take up on the demands for data sovereignty, data traceability and trust between the data space participants[8].

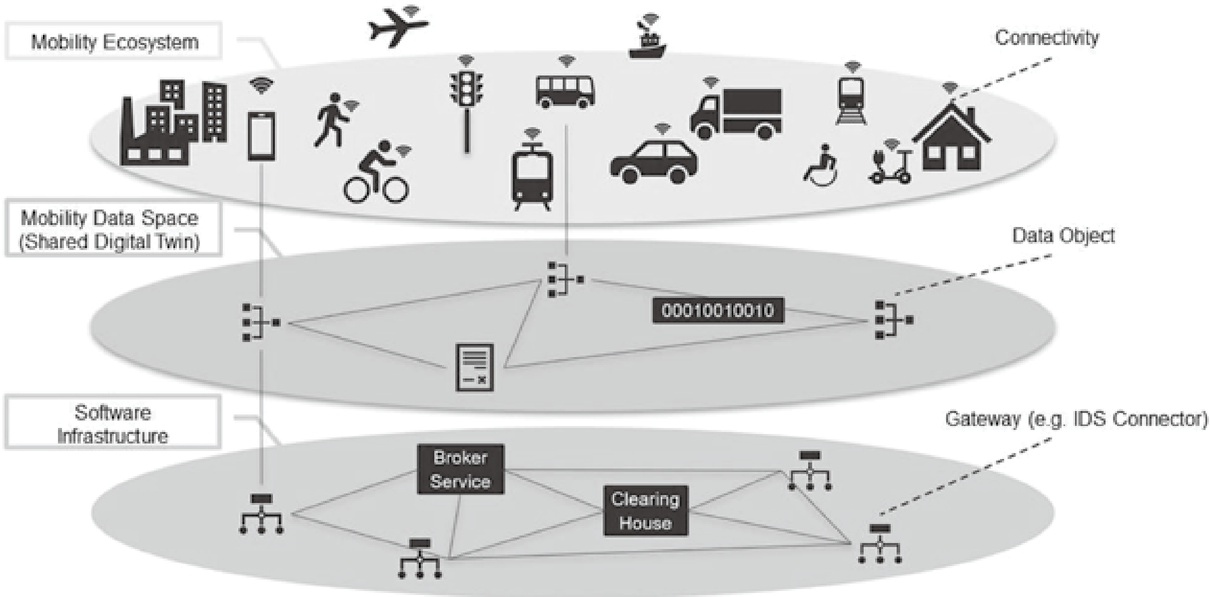
Thus, the concept of data spaces meets well with European requirements for management and use of data. Furthermore, its decentralized design matches with the nature of Industrie 4.0-based use cases and business models. The Internet of Things, as one key technology of Industrie 4.0, is decentralized as well. Cyber-physical systems (CPS) are processing data and even taking decisions on a local level without requiring a single central data store.

2.2 The Role of Data Spaces in Data Ecosystems

Data spaces play a central role when it comes to implementing data strategies in different sectors and domains.

Figure 2 shows the high-level architecture for a Mobility Data Space as it is currently built in Germany. The architecture consists of three layers. The top layer forms the mobility ecosystem in which ecosystem members (e.g. travelers, public transport providers, car sharing companies etc.) provide and consume end-to-end, inter-modal mobility services. These services aim at increased convenience for the traveler on the one hand side and at more efficient and effective management of traffic and passenger streams on the other hand.

According to the characteristics of data sharing in ecosystems as outlined in section 1.2, different data from multiple sources need to be accessed and combined to be able to provide these smart mobility services. The Mobility Data Space forms the middle layer of the overall architecture and aims at addressing this requirement by creating a shared digital twin of the various real-world objects in the mobility domain. In general, a digital twin consists of three parts, namely the physical items in the real world, their digital representations in the virtual space, and the bidirectional connections of data and information that links virtual and real objects[9]. For data to be shareable, data interoperability must be achieved. Thus, an unambiguous understanding about commonly used data objects is necessary among the members of the data ecosystem. Charging points for electric mobility, for example, have to be described by a consistent set of attributes (e.g. type, location, charging mode, charging



Source: acatech, Fraunhofer ISST, 2020.

Figure 2 Data Spaces in the Mobility Sector levels etc.) and attribute values.

The lower architecture layer specifies the software infrastructure that is required to support the creation, management and sharing of digital twin data. As mentioned above, a decentralized architecture design meets the requirements in various domains better than a central approach. However, a decentralized – or federated – architecture requires standard software components, which allow data providers and users to take part in the data space (and in the end be an ecosystem member). Furthermore, common services are needed to make sure data exchange and data sharing are possible among the different decentralized software components. Examples are broker services that connect data sources and data sinks and clearing houses that monitor data transactions.

2.3 Data Spaces Design Tasks

The architecture layers mentioned above serve as a blueprint for structuring the work when it comes to setting up projects to create a data space in a specific domain.

Figure 3 shows the "onion model" of data spaces design tasks. The onion core addresses the implementation of the software infrastructure for the data space. In the example, this software infrastructure is based on the IDS Reference Architecture Model[10]. The infrastructure system is made available by a provider organization that can be either a single entity or a consortium of multiple actors in the respective domain.

The second layer in the onion model addresses the commonalities of the data space including vocabularies, the integration of existing data platforms, and the specification of usage contract templates for the exchange and sharing of the data. The tasks in this category cannot be addressed by one individual member of the data ecosystem but have to be taken care of in a multilateral fashion.

In contrast to that, the outer ring of the onion model is dedicated to the individual use cases and related business models on top of the inner rings.

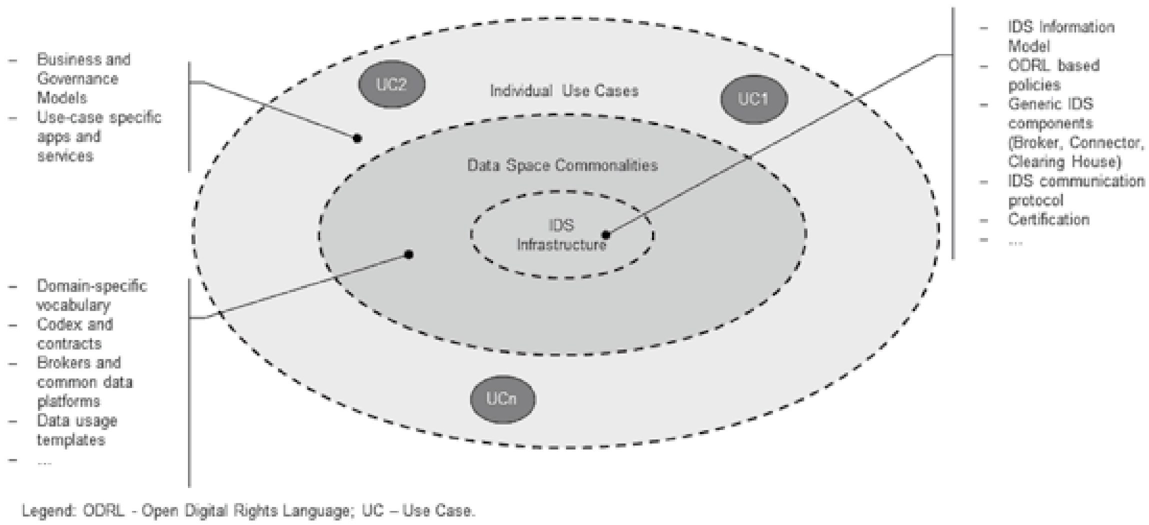
3. IDS and GAIA-X

3.1 International Data Spaces (IDS)

The IDS initiative has formed in 2015 with the ambition to set a standard for data sovereignty. A research project carried out by Fraunhofer formed the core of the initiative resulting in a first version of the reference architecture and first prototypes of the different software components (e.g. IDS Connector, Broker etc.).

Parallel to the research activities, the IDS Association was founded in 2016 as a non-for-profit organization in Germany. Over the years, the association has grown to more than 125 member organizations from more than 20 countries over the world.

The reference architecture and in particular the specification of the IDS Connector was taken up by standardization activities. DIN SPEC 27070 specifies requirements and a reference architecture of a security gateway for the exchange of industrial data and services[11].



Source: acatech, Fraunhofer ISST, 2020.

Figure 3 "Onion Model" of Data Space Design Tasks

As the IDS Association is non-for-profit and concentrates on developing the standard and forming an organizational platform for its member community, the standard itself is taken up by individual companies and consortia to build the software infrastructure which is needed – as outlined in section 2 – to implement data spaces.

3.2 GAIA-X

GAIA-X is an initiative started by the German and French governments to achieve data and cloud sovereignty in Europe. After a first presentation of the initiative's objective at the German Digital Summit in Dortmund in 2019, a community has formed to work on fundamental design principles and first architecture specifications for GAIA-X. Overall, GAIA-X "will support the development of a digital ecosystem in Europe, which will generate innovation and new data-driven services and applications. To this end, GAIA-X will enable interoperability and portability of infrastructure, data and services and establish a high degree of trust for users"[12].

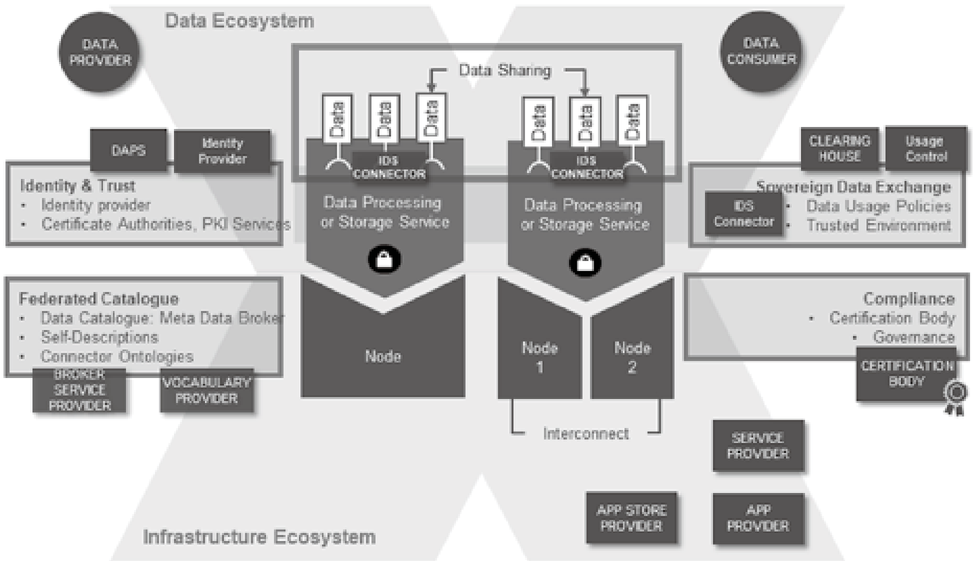
GAIA-X will institutionalize as an international non-for-profit organization according to Belgian law. A first milestone was reached on 15 September 2020 when the notarial deed for its incorporation was passed. Now, the association waits for the required royal decree by the King of Belgium to obtain its full legal personality.

Similar to the IDS Association, the GAIA-X association will not operate a monolithic cloud data center, but rather forms a standard for cloud and edge-cloud services to adhere to, in order to demonstrate its compliance to European values with regard to data and cloud sovereignty.

Thus, the deliverables of the GAIA-X association will be fourfold. First, it will specify the GAIA-X architecture of standards and related test criteria which allow to determine and test whether a cloud and/or edge-cloud services meets the GAIA-X standard. Second, the GAIA-X association will ensure the offering of fundamental services needed for the GAIA-X ecosystem. This may include the issuing of digital certificates or tokens as well as a federated catalog of GAIA-X compliant services. Third, the GAIA-X association will support the organization and management of an open source software community, which implements the GAIA-X architecture of standards and, in particular the federated services[13]. Fourth, the GAIA-X association will be prepared to support the creation of the data spaces as outlined in the European strategy for data (see above).

3.3 An Integrated View

Because of its history, the IDS reference architecture is more mature than the GAIA-X architecture specification. Both initiatives, though, share the same goals and are based on the same design principles, in particular decentralization



Legend: DAPS – Dynamic Attribute Provisioning Service

Note: This architecture viewgraph does reflect a Fraunhofer working status and is no official work result published by the GAIA-X initiative or the GAIA-X AISBL under incorporation.

Source: Fraunhofer, 2020.

Figure 4 Integrated Architecture View on GAIA-X and IDS and federation. Furthermore, the IDS Association is a founding member of the GAIA-X association in order to make sure that components of the IDS Reference Architecture are ready to be taken up the GAIA-X architecture of standards.

Figure 4 shows a first sketch of an architecture that integrates GAIA-X and IDS concepts. GAIA-X aims at addressing both data and cloud sovereignty and proposes four federated services, namely "Identity & Trust", "Sovereign Data Exchange", "Federated Catalogue", and "Compliance". The IDS Reference Architecture specifies software components that can be taken up by the federated services, namely the "IDS Connector", "Broker", "Clearing House", "Identity Provider", and "App Store".

The integration of GAIA-X and IDS on an architectural level has significant potential for a timely realization of the European data spaces.

4. Success Factors

Data spaces represent a relatively new class of information systems, in particular when to be created on an open community or even national or European level for an entire domain.

Because of its technical novelty and its federated design, it brings about a couple of challenges that need to be overcome to be successful.

First, from an organizational or institutional perspective there is a difference between the organization that implements and the organization that operates the fundamental infrastructure services required for data spaces. GAIA-X federated services, for example, must be there before individual cloud and edge cloud services can evolve which then form the foundation for a data ecosystem. The individual cloud services and IDS Connector use cases, though, follow specific business case needs and, thus, are not part of the infrastructure. The operating model of the latter, though, follows the general principles well known for "traditional" national infrastructures such as motorway and railway systems. Therefore, cooperative and public-private partnership models are discussed.

Second, and related to the question of operating models, the funding and financing model must be clarified. According to the nature of an infrastructure, combined financing models that include both public and private funding sources seem useful to for the task at hand.

Third, incentive systems are needed for companies to share their data. At present, many businesses are reluctant when it comes to sharing their data. While the benefits of exchanging and sharing data are understood, many companies fear the loss of sensitive data and the lack of transparency about what happens to their data after being shared[14]. A simple, but promising design principle for data ecosystem use cases is "quid pro quo". Companies seem to be much more willing to share data when data sharing is no one-directional endeavor, but instead is based on bi-directional exchange of data.

This is very much in line with the concept of business ecosystems in general which can be defined as a multilateral form of organizing to achieve an innovation objective that cannot be achieved by any individual ecosystem member on its own.

---

[1] Bundesministerium für Wirtschaft und Energie (BMWi): Kollaborative datenbasierte Geschäftsmodelle. Berlin, 2020.

[2] Bundesregierung: Eckpunkte einer Datenstrategie der Bundesregierung. Berlin, 2019.

[3] Ministry of Economic Affairs and Climate Policy: Dutch vision on data sharing between businesses. Amsterdam, 2019.

[4] European Commission: A European strategy for data. Brussels, 2020.

[5] European Commission, Ministers responsible for Telecommunications/ Digital Policy: Building the next generation cloud for businesses and the public sector in the EU. Baden-Baden, 2020.

[6] Borjigin, C. et al.: Dataspace and its application in digital libraries. The Electronic Library 31 (6), 2013, pp. 688-702.

[7] Franklin, M. et al.: From Databases to Dataspaces: A New Abstraction for Information Management. SIGMOD Record 34 (4), 2005, pp. 27-33.

[8] Otto, B.; Jarke, M.: Designing a multi-sided data platform: findings from the International Data Spaces case. Electronic Markets 29 (4), 2019, pp. 561-580.

[9] Grieves, M.; Vickers, J. (2017): Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: F.-J. Kahlen, S. Flumerfelt, A. Alves (Eds.), Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches (pp. 85-113). Cham, Switzerland: Springer International Publishing.

[10] IDS Association: Reference Architecture Model, Version 3.0. Berlin, 2019.

[11] DIN: Anforderungen und Referenzarchitektur eines Security Gateways zum Austausch von Industriedaten und Diensten (DIN SPEC 27070). Berlin, 2020.

[12] Bundesministerium für Wirtschaft und Energie (BMWi): GAIA-X: The European project kicks off the next phase. Berlin, 2020.

[13] *See* Bundesministerium für Wirtschaft und Energie (BMWi): GAIA-X: The European project kicks off the next phase. Berlin, 2020.

[14] PricewaterhouseCoopers GmbH: Datenaustausch als wesentlicher Bestandteil der Digitalisierung. Düsseldorf, 2017.