Data spaces for data ecosystems

Boris Otto[1,2]

[1]*TU Dortmund University, Joseph-von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany*
[2]*Fraunhofer ISST, Emil-Figge-Str. 91, 44227 Dortmund, Germany*

Abstract

The keynote talk motivates the sharing of data within ecosystems as a prerequisite for data-driven innovation and proposes data spaces as an appropriate data infrastructure pattern in this regard. It puts the current activities concerned with building and scaling data spaces in the context of the European strategy for data which calls for the establishment of common European data spaces. Furthermore, the talk introduces conceptual and technical foundations of data spaces and points to recent developments in practice, such as Gaia-X and the IDS Association.

Keywords

Data space, data ecosystem, data sharing

1. The Role of Data Spaces in the European Data Economy

Many data-driven innovation scenarios require exchange and sharing of data among many different partners within an ecosystem. Catena-X is an example from the automotive industry which is characterized by distributed value creation across the entire production and supply network, demands for supply chain transparency from the original equipment manufacturer (OEM) over tier 1 and tier 2 suppliers to the raw material suppliers and shortage situations of components and parts (e.g. semiconductor components). Catena-X aims at end-to-end data value chains to allow for every ecosystem member to cope with current challenges such as carbon footprint transparency and supply chain act compliance.

A second example can be found in the mobility domain. End-to-end inter-modal mobility services can be only achieved if the different service providers in this case share information about the mobile traveller, their preferences, payment information, about time-tables and context information such as weather and traffic.

In fact, ecosystems emerge in situations where innovation cannot be achieved by one company alone, but where different data need to be used and re-used collaboratively. Thus, ecosystems can be understood as a multilateral form of organizing for joint customer innovation. They balance the viability of the ecosystem at large and of its individual members. At present, data ecosystems emerge in different domains such as healthcare, mobility, and manufacturing.

Data spaces provide data infrastructures for data ecosystems and play an important role for the implementation of the European data strategy which calls for the establishment of common

European data spaces. Apart from that, data spaces form a layer in an overall architecture stack digital architecture stack and, thus, are important in the ongoing debate regarding technology sovereignty in Europe.

To differentiate data spaces from data ecosystems on a conceptual level, applying an architecture approach is useful. Typically, private data from different data providers need to be combined with context (often open) data. Examples can be found in healthcare (personalized medicine), smart cities (traffic management, multi-modal mobility services), and manufacturing (collaborative supply chain management, end-to-end supply chain transparency). Thus, data sharing enables co-opetition in ecosystems when every individual member gives something to gain something. With regard to data sharing in ecosystem, it is clear that a balance is required between using the data and protecting the data. In this context, data spaces represent a promising data integration approach as they embrace a federated data architecture and typically come with measures which foster data sharing while ensuring trust and data sovereignty among participants.

2. Data Spaces Foundation

Research on data spaces has its roots in semantic web and Linked Data research. In general, a data space is a distributed integration concept which does not require physical data integration or a common schema. As mentioned above, data spaces are seen as a promising approach to support data ecosystems which results in a set of business requirements. Among these are, for example, support of different data ecosystem roles (such as data provider, data user, and data sharing service intermediary), traceability of data in the ecosystem as well as policy management, and trust among participants.

Gaia-X and the International Data Spaces (IDS) are initiatives kick-started in Europe, which aim a setting de-facto standards for data spaces and, hence, supporting European regulation (e.g. Data Governance Act and Data Act). The IDS Reference Architecture Model (RAM) envisages a set of essential services to support data spaces. Among these are a broker service, a clearing house and an app store service. Apart from that, the IDS RAM identifies the so-called IDS connector as a key component. It provides access to data sources, manages policies which constrain the use of the data and support the exchange of the data between data provider and data user as well as to the various essential services. Policies are articulated as rules in the IDS RAM. Typical data policies constrain the use of data for a certain period of time, allow/prohibit forwarding of data and determine the number of read accesses to the data by a data user.

In this context, data sovereignty can be understood as the capability of a legal entity or natural person to be self-determined regarding their shared data. Interoperability, traceability, and enforcement of usage policies allow for executing data sovereignty in data ecosystems.

Gaia-X is a non-for-profit initiative aiming at setting de-facto standards for data sovereignty in the cloud. It envisages four so-called federation services, namely "identity and trust", "sovereign data exchange", "federated catalogue", and "compliance". Gaia-X and the IDS Association work together in the Data Spaces Business Alliance (DSBA) on conceptual consistency and architectural convergence.

Both initiatives aim at three deliverables, namely specifications, open-source software, and

compliance/certification tools.

3. Outlook

Today, the European data economy is characterized by the genesis and emergence of individual, open data ecosystems. To achieve the vision of common European data spaces, however, interoperability and sharing of data not only within individual data spaces but between them is required.

Apart from that, critical mass must be achieved when it comes to adoption of the fundamental architecture building blocks and software components. The EDC project is an open-source project hosted by the Eclipse Foundation aiming an providing open-source implementations of the most important data spaces components. It is coordinated by Fraunhofer and supported by a number of large industrial partners.

Finally, the EU Data Spaces Support Centre will define common building blocks for data spaces as a recommendation for the various data ecosystems which have already been started or will be started soon.