DIO

A SHORT INTRO INTO DATA SPACES

DIO

Data Intelligence Initiative
Hintere Zollamtstraße 17 / 3rd Floor
1030 Vienna

DATA SPACE - BASICS

A data space, also referred to as a data room, refers to a type of trusted data relationship between partners and organizations that apply the same standards and rules for storing, processing and sharing their data. However, what is most important for the concept of a data room is that the data is not stored centrally, but distributed at the respective source and is therefore only used collectively when needed in the context of common use cases. Data rooms focus on domains (economic sectors, industrial sectors or other specialist fields of application) and make metadata available for potential innovative applications while maintaining data sovereignty, i.e. the greatest possible control and mastery over one's own data. Domain-specific data spaces can also connect (federate) with other data spaces, such as a data space for agriculture and forestry and one for energy or the EU Green Deal.

Sovereignty

Sovereignty means that participants retain full control over their data. They decide for themselves who has access to their data, how it is used and under what conditions it may be shared. This data sovereignty is essential to create trust between stakeholders and ensure that data is only used in accordance with the interests of the data owners.

Decentralization

Data centrality means that data is not stored or managed centrally in one place, but distributed across different participants and systems. Each actor retains control over their own data and only shares it as required and under defined conditions. This structure promotes the autonomy of the participants, reduces dependencies on central instances and increases the resilience of the overall system.

Trust

means the participants' trust in the integrity and security of the shared data. This includes the certainty that the data originates from reliable sources, that data protection and data sovereignty are maintained and that the infrastructure is secure against unauthorized access. This trust is crucial to enable the open and secure exchange of data between different stakeholders.

Identity

Every entity involved, whether human, machine or organization, needs a unique digital identity to ensure that only authorized actors can access and share certain data. These identities make it possible to track activities within the data space, build trust and ensure compliance with data protection and security standards.

- A data space architecture consists of various technical components that together ensure that data can be exchanged and traded transparently, securely and reliably within a data space. This modular structure offers developers the opportunity to customize data spaces and their access and use.

- On the one hand, there is the option of using connector frameworks as the basis for point 1 in order to incorporate extensions and implement specific solutions. Many of these connector frameworks are available as free and open source software (FOSS), for example the Eclipse Data Space components and the FIWARE ecosystem with the TRUE Connector.

- For detailed information on various connector implementations, please refer to the connector report of the International Data Space Association (IDSA), which is published at regular intervals.

- When developing and implementing technical components, common standards are essential to ensure interoperability. One standard that is currently in the development phase is the Data Space Protocol.

DATA SPACE

There are different roles and functions within a data space, and here is a brief overview of the most important ones based on the IDSA:

1

Connector

The connector enables companies to access the IDS ecosystem in accordance with the guidelines of the Reference Architecture Model and the IDS certification criteria. The dataspace protocol regulates data exchange, usage agreements and data access in dataspaces for technical interoperability.

2

Identity Provider

The exchange between companies regarding identity management takes place via the connector, which uses an X509v3 certificate. Identity attributes are transferred via dynamic tokens. An attribute server manages the connector properties and issues tokens as required, which makes identity management more flexible.

3

Clearing House

The clearing house logs data exchange and enables the billing of fee-based usage and quid pro quo transactions. It stores transactions, participants and contracts in a verifiable manner and serves as a backend for logging services in various applications.

4

Information model

The model describes actors and resources in a data space, serves as a schema for self-disclosures, message headers and metadata of data resources. It is freely available and is being further developed on GitHub.

5

App Store

Connectors use different app types: Apps

- developed in-house for the connector
- Third-party apps from the App Store Apps
- from the data consumer connector

Apps are divided into categories: System adapters for

- system access and data model customization
- Intelligent data apps for data processing Other
- apps with specific functions

6

Metadata Broker

The search for data and components requires decentralized solutions. The metadata broker serves as a connector that provides interfaces for communication with other connectors. It indexes self-descriptions, enables search functions such as full-text search and SPARQL queries, as well as services such as heartbeats for detecting inactive connectors and quality of service metrics.

7

Usage Control

The protection of sensitive information and compliance with data protection regulations are crucial for companies. Usage control and connectors play an important role by automatically integrating usage restrictions and metadata. Usage rights are based on ODRL and Provenance Tracking ensures transparency.

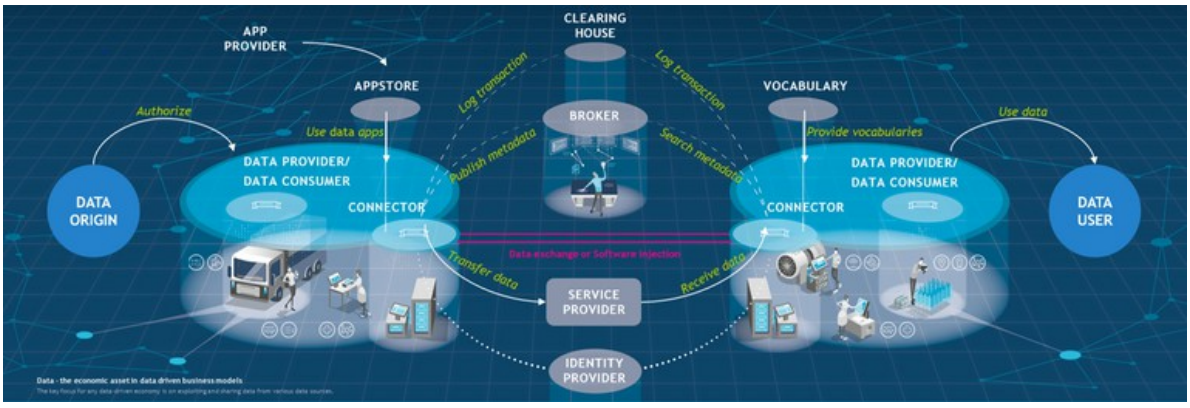

Fig. 1: Data driven Economy in Data Spaces
(https://www.dataspaces.fraunhofer.de/de/software.html)

DATA SPACE USECASES

The DIO uses a use case canvas to specifically define use cases and cover all relevant aspects from the outset. If a data space is already at an advanced stage and several use cases have already been implemented, the use case canvas can be adapted more specifically to the respective requirements.

The IDSA canvas is used within the DIO.

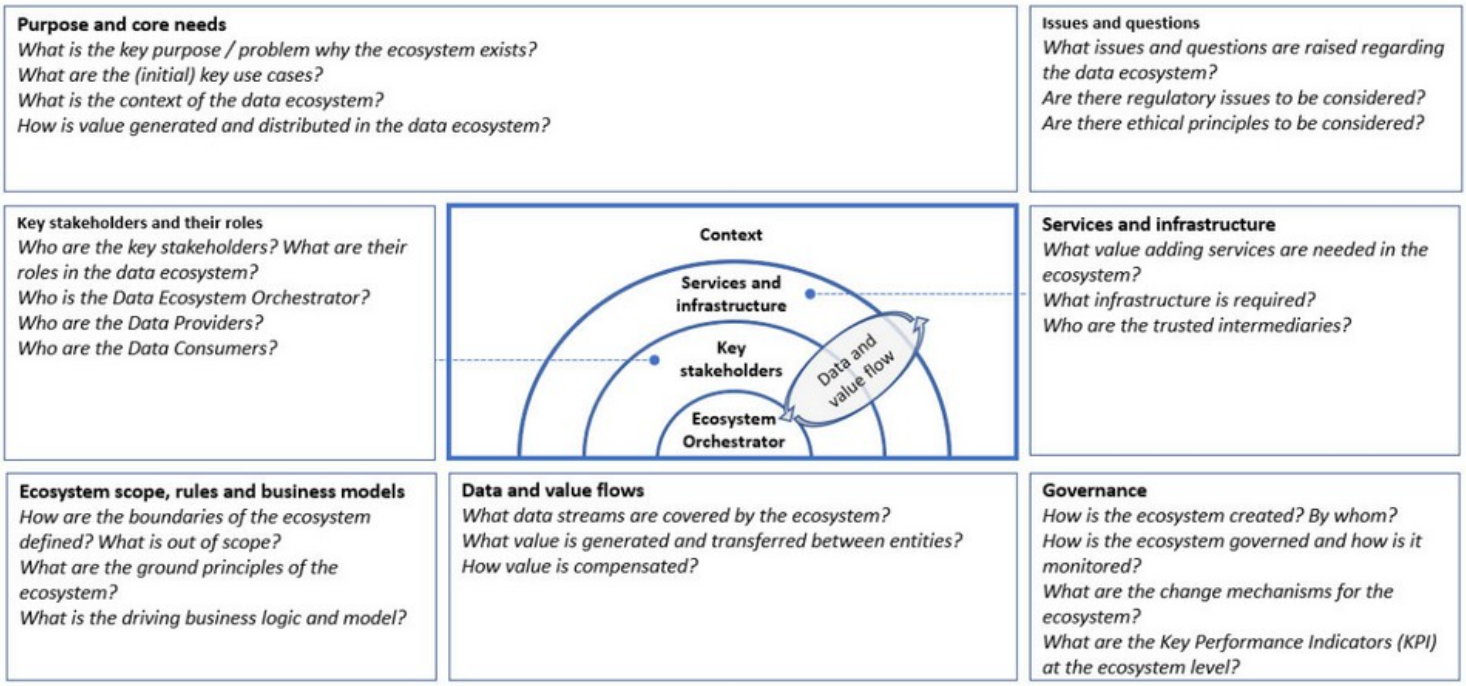

Fig. 2: IDSA Use Case Canvas
(https://docs.internationaldataspaces.org/ids-knowledgebase/v/how-to-build-data-spaces/2_define-your-use-case)

Example of data space circular economy

Ten ideas for improving the circular economy were initially identified within the Circular Economy Data Space. The idea of a smart recycling platform was defined in more detail here. The aim is to define a digital platform that shares real-time data on the availability and demand for recyclable materials between stakeholders in order to promote the circular economy.

CONCEPTUALIZATION DATA SPACE

Phase 1: Initiation and planning

- Identification and involvement of stakeholders
  - Identification of key stakeholders from various industries, the public sector and civil society.
  - Holding initial meetings to discuss the project objectives and benefits.
- Definition of the project scope and integration of data spaces
  - Definition of specific objectives, deliverables and consideration of data spaces and decentralized data exchange in the project scope.
- Resource allocation
  - Determination of the required financial, human and technological resources
  - Development of the budget plan
- Risk assessment
  - Identification of potential risks and development of mitigation strategies.
- Development of a project schedule
  - Creating a detailed schedule with milestones

Phase 2: Development of use cases and collaboration

- Workshop series, including data space architecture
  - Organization of workshops for knowledge exchange and joint brainstorming, with a focus on data spaces and decentralized data exchange.
- Selection of use cases with a focus on data spaces
  - Evaluation and selection of the most promising use cases for development, especially those that want to use Data Spaces.
- Development of a project plan for each use case
  - Development of thematic project plans for each use case, including specific objectives, resources and timelines.
- Cross-industry collaboration
  - Regular meetings and collaboration sessions between stakeholders.

Phase 3: Implementing and testing

- Implementation of data spaces in solutions
  - Integration of data spaces and decentralized data exchange mechanisms into the developed solutions.
- Piloting and evaluation of decentralized data exchange
  - Review of the effectiveness and security of data exchange in the pilot projects.
- Feedback and iteration
  - Obtaining feedback and refining solutions based on the feedback.

Phase 4: Scaling and productivity

- Preparation and scaling
  - Preparation of case studies and reports
  - Planning for scaling successful use cases Going live
  - Establishment of a business plan and activation of future use cases.

AUTHORS:

Anna Neureiter

Günther Tschabuschnig

Data Intelligence Initiative
Hintere Zollamtstraße 17 / 3. Floor
1030 Vienna