How to build secure Dataspaces with (F)OSS technologies in the context of IDS and Gaia-X

Markus Spiekermann
Head of Department Data Business, Fraunhofer ISST
Project Lead of Eclipse Dataspace Connector

- *Introduction*
- *Business Ecosystems*
- *Data Spaces*
- *Eclipse Dataspace Project*
- *Key Takeaways*

Motivation

»A real data economy, on the other hand, would be a powerful engine for innovation and new jobs. And this is why we need to secure this data for Europe and make it widely accessible. ***We need common data spaces*** *- for example, in the energy or healthcare sectors. This will* ***support innovation ecosystems*** *in which universities, companies and researchers can access and collaborate on data. And it is why we will build a European cloud as part of NextGenerationEU - based on* ***GaiaX****.«*

Business Ecosystems

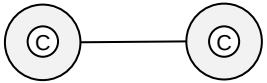
Business Ecosystems

- *Cluster organizations from various interests (e.g. domain)*
  - *Including service provider and operating companies*
- *Enable collaboration for innovation and business models*
- *Elaborate on future requirements and challenges to be addressed*
- *Define common governance rules with democratic structures*
- *Require openness for new participants and technology*

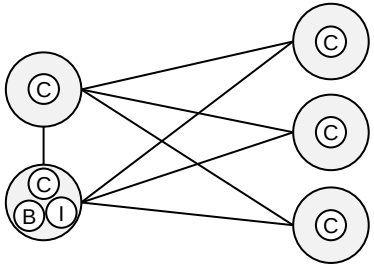



Images from freepik.com by tawatchai07 and aleksandarlittlewolf

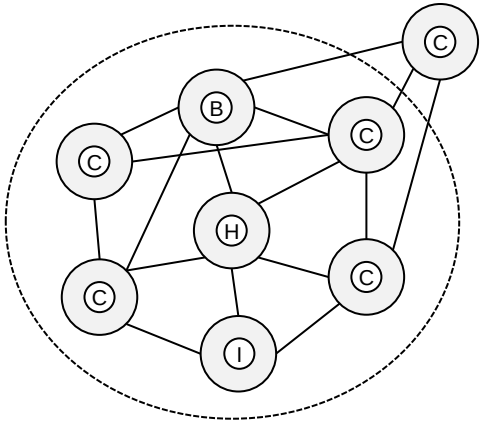
Change of Data Exchange and Sharing



Bilateral data exchange



Closed group data exchange



Open and dynamic data exchange

Mobility Data Space

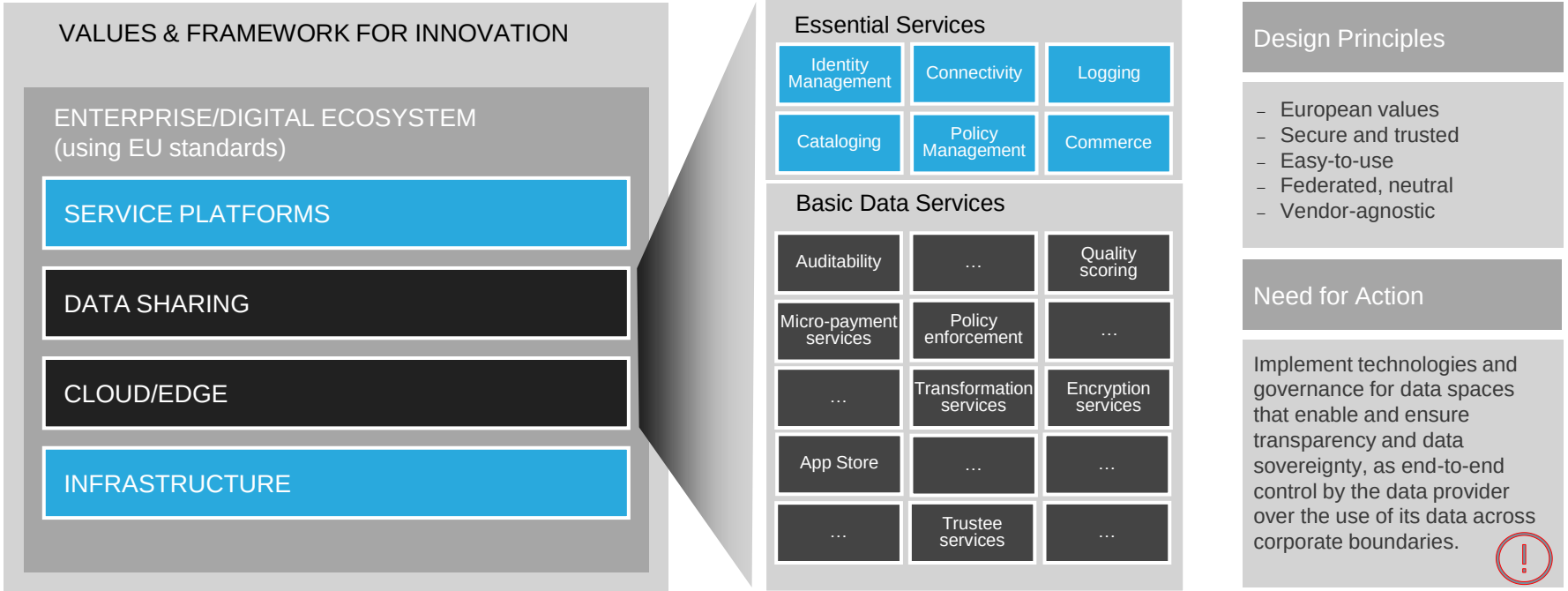
- *Initiated by the German federal government's "Concerted Action on Mobility" committee in 2019*
- *Data sharing community to build the future of mobility*
- *Promotes forward-looking mobility services*
- *More than 200 stakeholders of German mobility landscape, science, business and government*
- *20+ use cases that were presented on the ITS Worldcongress 2021*
- *Productive operation planed from early 2022*

Catena-X Automotive Network

- *Founding of Catena-X Automotive Network e.V. took place on 07.05.2021.*
- *Alliance for secure and standardized data exchange along the automotive value chain*
- *Offer network and technologies for collaboration and innovation*
- *Ensure the economic viability of all network partners*
- *Technical components and services incl. transfer and scale out*
- *Initial use cases, e.g. Traceability, CO_2 Footprint, Circular Economy, Demand and Capacity Mgt.*
- *Consortium of Industry, technology and platform experts*

Data Spaces

What does a Data Space bring to the Table?



Design Principles

- European values
- Secure and trusted
- Easy-to-use
- Federated, neutral
- Vendor-agnostic

Need for Action

Implement technologies and governance for data spaces that enable and ensure transparency and data sovereignty, as end-to-end control by the data provider over the use of its data across corporate boundaries.

International Data Spaces

	IDS	Gaia-X
Rationale	How to keep data sovereignty in data sharing?	How to keep sovereignty over data and services in the cloud?
Institutionalization	Non-for-profit registered association according to German law	Non-for-profit registered association according to Belgian law
Members	130+	310+
Data Space Scope	Within data spaces	Within data spaces Across data spaces
Key Deliverables	1) Specification 2) Certification 3) OSS Community	1) Specification 2) Labelling and Compliance
Geographical Focus	International	International w/ strong European »flavor«
Architecture Design	Federated	Federated
Key Architecture Elements	– IDS Connector – IDS Essential Services (e.g. Clearing House, Broker etc.)	– Gaia-X Registry – Gaia-X Trusted Framework – Gaia-X Federation Services
Certification Strategy	– Component Certification – Certification of Operations Environment (based on ISO 27001)	– Self-Certification based on Self-Descriptions – Compliance Service for claim validation

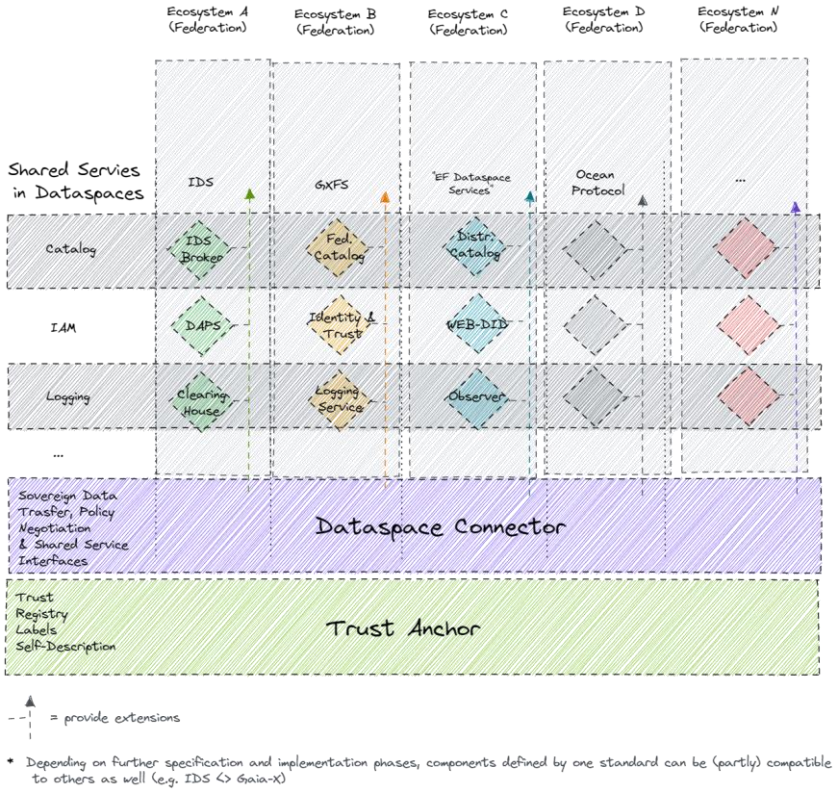
New challenges arise!

- *Catena-X, Mobility Data Space, and other initiatives present various sharing challenges*
  - *Support for data flow and transfer protocols to handle diverse data types*
    - *Push, Streaming, Large Volumes, Realtime, Code2Data*
  - *Cataloging across many providers*
  - *Policy management that traverses multiple infrastructure layers*
  - *Need for extensibility and modularity to accommodate diverse needs and use cases*
- *Different architectures and implementations for data space services*
  - *Organizations participate in various data spaces*
  - *Connection and interoperability with multiple data spaces*
  - *Identity across multiple jurisdictions*

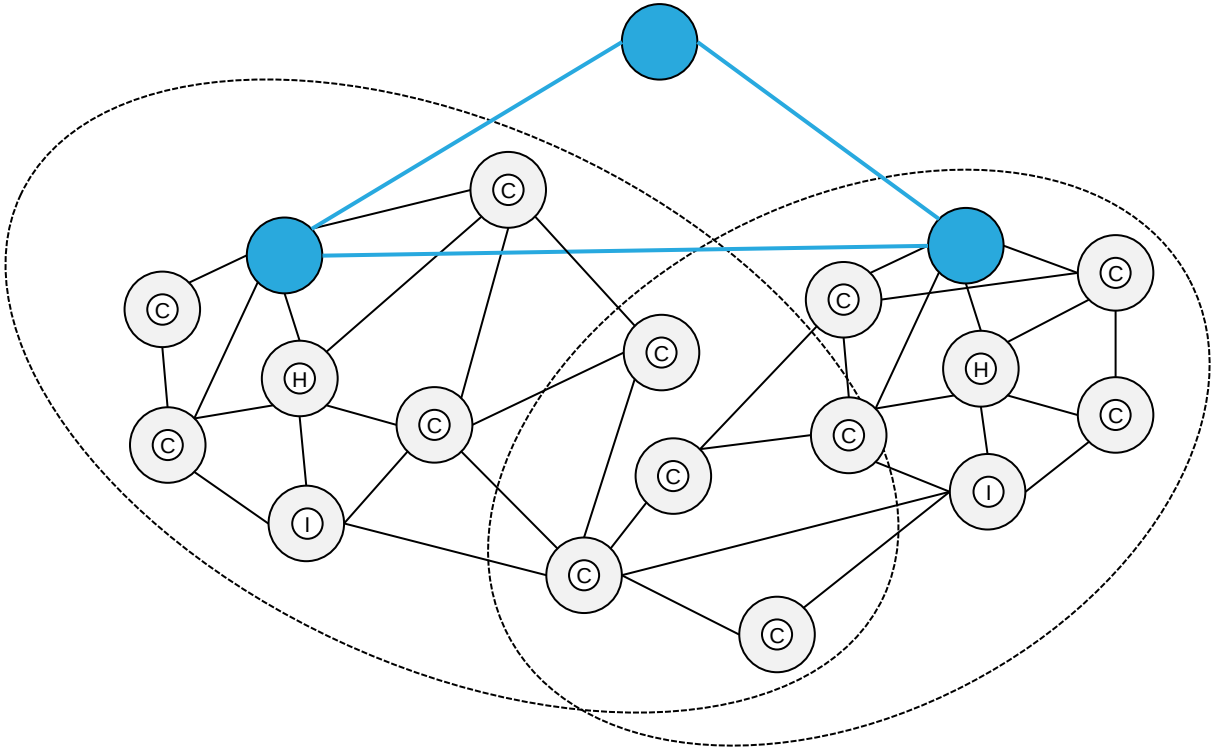
We need a decentralized design

- *Participants of a dataspace must have* **control** *over their data they share with whom*
- *Autonomy starts with controlling* **identity**, *if you are not in control of your identity you can't act autonomous*
- *Participants need to decide who they* **trust** *on in a case-by-case basis*
- *Participants in a dataspace must be based on* **rules** *that apply to everyone*
- *No central control system can make arbitrary* **decisions** *on individual participation*
- *Decentralized systems are* **resilient** *and provide higher* **availability**
- *No central system that holds the keys to the entire federation – to improve the* **security**
- **Interoperability** *of heterogenous environments with many different technologies and operating models*
- *Transitive trust based on common* **trust anchors**

Big picture of dataspace technologies



Implications of Ecosystems and Data Spaces



Eclipse Dataspace Project

Eclipse Dataspace Connector

- *Stared in June 2021*
  - *Official Eclipse Foundation project*
  - *Apache 2.0*
- *12 active committers*
  - *43 contributors*
  - *>750 Pull requests*
- *Based on Java*
  - *Minimal dependencies*
- *Platform architecture*
  - *Scalable, Highly Available, Customizable*

EDC Architecture & Components

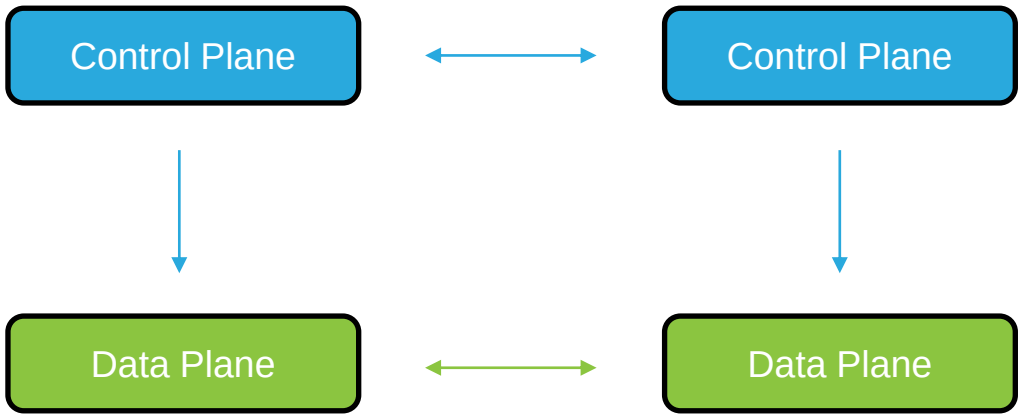
Architecture

- *Separation of Control and Data Plane*
- *Extensible through Java SPI*
- *Acts as an orchestrator for data transfer*
- *Asynchronous processing for maximum scalability*
- *Decentralized Identity Management with customizable Trust Anchors*

Components

- *EDC Control Plane*
- *EDC Data Plane*
- *Federated Catalog Node*
- *Federated Catalog Crawler*
- *Policy Management*
- *Data Asset Management*
- *Identity Hub*

Control Plane and data Plane



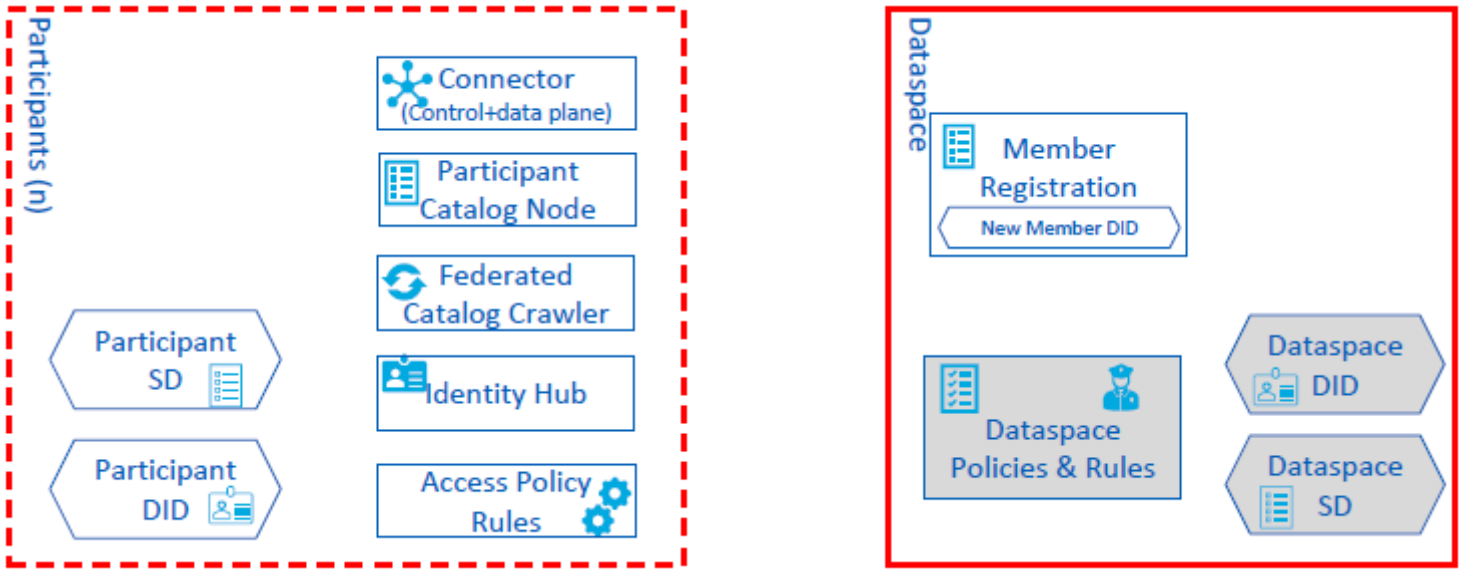
- Verification
- Contract negotiation
- Oversee policy enforcement
- Manages provisioning

- Moves bits
- Big Data
- Streaming
- Events

Minimum Viable Dataspace

- *The Minimum Viable Dataspace (MVD) is a sample implementation of a dataspace that leverages the Eclipse Dataspace Connector (EDC).*
- *The main purpose is to demonstrate the capabilities of the EDC,*
- *make dataspace concepts tangible based on a specific implementation, and to serve as a starting point to implement a custom dataspace.*
- *The MVD allows developers and decision makers to gauge the current progress of the EDC and its capabilities to satisfy the functionality of a fully operational dataspace.*
- *serves the purpose of demonstrating how decentralization can be practically implemented.*

Minimum Viable Dataspace



Contributors, Committers & Partners

- *Fraunhofer*
- *T-Systems International*
- *BMW*
- *Mercedes-Benz*
- *SAP*
- *Amadeus*
- *Microsoft*
- *ZF Friedrichshafen*
- *Amazon AWS*
- …

Key takeaways

- *The EDC is* **completely FOSS** *supported by various companies*
- *The EDC (through Eclipse Foundation) has clear and* **accepted governance** *structure and community processes*
- *The EDC is* **more than connecting a database**
- *The EDC manages* **data transfer and flow inclusive management of contract and policy management** *in cloud-native environments*
- *The EDC follows a* **modular system** *to serve as facilitator*
- *We have a good foundation already* **present on Github**
- *There is still work to do*
- *We* **welcome everyone** *to drive the idea and grow the community*

Further information and contact

General project information:

- *https://projects.eclipse.org/projects/technology.dataspaceconnector*

Github-Repository:

- *https://eclipse-dataspaceconnector.github.io/DataSpaceConnector/#/*
- *https://github.com/eclipse-dataspaceconnector/DataSpaceConnector*
- *https://github.com/eclipse-dataspaceconnector/MinimumViableDataspace*

Mailing list:

- *dataspaceconnector-dev@eclipse.org*
- *dataspaceconnector-community@eclipse.org*

YT-Channel

- *https://www.youtube.com/channel/UCYmjEHtMSzycheBB4AeITHg*

How to build secure Dataspaces with (F)OSS technologies in the context of IDS and Gaia-X

Markus Spiekermann
Head of Department Data Business, Fraunhofer ISST
Project Lead of Eclipse Dataspace Connector