Eclipse Dataspace Connector – How to build Data Spaces

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

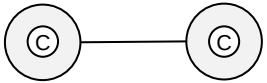
- *Introduction*
- *Data Spaces*
- *Business Ecosystems*
- *Eclipse Dataspace Connector*
- *Key Takeaways*

Motivation

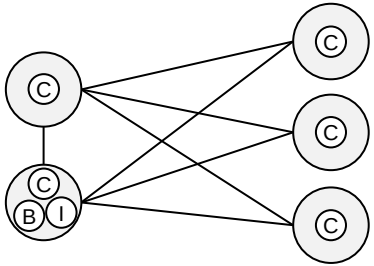
»A real data economy, on the other hand, would be a powerful engine for innovation and new jobs. And this is why we need to secure this data for Europe and make it widely accessible. ***We need common data spaces*** *- for example, in the energy or healthcare sectors. This will* ***support innovation 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****.«*

ECLIPSE
INCUBATION

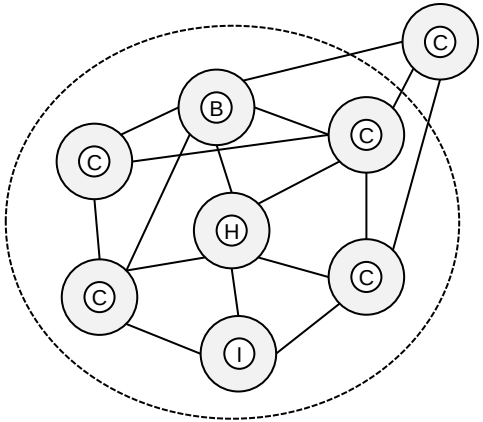
Change of Data Exchange and Sharing



Bilateral data exchange



Closed group data exchange

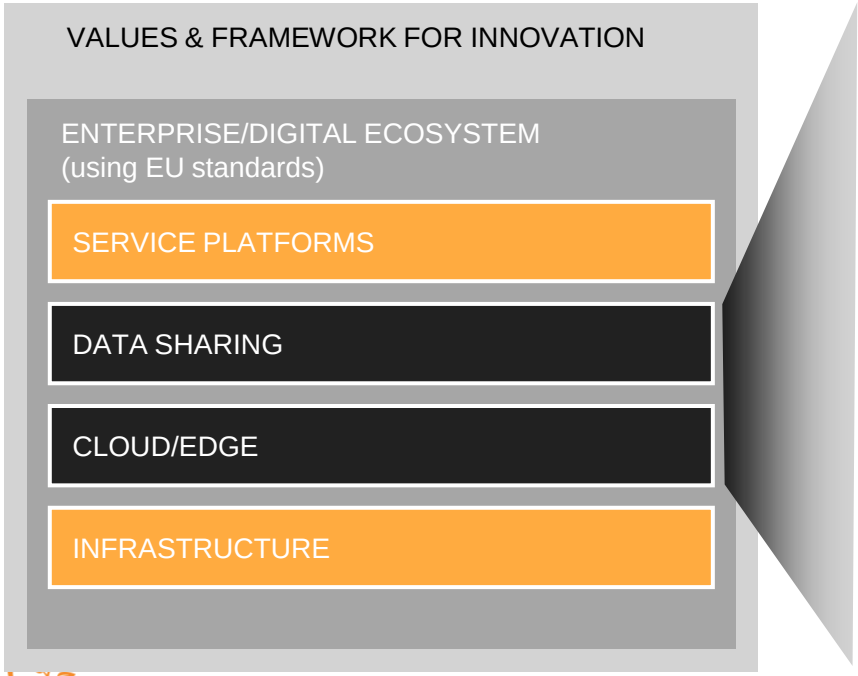


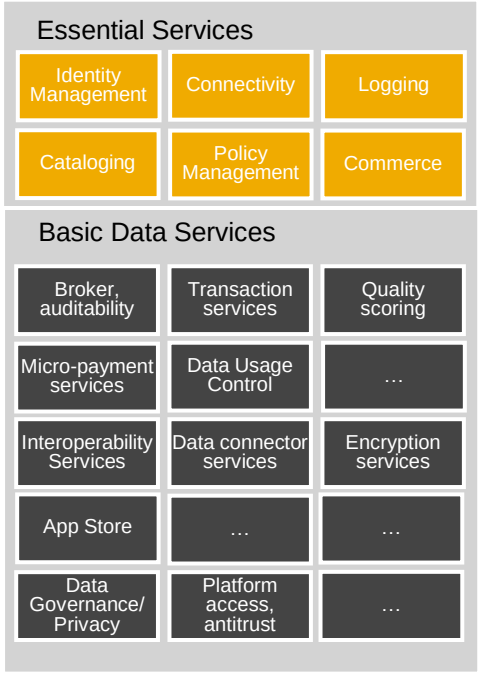
Open and dynamic data exchange

ECLIPSE
INCUBATION

Data Spaces

What does a Data Space bring to the Table?





Design Principles

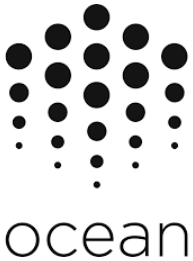
- European values
- Secure and trusted
- Easy-to-use
- Federated, neutral
- Vendor-agnostic

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.

ECLIPSE
INCUBATION

Landscape of Architectures and Services

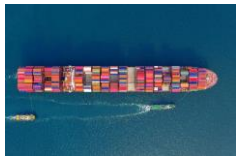


And others to follow…

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*
  - *Adapts the concepts of data spaces and their reference architectures*

Images from freepik.com by tawatchai07 and aleksandarlittlewolf

ECLIPSE INCUBATION

Mobility Data Space

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*
- *Adapt the IDS RAM and available open source implementations of components*
- *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*

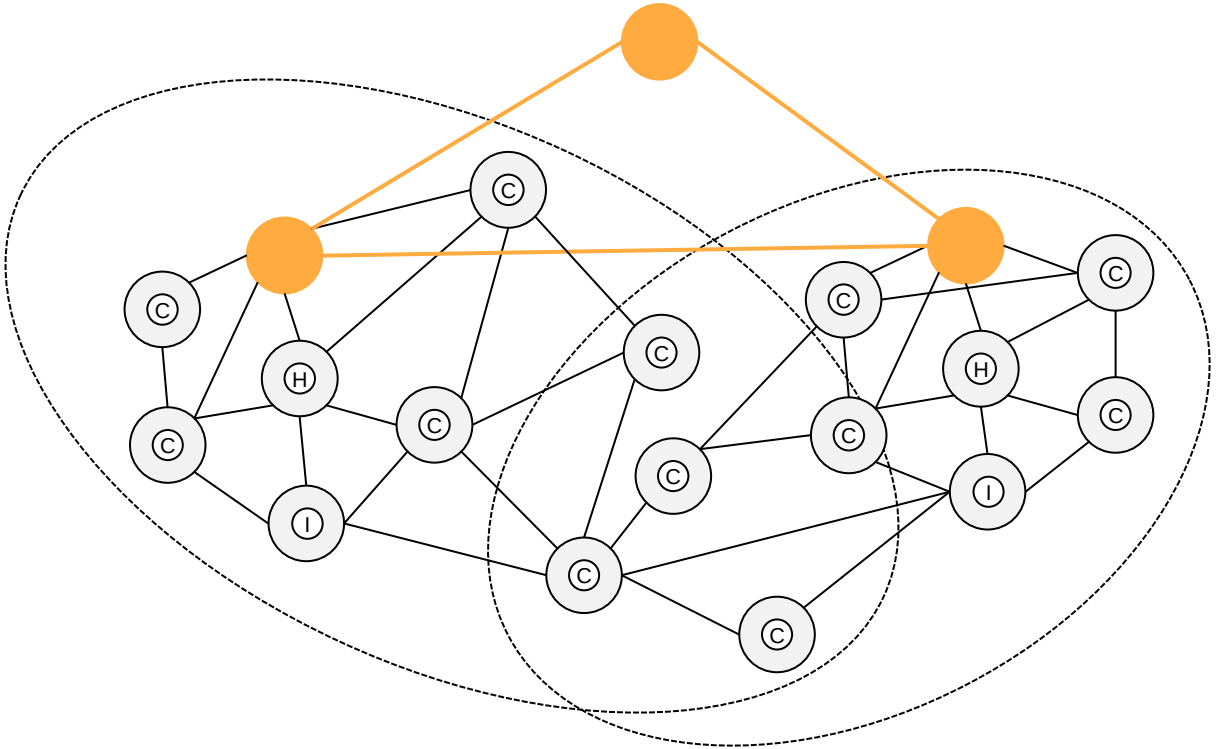
ECLIPSE
INCUBATION

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*
- *Connect to cross-industry networks “built a Gaia-X compliant IDS-system”*
- *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*

ECLIPSE
INCUBATION

Implications of Ecosystems and Data Spaces



ECLIPSE
INCUBATION

New challenges arise!

- *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*
- *Catena-X, Mobility Data Space, and other initiatives present new data sharing challenges*
  - *Support for data flow and transfer protocols to handle diverse data types*
    - *Push, Streaming, Large Volumes, Realtime*
  - *Cataloging across many providers*
  - *Policy management that traverses multiple infrastructure layers*
  - *Need for extensibility and modularity to accommodate diverse needs and use cases*

ECLIPSE
INCUBATION

Eclipse Dataspace Connector

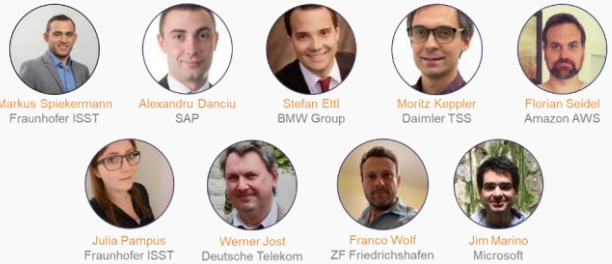
Eclipse Dataspace Connector

Overview | Downloads | Who's Involved | Developer Resources | Governance | Contact Us

Existing open-source projects address the technical challenges of cataloguing and transferring data for a wide range of use cases. However, there is no open-source effort aimed at providing an interoperable, cross-organization framework for data sharing that is built on a common identity model and uniform policy enforcement. This project will integrate with existing data exchange technologies and provide these missing pieces to create a system for data sharing where each organization is able to exert control over how its shared data is used.

A data-sharing system requires a protocol implementation for policy enforcement among participants. The Eclipse Dataspace Connector will implement the International Data Spaces standard (IDS) as well as relevant protocols associated with the GAIA-X project. However, the connector will be extensible so that it can support alternative protocols.

This project will provide implementation and use case feedback to IDS and GAIA-X.



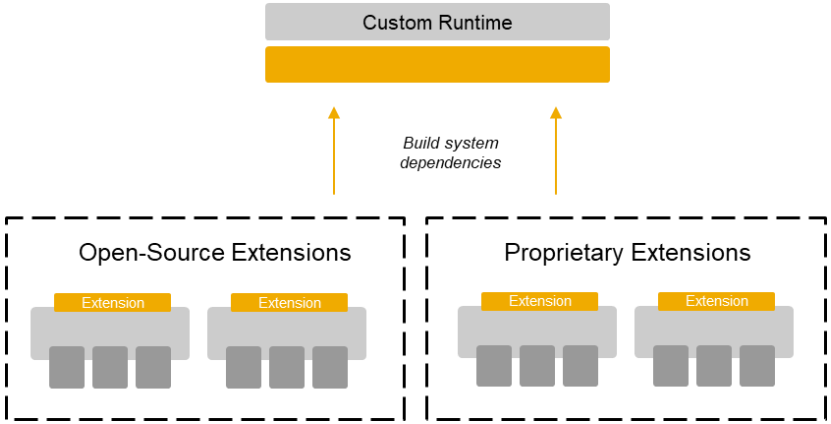
ECLIPSE
INCUBATION

Features

- *Based on a simple modularity system*
- *Separate control and data planes*
- *System is asynchronous and highly available*
- *Transfer processes are fully auditable*
- *Eliminate single points of failure*
- *Cloud aware policy enforcement and projection*
- *System security*

The Connector is a Coordinator

- *Modules are assembled into a runtime*
- *Data cataloging handled by external system*
  - *e.g., Apache Atlas*
  - *Modularity allows this to be substituted*
- *Data transfer and storage performed out of band*
  - *Modularity allows this to be substituted or augmented*
    - *Ability to add multiple transfer mechanisms to support diverse data types*
  - *Solves key aspects of how to handle streaming, large data transfer, and hyper scaling*



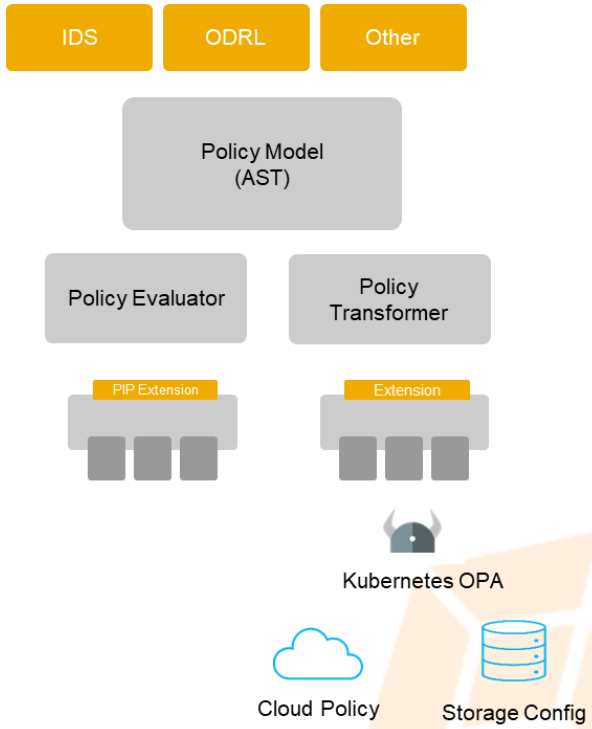
ECLIPSE INCUBATION

The Connector is a Coordinator

- *Processes are modeled as persistent state machines*
  - *Nearly symmetric on the client connector and provider connector*
  - *Includes states for initiation, receipt, resource provisioning, transfer, and termination*
    - *State transitions happen asynchronously after a defined task is completed*
      - *For example, provisioning cloud storage for data*
    - *Tasks must be idempotent for reliability*
  - *Processes are persisted to a store*
    - *Can be a highly available database with geographic fail over*
    - *Or, as simple as an in memory Map*
- *Provides a foundation for full observability and data audit*
  - *Metadata about each state is recorded*
  - *Observability can be achieved by correlating the process id with separate telemetry data*

Cloud-Aware Policy Enforcement

- *Policy engine that is cloud aware*
  - *Not limited to the connector*
  - *Can be embedded in other processes*
- *Horizontal and vertical enforcement*
- *Parses policy syntax into an internal Policy Model AST*
- *Evaluators and transformers to enforce policy*
  - *Evaluators can make policy decisions, e.g. is a connector authorized*
  - *Transformers can create and deploy policy to different levels*
    - *OPA, storage, etc.*



ECLIPSE INCUBATION

Status-Quo and Roadmap

- *Official Eclipse project since June 2021*
  - *Initial contribution has been made, checked and approved by Eclipse Foundation*
- *Actual results*
  - *Provide assets*
  - *OAuth2 and SSI authentication*
  - *Data flow, transfer and resource provision (via state machine, asynchronous and observable)*
  - *IDSA compliant self-description and contract offers*
  - *Provide samples for different use cases*
- *What's next (End of 2021)*
  - *Participation on the 2nd Gaia-X Hackathon 02/03 December*
  - *IDSA-based messaging and Control APIs*
  - *Policy negotiation and contract-aware data transfer*
  - *Release of EDC 0.0.1*

ECLIPSE
INCUBATION

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*

ECLIPSE
INCUBATION

Further information and contact

General project information:

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

Github-Repository:

- *https://github.com/eclipse-dataspaceconnector/DataSpaceConnector*

Mailing list:

- *dsconnector-dev@eclipse.org*

YT-Channel

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

ECLIPSE
INCUBATION