DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Data sharing experiences at ITI

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION

The challenge

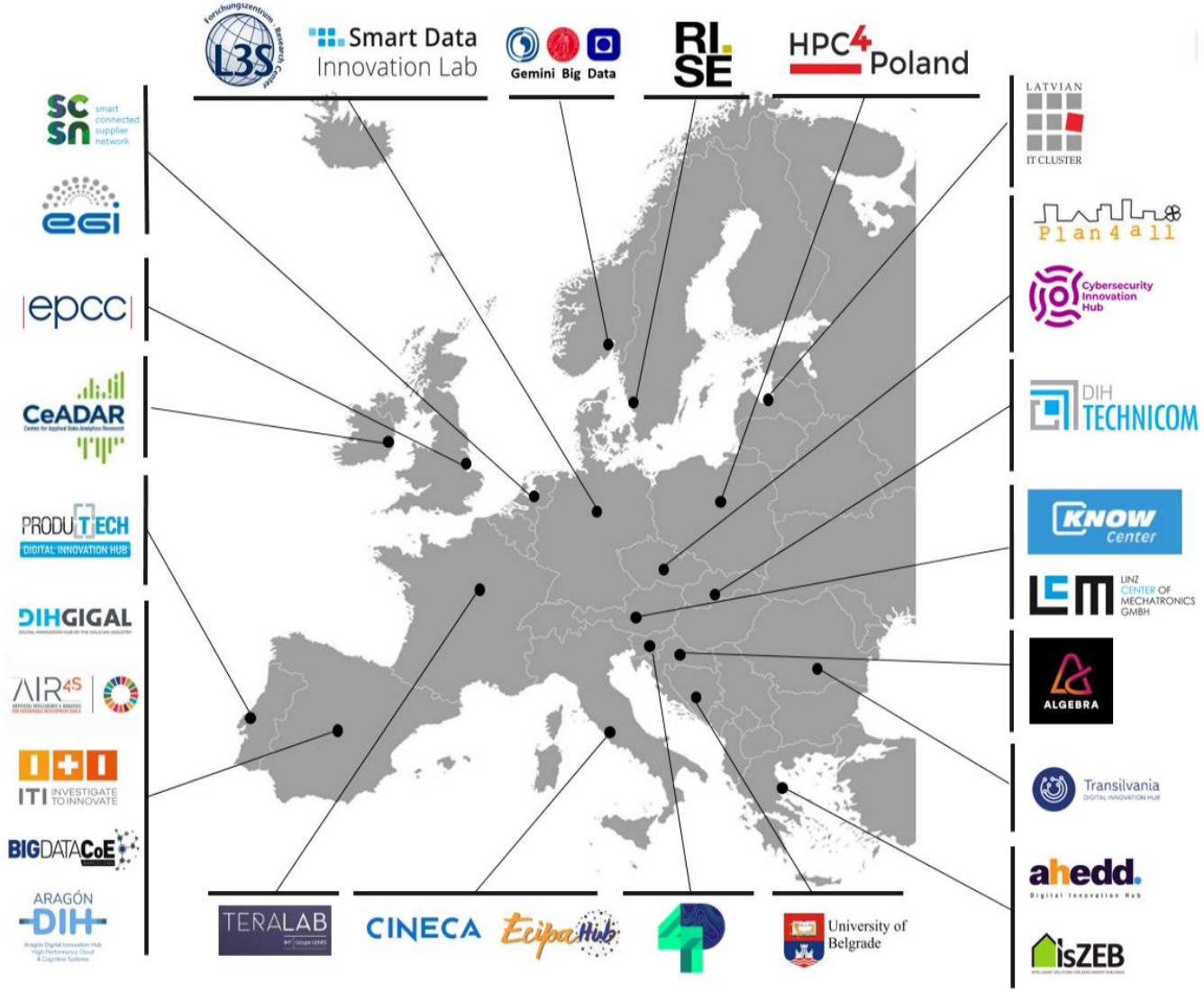
What we wanted to achieve at the begining of the project

Overview

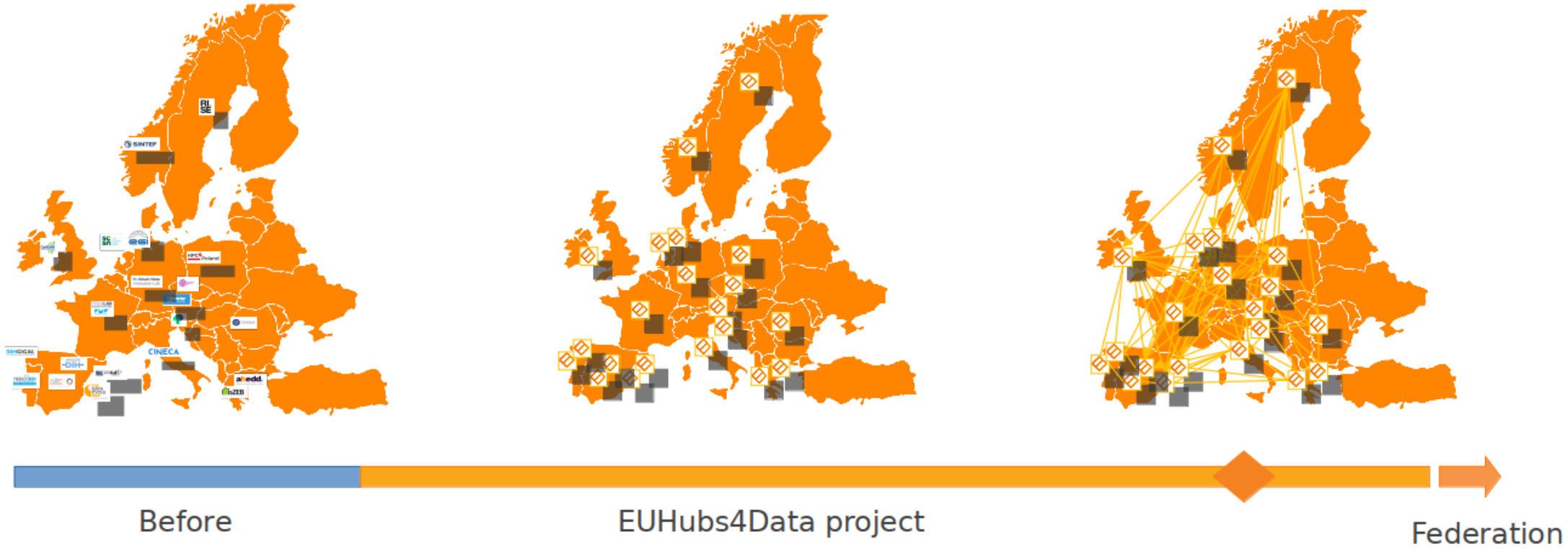
EUH4D

=

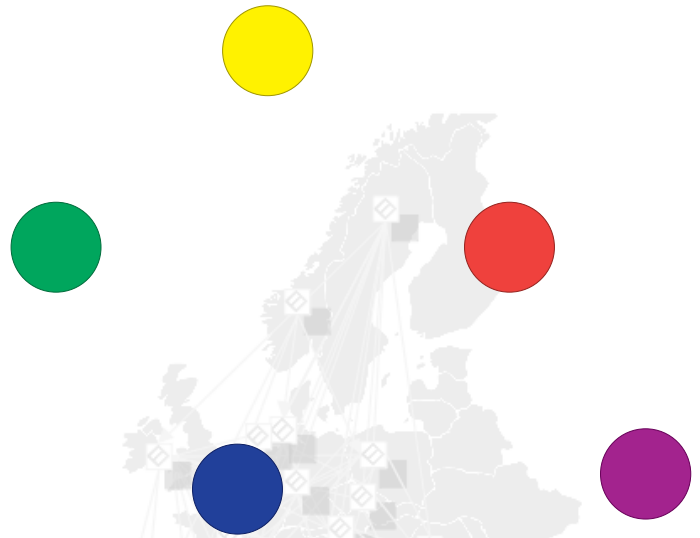
European Federation of Data Driven Innovation Hubs



Starting at 2020, this project will result on a federation of Hubs at the end of 2023

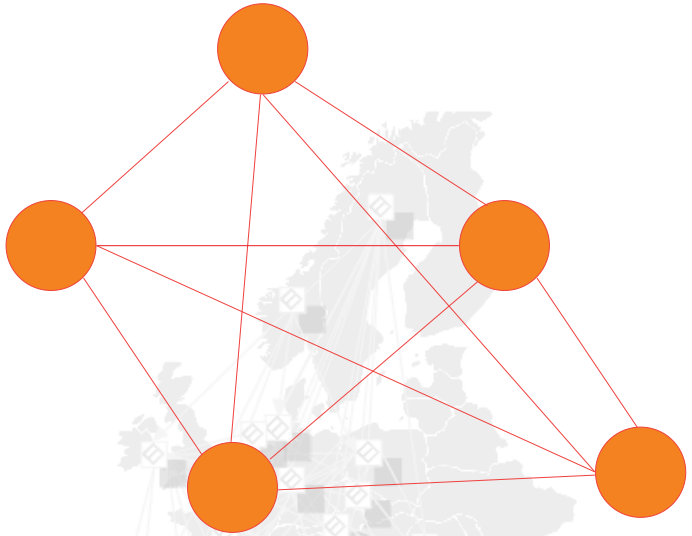


Before:
Hubs offer support to SMEs in their innovation processes giving them access to their **own** technological services, trainings, consultancy and data.



The goal:
Hubs **work as a group:**

- Have a **global common catalogue** of services and datasets
- **Share** data
- Providing **local support** to companies offering resources from other Hubs



Before

EUHubs4Data project

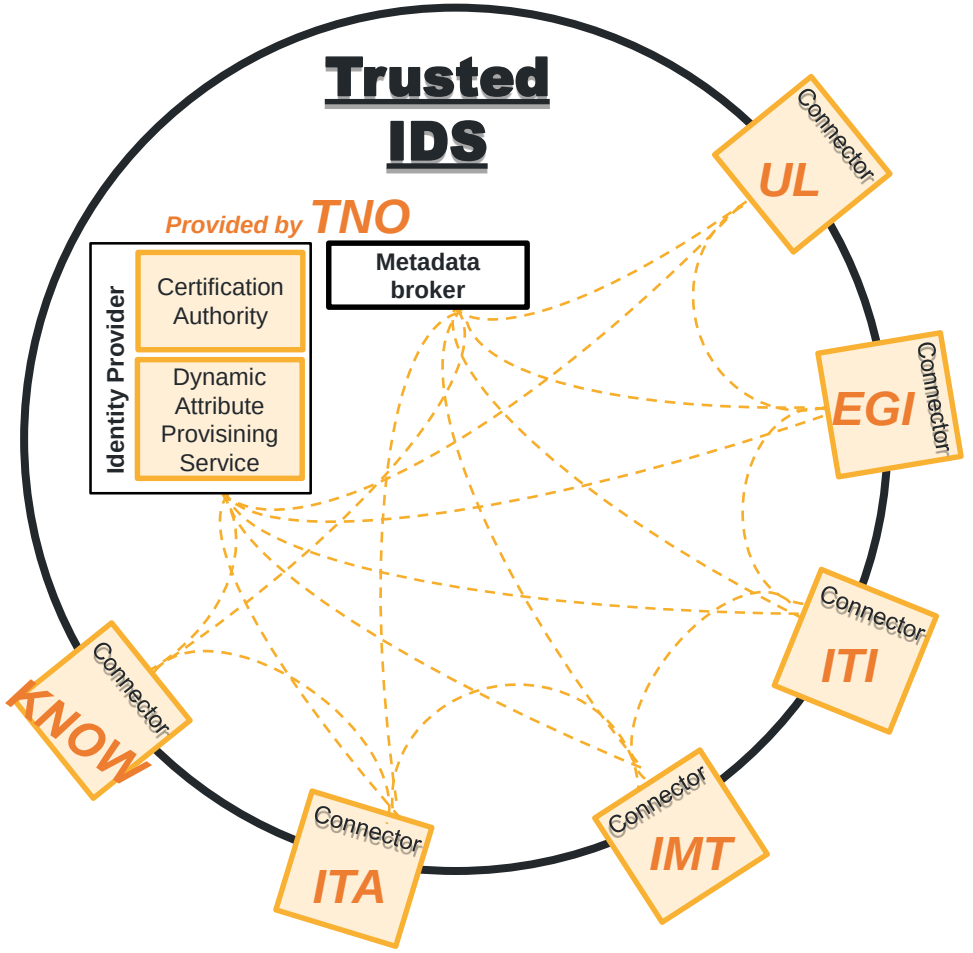
Federation

The IDS framework deployment at EUHubs4Data

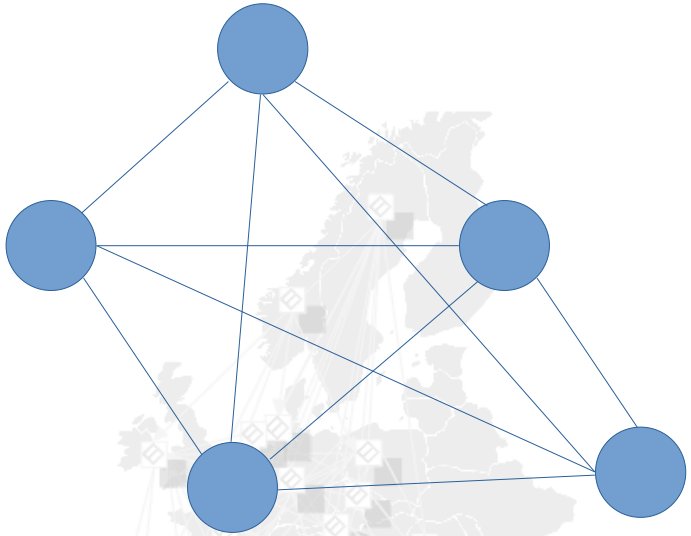
IDSA RAM components used:

- Identity provider
- Connectors
- Metadata broker

This deployment builds a **common Dataspace** where all the Hubs participate



We use **IDS DataSpace Connector** for interoperability between the Hubs.

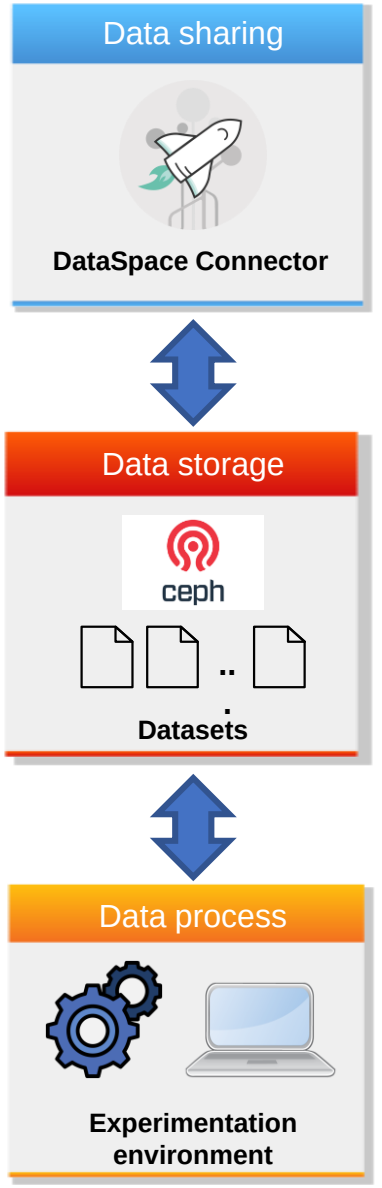


- This is the technology used to establish a **trusted environment and consume assets** (services and datasets) published by other partners.
- Then, the **connectors** are used to negotiate **policies** (conditions of use) and to **consume** such resources

ITI components

Three different levels:

1. The **DataSpace connector** is used to share data between the partners of the **Federation / DataSpace.**
2. The shared data is stored in a **Ceph** system. The **datasets are available** here to be shared with other Hubs (1) or to be used internally (3).
3. There is an **experimentation environment** for researchers where data is **accessed and processed** using different resources and tools: GPUs, Jupyter Notebook, etc.



Use case

How this is being used

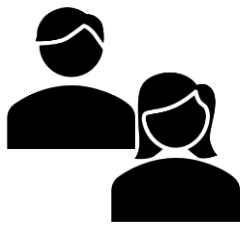
Support SMEs with their innovation/experimentation processes

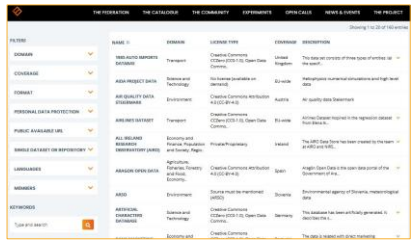
SMEs

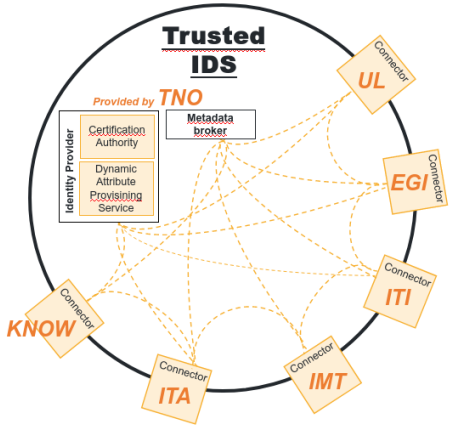
EUHubs4Data Catalogue

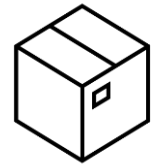
Interoperability

Output







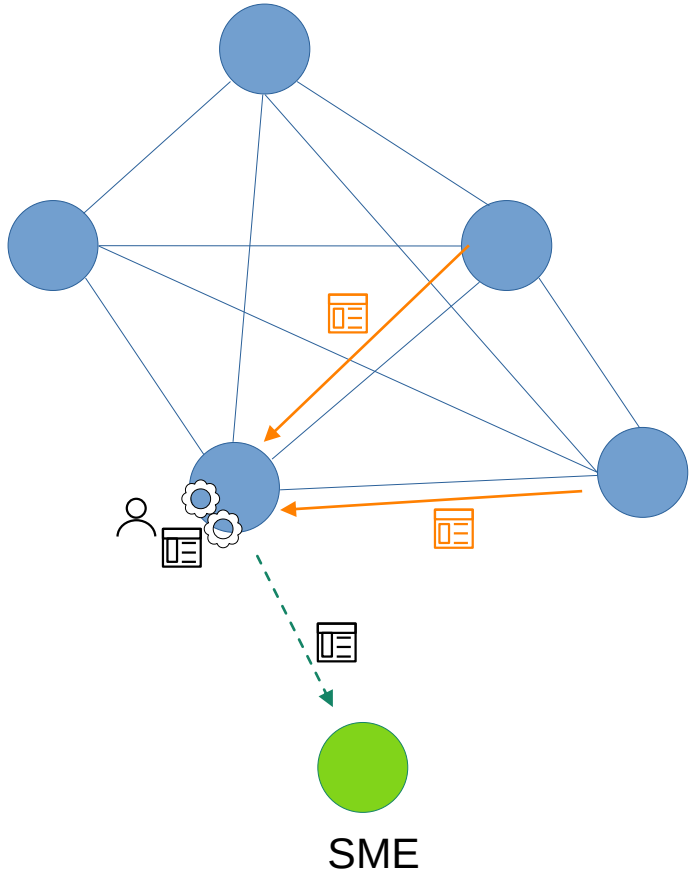


Current approach

Interoperability between Hubs.

- The **data should be shared** between members of the organization.
- The data is **processed within the infrastructure** of another member of the federation.

Tested internally by EGI, IMT and ITI **and externally** with DataPorts project.



Experiments in EUHubs4Data

Our initial **idea** was to use the **experiments** to test our interoperability

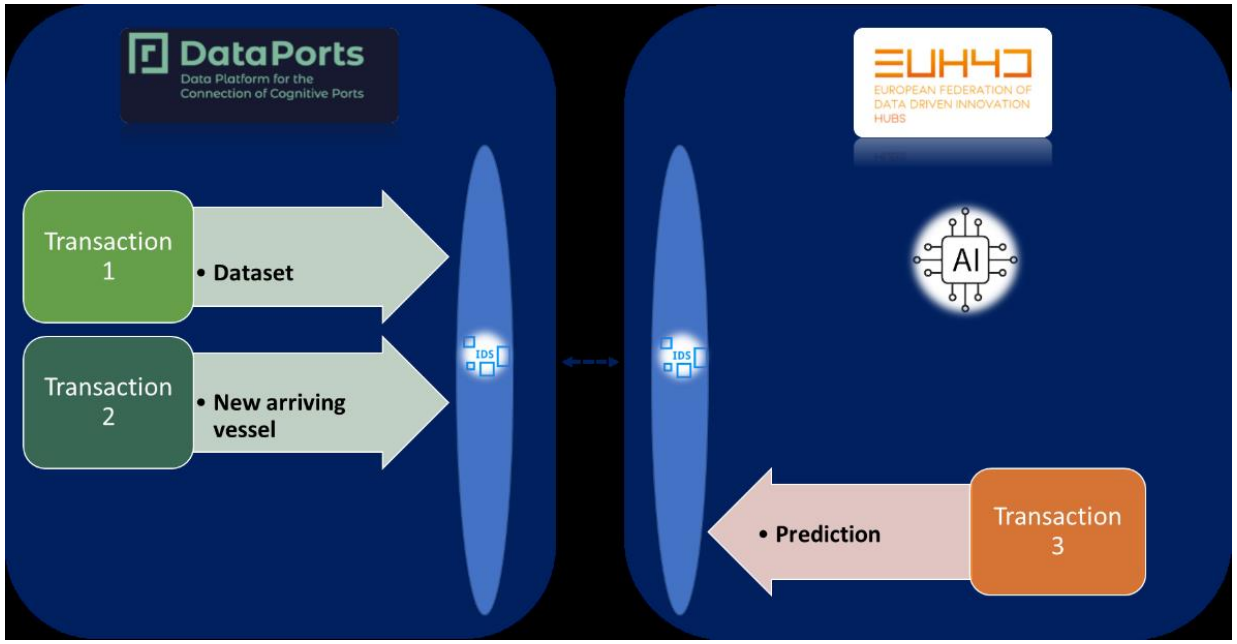
- We had three open calls to do it.
- But:
  - **At the same time** we were deploying our components.
  - It **was possible for the 3rd** open call **but very risky** for the SMEs.
  - Experiments are **9-month** projects.

DataPorts scenario

Valencia Port wants to get some **advanced analytics based on their data**, but the **services** they want to use are **inside the EUH4D federation** and are not reachable through the DataPorts data space.

A **technology provider inside EUH4D offers an AI model** that will suit Valencia Ports' needs and provide valuable predictions.

This scenario assumes that some exchange of data has already taken place. So, **the technology provider has built models capable of predicting vessel departures** based on data from Valencia Port.

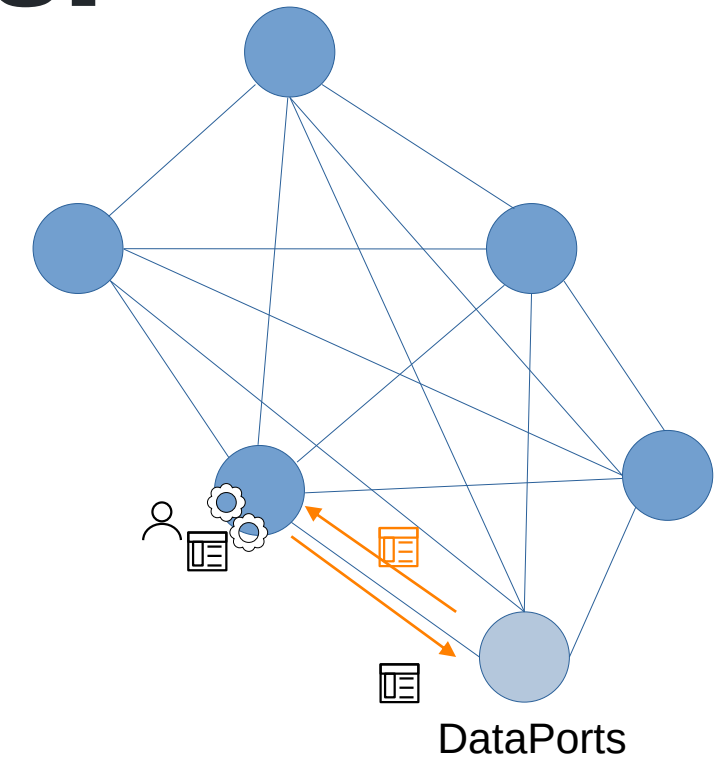


DataPorts as a member

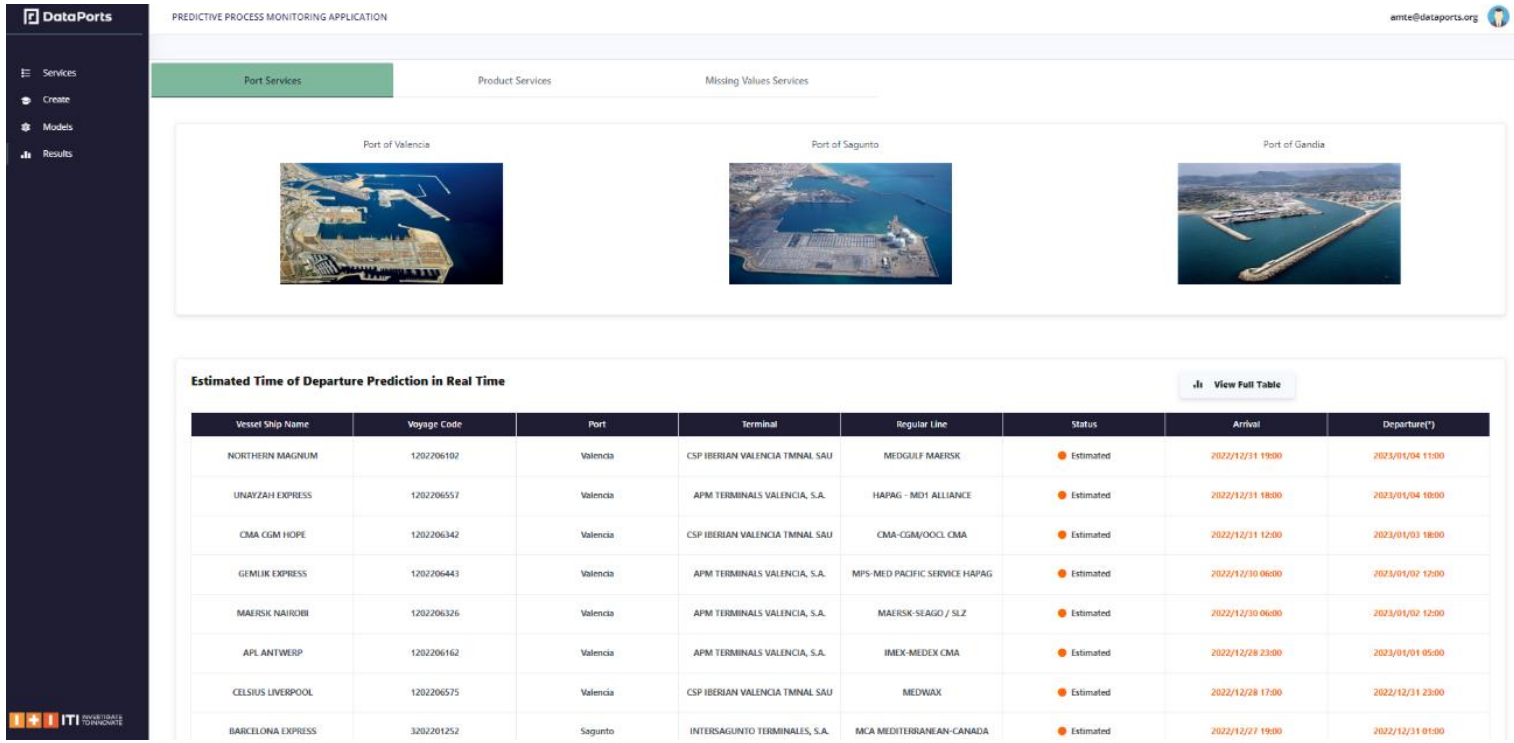
The scenario will make use of sovereign and **trusted data sharing between** two participants within the EUH4D Federation:

DataPorts has registered as EUHUBS4Data **participant in the trusted IDSA network** :

- Connector deployment
- Connector identity registry
- DataPorts and EUHubs4Data share data in the trusted DataSpace.



DataPorts departure predictions



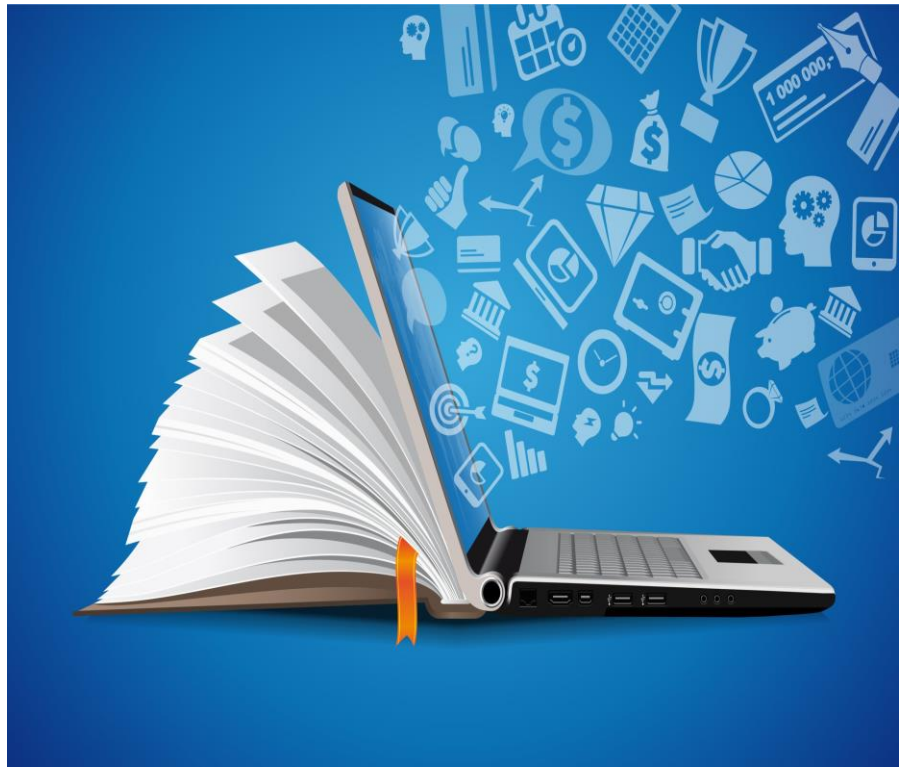
- DataPorts acts as a Data Owner/Provider and ITI as Data Consumer/User
- Data is processed by another member of the EUHubs4Data DataSpace to derive a new offer with additional information (departure predictions)
- New offer at ITI is available to be consumed from DataPorts

Lessons learned

About the data sharing and the use of IDS components

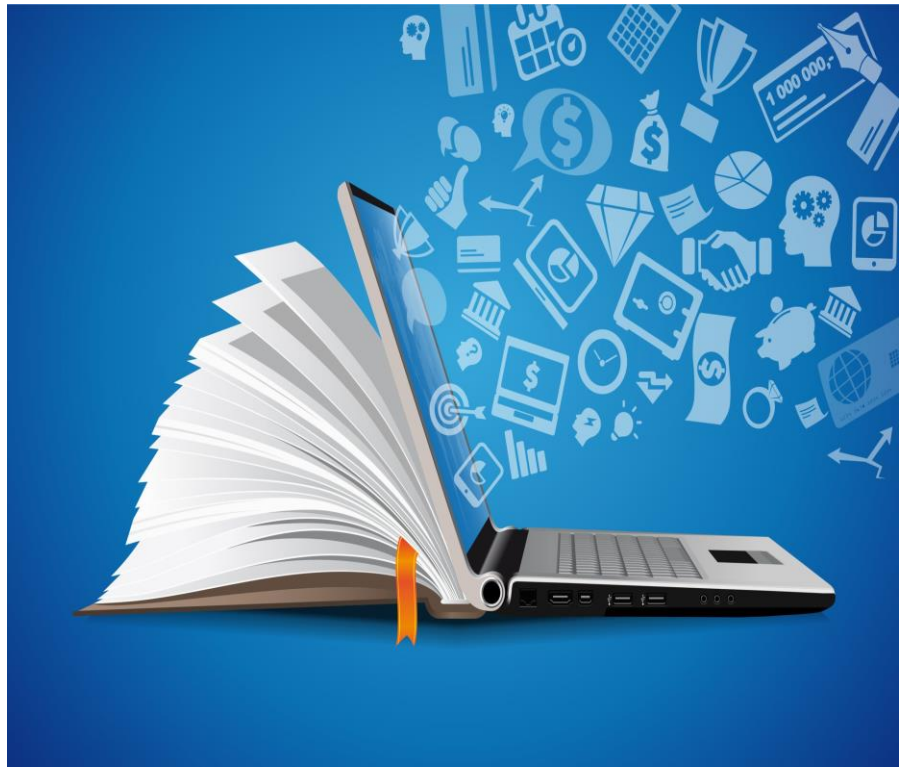
What we have learned 1/2

- There is an important **learning curve** to really understand the use case.
- Regarding the technology, It was difficult to **choose the right option (connector)**
  - **Connectors are being developed** and including new features.
    - Some **stability** is needed.
    - We experienced **breaking changes** using DSC v5, 6 and 7.
  - Now most of the **DSC developers have moved to EDC.**



What we have learned 2/2

- You don’t need connectors to exchange data but the deployment **makes sense if**:
  - A **trusted environment** is needed.
  - **Data sharing policies** are implemented
- Data and Dataspace **governance is a priority**. To create a Dataspace is not only a technical work.
  - The **rules to control the relationship between members** should be defined as soon as possible.
  - They will **determine which policies should be implemented**



DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Thank you!

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION