DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

# The EGI DataHub Adapter for IDS Connector

Levente Farkas (EGI Foundation)

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION

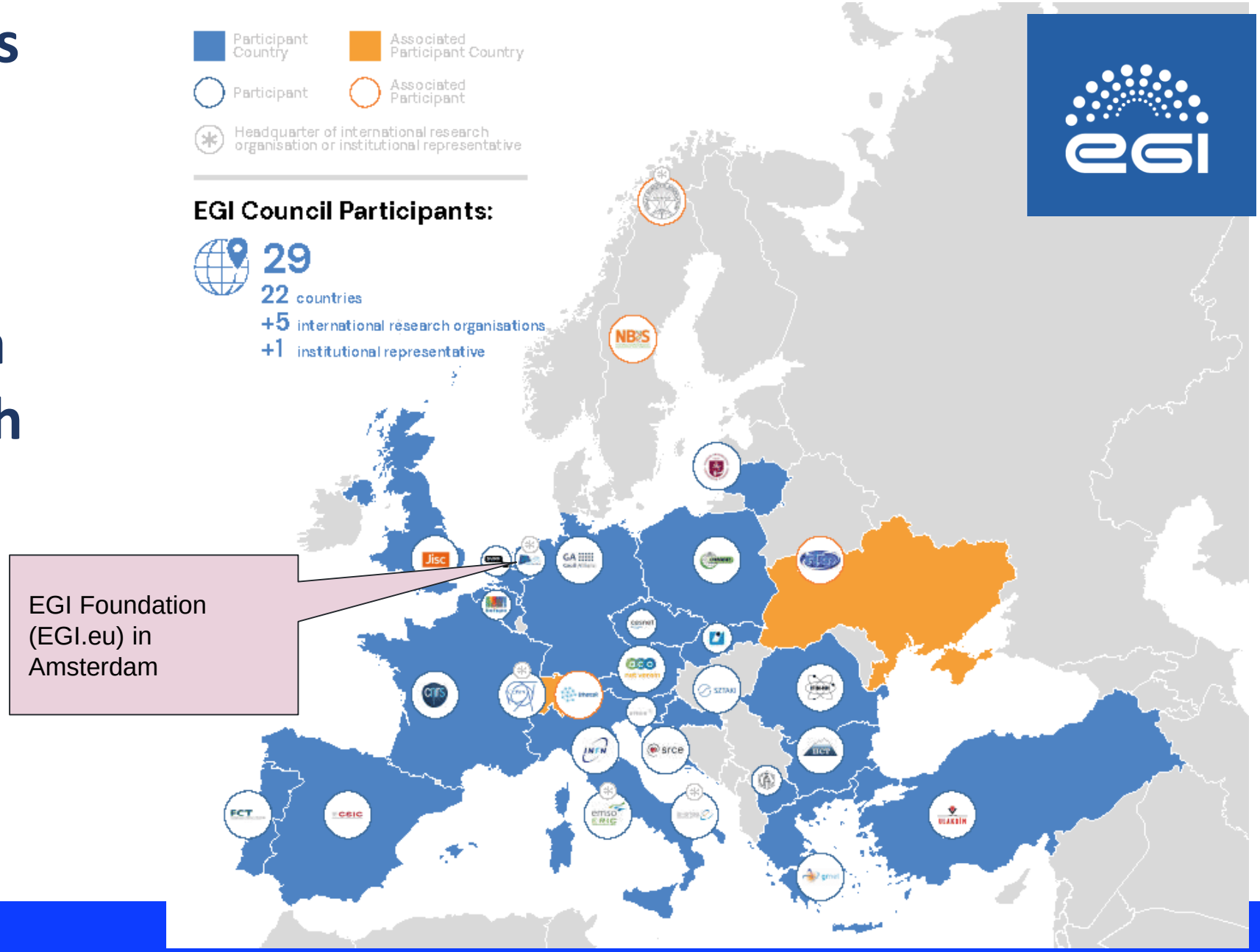
# Outline

- The EGI Federation and the EGI DIH
- The EGI DataHub
- The EGI DataHub adapter for IDS Connector

# The EGI Federation is an international e-infrastructure that provides advanced computing and data analytics for research and innovation.

**BDVA** member since March 2019

EUHUBS4DATA



# EGI Digital Innovation Hub

It aims at:

- Help and support companies in the **adoption/ test/ validation** of advanced computing technologies to become more **competitive**
- Provide an **open and innovative environment** for experimentation, training and networking.
- Increase the **innovation** among the EGI Federation members
- Facilitate the **relationship** between EGI community and business sectors for a wider economic and societal **impact**

# Current Industry relationships

Examples of focus

## Current industry engagement

Typically start-ups and SMEs benefit the most from piloting, proof concepts, consultancy and visibility activities

Established connections with companies or larger entities

## Projects with Industry partners

# EGI Service Catalogue

https://www.egi.eu/services

### Cloud Compute

Run virtual machines on demand with complete control over computing resources

### Cloud Container Compute

Run Docker containers in a lightweight virtualised environment

### High-Throughput Compute

Execute thousands of computational tasks to analyse large datasets

### Software Distribution

Publish and access software efficiently across multiple sites

### Notebooks

Create interactive documents with live code, visualisations and text

### Replay

Reproduce and share research on a notebooks-based platform

### FitSM Training

Learn how to manage IT services with a pragmatic and lightweight standard

### ISO 27001 Training

Learn how to manage and secure information assets

### Training Infrastructure

Dedicated computing and storage for training and education

### Online Storage

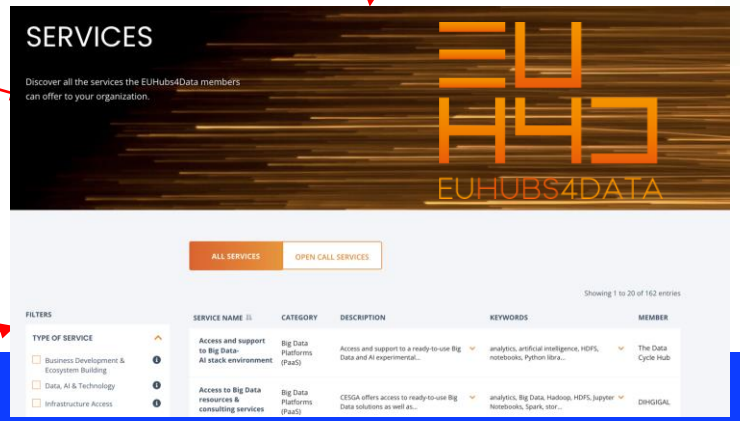
Store, share and access your files and their metadata on a global scale

### Data Transfer

Transfer large sets of data from one place to another

### DataHub

Access key scientific datasets in a scalable way



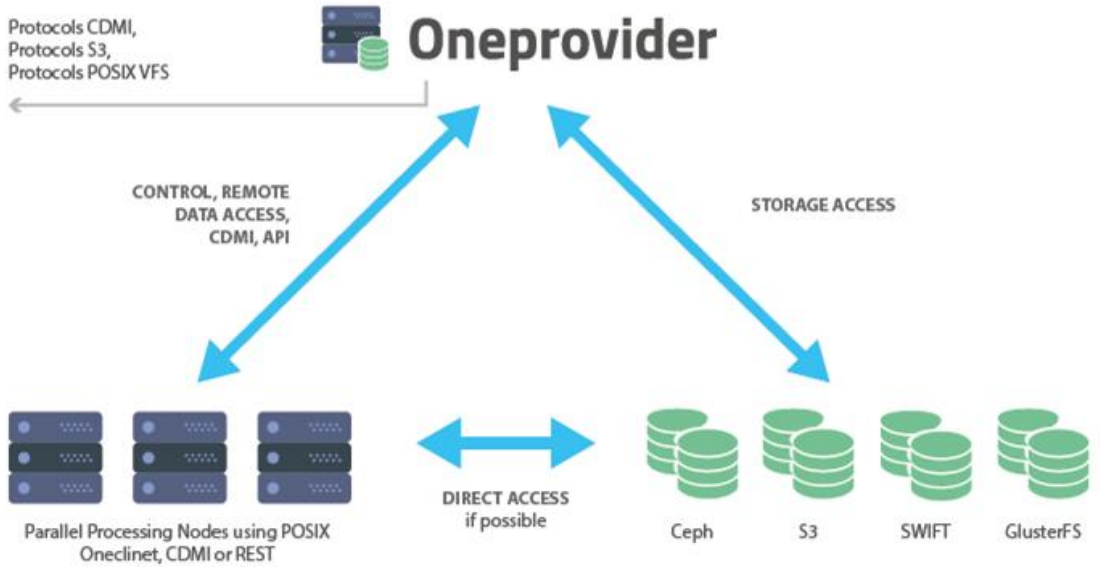

**ISO 9001**
Certified
Quality Management System
www.tuv-sud.com/ms-cert

**ISO 20000**
Certified IT Service
Management System
www.tuv-sud.com/ms-cert

**Supporting 6 SMEs selected during the EUHubs4Data project Open calls**

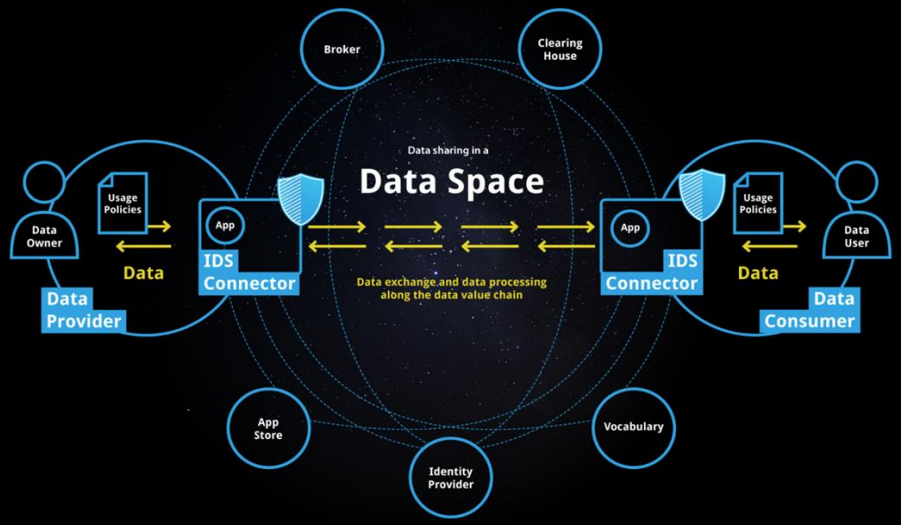
# EGI DataHub: Main functionalities

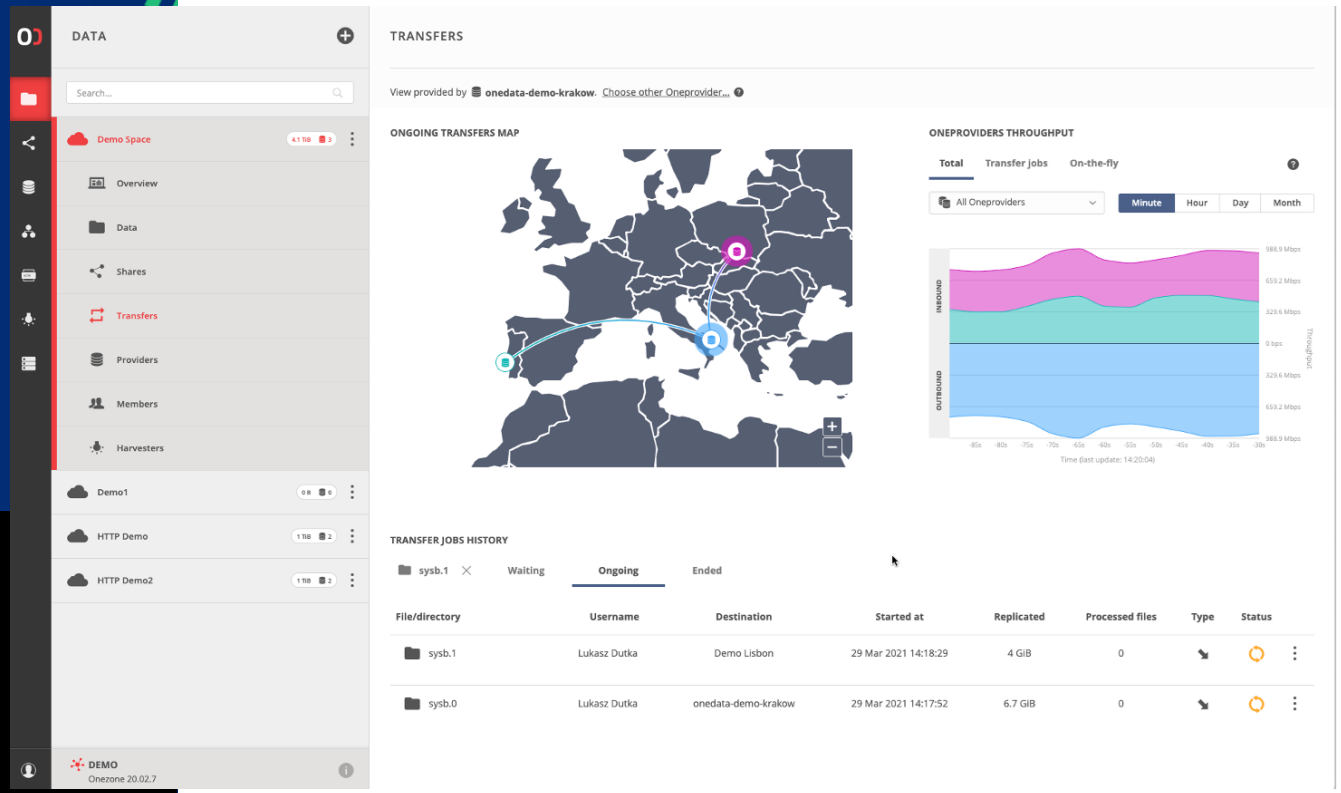
- **Data-collections aggregation service**
  - Aggregates Data providers
- **Opendata Repository**
  - Easy sharing data for anonymous access
  - PID/DOI minting
- **Metadata indexing service**
- **Platform for hybrid cloud data management and processing environment**
  - POSIX access from compute nodes



EUHUBS4DATA

# Enable data sharing and data interoperability in the EUHubs4data Federation via IDS



# EGI DataHub Adapter

https://egitlab.iti.es/euhubs4data/egi-datahub-connector

- **Published in the 2023 IDSA Data Connector report**
  - https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Data-Connector-Report-2.pdf
- **Is a collection of containers:**
  - **An IDS connector**
  - **An IDS connector UI**
  - **An adapter that parses an EGI DataHub space and populates the connector with its content**
  - **Telemetry components**

**2.2.3 EGI DataHub connector**

| | |
|---|---|
| Name of the connector | EGI DataHub connector |
| Logo of the connector or company logo | |
| Maintainer (company name) | EGI Foundation |
| Short description | Policy based access via IDS to multiple storage backends supported by EGI DataHub (e.g. S3, Swift, NFS, GlusterFS, etc). It is based on the Dataspace Connector available open source. |
| Maturity level | TRL 4-5 |
| License type | Planned to become open-source. License type unknown. |
| Features | • All usage control supported by IDSA Data Space Connectors (9 in total)<br>• Information model version 4.2.7<br>• Protocol supported: https |
| Adoption | EUHubs4data |
| External resources | |

# EGI DataHub Adapter

https://egitlab.iti.es/euhubs4data/egi-datahub-connector

- **Simple REST API**
  - Publishes a space from EGI DataHub as a catalog to an IDS connector
  - For each file it creates:
    - 1 resource, with 1 representation, with 1 artifact
    - 1 contract, with 1 usage policy for each resource
  - Artifacts get access URLs to the associated file with an embedded temporary access token
  - File path is saved in a custom attribute on the artifact, allowing for reconstruction of the folder structure
  - You can specify the usage policy, default is "*free to use*"
  - Supported policies are those supported by the DataSpace connector

# Customizing the DataHub Adapter

https://egitlab.iti.es/euhubs4data/egi-datahub-connector

- Can publish the content of any data store, if:
  - It is organized in a hierarchical fashion
  - Can store key/value pairs on files/objects
- Implement the Java interface DataStore
- Instantiate this class as your backend

# Summary

https://www.dataspace-connector.io

## Achievements

- Standardized access to EGI DataHub
- Publish data from EGI DataHub respecting data sovereignty
- Include datasets stored in EGI DataHub in the EUHubs4Data catalogue
- Can easily be reused for other backends

## Roadmap

- Add sovereignty feature to DataHub (OneData)
- Add support to publish datasets in EOSC Marketplace

THE DATA IS OUT THERE

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

# Thank you!

EUHUBS4DATA

BDV BIG DATA VALUE ASSOCIATION