



ONEDATA

summer webinars 2026

Chapter 1: General introduction to Onedata

Lukasz Opiola

25.06.2026

AGENDA & GOALS

- About the people behind Onedata
- Get a general understanding of the platform
- Learn about typical use cases
- Take a first look at the Web user interface
- A teaser for upcoming webinars



ONEDATA @ CYFRONET AGH, KRAKOW, POLAND

ONEDATA



13 years of devoted development



Distributed data management platform
for large-scale problems and HPC applications

An ecosystem for collaborative data sharing
across organizational domains



National HPC and supercomputing centre in Poland
Operating the largest supercomputers since 1975

Operator of PLGrid (6 Polish supercomp. sites)
and major EU research infrastructures, such as: EOSC, EuroHPC, Gaia AI Factory

Provides resources for science and industry
HPC, AI, storage, and networking services



OUR SUPPORTERS AND PARTNERSHIPS



- Onedata is developed primarily within grants and collaborations
- With every project, its open-source codebase evolves further
- New developments build a stronger base for future projects
- Hall of fame:

eOSC Node | Poland

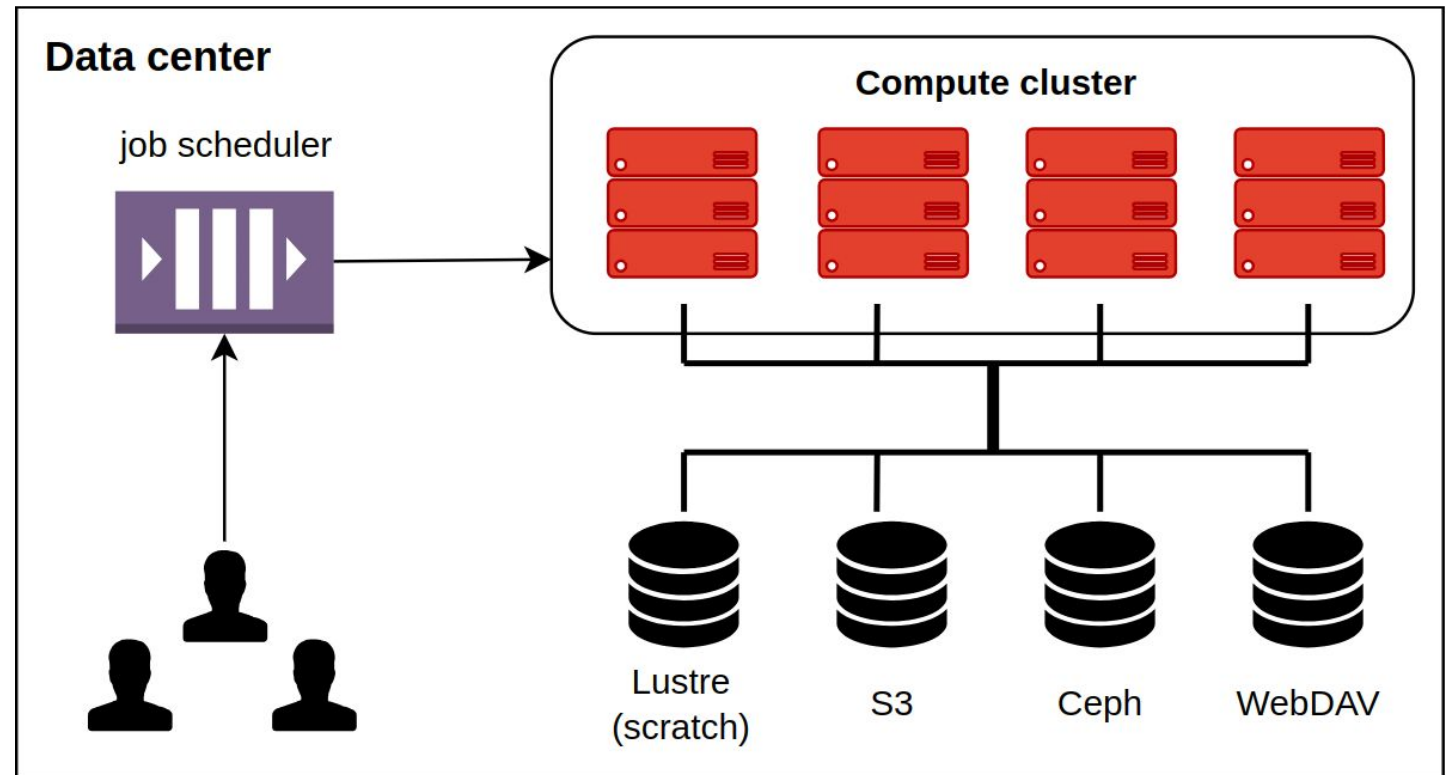


HOW IT STARTED

DATA ACCESS WITHIN A DATA CENTER

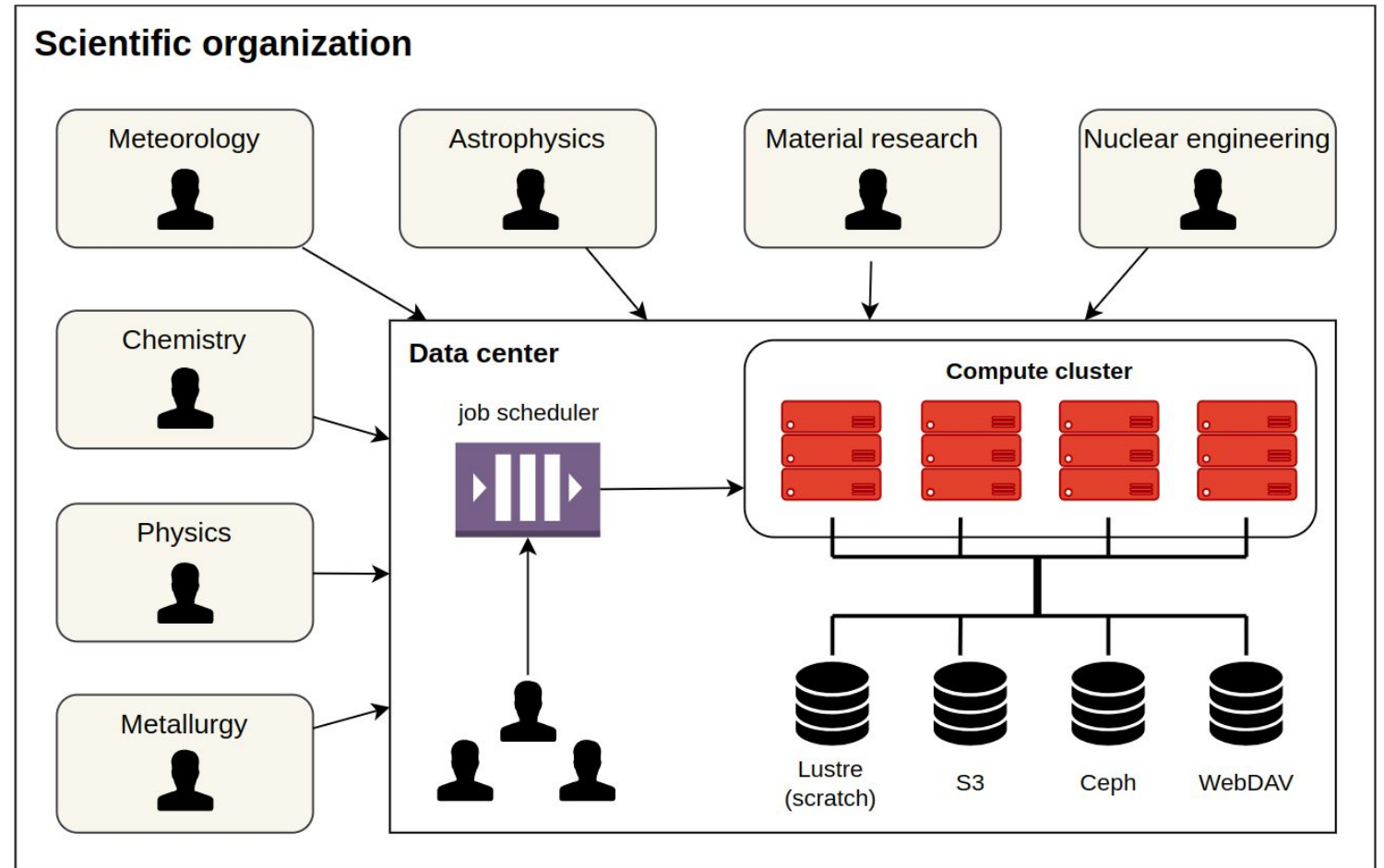


- heterogeneous storage systems
- manual data management
- **need for unified data access**



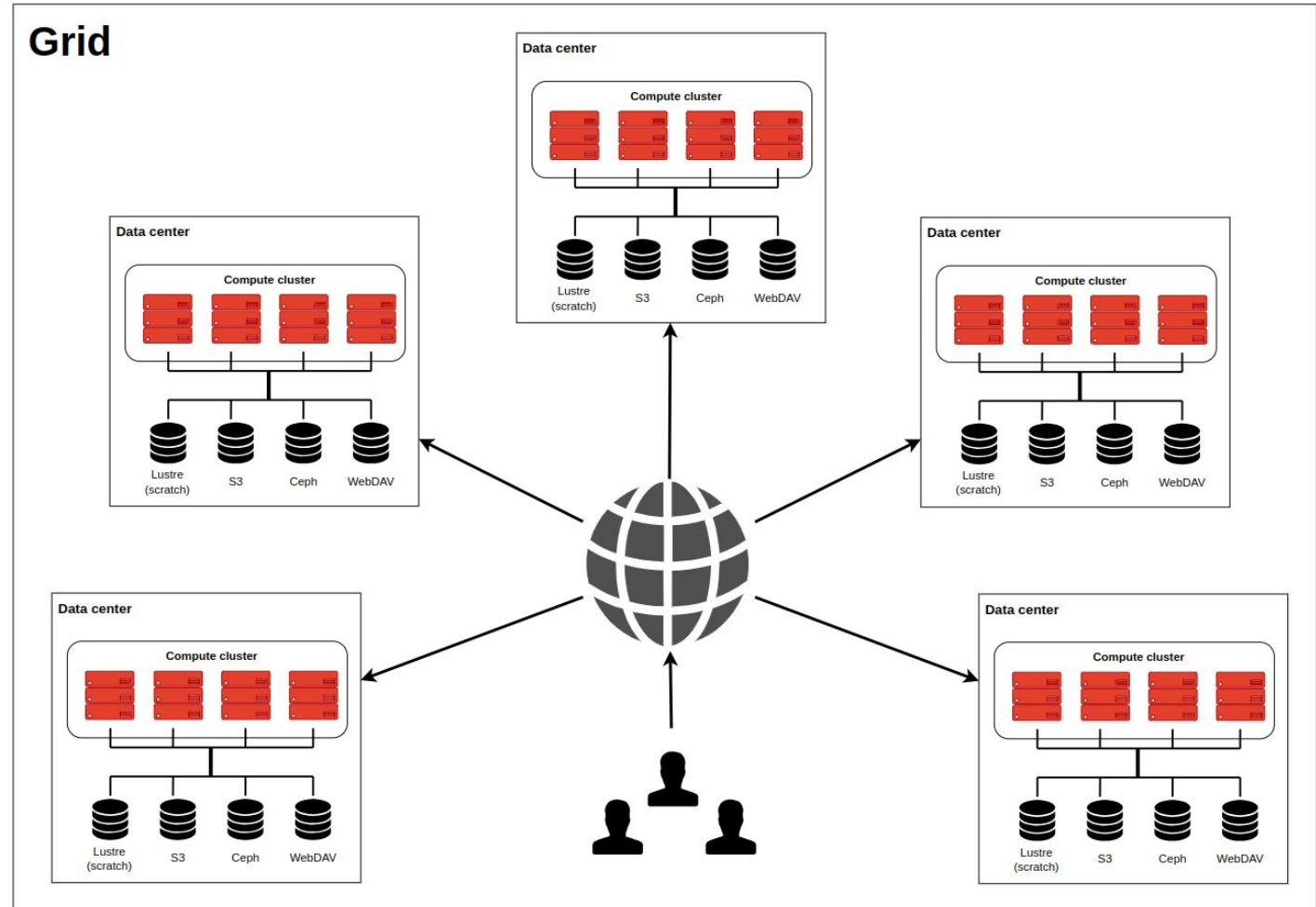
OTHER SCIENCE FIELDS WANT THEIR SHARE TOO

- heterogeneous storage systems
- manual data management
- required IT skills
- need for a user-friendly approach



DATA ACCESS WITHIN A GRID

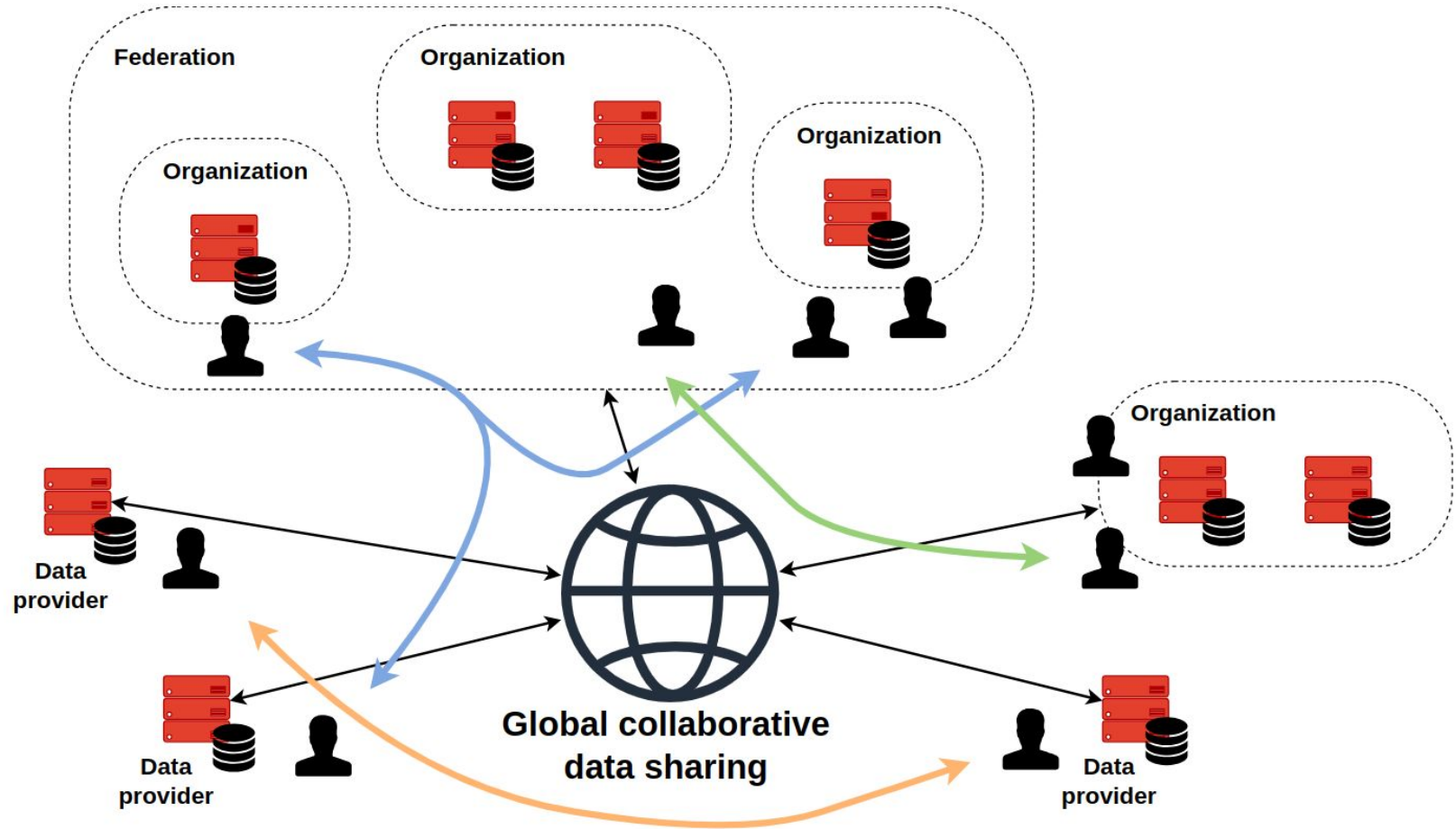
- heterogeneous storage systems
- manual data management
- required IT skills
- geographical data distribution
- troublesome data transfers
- different access control policies
- **Need for suitable distributed data management tools**



OUR MAIN GOAL — GLOBAL COLLABORATIVE DATA SHARING

Generic approach

- unified access across autonomous data providers
- data sharing between organizational domains
- trust-driven approach with privacy and security guarantees



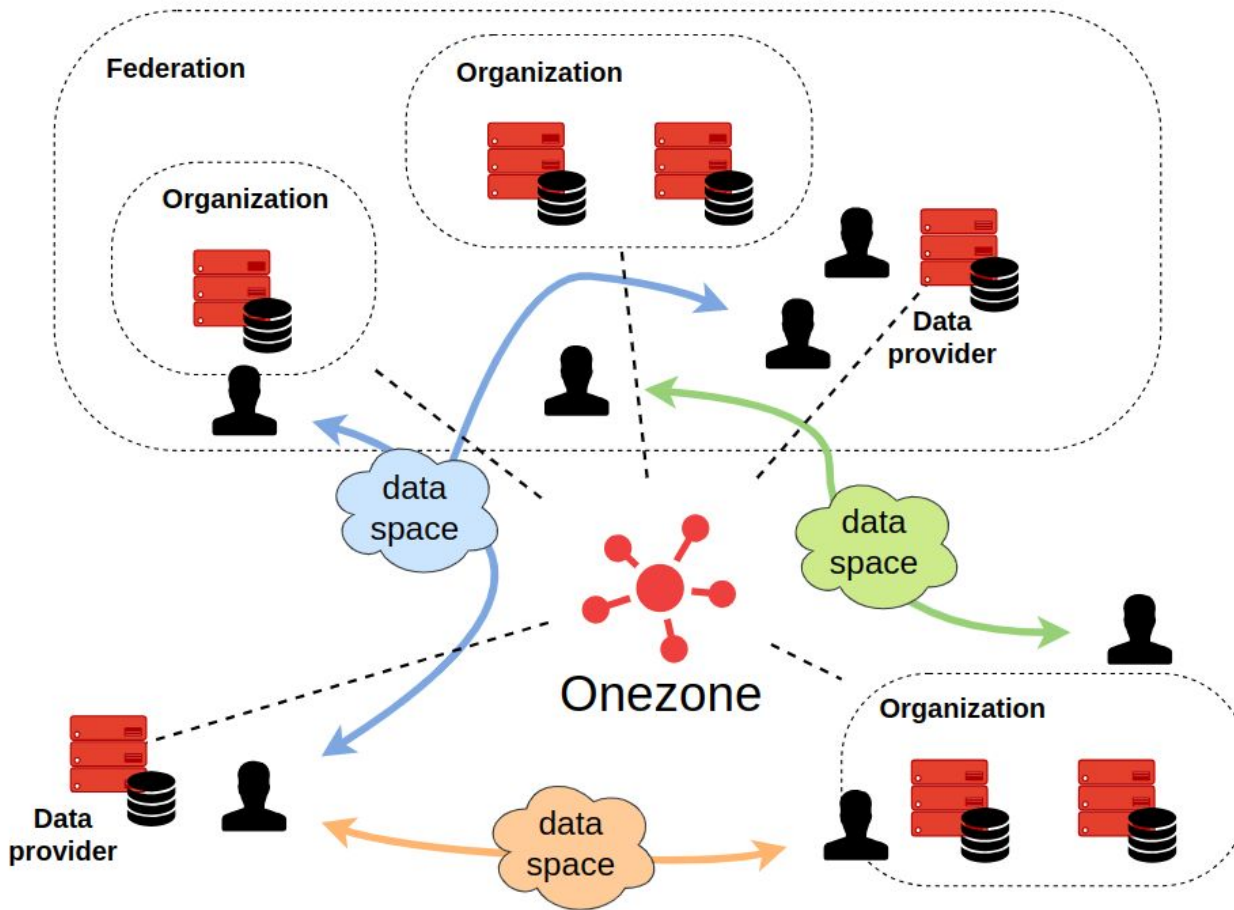
ONEDATA INSTANCES

- Between 5 and 10 active Zones in Poland and EU (depending on project lifecycles).
- Several instances not maintained by us.
- DataHub (**on the map**), long haul project:
 - **18** sites (Oneproviders)
 - **2255** data spaces
 - **~2.16PB** total storage size
 - **800+** users
- Archive for Polish National Museums:
 - **5PB** of data — the current phase
 - **10PB** of data — target scale
 - **~100M** files



Main concepts

UNIFIED ACCESS TO GEOGRAPHICALLY AND ORGANIZATIONALLY DISTRIBUTED DATA



Onezone

- central coordinator
- implements AAI
- entry-point to the system

Oneprovider

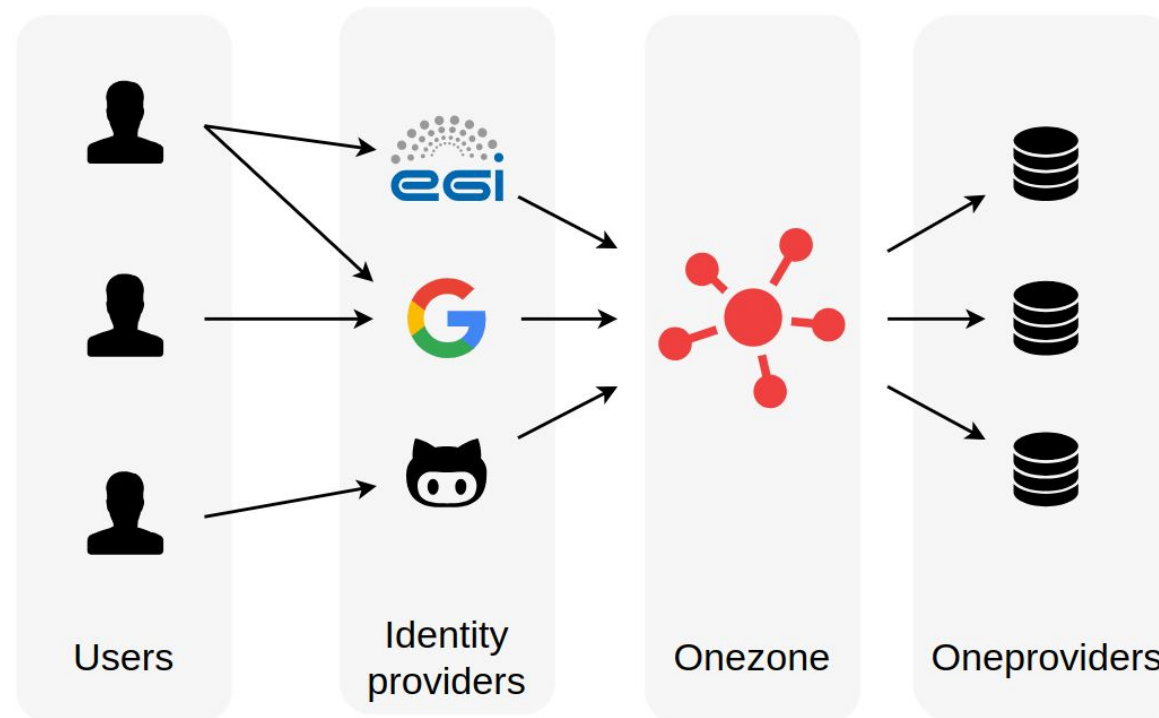
- installed at a data provider
- realizes data access & mgmt
- virtualizes physical storage

Open-source software stack

Take it and use it, **retaining control over your user data and storage resources.**

USERS

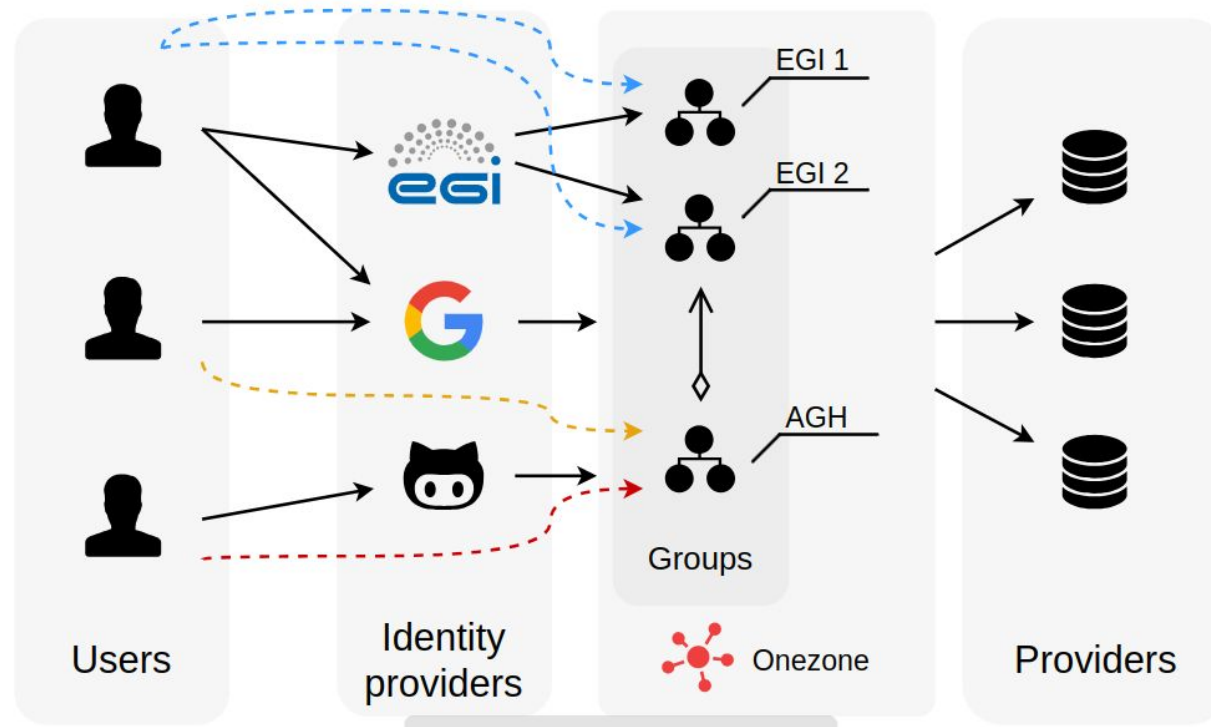
Users gain access to a Onedata platform by authenticating to the **Onezone service**. Authentication is done using **identity providers**, offering social and federated Single Sign-On (SSO) services — currently supported are **OpenID**, **OpenID Connect**, and **SAML 2.0**. Onezone admins configure the allowed IdPs.



GROUPS

A user can be a member of multiple groups. Groups can be organized in hierarchical structures.

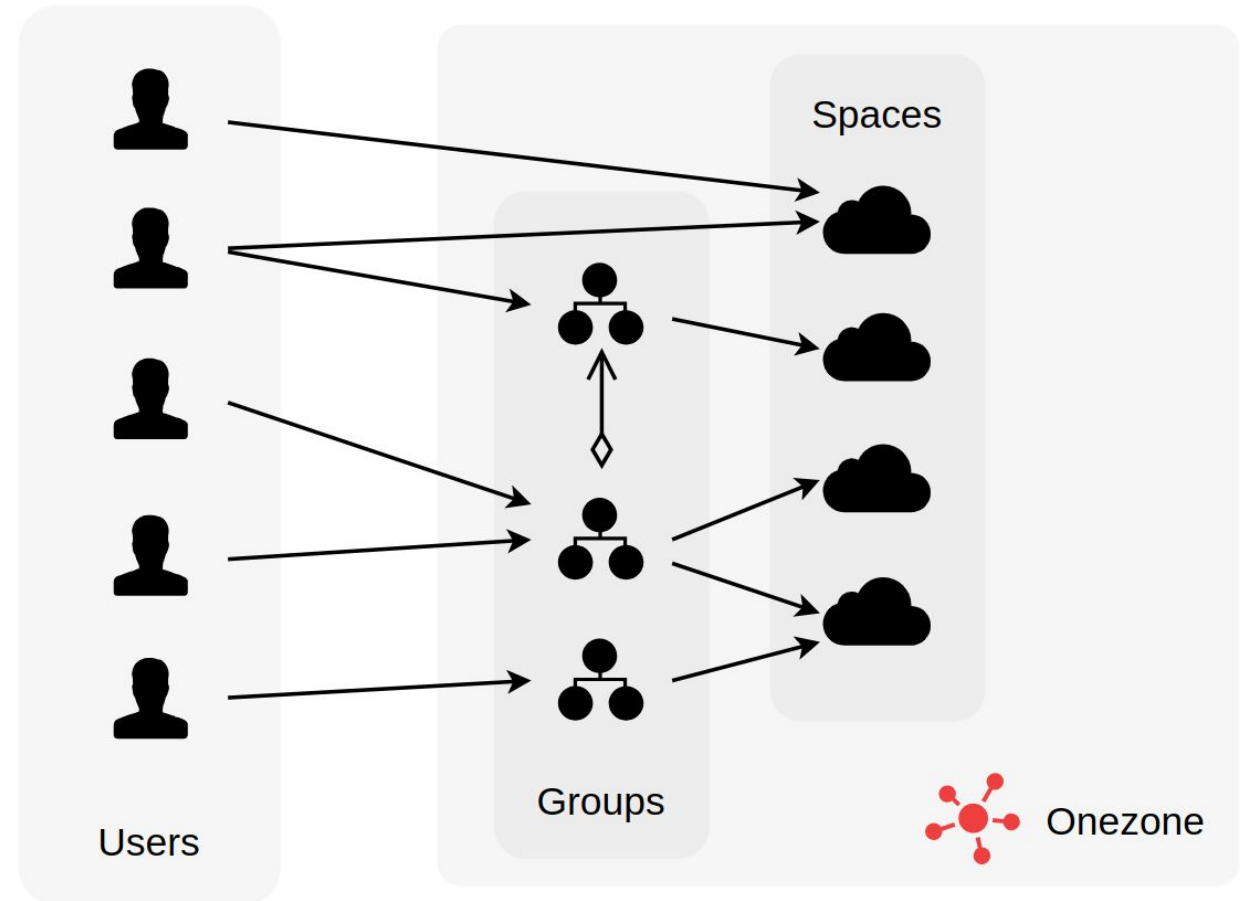
With **entitlement mapping**, user's entitlements (organizational group memberships recognizable by given IdP) are transmitted in SAML assertions or OpenID user info and then mapped to Onedata groups upon login.



SPACES

All the data stored in Onedata is organized into **Spaces**.

- Space is a **virtual, logical data volume** with a **POSIX-like filesystem**.
- **Physical data** of a single **logical file** can be **distributed**.



Web UI

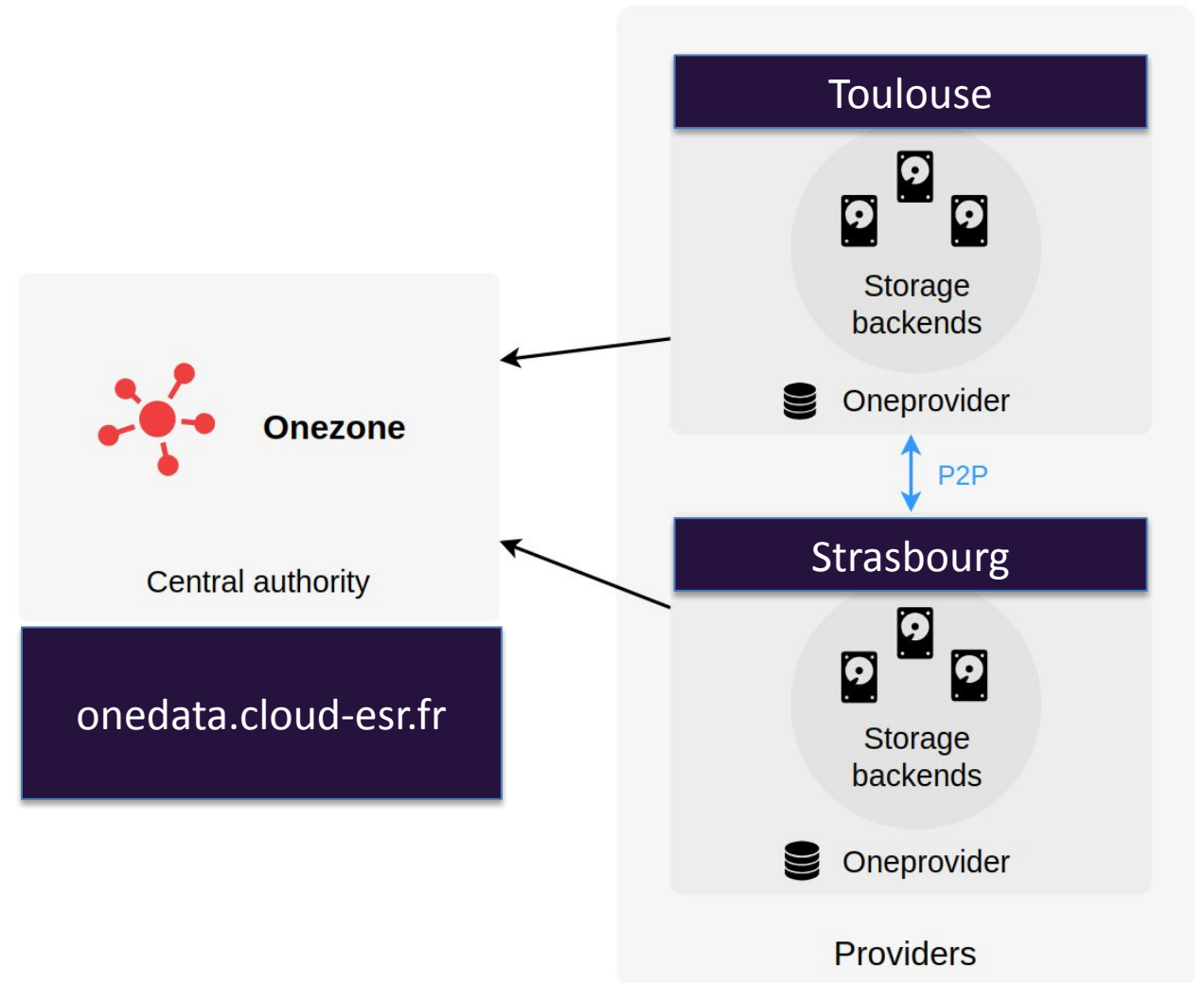
<https://onedata.cloud-esr.fr>

main concepts

PROVIDERS

Oneprovider is a piece of software installed at a data center.

- Manages **physical data** on underlying **storage backends**.
- **Synchronizes metadata** with other providers in a **P2P** manner.
- Handles **data access** requests.
- Subject to a Onezone.



Web UI

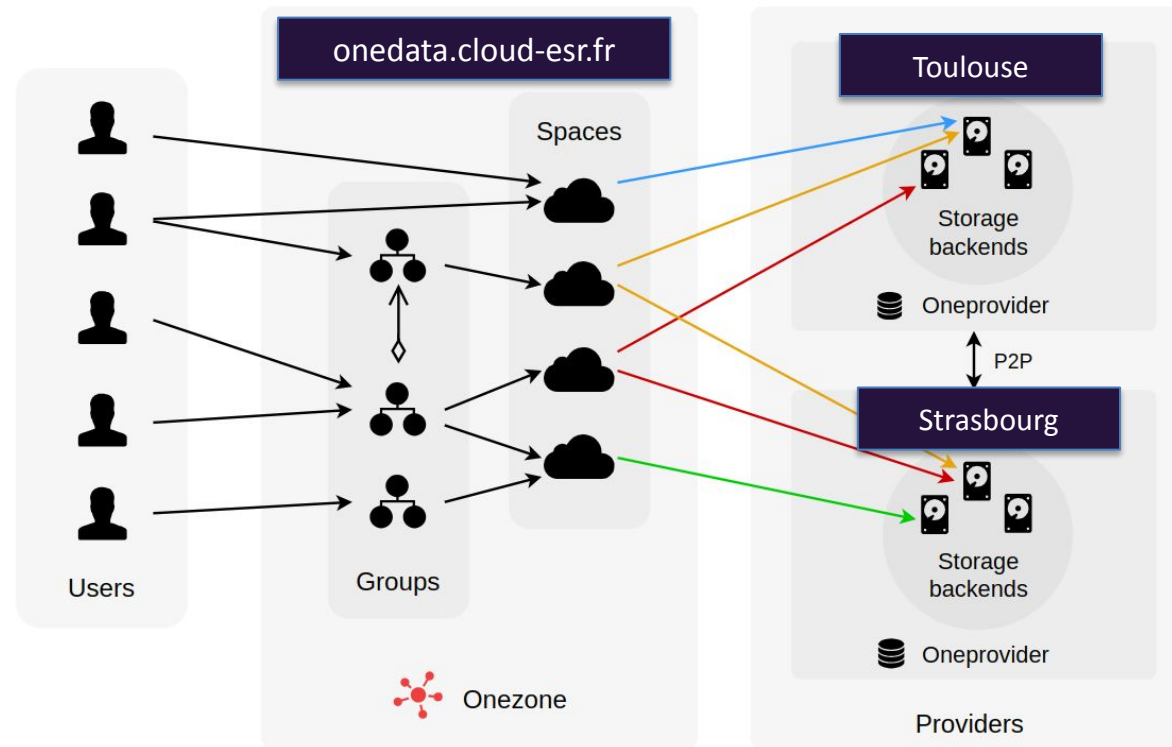
<https://onedata.cloud-esr.fr>

providers

SPACE SUPPORT

Spaces can be seen as **virtual directories or volumes**, which contain a **POSIX file tree**, while the underlying **physical data is distributed** across multiple **storage backends**, governed by providers.

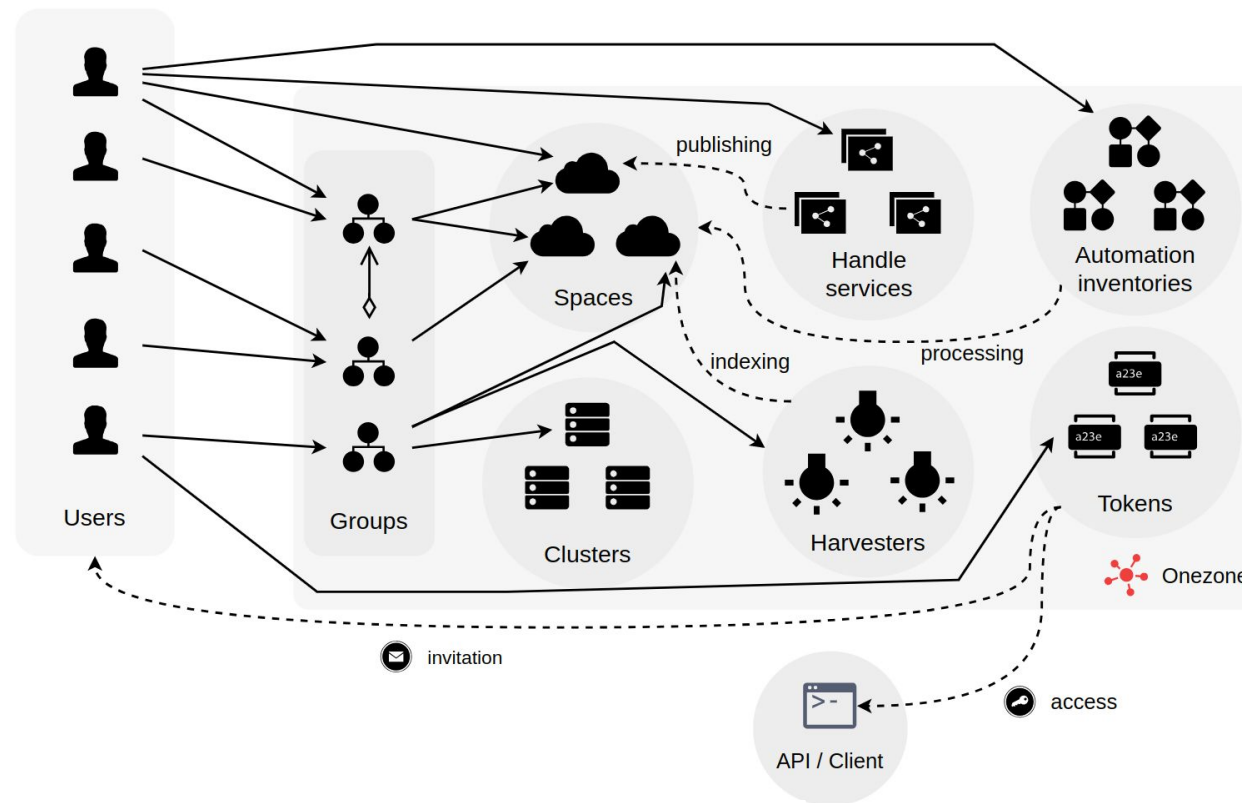
A provider grants **support** for a space, declaring a certain **storage quota**.



ONEDATA MEMBERSHIP MODEL

Groups in Onedata are fundamental to the **membership model** that regulates who can **access certain resources**.

Users and groups can become **members** of **Onedata resources**. A member essentially **gains access** to the resource, but the scope of the access is regulated by **granular privileges**.



Web UI

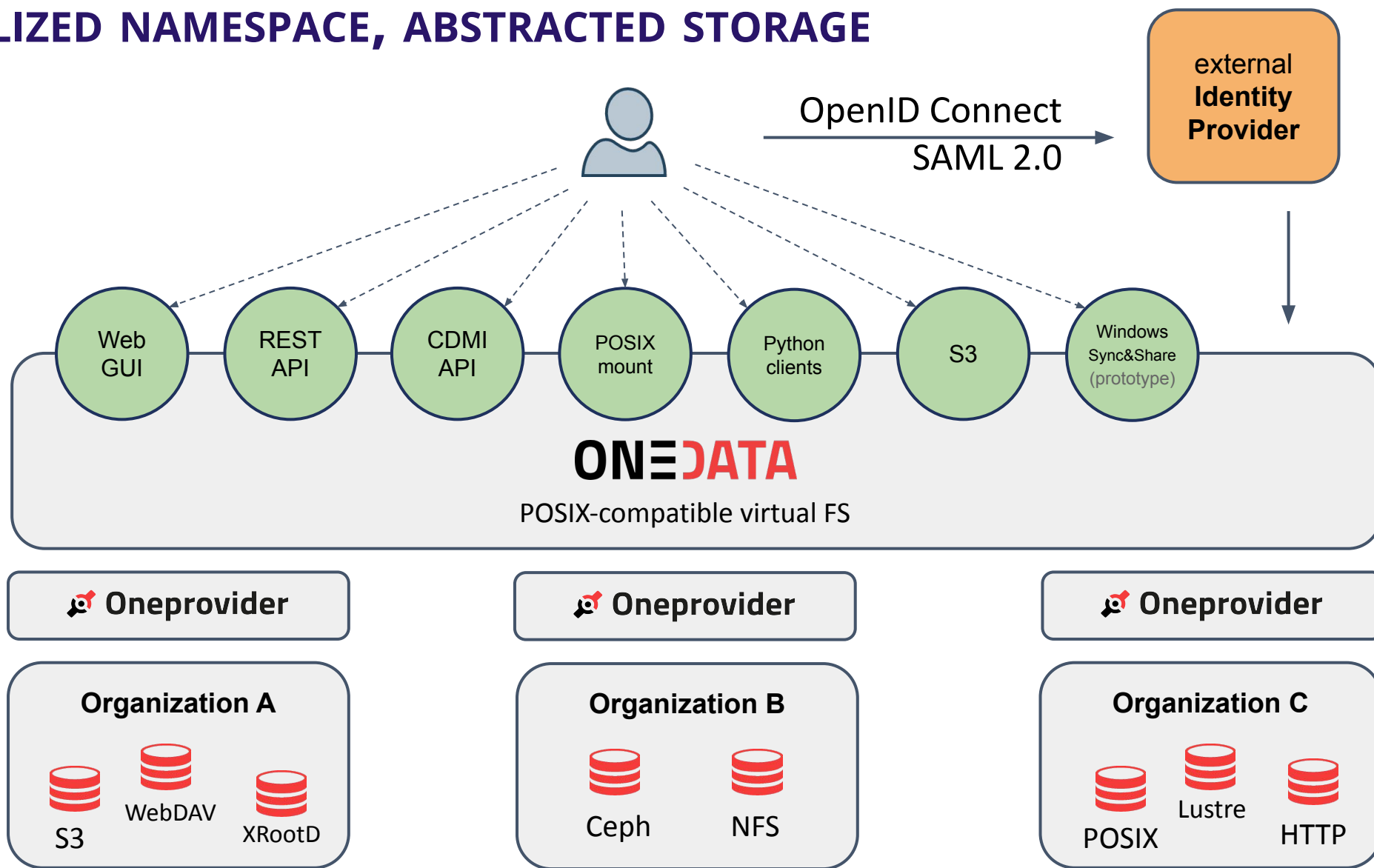
<https://onedata.cloud-esr.fr>

members and privileges

Logical data

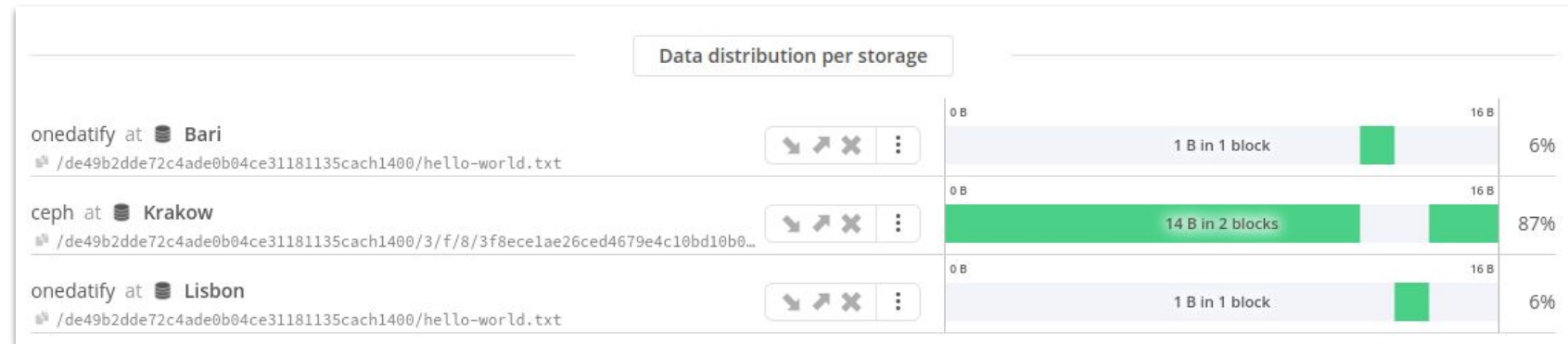
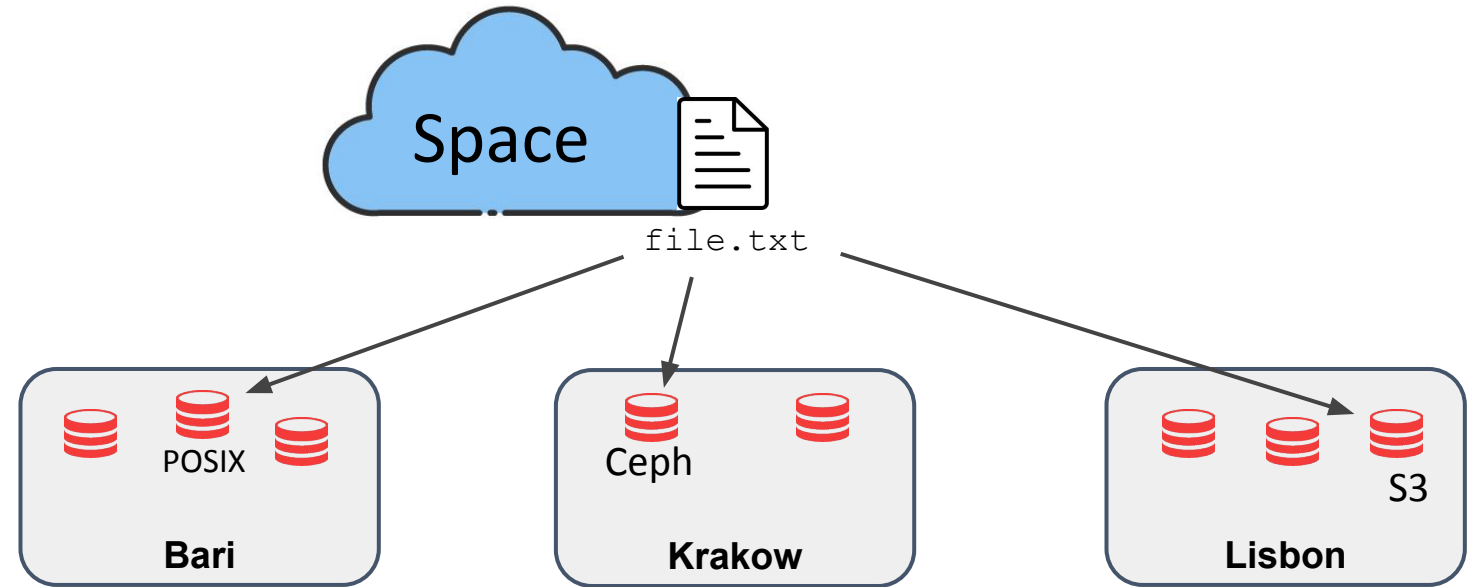
replication, transfers

VIRTUALIZED NAMESPACE, ABSTRACTED STORAGE

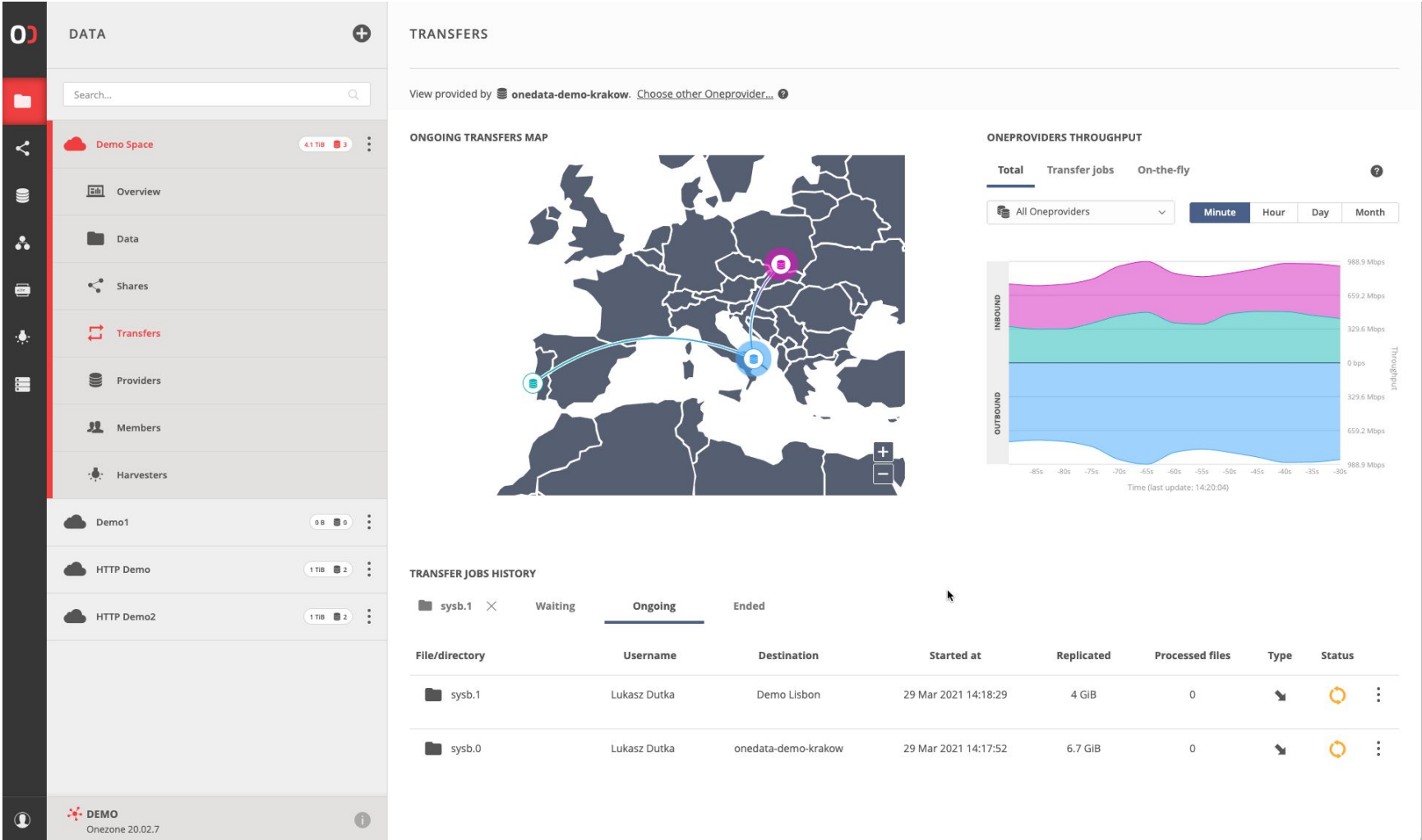


REPLICA MANAGEMENT

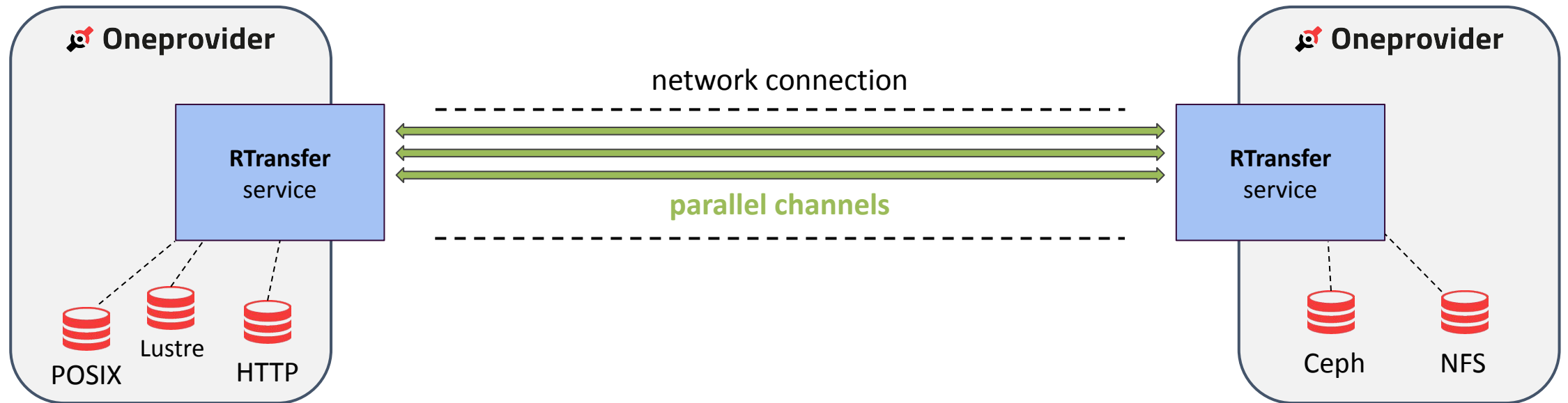
- Chunk-based replica management
- Missing chunks delivered on-the-fly
- Handled automatically, but controllable manually -
prestage, transfer, migrate, define QoS rules
(policy-based replica mgmt)



FLEXIBLE, HIGH-THROUGHPUT TRANSFERS



THE RTRANSFER PROTOCOL



Web UI

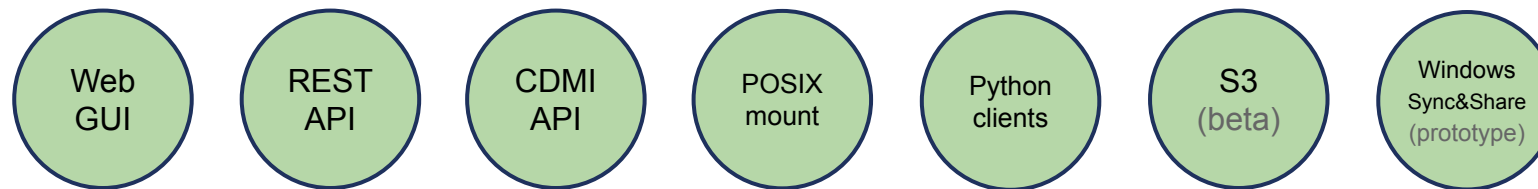
<https://onedata.cloud-esr.fr>

data distribution, transfers

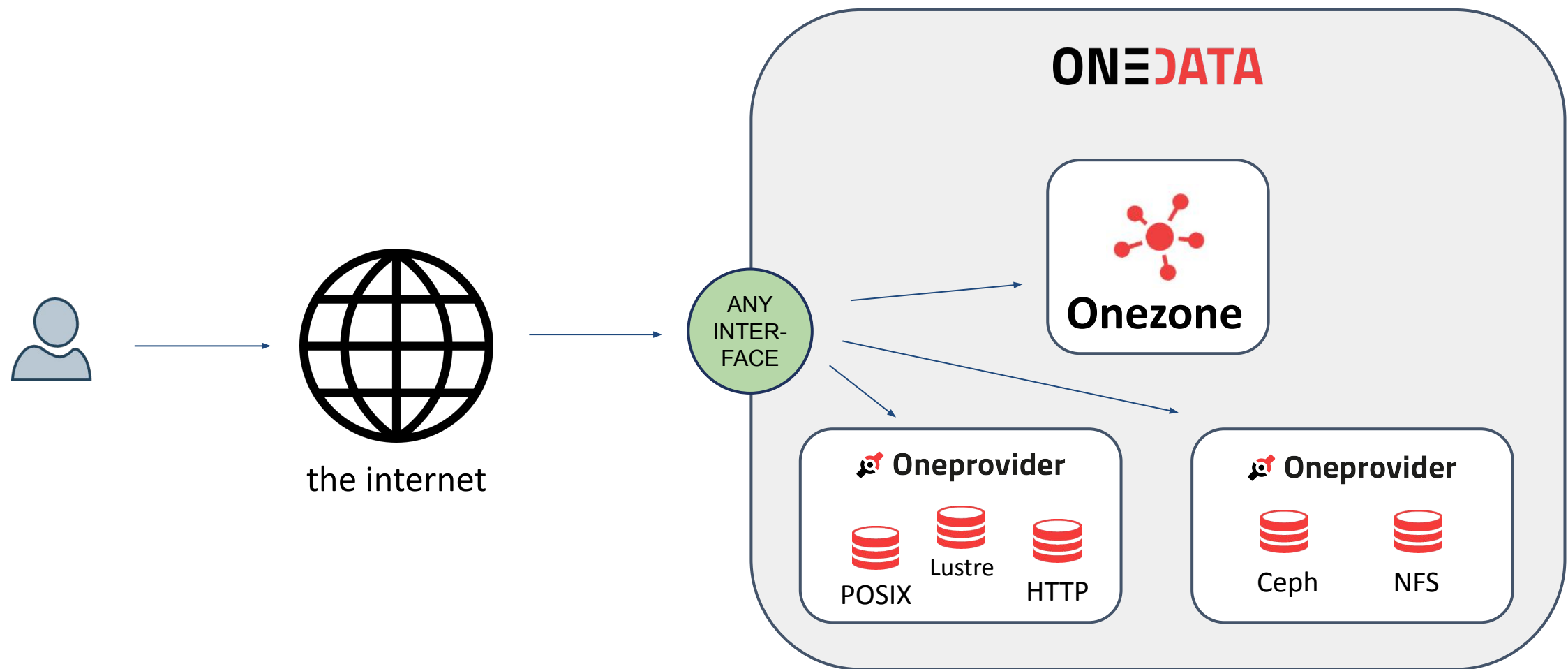
Data access

VIRTUALIZED DATA ACCESS

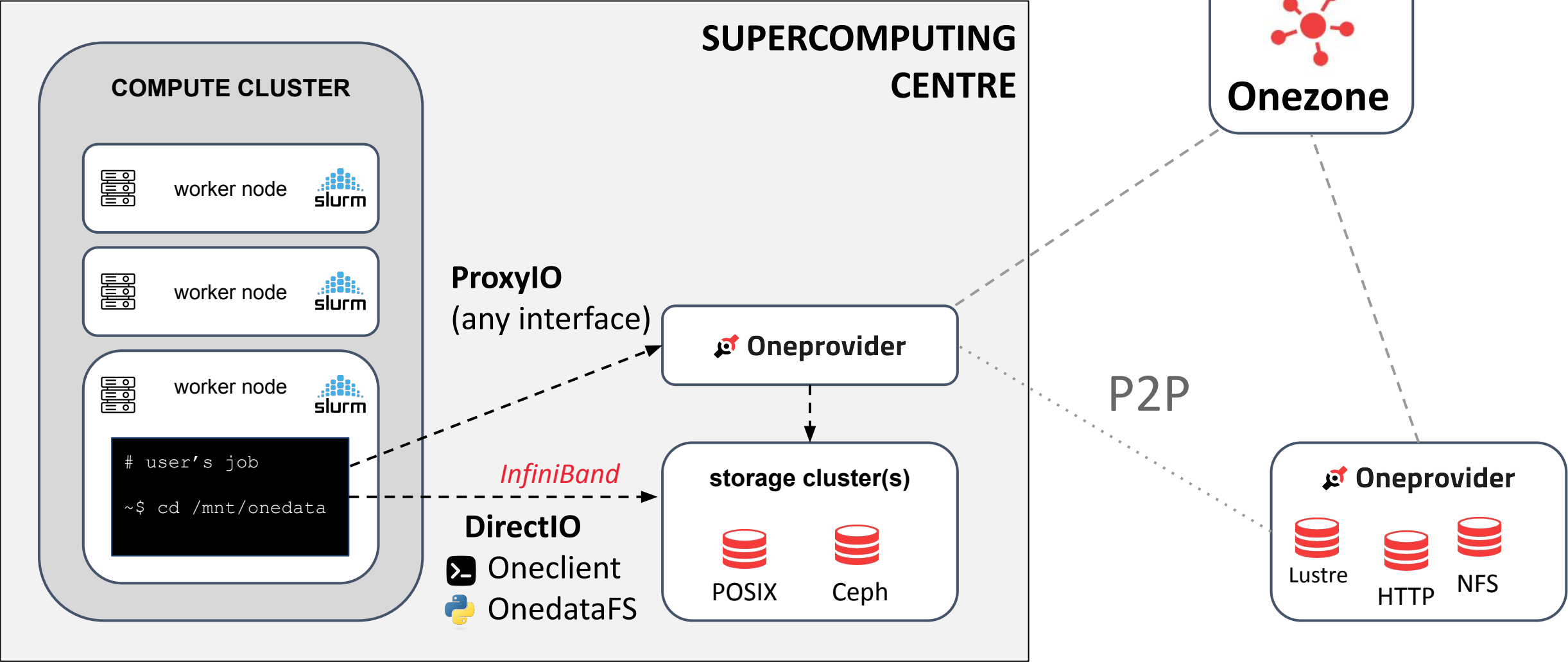
- **Unified** view on the data, regardless of the interface.
- **Anytime / anyplace** data access with any interface.
- Support for a wide range of **use cases**, including personal use, enterprise setups and scientific ecosystems.



USE CASE A) ACCESS FROM A PERSONAL LAPTOP

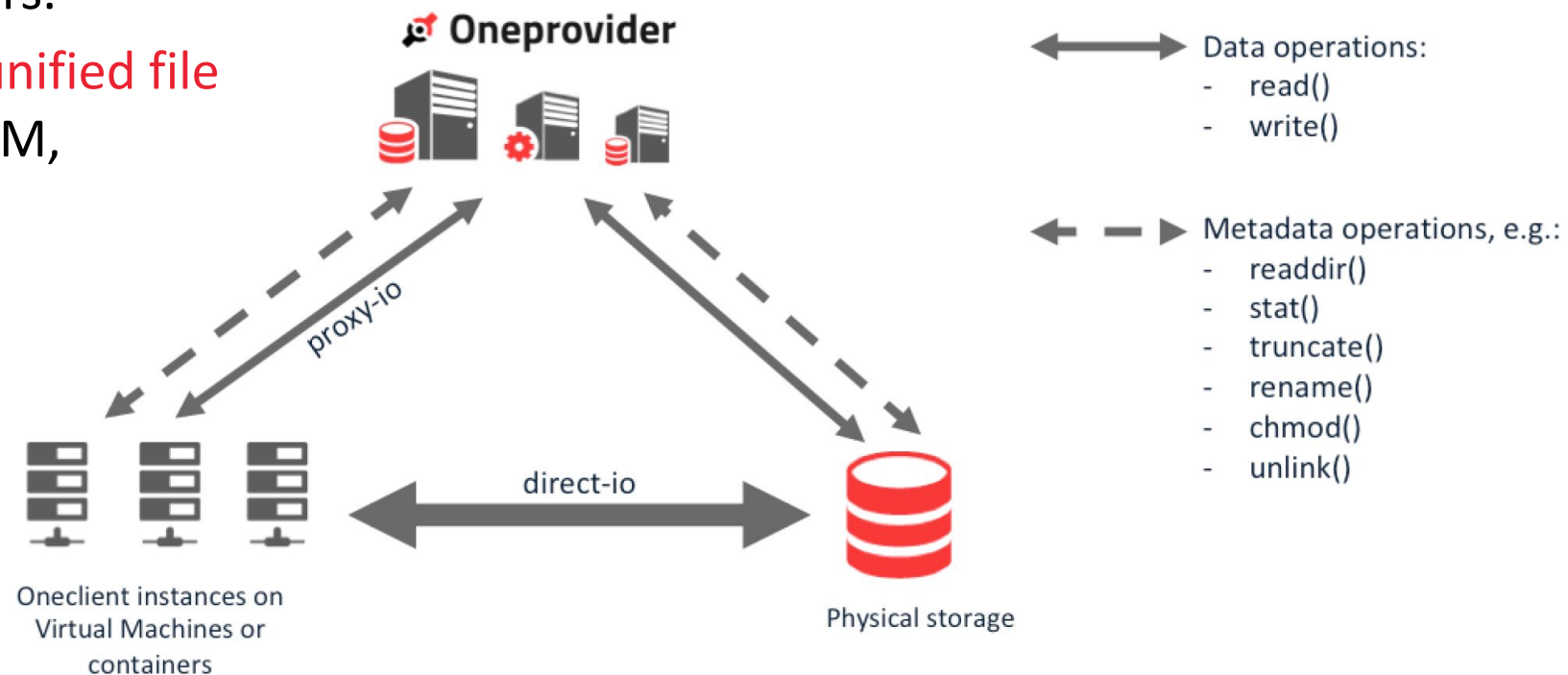


USE CASE B) COMPUTE CLUSTER WITH SLURM

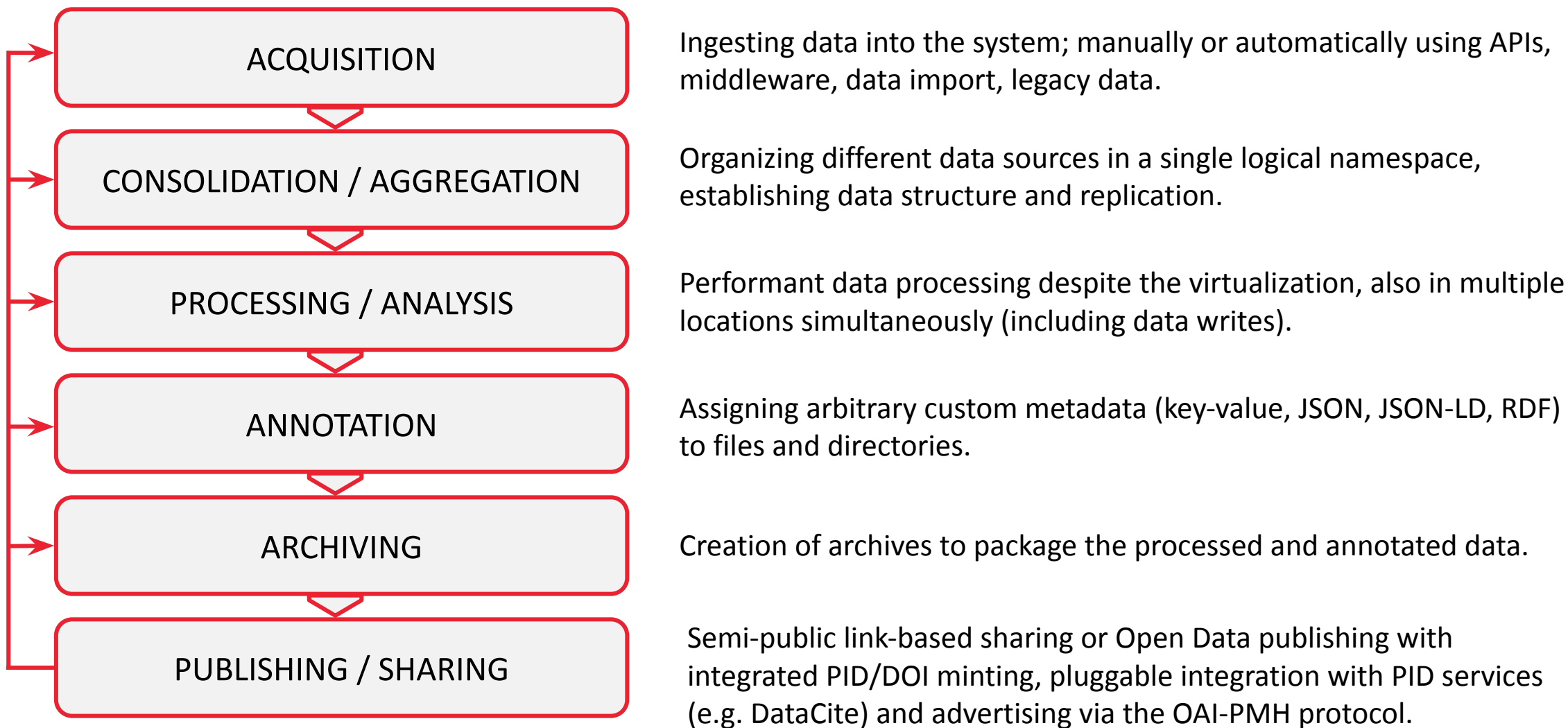


DIRECT STORAGE ACCESS FOR HPC — DIRECTIO

- **Oneclient** - Linux **POSIX** mount supporting most of the POSIX operations (C++).
- **OnedataFS** - **Python** `pyfilesystem` plugin reusing the C++ storage drivers.
- All your data accessible in a **unified file system** mountable on a PC, VM, container, or Cloud/Grid worker node.



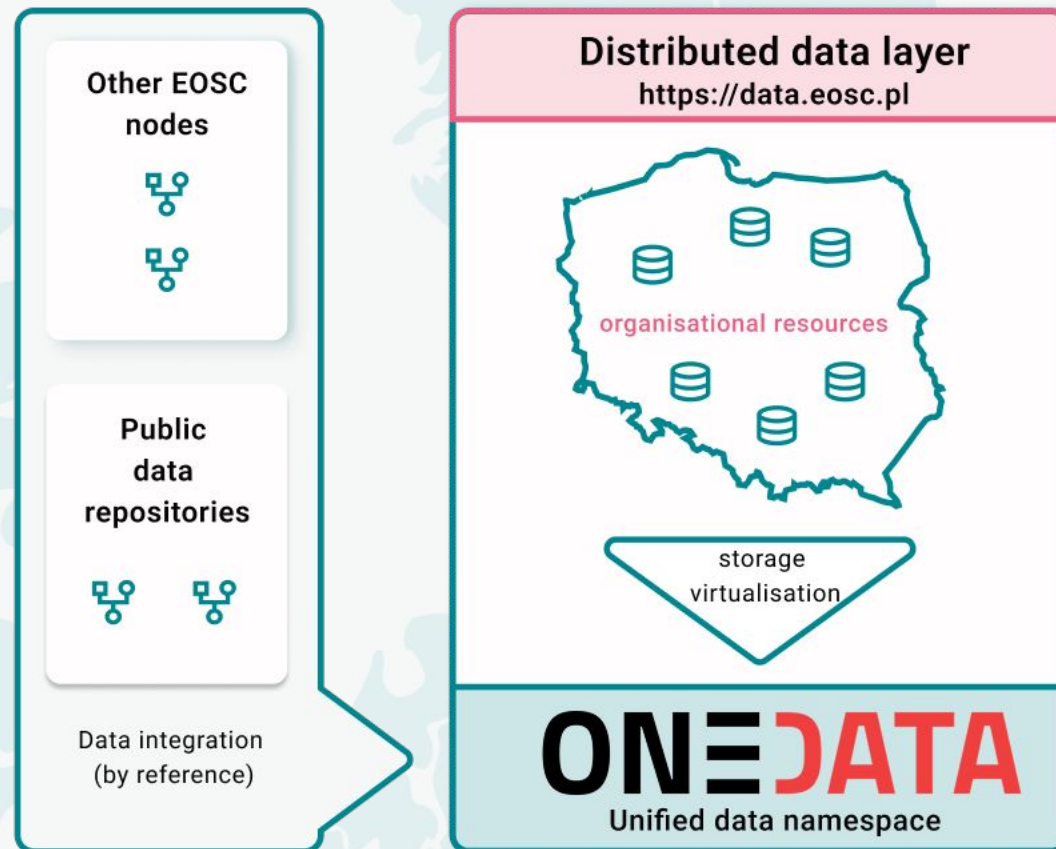
EXEMPLARY DATA LIFECYCLE IN ONEDATA

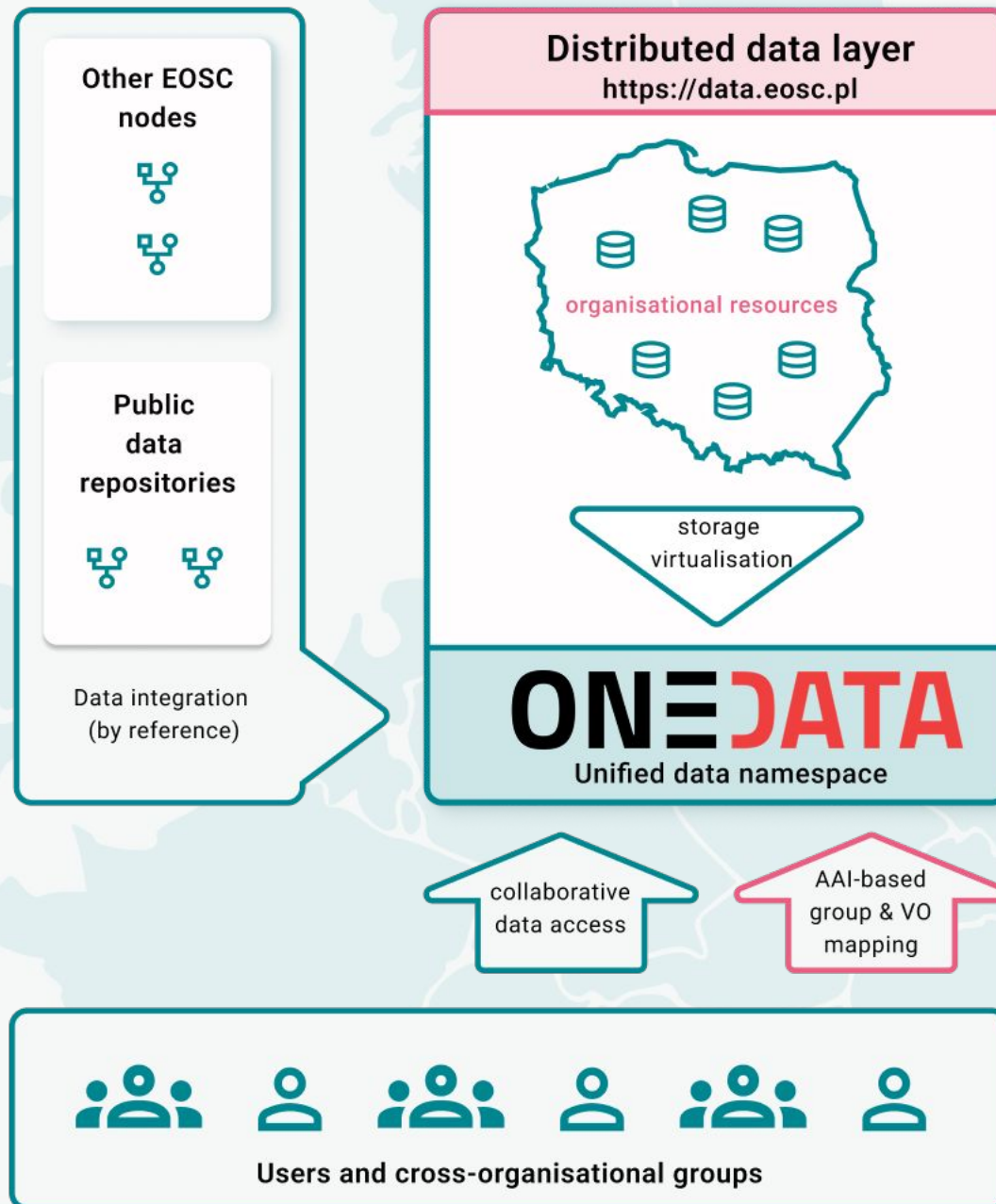


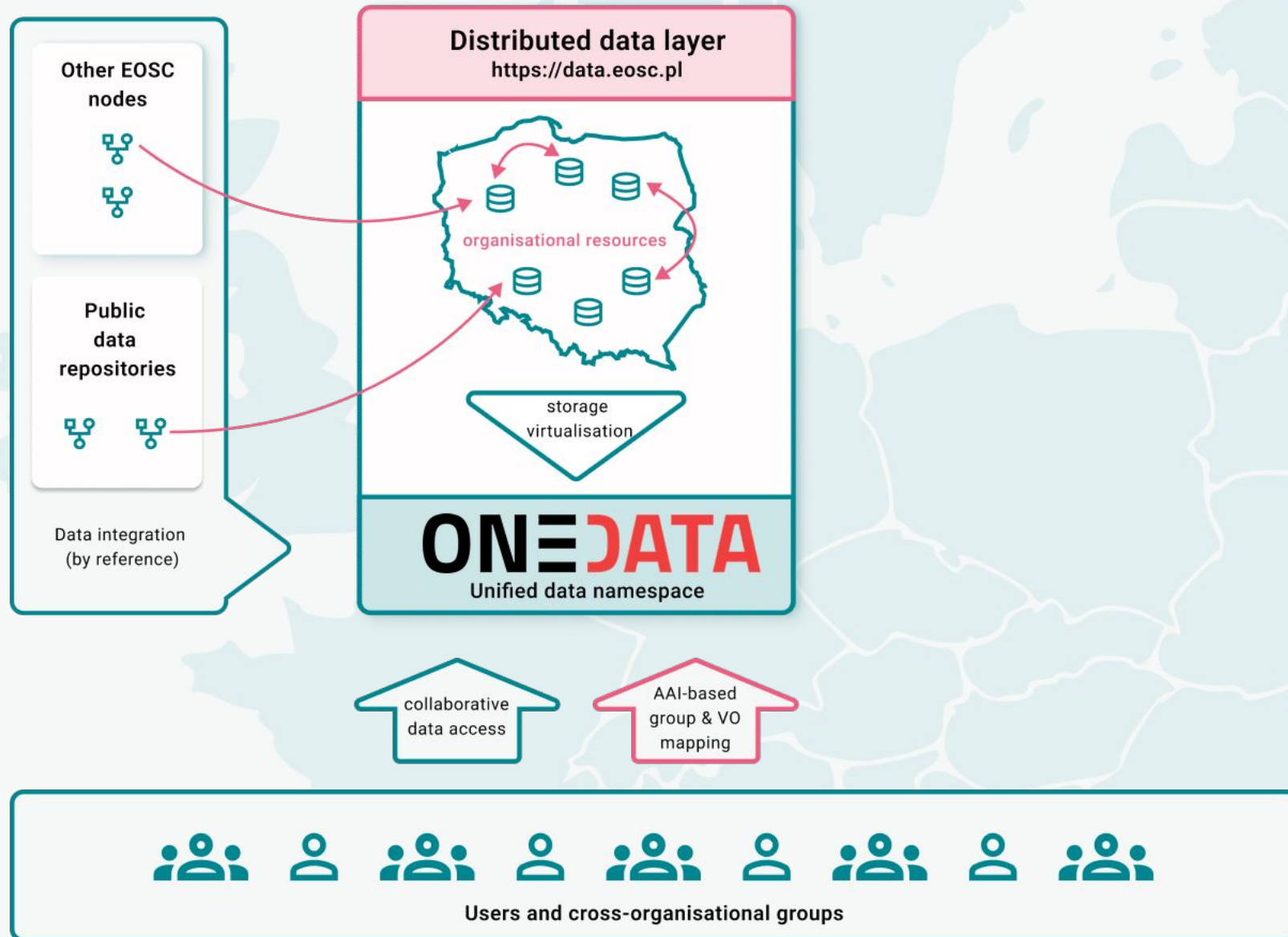
Use case from
eosc.pl

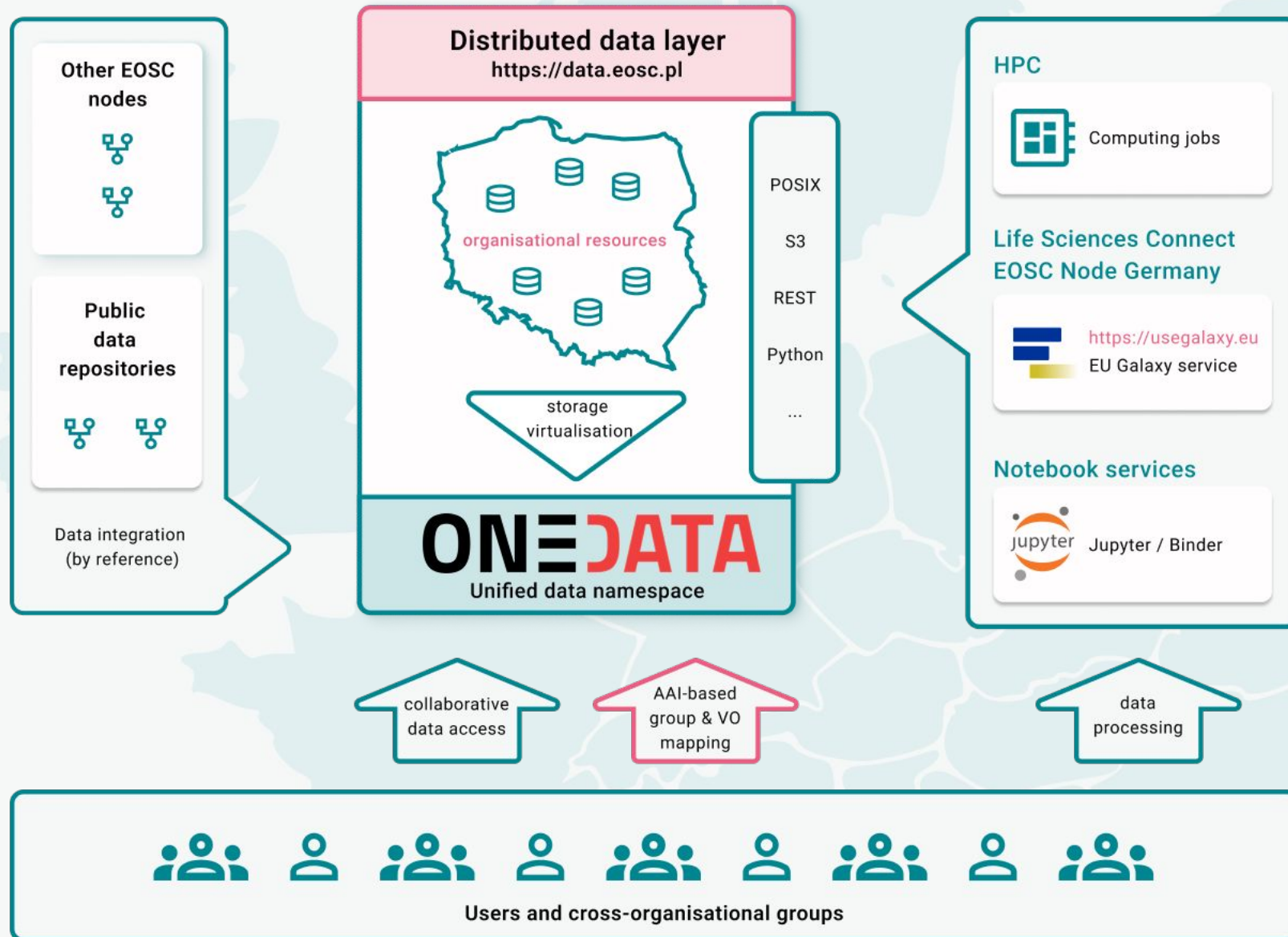
Onedata in EOSC.PL

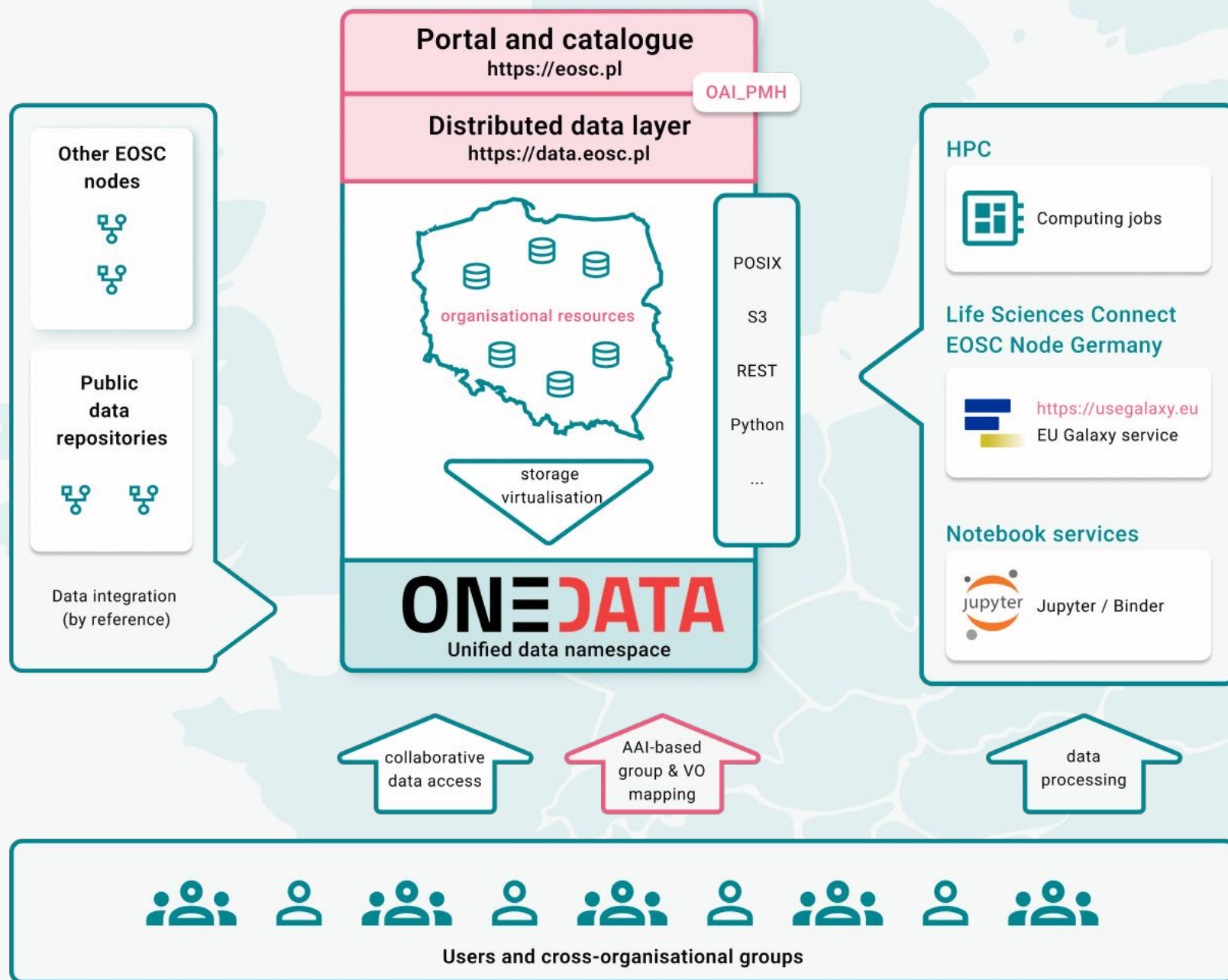


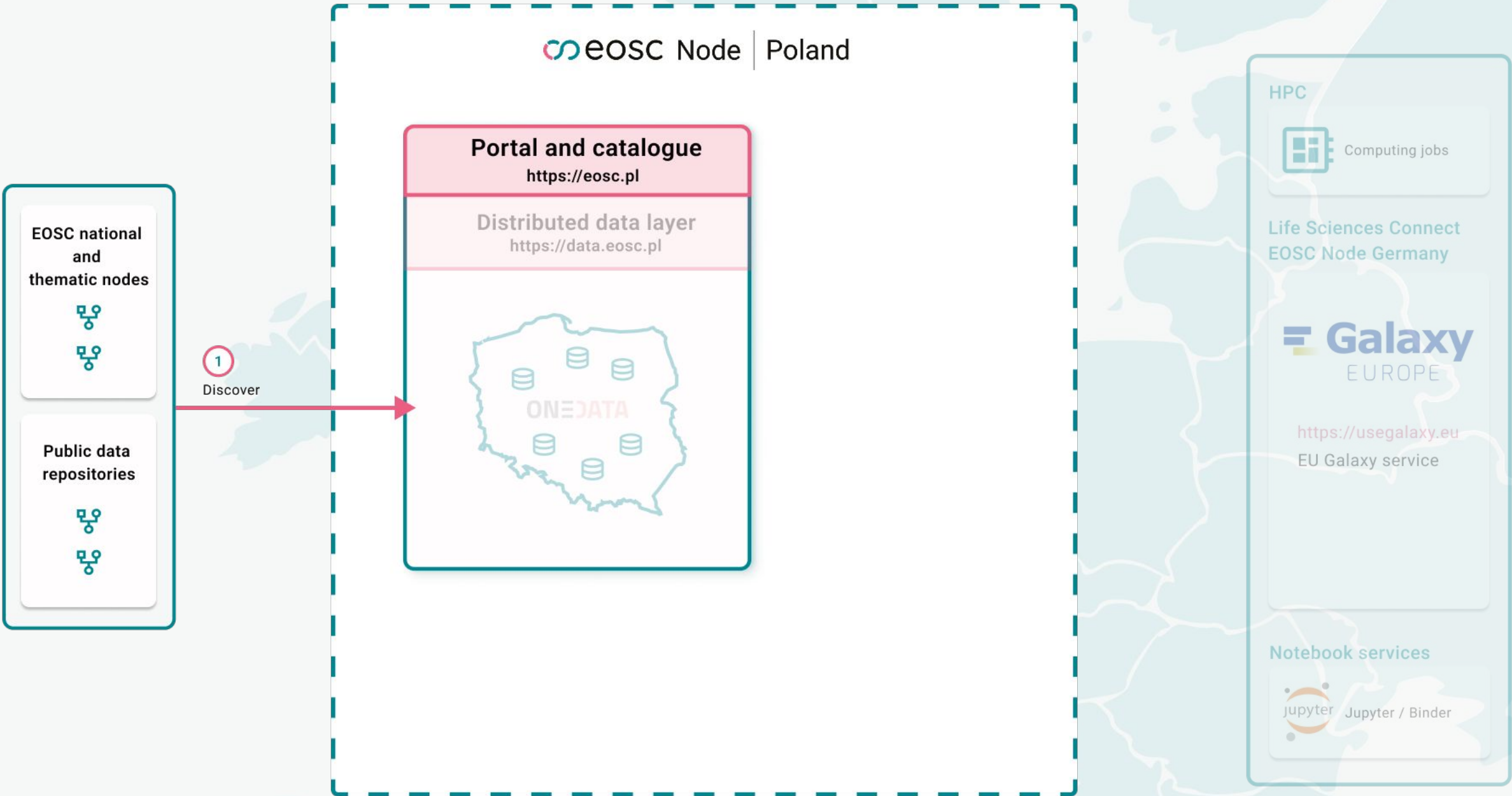


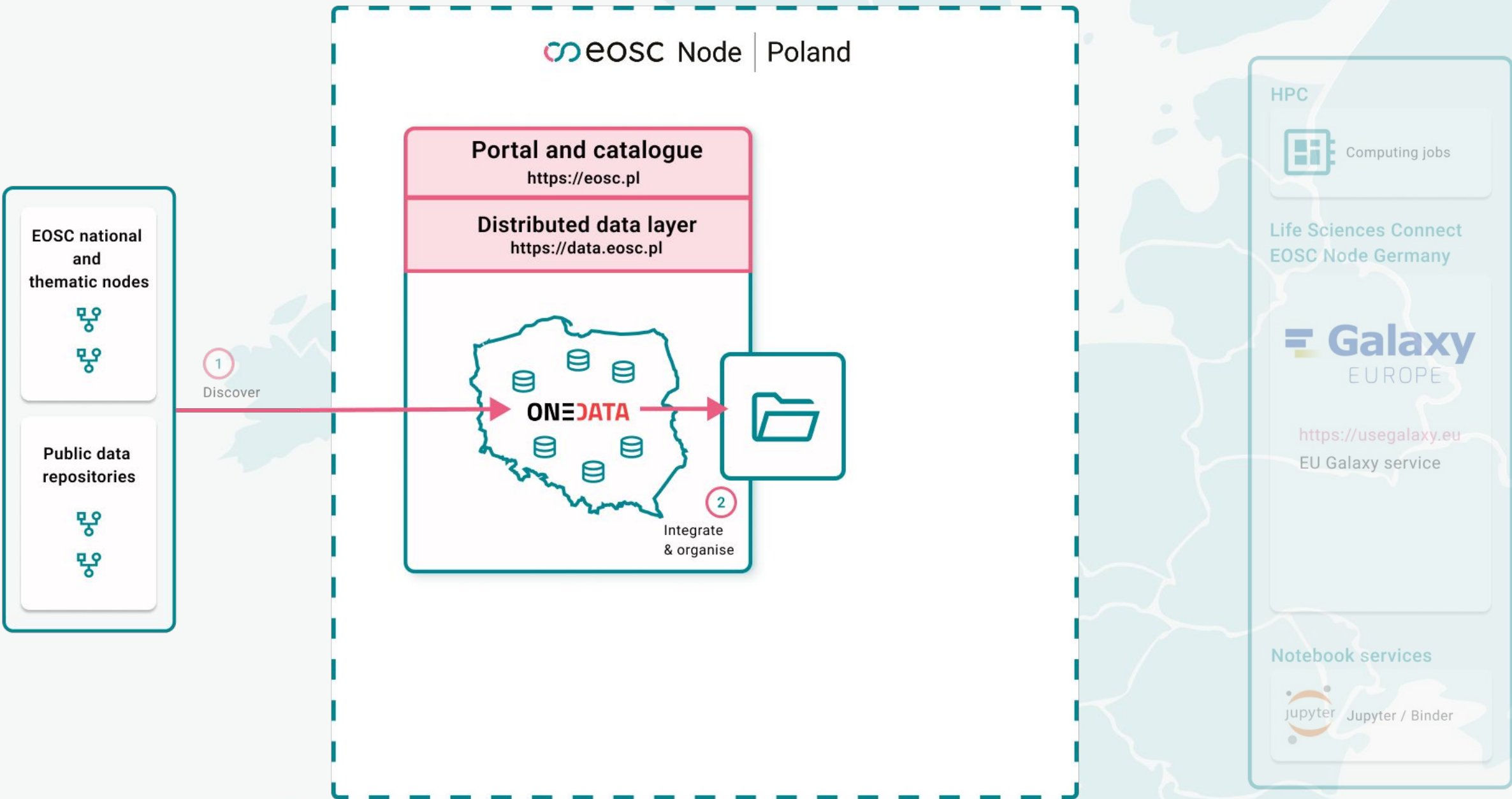


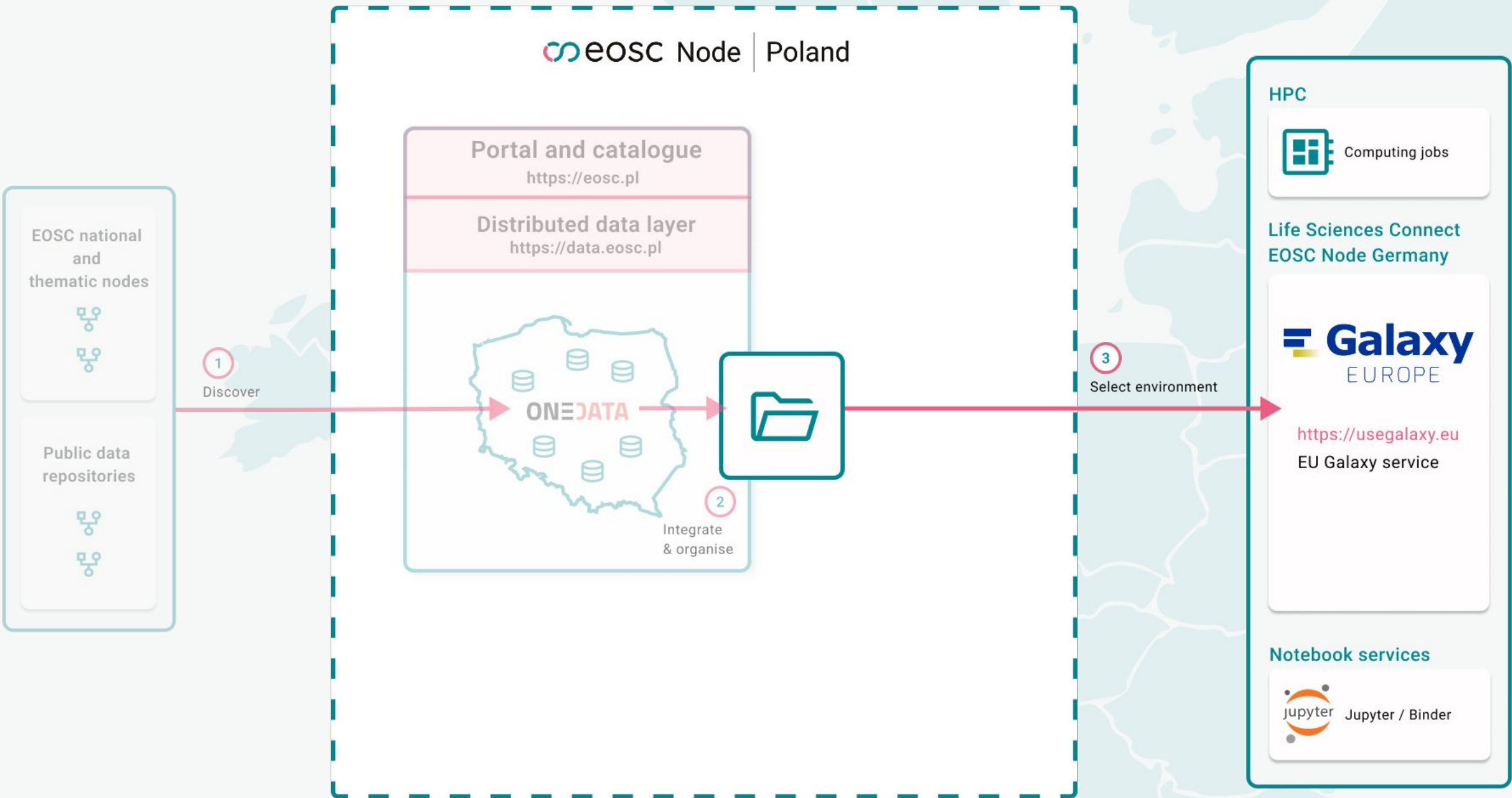


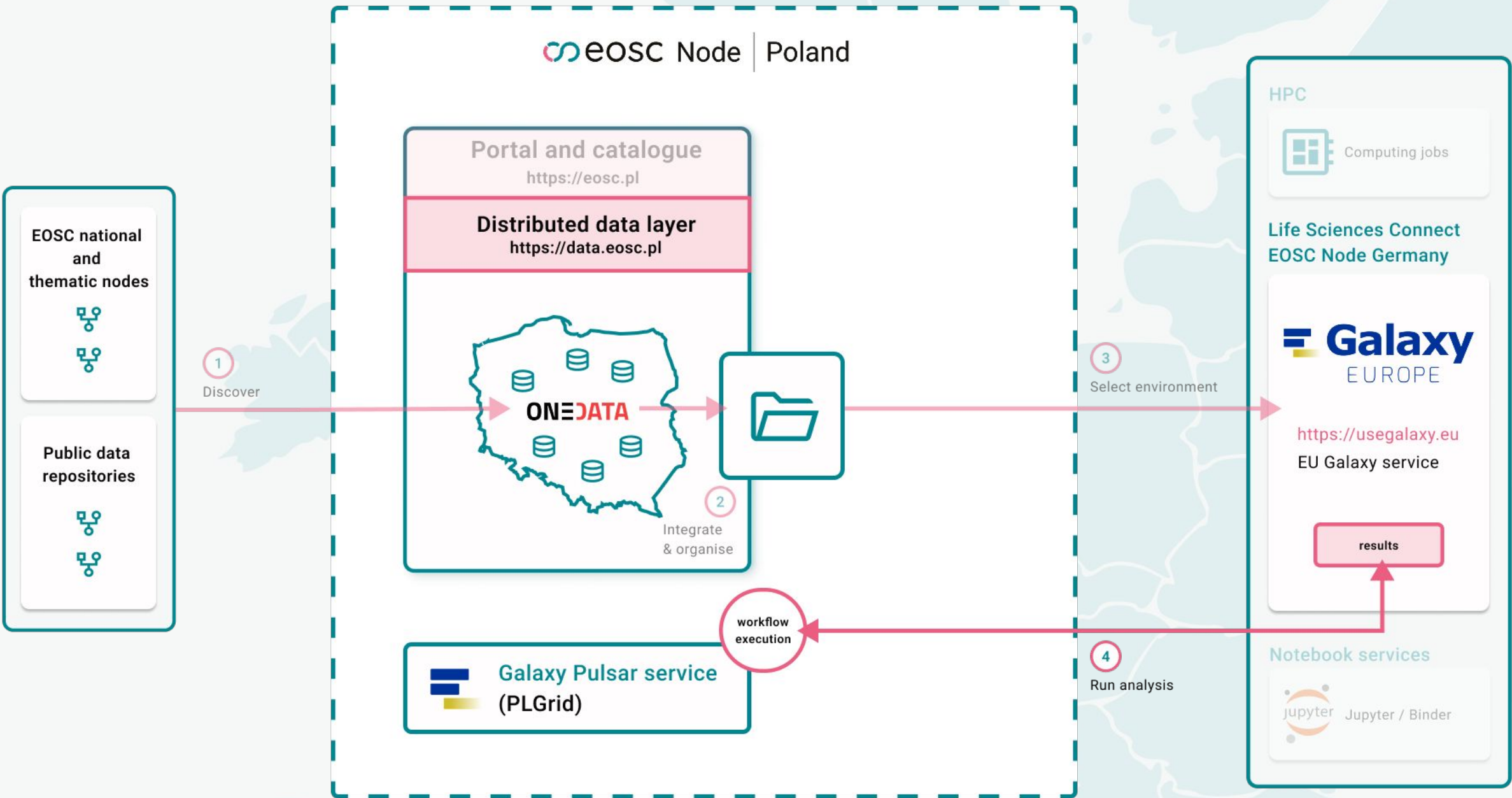


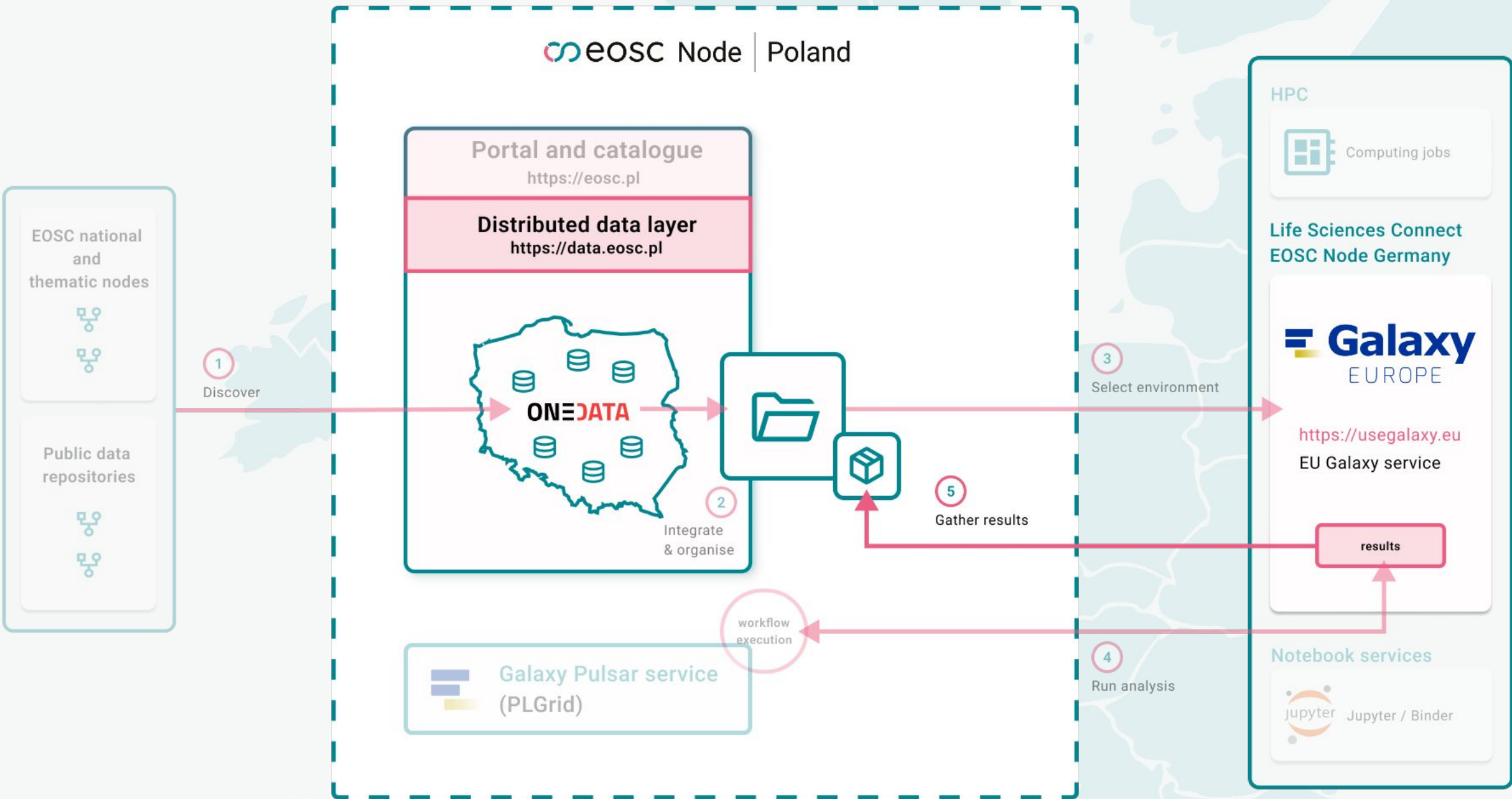


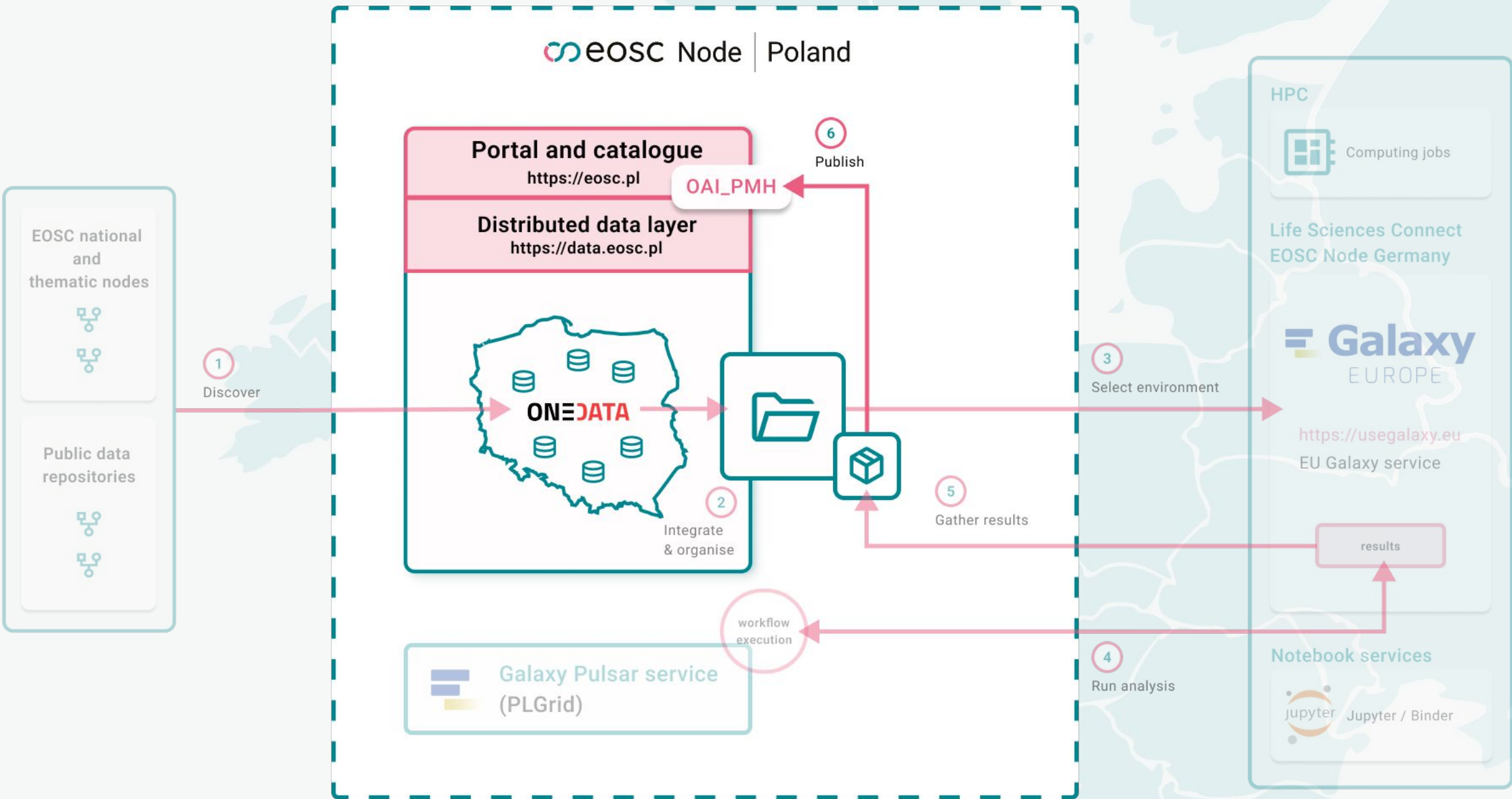


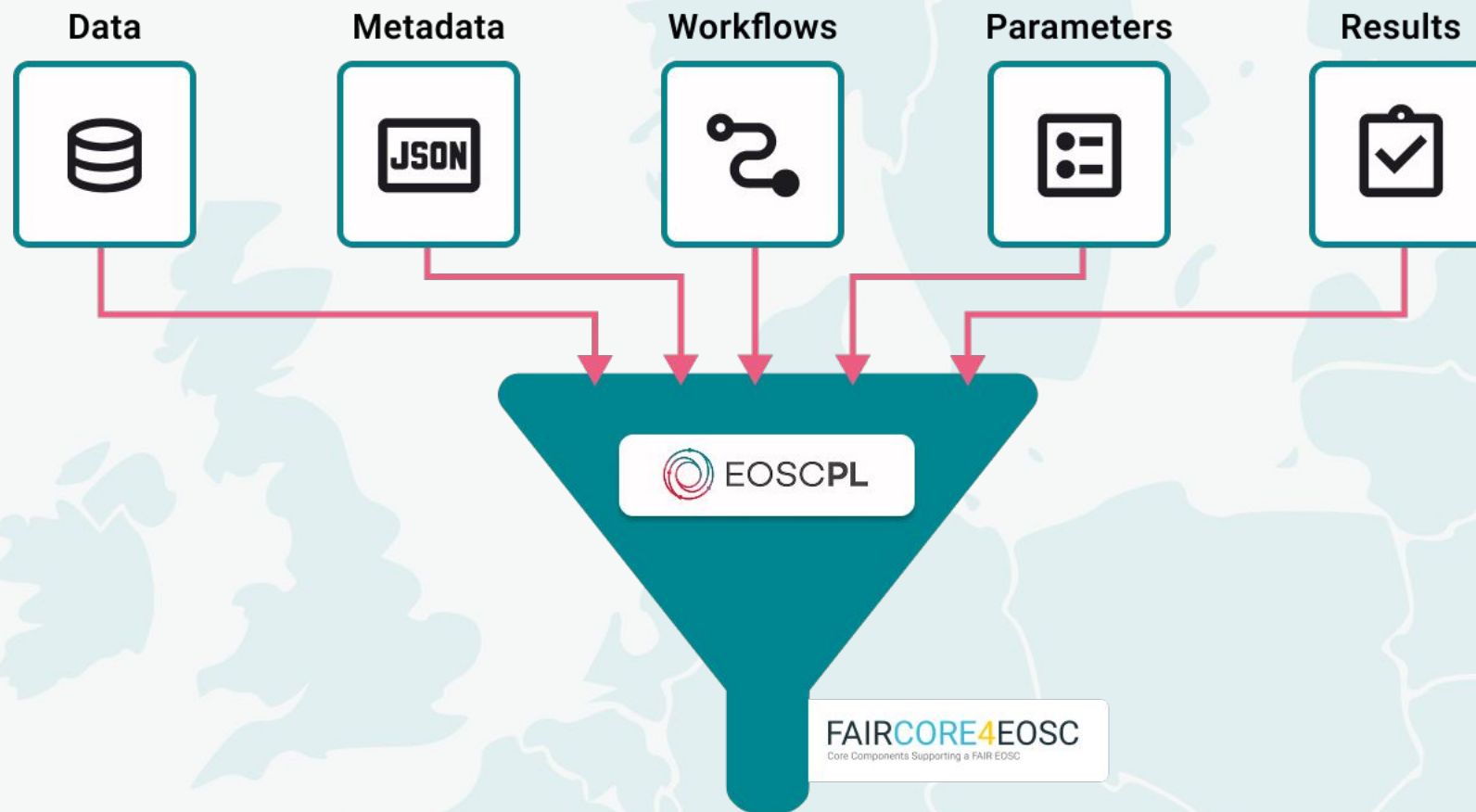










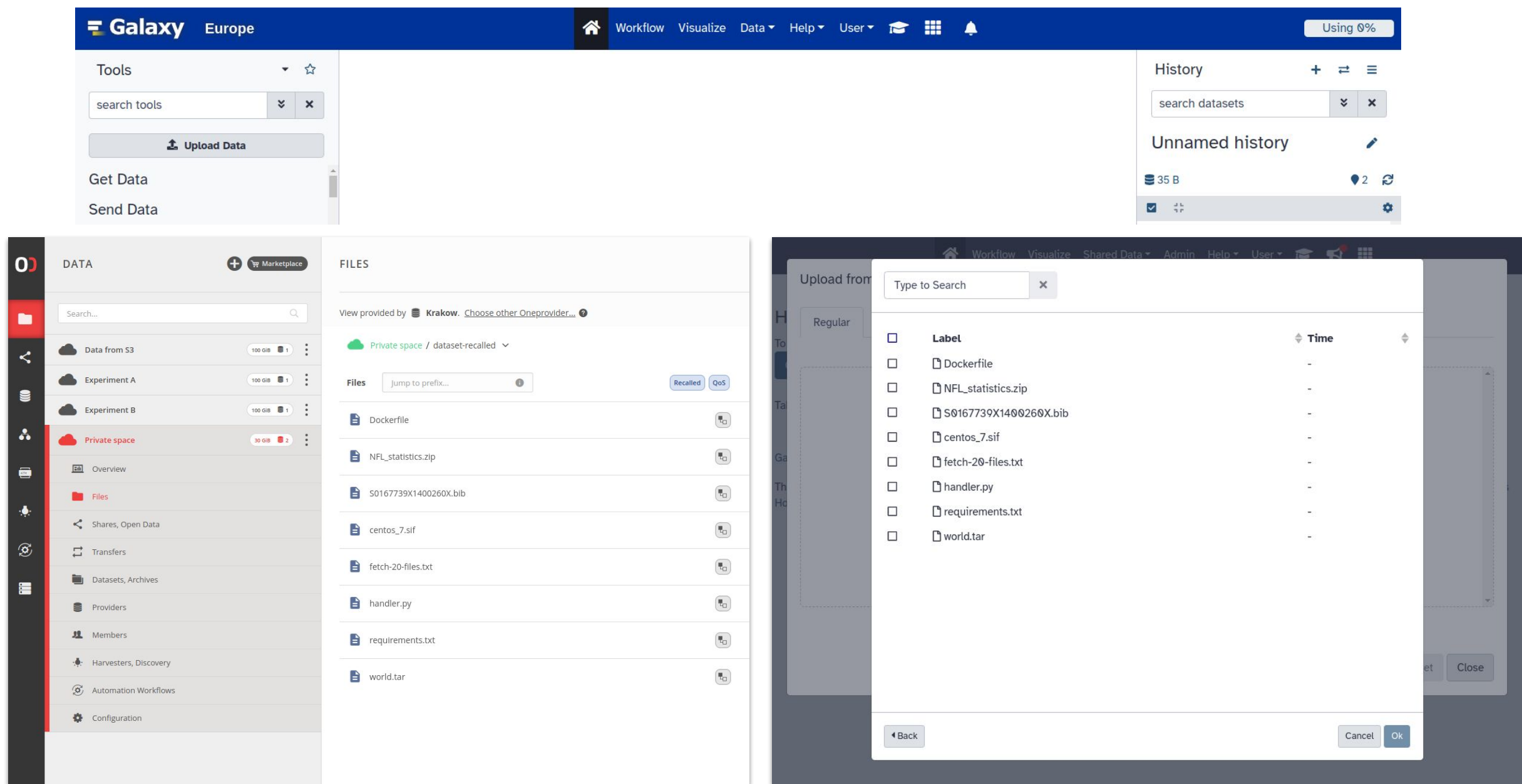


RO-Crate Research Package



- Open Science
- Data reuse
- Reproducibility

PID: <http://hdl.handle.net/21.15131/pLrlicyt>

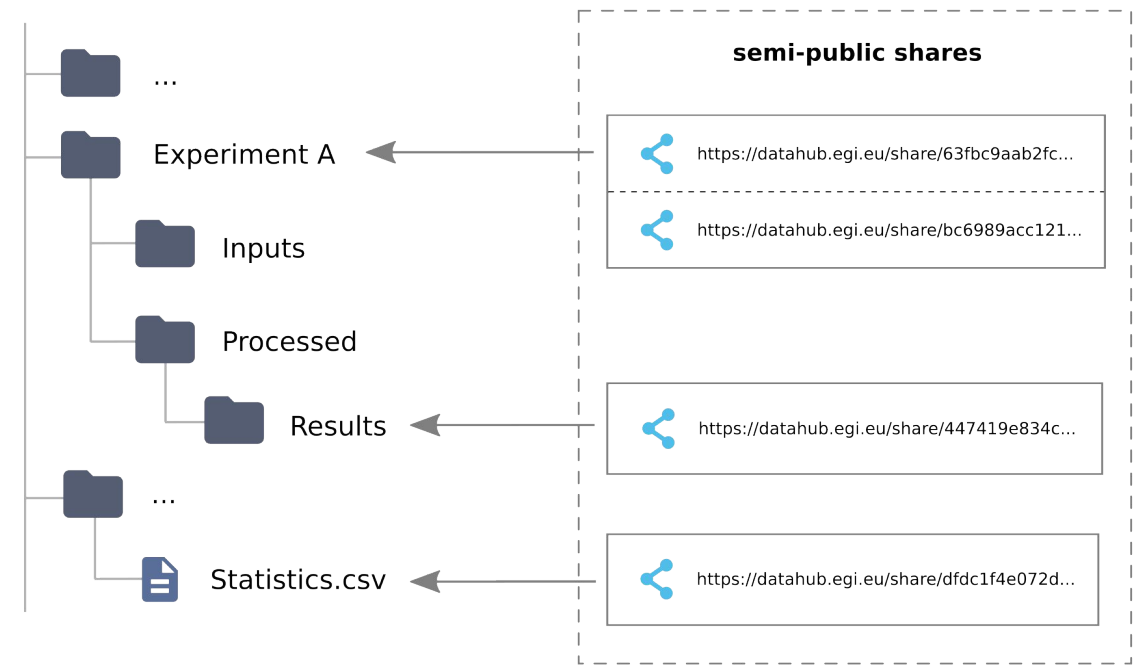


Data publishing

Shares, Open/Public Data

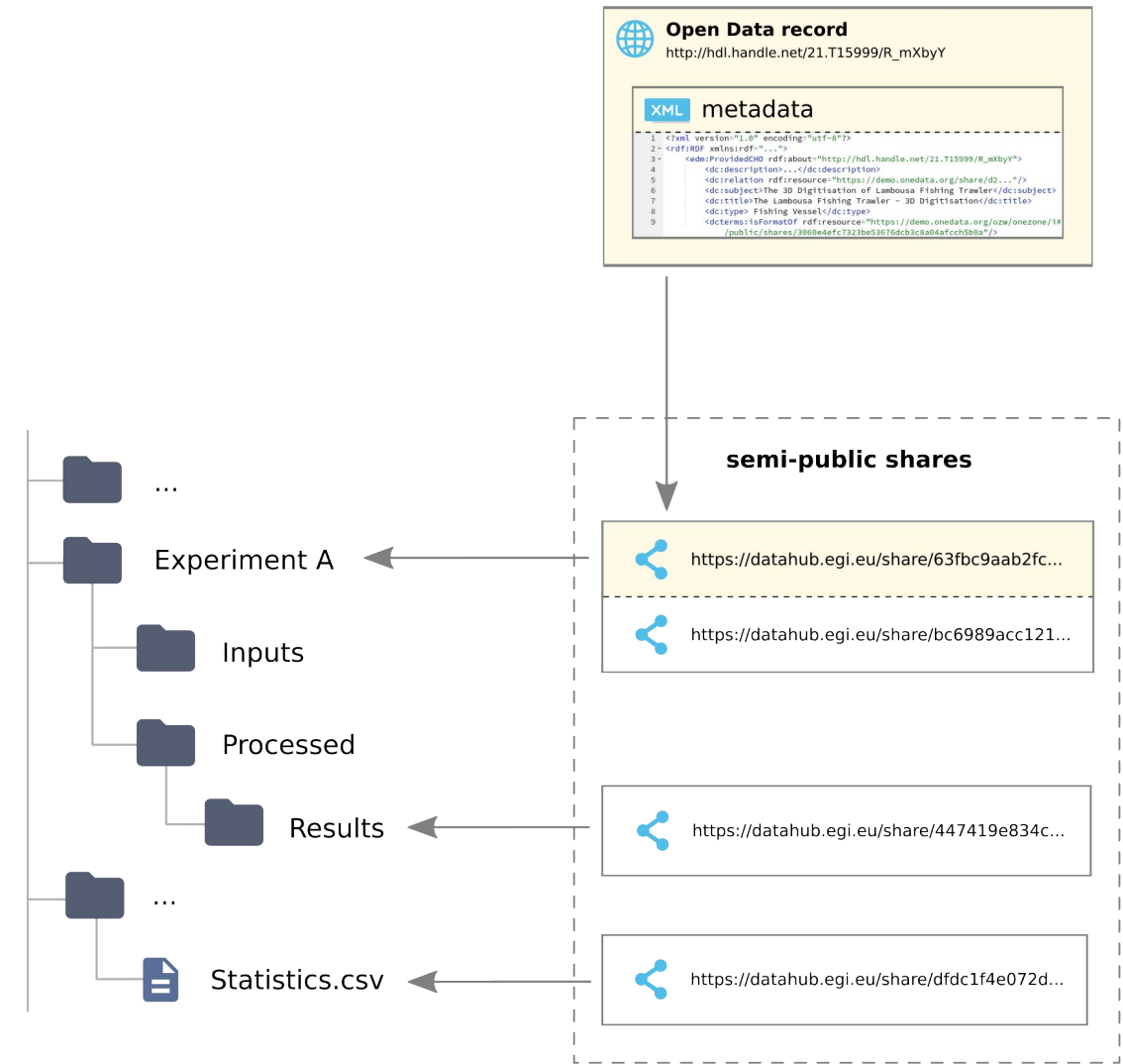
DATA PUBLISHING - SHARES AND OPEN DATA

- **Share — a semi-public link**
 - Allows unauthenticated, read-only access
 - The link is impossible to guess
 - Description in Markdown format



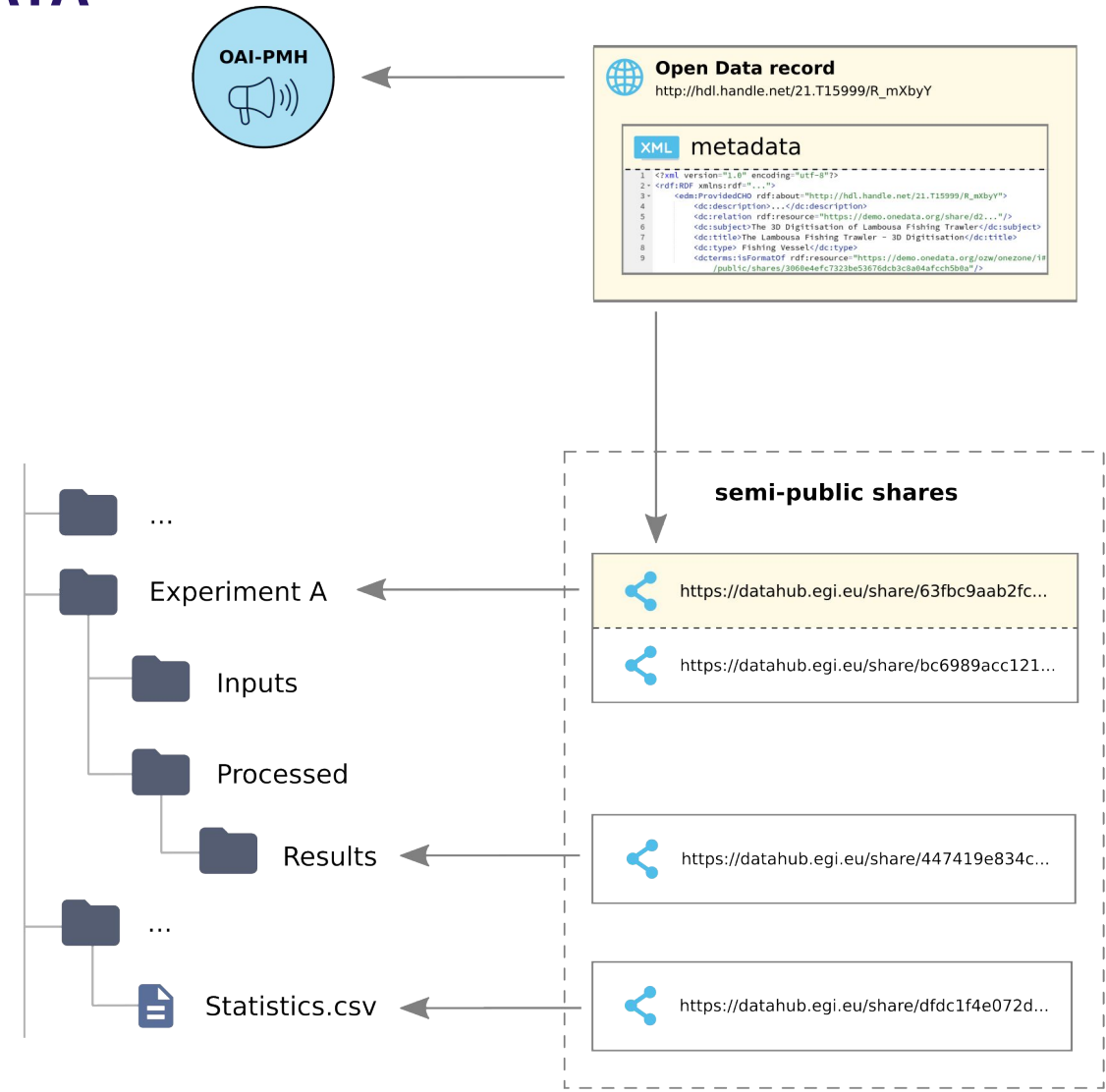
DATA PUBLISHING - SHARES AND OPEN DATA

- **Share — a semi-public link**
 - Allows unauthenticated, read-only access
 - The link is impossible to guess
 - Description in Markdown format
- **Open Data record — Share extension**
 - Built-in PID/DOI minting
 - (extensible plugins for different providers)
 - Public and findable link
 - Standardized XML metadata (DC, EDM, ...)
 - The underlying Share is retained



DATA PUBLISHING - SHARES AND OPEN DATA

- **Share — a semi-public link**
 - Allows unauthenticated, read-only access
 - The link is impossible to guess
 - Description in Markdown format
- **Open Data record — Share extension**
 - Built-in PID/DOI minting
 - (extensible plugins for different providers)
 - Public and findable link
 - Standardized XML metadata (DC, EDM, ...)
 - The underlying Share is retained
 - **Exposition for discovery via OAI-PMH**



THANK YOU

Visit our website: <https://onedata.org>

EOSC Symposium demo: <https://www.youtube.com/watch?v=826burTjOEU>

EOSC Symposium demo dataset: <http://hdl.handle.net/21.15131/pLrlicyt>