



EGI Data Transfer Activities

Andrea Manzi, *Data Solutions Manager at EGI Foundation*

- **EGI intro and EGI Data Transfer**
- **FTS and OIDC in EGI**
- **EOSC Data Transfer**
- **interTwin project and its Data Lake**



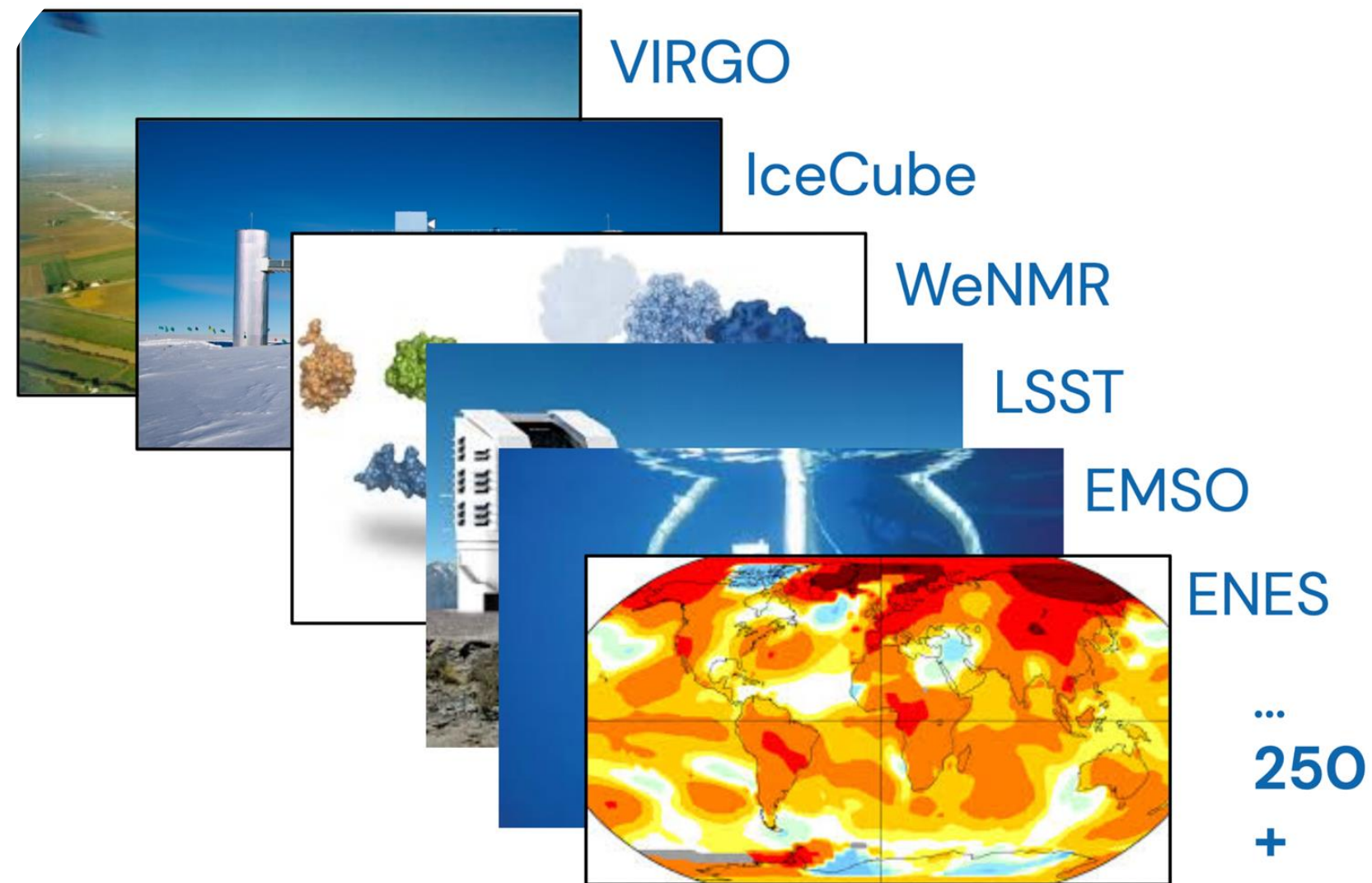
EGI intro and EGI Data Transfer

The EGI e-Infrastructure is an international Federation



2010

From the high-energy physics compute grid (WLCG)

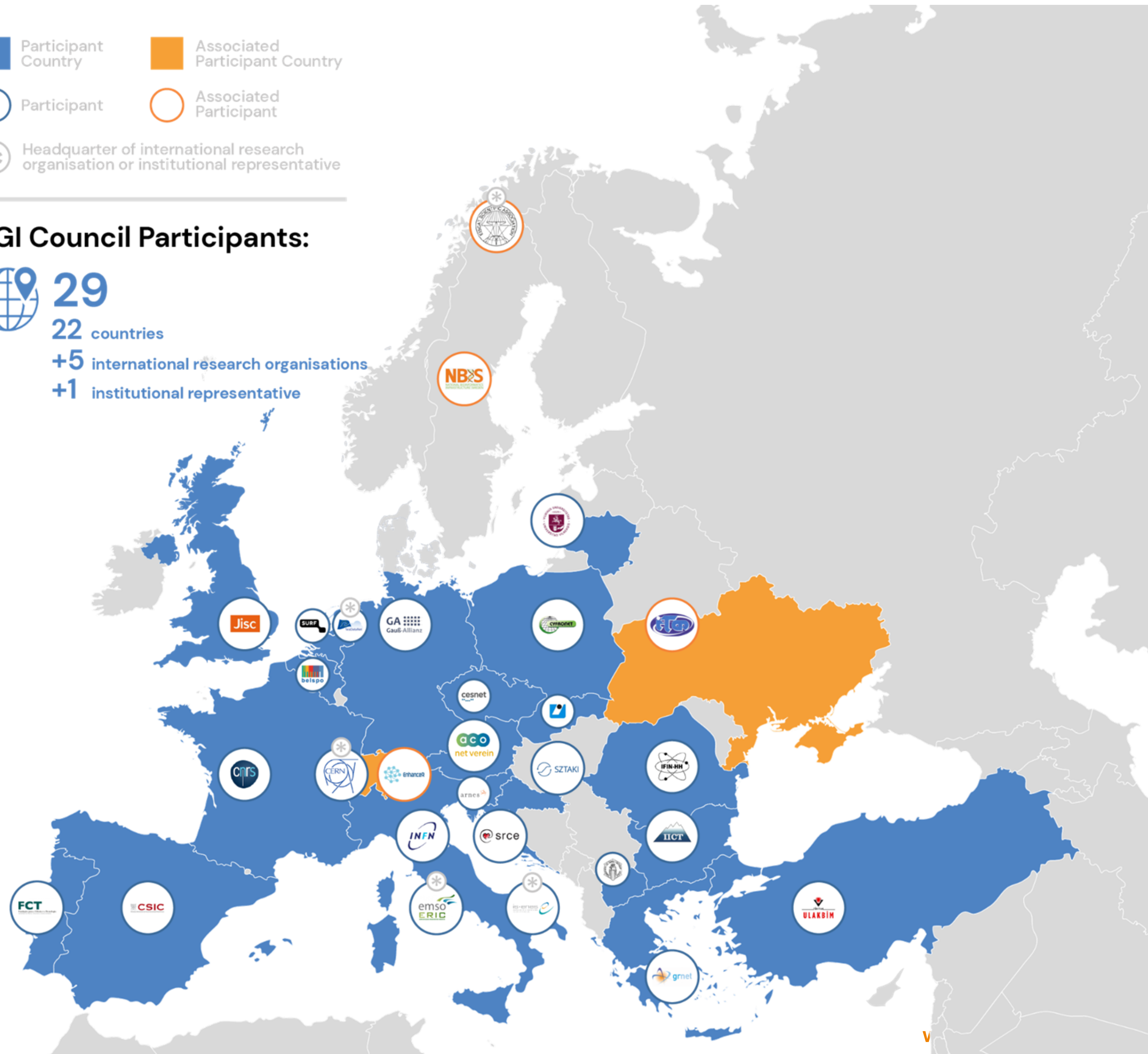


2023

To a multi-disciplinary, multi-technology infrastructure



EGI Council Participants:



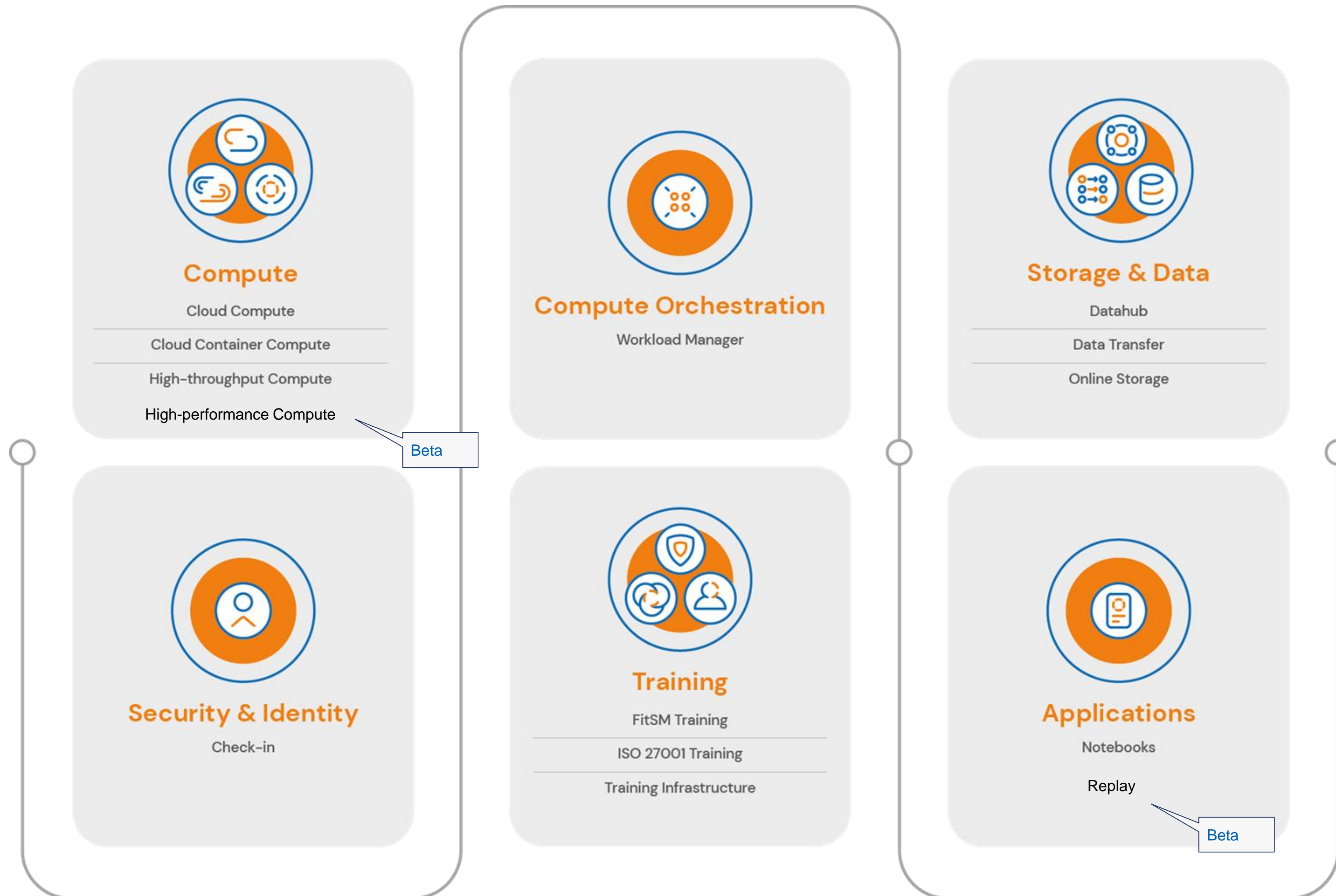
Mission of the EGI Federation

Deliver open solutions
for advanced computing and data
analytics in research and innovation

Mission of the EGI Foundation

Enable the EGI Federation
to serve international research and
innovation together

EGI Services for Research



Service catalogue:
<https://www.egi.eu/services/>

User documentation:
<https://docs.egi.eu>

EGI Data Transfer



- The EGI Data Transfer, based on the FTS
 - <https://docs.egi.eu/users/data-transfer/>
- Available in EOSC marketplace
 - <https://marketplace.eosc-portal.eu/services/egi-data-transfer>
- 2 FTS installations supporting EGI communities
 - UKRI-STFC
 - 3 VOs outside WLCG
 - CERN (Public instance) + WebFTS (X509 based)
 - Piloting AAI integration+HTTP TPC and new EC Projects (see later)



FTS and OIDC in EGI

- **Federated Identity Management service that makes it easy to secure access to federated services and resources**
- **Allow users to have single sign-on to services through Home Organisation login (eduGAIN), or social media (google, ORCID, Facebook etc)**
- **Support standard and open technology (SAML, OpenID Connect/OAuth 2.0, X.509)**
- **Authentication & Authorization**
- **Aggregation and harmonisation of authorisation information (Virtual Organisations/groups, roles, assurance) from multiple sources**

FTS and EGI-Check-in integration

- FTS Public at CERN is configured with all the instances of EGI Check-in
 - Dev/Demo/Production
- N.B. EGI Check-in supports AARC token profile vs WLCG token profile
- FTS @UKRI under configuration (talk by Rose yesterday)
- Piloting transfers via EGI Check-in
 - dCache (already supporting EGI Check-in)
 - HTTP TPC
 - S3
 - S3 <-> dCache HTTP TPC or * <-> S3 streamed



- **OIDC-only WebFTS maintained in our github**
<https://github.com/EGI-Federation/webfts>

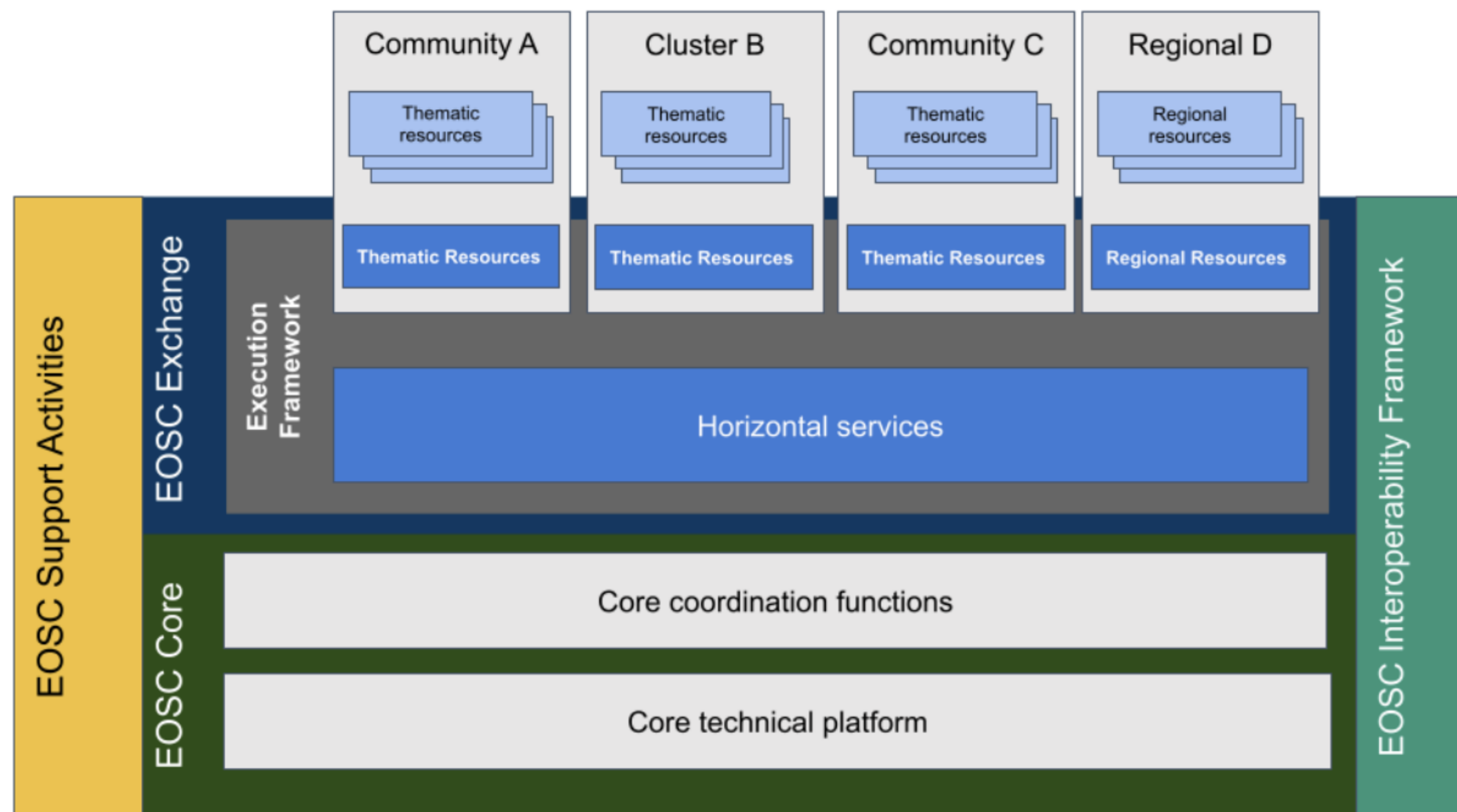
- Support for HIFIS Infrastructure in Germany, commissioning their Data transfer service
 - <https://www.hifis.net/doc/core-services/fts-endpoint/>
 - Based on:
 - Helmholtz AAI based on Unity
 - WebFTS @EGI
 - CERN FTS Public instance
 - Storage endpoints
 - dCache
 - Lightweight Apache endpoint (Passive HTTP endpoint)
 - <https://codebase.helmholtz.cloud/hifis/backbone/transfer-service/apache-fts-endpoint>



EOSC Data Transfer

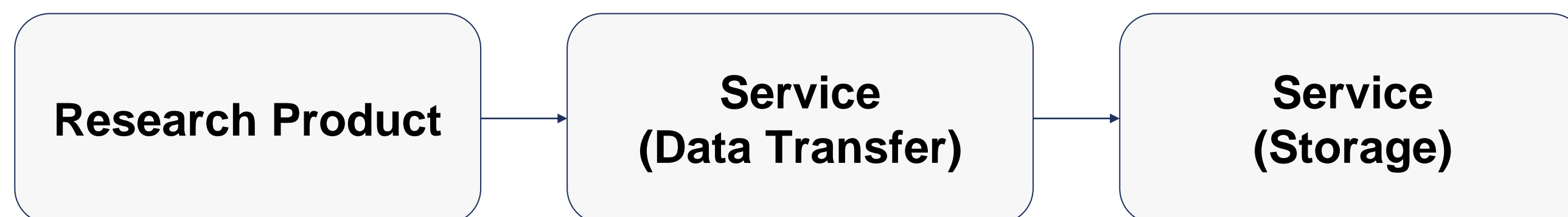
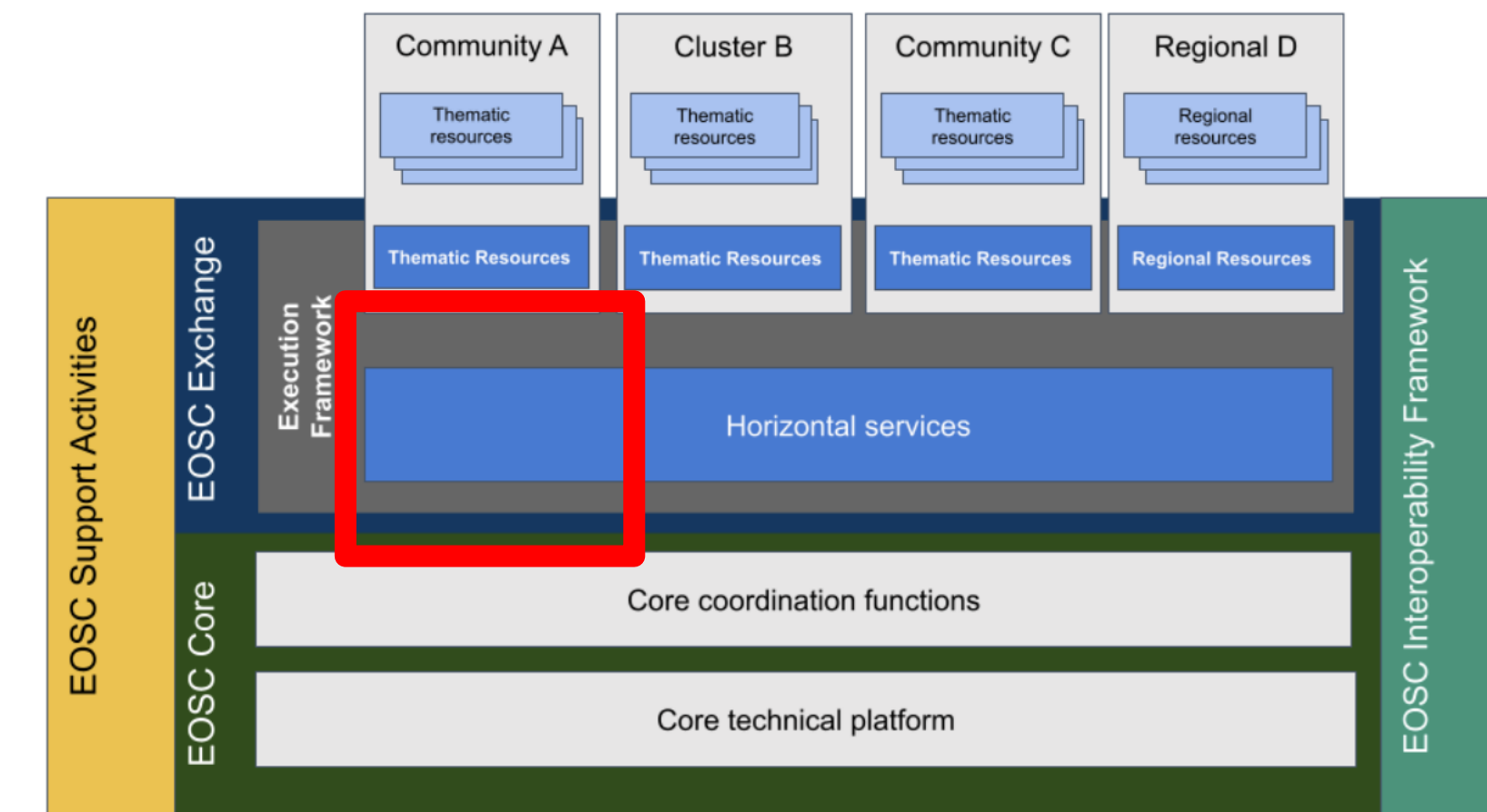
EOSC and EOSC Future

- EOSC Future is an EU-funded H2020 project that is implementing the European Open Science Cloud (EOSC). EOSC will give European researchers access to a wide web of FAIR data and related services



EOSC Data Transfer as EOSC Horizontal service

- Enable transfer of a Research Product (such as a dataset) from its location at the Data Repository available via DOI, to a storage Resource accessible by the User
- Use service composability by defining guidelines in the EOSC Interoperability framework to be implemented by a Data transfer service



- **Data Transfer Service Interoperability Guidelines and Proxy in development**
<https://github.com/EGI-Federation/eosc-future-data-transfer>
- **Currently supported sources, transfer systems, and protocols:**

Source types	Transfer systems	Storage protocols
<ul style="list-style-type: none">• Zenodo records• EUDAT B2SHARE records• Any URLs that resolve to Zenodo/ B2SHARE records• Signposting URLs <link>	<ul style="list-style-type: none">• EGI Data Transfer Service (FTS Public@CERN)	<ul style="list-style-type: none">• WebDAV• S3

EOSC Data Transfer Proxy and GUI

- Proxy Developed by EGI Foundation while the GUI integrated with the EOSC Explore is developed by OpenAIRE
- Proxy is composed of
 - Abstract interface for Data transfer service
 - o any data transfer service could be mapped
 - Implementation of the interface for FTS
 - o Based on Quarkus Java framework
 - o Invoking FTS REST API using EGI Check-in
 - Automatic configuration of S3 storage via Admin API
(X509 certificate used ATM, as this is not available via OIDC)



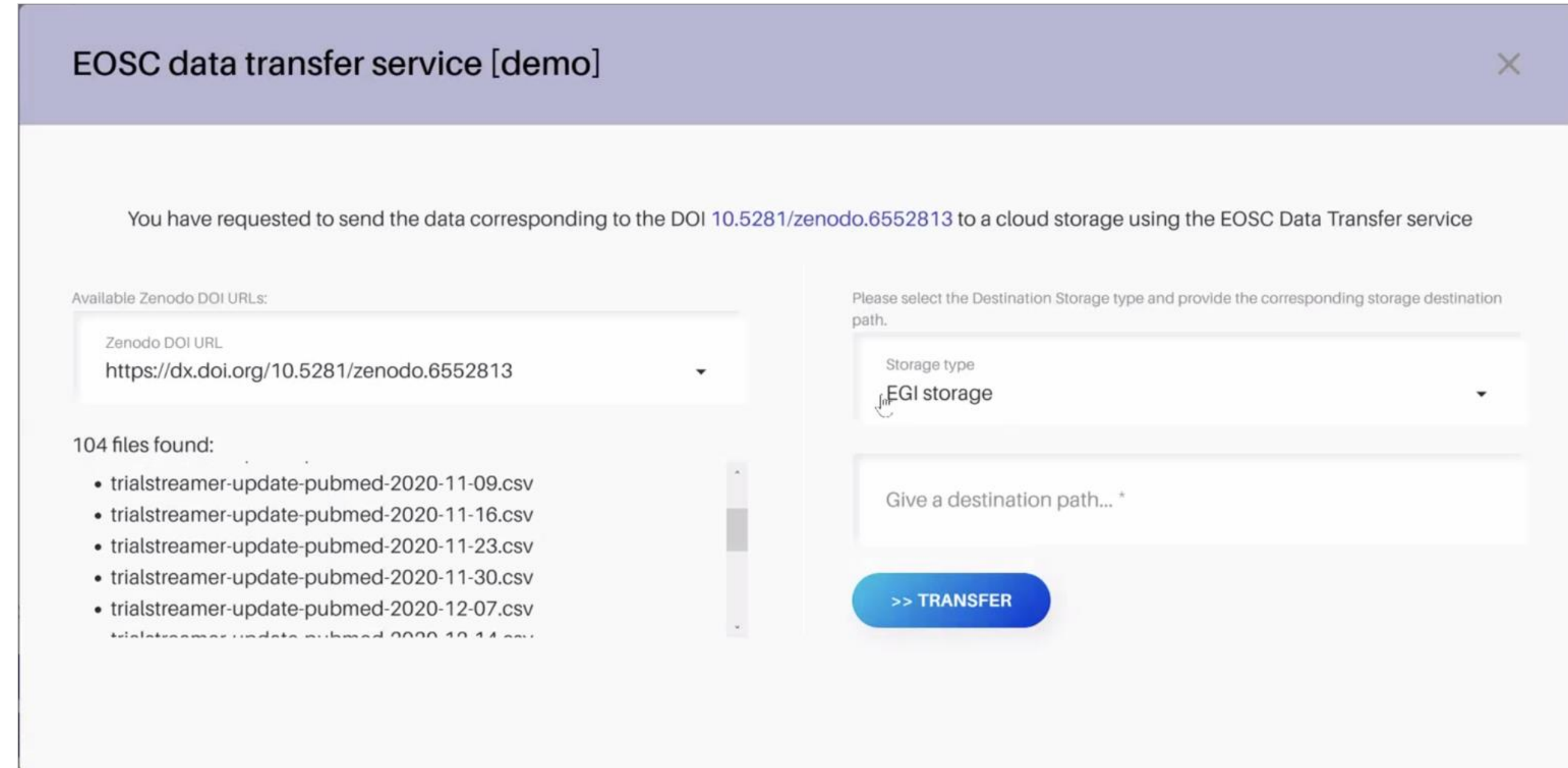
025166931789a0f57793a6092726c2ad89387a4cc167e7c63c5d85fc91021d18@egi.eu

ACMOVA5

AtFBbds7f

EOSC Data Transfer Step by Step

- Use EOSC EXPLORE to find a dataset (225M entries)
 - Click on Data Transfer Icon
 - DOI is parsed to retrieve all versions of the dataset
 - Display list of files for selected version



The screenshot shows a web interface titled "EOSC data transfer service [demo]". It contains a message: "You have requested to send the data corresponding to the DOI 10.5281/zenodo.6552813 to a cloud storage using the EOSC Data Transfer service". Below this, there are two main sections. The left section, titled "Available Zenodo DOI URLs:", contains a dropdown menu with the selected URL "https://dx.doi.org/10.5281/zenodo.6552813". Below the dropdown, it says "104 files found:" and lists several CSV files, including "trialstreamer-update-pubmed-2020-11-09.csv", "trialstreamer-update-pubmed-2020-11-16.csv", "trialstreamer-update-pubmed-2020-11-23.csv", "trialstreamer-update-pubmed-2020-11-30.csv", and "trialstreamer-update-pubmed-2020-12-07.csv". The right section, titled "Please select the Destination Storage type and provide the corresponding storage destination path.", contains a dropdown menu for "Storage type" with "EGI storage" selected. Below this is a text input field labeled "Give a destination path... *". At the bottom right of the form is a blue button labeled ">> TRANSFER".

- Specify destination storage service
 - Browse the destination to pick destination path or enter it manually
- Start Transfer
 - Dashboard to monitor transfer progress and success (under development)

EOSC Data Transfer Next steps

- By end of April we will deploy a new version, available from the EOSC Portal that supports transfers to S3 storages
- Production version to be released in September at the end of the EOSC Future project.
 - Possibly integrated with EOSC AAI Federation
- EGI and CERN partners in a new project proposal (EOSC Beyond) to continue the support and extend the functionalities
 - Support for more storage types and protocols
 - Supporting for different Auth methods at source and destination
 - Better integration with EGI Check-in



interTwin project

General Information

Duration

36 months

Period

1.09.22 -
31.08.25

interTwin overall objective

Co-design and implement the prototype of an interdisciplinary Digital Twin Engine.

Digital Twin Engine

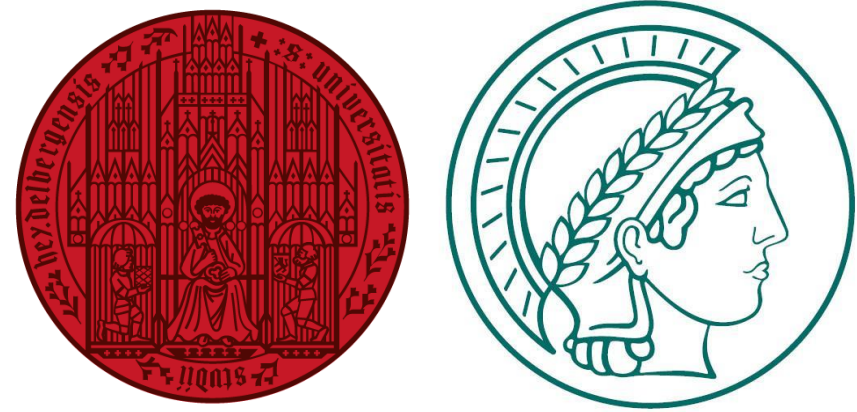
Open-source platform based on open standards offering the capability to integrate with application-specific Digital Twins.

Digital Twin

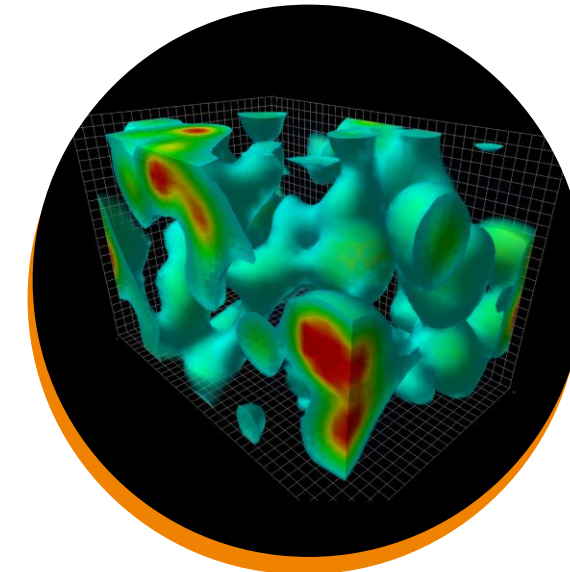
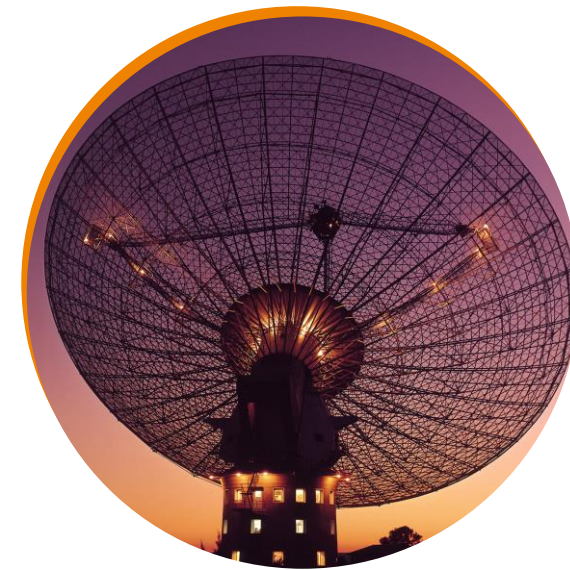
A digital twin is a virtual representation of an object or system updated from real-time data, and uses simulation, machine learning and reasoning to help decision-making.



Physics domain DTs

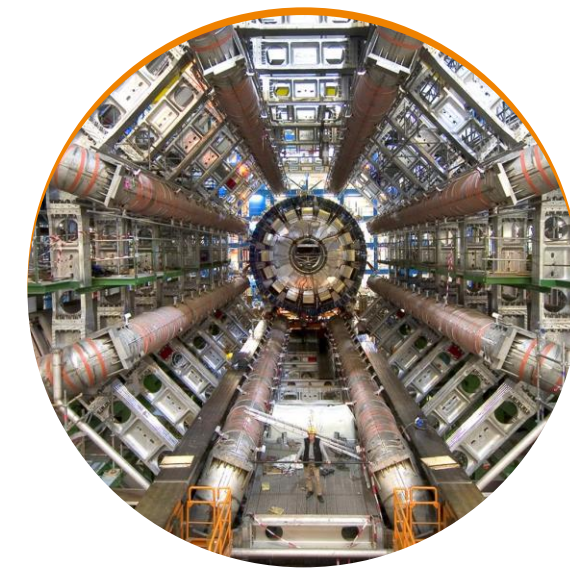
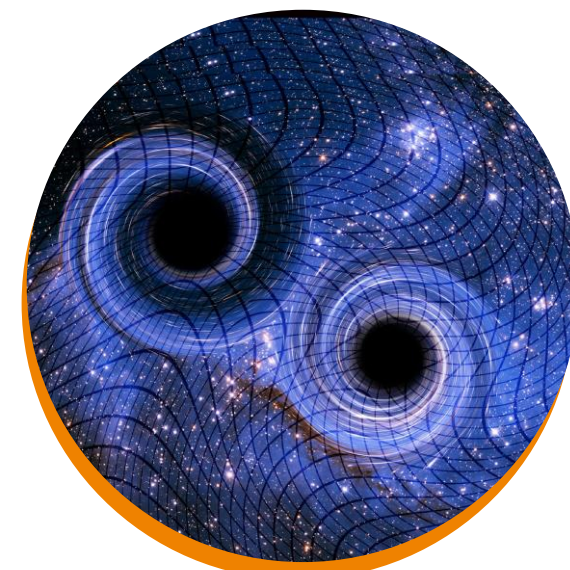


Radio Astronomy

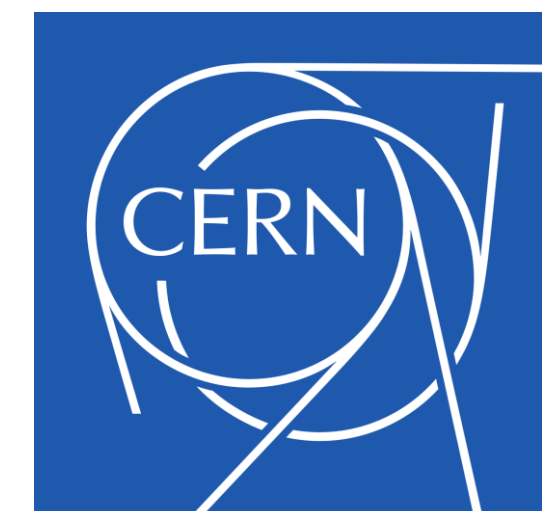


ETH zürich

Quantum Field Theory



High Energy Physics



Gravitational Wave Astronomy

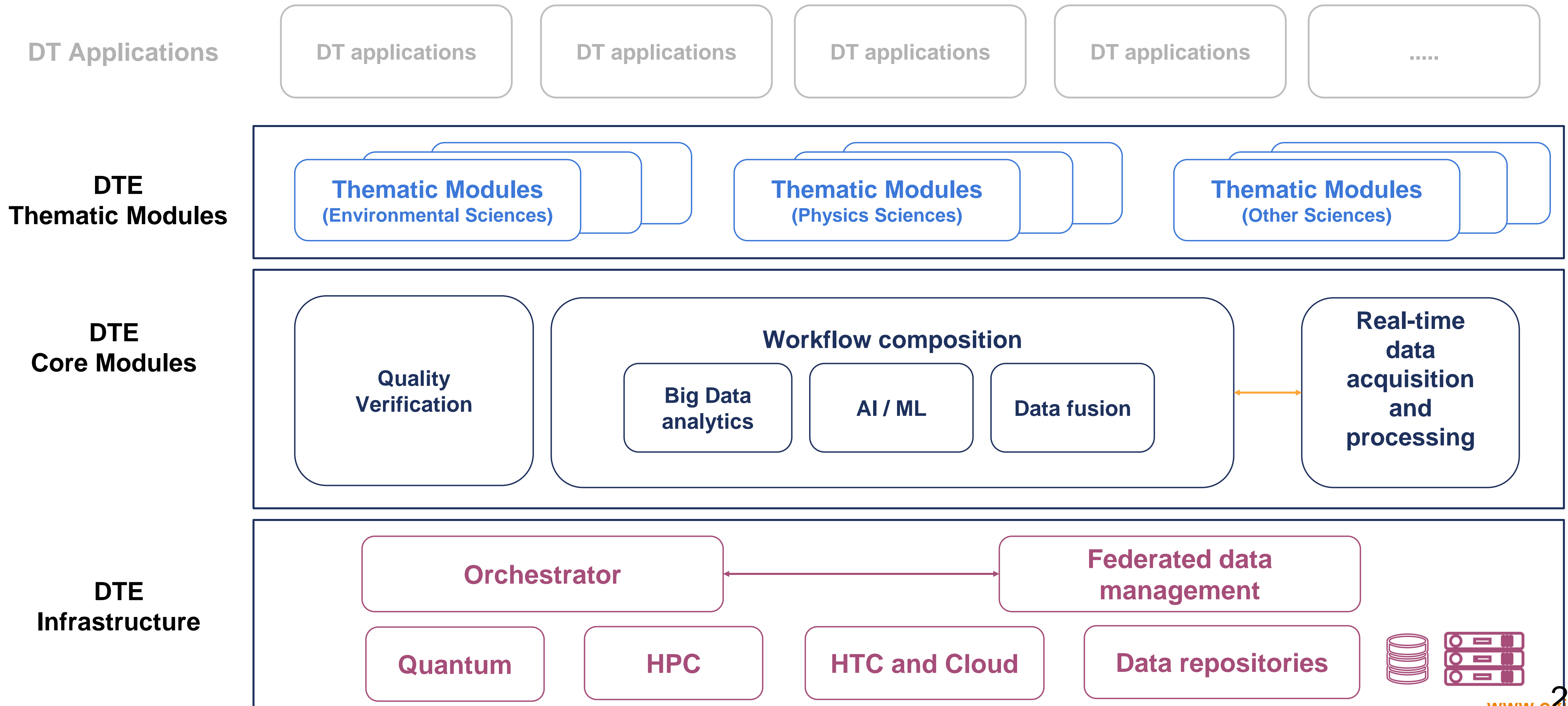


Climate change predictions, impact and early warning for extremes events DTs

DT	Geographical region of Interest
Tropical Storms change in response to climate change	Indian and Pacific Ocean
Wildfires risk assessment in response to climate change	Europe
Flood Early Warning in coastal and inland regions	Selected European regions, Philippines
Alpine droughts early warning	European Alps
Extreme Rainfall events change in response to climate Change	Europe
Flood Climate impact in coastal and inland regions	Selected European regions, Mozambique



interTwin components

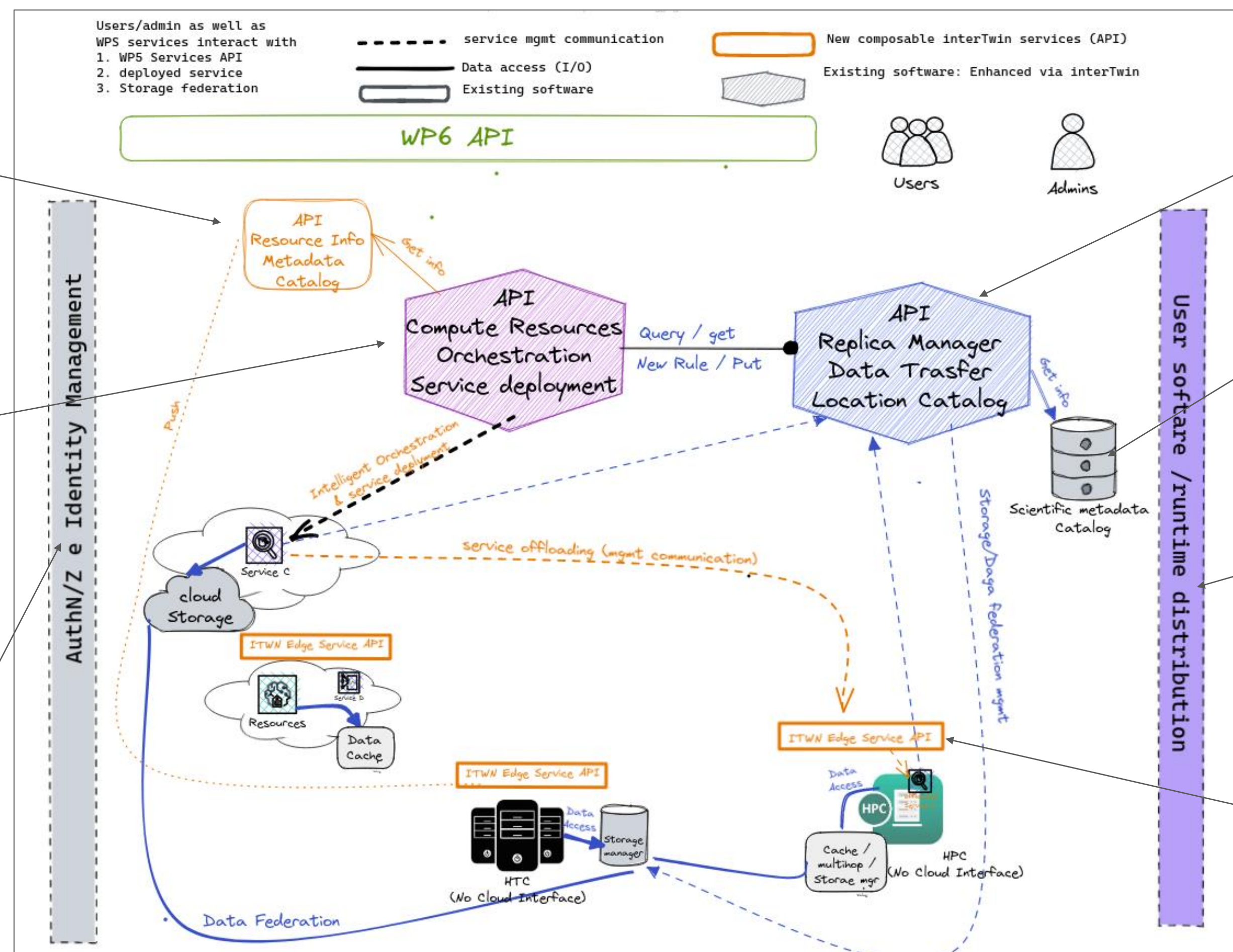


DTE Infrastructure

Work In Progress

Indigo PaaS Orchestrator

CheckIN and
Any other oidc authz
provider (INDIGO-IAM)



Rucio + FTS

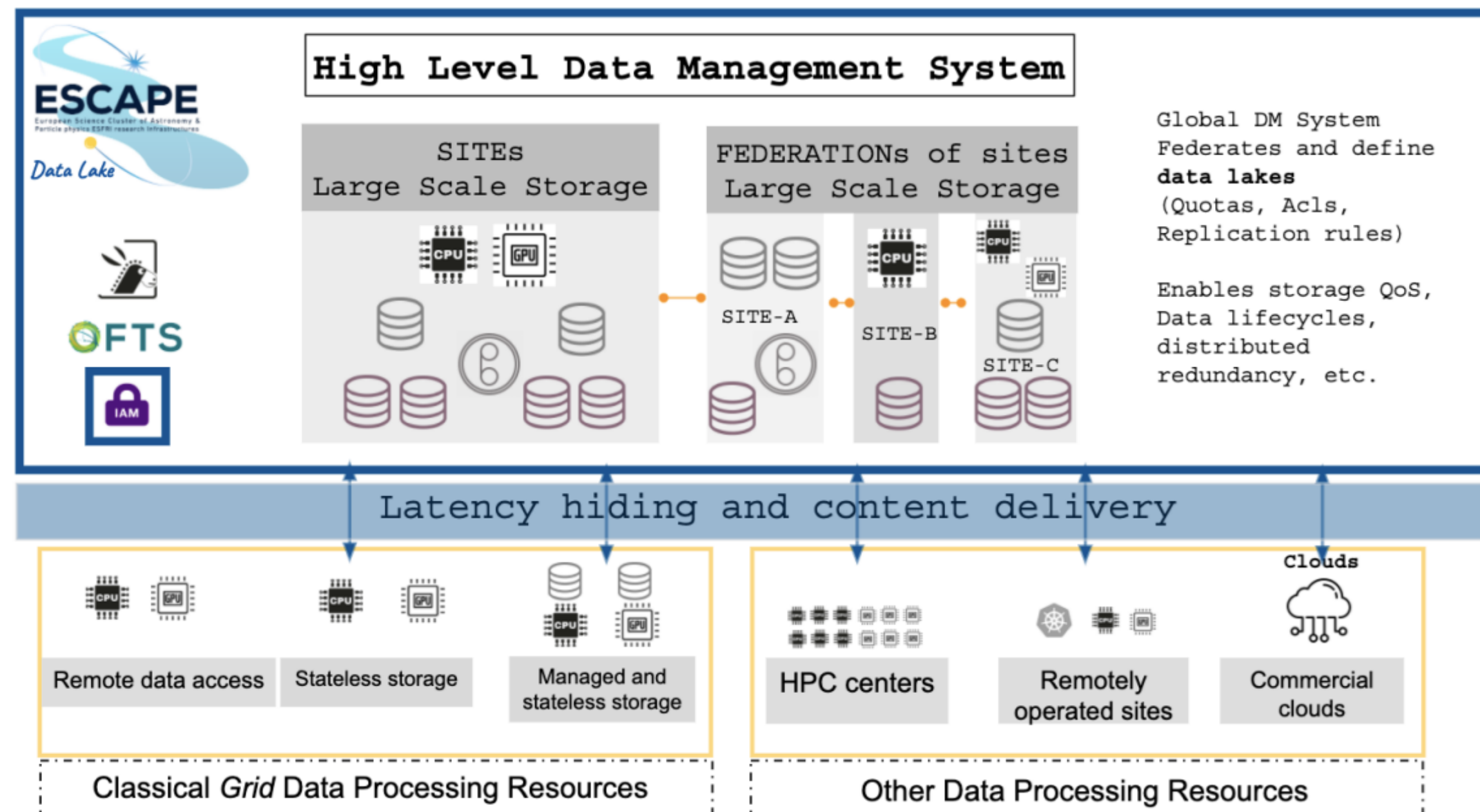
Domain
Specific

CVMFS

To be developed

interTwin Federated Data Management

interTwin Data management based on the ESCAPE H2020 project Data Lake blueprint



- Main challenges
 - Hybrid infra
 - HPC (also EuroHPC Vega)
 - HTC
 - Cloud
 - Requirements from a new domain (Environmental science)
- Started setting up the first testbeds with central components
 - Rucio at DESY
 - FTS Public at CERN
 - EGI Check-in Demo
 - dCache storages

- EGI is expanding the activities related to Data Transfer using FTS thanks to new projects and demand from communities
- The EOSC Data Transfer capability has seen already quite interest and could potentially become one of the key horizontal services in EOSC
- In particular with the interTwin project, EGI plans to kickstart more widespread use of distributed hybrid workflows and make cooperation between different scientific fields and communities easier in the future



Contact us

Thank you!

Get in touch with us

www.egi.eu



This work is partially funded by the EU research and innovation programme