EGI collaboration with IS-ENES

Tiziana Ferrari

Tiziana.Ferrari@egi.eu

18 Jan 2023, IS-ENES GA

EGI Federation

European flagship digital infrastructure for data-intensive scientific computing

77,000 researchers worldwide

Why a federation?

- Support science at international scale
- Build an hyperscale compute facility for research
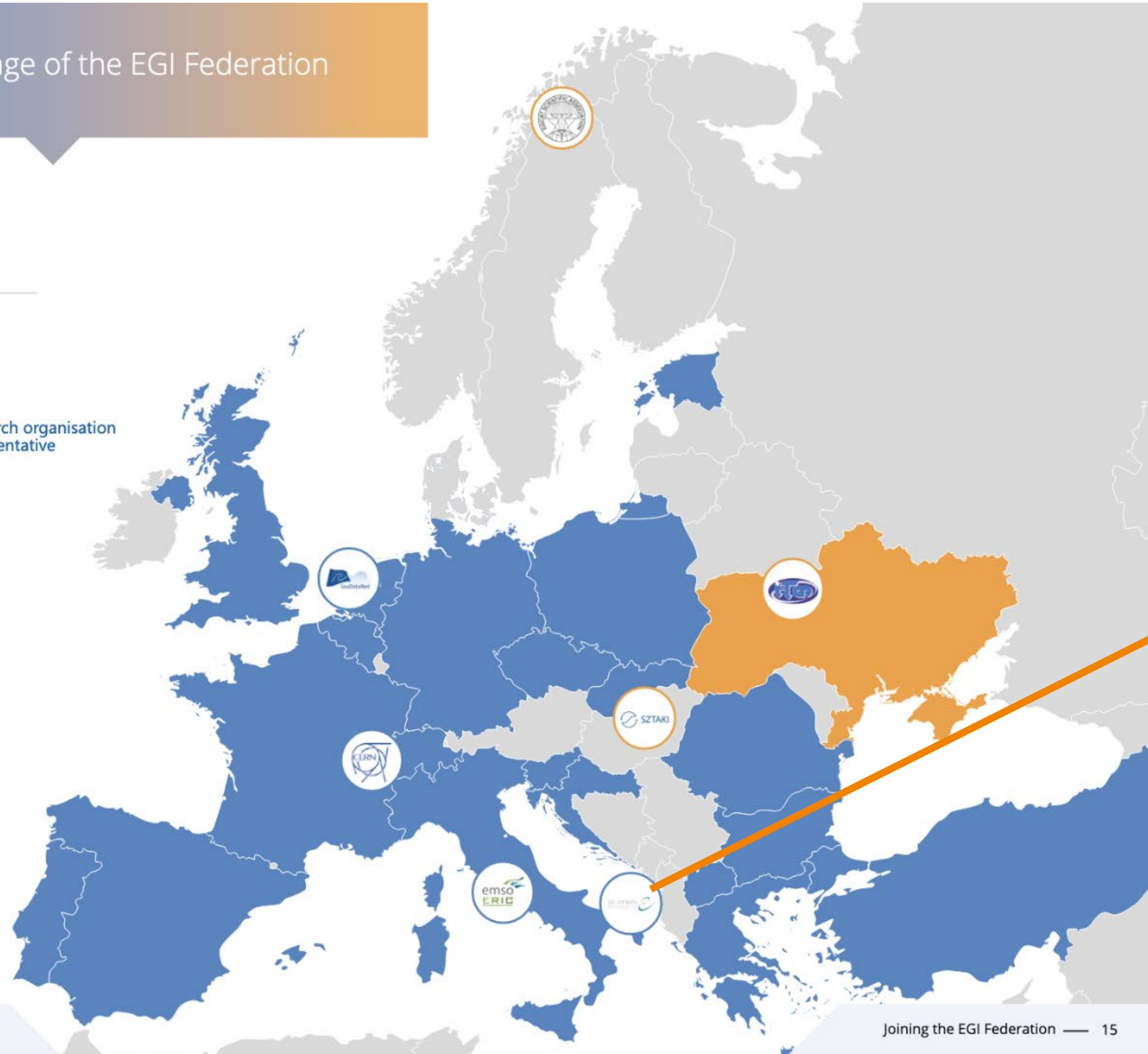- Invest nationally, access globally
- Bring computing to the data

Leaflet | Map data © OpenStreetMap
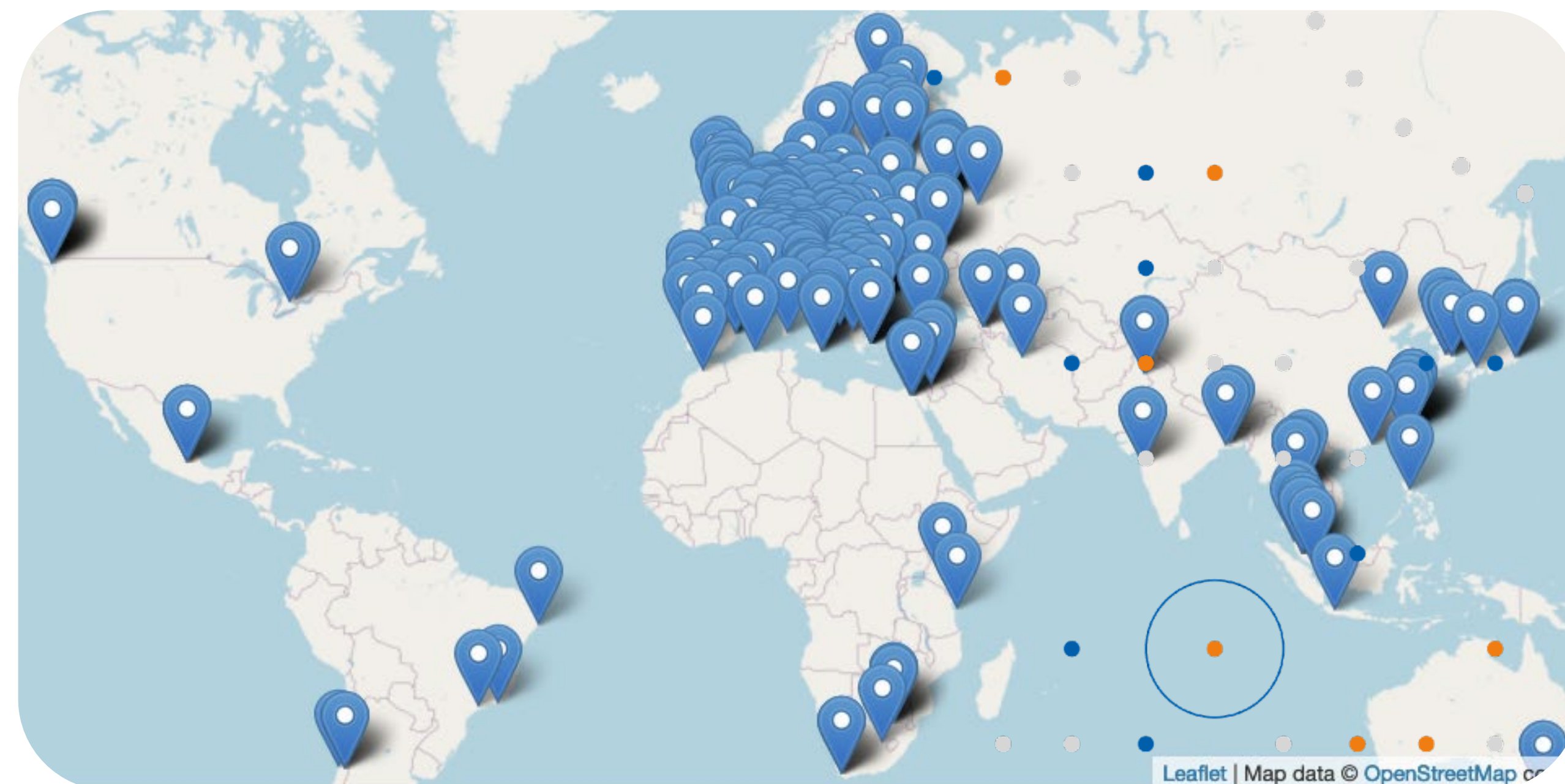
A membership organization
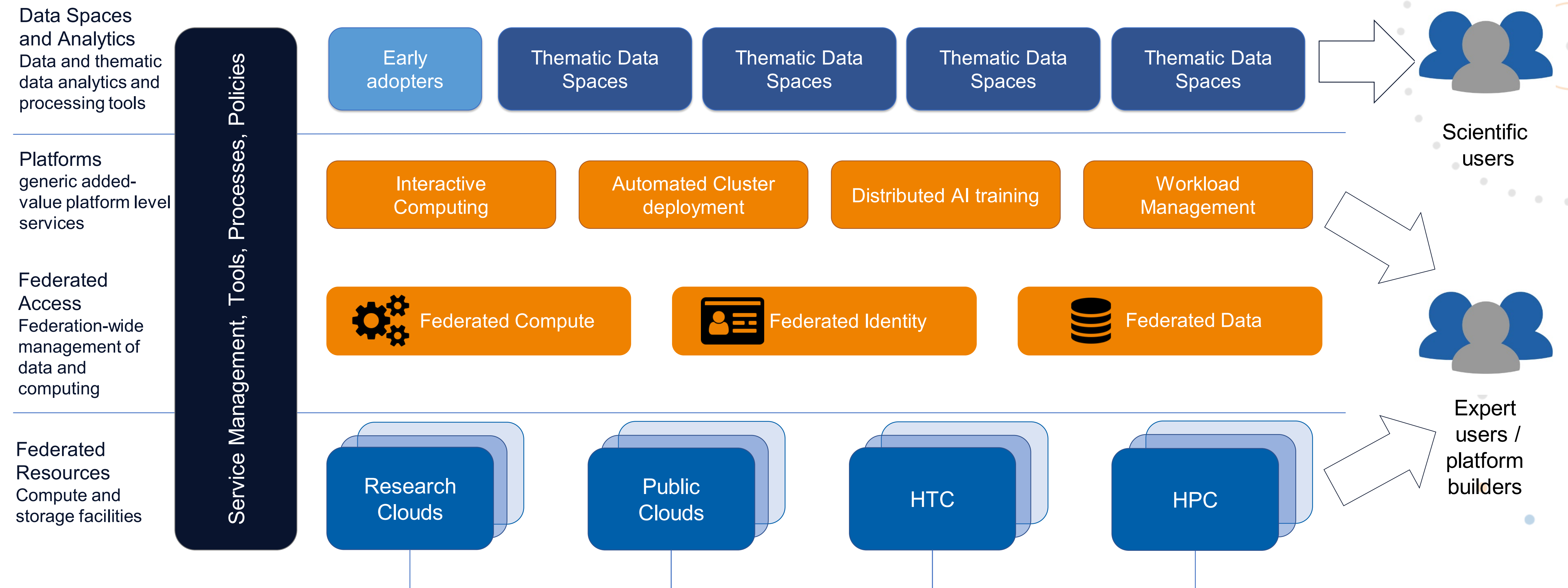

Geographical coverage of the EGI Federation

Participants

Associated participant

EGI Council Participants:

27
21 countries
+ 5 international research organisation
+ 1 institutional representative

IS-ENES is member of the EGI Council since July 2020 through CMCC

Technical Architecture

Data Spaces and Analytics
Data and thematic data analytics and processing tools

Platforms
generic added-value platform level services

Federated Access
Federation-wide management of data and computing

Federated Resources
Compute and storage facilities

Service Management, Tools, Processes, Policies

| Early adopters | Thematic Data Spaces | Thematic Data Spaces | Thematic Data Spaces | Thematic Data Spaces |

Scientific users

| Interactive Computing | Automated Cluster deployment | Distributed AI training | Workload Management |

| Federated Compute | Federated Identity | Federated Data |

Expert users / platform builders

| Research Clouds | Public Clouds | HTC | HPC |

⇒ Foundation of the EOSC Compute Platform - http://go.egi.eu/egi-ace

⇒ Governance, Architectures and Business Models for Data and Cloud Federations: the EOSC and GAIA-X Case Studies - https://zenodo.org/record/5081865

Collaboration

- **R&D** of technical solutions of common interest for federated data access and analysis, data spaces and digital twins
- **Collaboration with other research communities**
- **Access to federating services of the EGI Federation**, e.g. Check-in
- **Projects** for integration and technical support of IS-ENES data analytics tools

EGI Community and IS-ENES collaborative projects

2015-2017 INDIGO - DataCloud
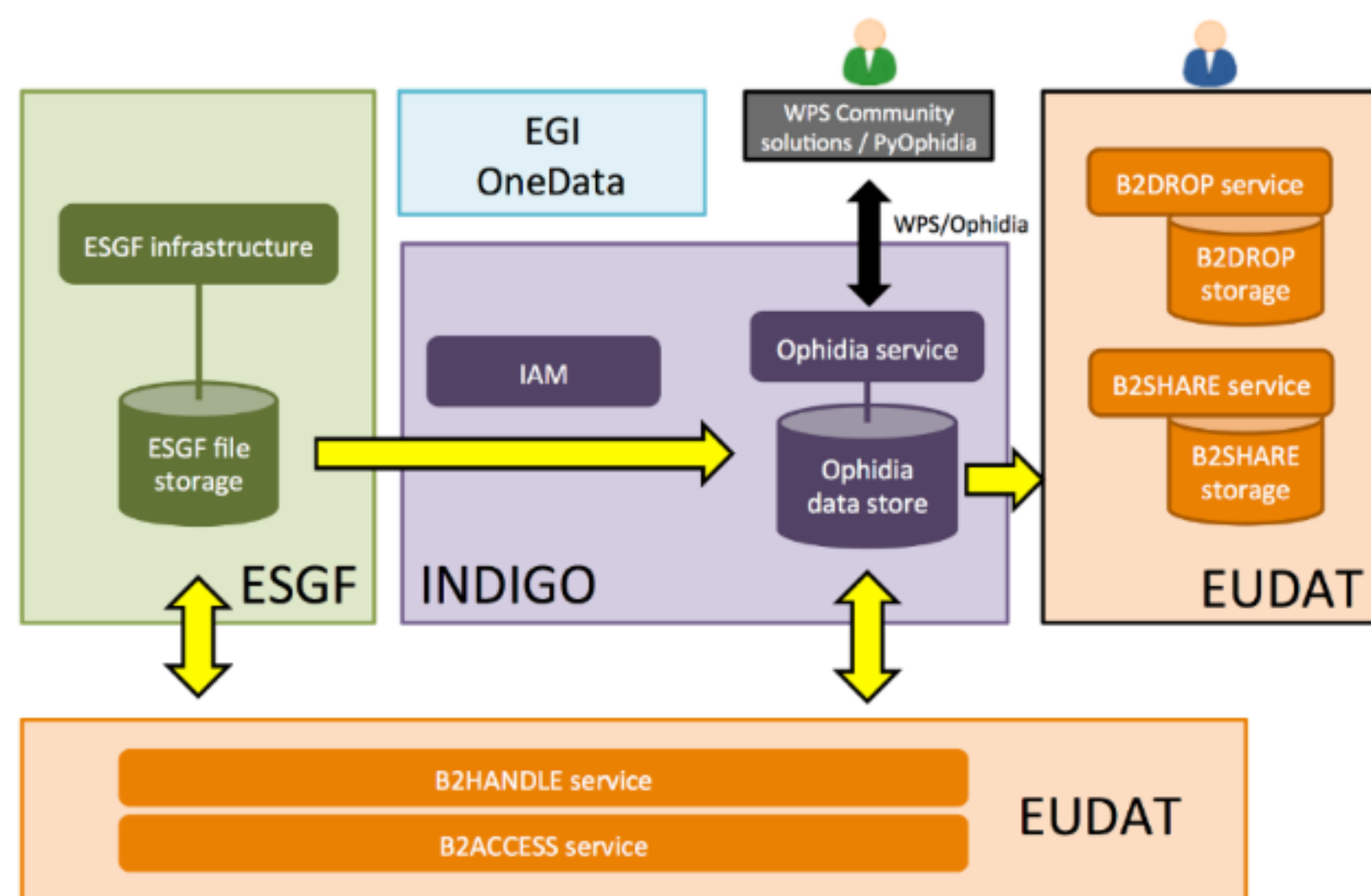Better Software for Better Science

2018-2021 EOSC-hub

2018-2021 EGI-ACE

2022-2025 interTwin

Climate Data Analytics in the cloud: EOSC-Hub and INDIGO-DataCloud

Big Data Challenges, Research, and Technologies in the Earth and Planetary Sciences

A workshop to be held Monday December 5th at the 2016 IEEE International Big Data Conference

A distributed EU/US testbed for analytics workflows

S. Fiore, M. Plóciennik, et al.: Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system. BigData 2016: 2911-2918

EGI reusable open components for the federated data access

Federated Trust and Identity: Check-in

A standards based and interoperable Authentication and Authorization service

Federated Data: DataHub

Replication of scientific data to/from distributed providers with transparent access for users

Federated access to applications

AppDB for community-based VM image sharing, and CVMFS for scalable software distribution across the federation

Towards a European compute continuum with IS-ENES use cases

1 **HPC as a Service**

Virtual infrastructures on HPC facilities

2 **Access via ssh + federated identity**

Enable used of OpenID Connect to access ssh login nodes of HPC systems + integration of accounting and monitoring

3 **HTC middleware/Gateways**

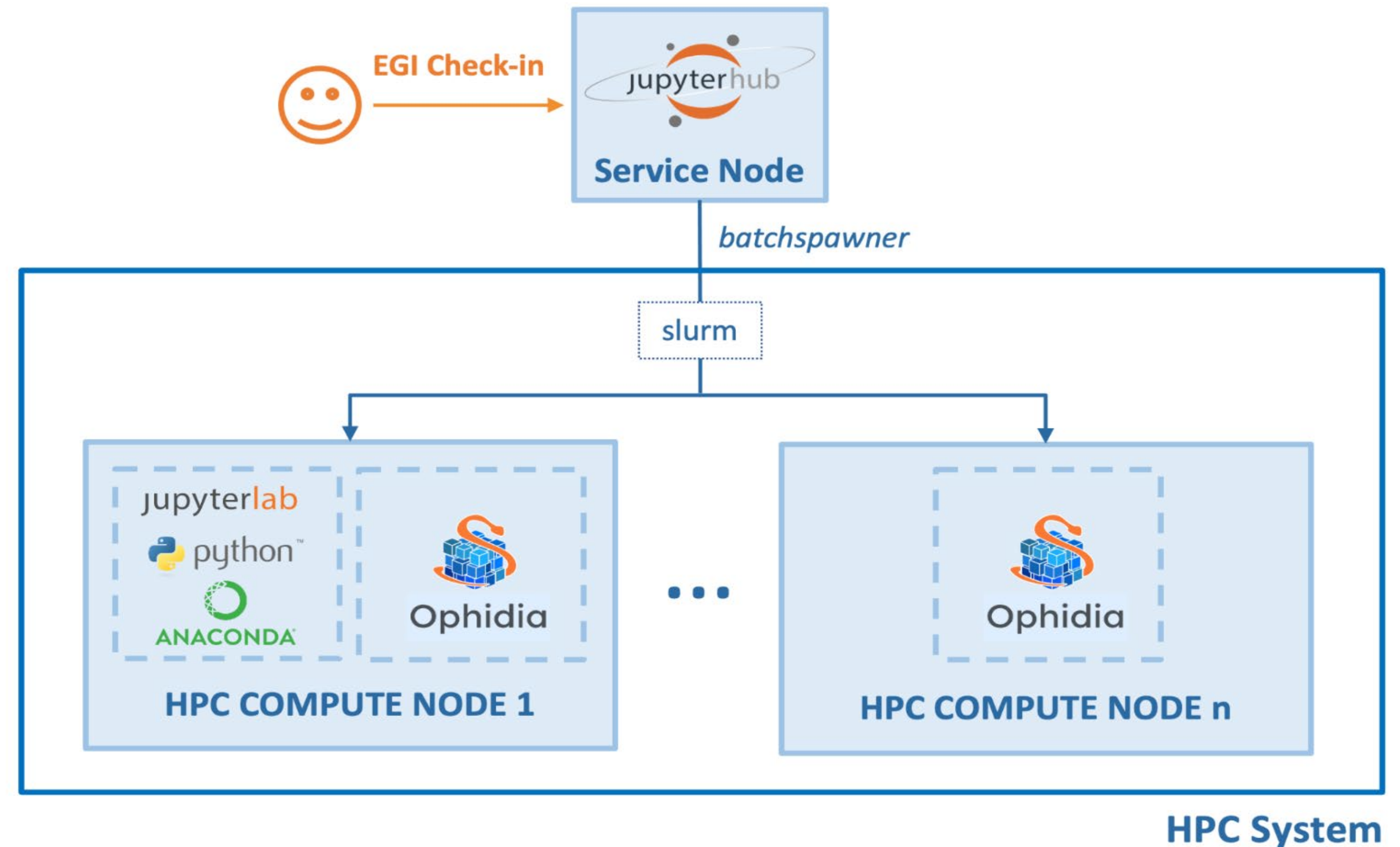Reuse HTC middleware so HPC systems can be integrated into existing research workloads.

ENES Scenario 1
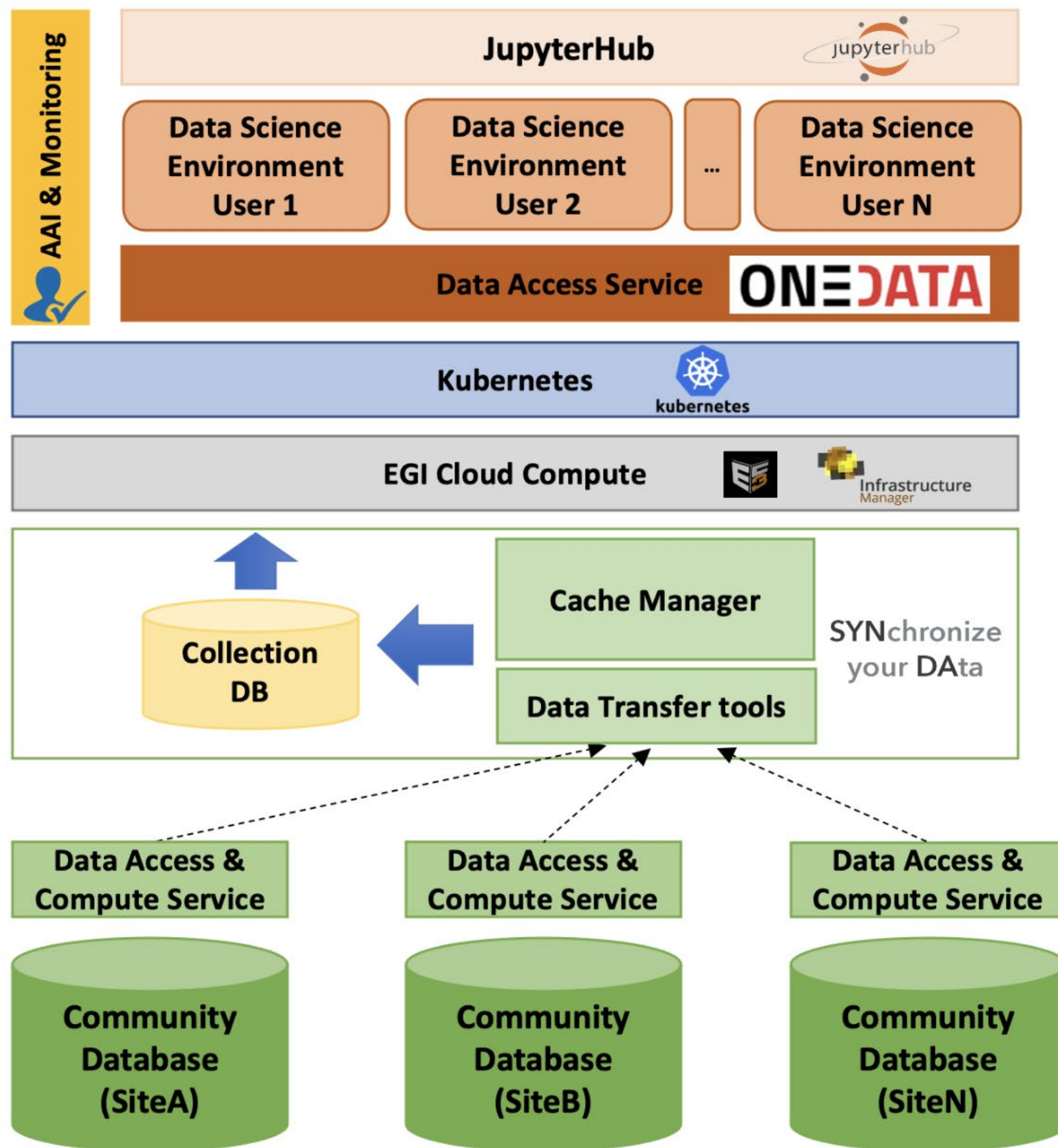
Scenario 1: experienced users

- Access HPC with ssh-oidc
- All software deployed as user, no additional privileges needed
  - Data movement from/to HPC using DataHub, user-side fuse mount of remote datasets
  - Ophidia Big Data Analytics framework deployed as containers with udocker (container runtime)
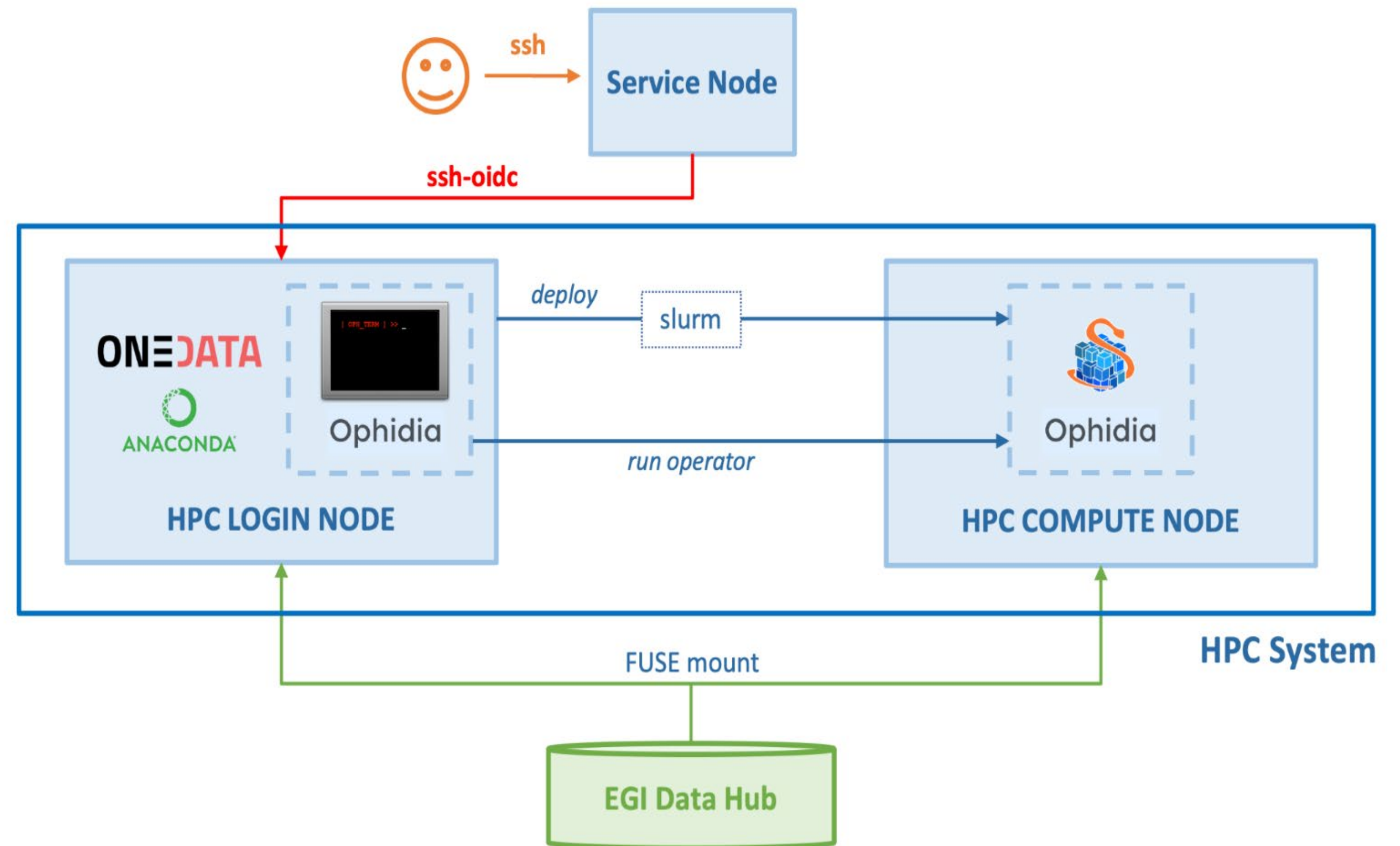
Scenario 2: non-expert users

- No direct interaction with the HPC system for users, JupyterHub as a gateway

- Users login with Check-in, fine-grained authorization based on user group membership

- User-level Jupyter servers running as containers with udocker

ENES Data Space in EOSC

Data science environment for climate data analysis. Single entry point for:

- Datasets
- Storage & Compute resources
- Data Science Software Stack
- IDE & Applications
- SaaS/PaaS approach

A multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, enabling them to better conduct their work

> Builds on existing infrastructures and services supported by the European Commission, Member States and research communities.
> Brings these together in a federated 'system of systems"
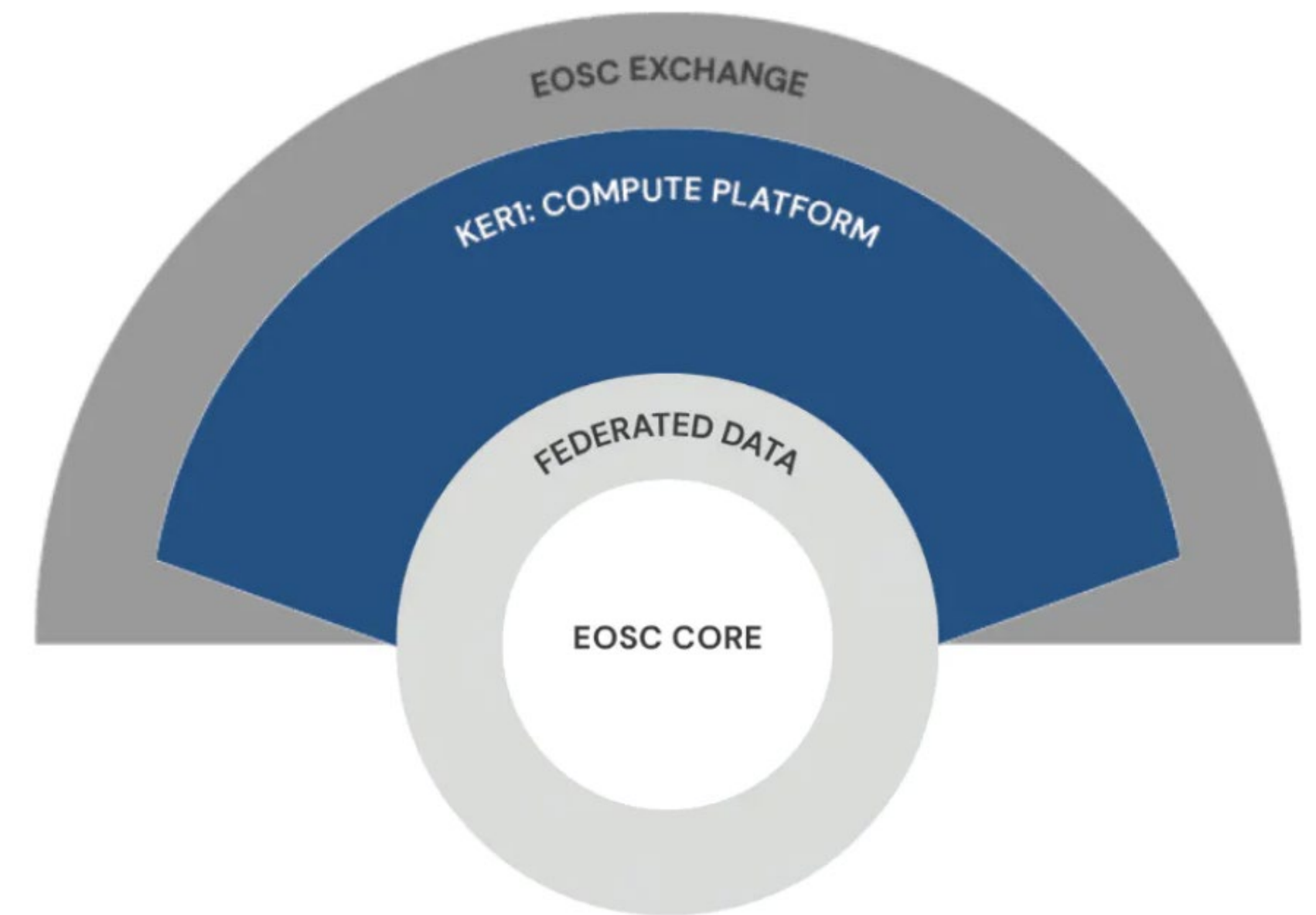
Data Space Building Blocks

- Data Space: "a federated data ecosystem within a certain application domain and based on shared policies and rules", Data space design principles

Source: OpenDEI project
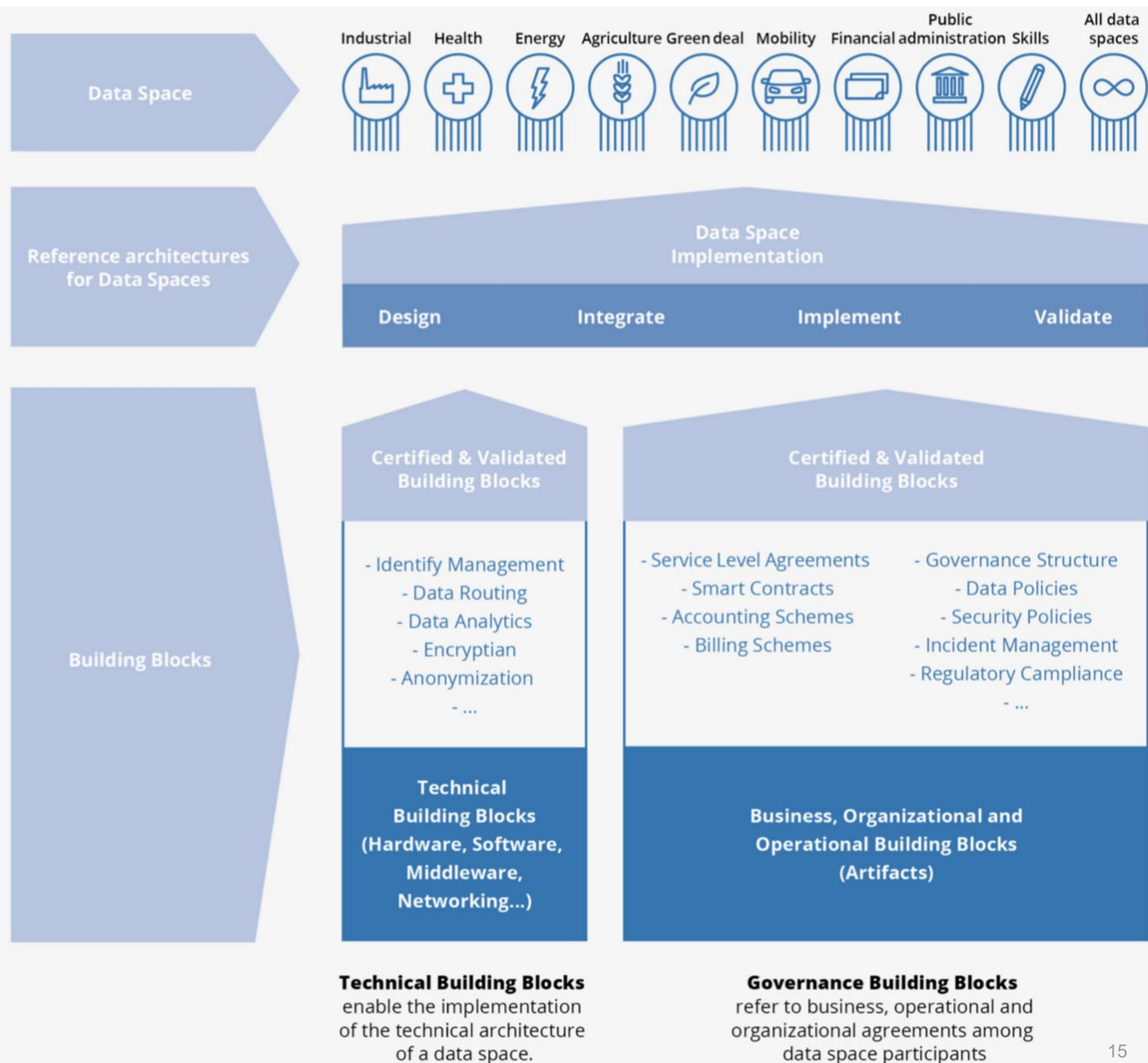
IS-ENES as member of the EGI Federation has been contributing to scientific computing innovation for the benefit of many scientific communities beyond climate modelling