Welcome to EGI 2024!

70+
Scientific Communities and Projects represented

More than
350
Attendees for the 5 days (including side meetings)

50+
Sessions, side meetings and trainings

123
Abstracts

26
Posters

16
Demonstrations

#EGI2024

Towards 2030:
Challenges and Opportunities for Data-Intensive Computing

Tiziana Ferrari

01-10-2024

Director, EGI Foundation

EGI 2024, Lecce

TLP: WHITE Public

Outline

Subtitle

- **On data-intensive computing in Europe**

  - **EGI Federation and its state of the art**

- **The road ahead of us: towards 2030**

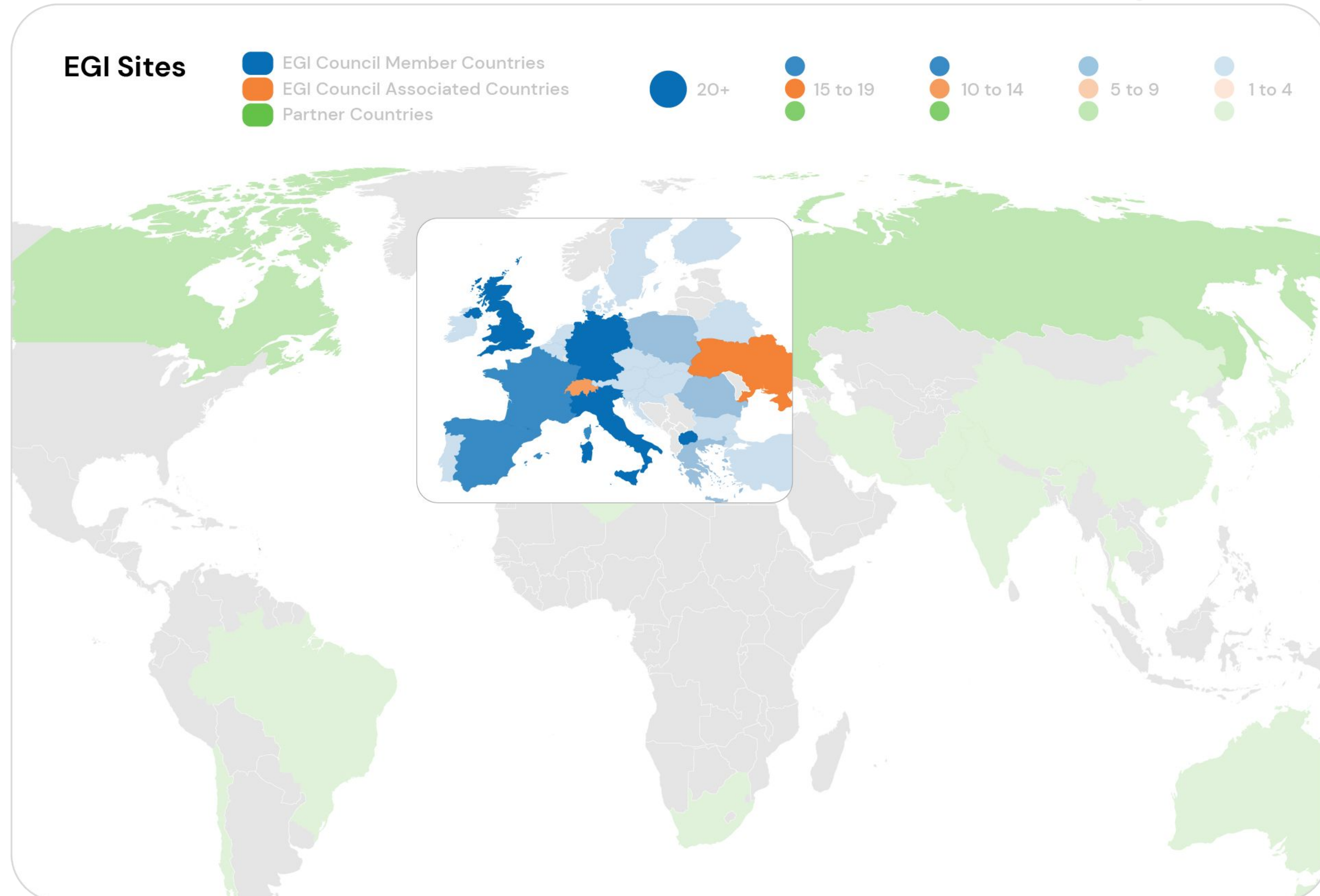- **Role of EGI in the European Open Science Cloud**

Data-Intensive Computing and Open Science

Global Open Research Commons - Essential Elements

GOVERNANCE STRUCTURES

RULES OF PARTICIPATION AND ACCESS

ENGAGEMENT

HUMAN CAPACITY

INTEROPERABILITY & STANDARDS

SUSTAINABILITY

COMPUTE, STORAGE, NETWORK, AAI

RESEARCH OBJECTS

SERVICES & TOOLS

- **Collaboration is the hallmark of modern science**
- **The Open Science Commons embodies this vision**

The Global Open Research Commons International Model Report, Version 1; Payne,K., Corrie,B., Crawley,F., Harrower,N., Macneil,R., Maxwell,L., Sansone,S.–A., Treloar,A., Woodford, C., Åkerström, W.N.

EGI contribution to Open Science 2004–2024

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

EGI Sites

- EGI Council Member Countries
- EGI Council Associated Countries
- Partner Countries

20+ | 15 to 19 | 10 to 14 | 5 to 9 | 1 to 4

EGI in Numbers (2023)

17% growth
82M
Cloud CPU hours consumed

2.8% growth
37
EU funded projects

13% growth
94,200
Users

63% growth
2,875
Enabled articles [1]

-1.4% growth
7.0B
HTC CPU hours consumed

2022 | 2023

+100,000 users in 2024!

Key Numbers

95,000
Total number of users

+10.200
New users in 2023

Top 5 Cloud Communities

WeNMR	NBIS	Biomed	BioISI	ENVRI
41K	21K	1.5K	1K	967

By number of registered users

Top HTC Communities

Atlas, CMS, ALICE, LHCb, Belle II, Virgo

Essential Partners and the Largest Adopters

Research infrastructures (RI) and research communities

13 new scientific communities

23 RIs from the ESFRI roadmap

49 RIs of pan-European scope using our services

1 new RIs engaged in 2023

Physical Sciences & Engineering

Landmarks
CTA, ELI ERIC,
HL-LHC
SKAO
European XFEL

Projects
KM3NeT 2.0

Data Computing & Digital RIs

Projects
SoBigData++
EBRAINS
SLICES

Environment

Landmarks
ACTRIS ERIC
EPOS ERIC,
Euro-Argo ERIC
IAGOS
ICOS ERIC

Projects
DANUBIUS-RI
DiSSCo
eLTER RI

Social & Cultural Innovation

Landmarks
CLARIN ERIC
DARIAH
CESSDA ERIC

Projects
E-RIHS
OPERAS

Health & Food

Landmarks
ELIXIR
INSTRUCT ERIC
BBMRI
EU-OPENSCREEN ERIC

Projects
EMPHASIS
METROFOOD-RI

ESFRI research infrastructures supported by EGI

■ New RIs engaged in 2023
■ EGI Federation member

ENVRI-Hub

The Data Portal of the European Environmental Research Infrastructures (ENVRI service catalogue, science demonstrators – Jupyter Notebooks and integrated services for computing, storage and Trust and Identity Management)

A joint IT integration effort of

- **ESFRI Landmarks (ACTRIS, AnaEE ERIC, EPOS ERIC, EuroArgo ERIC, IAGOS AISBL, ICOS ERIC, LifeWatch ERIC)**
- **ESFRI Projects (eLTER), SeaDataNet and**
- **EGI Federation**

EGI Annual Report 2023

Propelling Research and Innovation
Through Collaboration and Support

The road ahead of us

Towards 2030

1/The crisis of accessibility to computing and storage services

1/

Increasing demand of compute and storage capacity driven by the growing amount of research data produced by national and European observatories

2/

Constant research budgets

Moore's Law may end sometime in the 2020s due to approaching physical limits

3/

No or limited integrated access to HPC, HTC and Cloud critical infrastructure at European level

EGI infrastructure in 2024

HTC Federation

EGI Sites 20+ 15 to 19 10 to 14 5 to 9 1 to 4

Cloud Federation

EGI Sites
- HTC
- Cloud
- HTC / Cloud
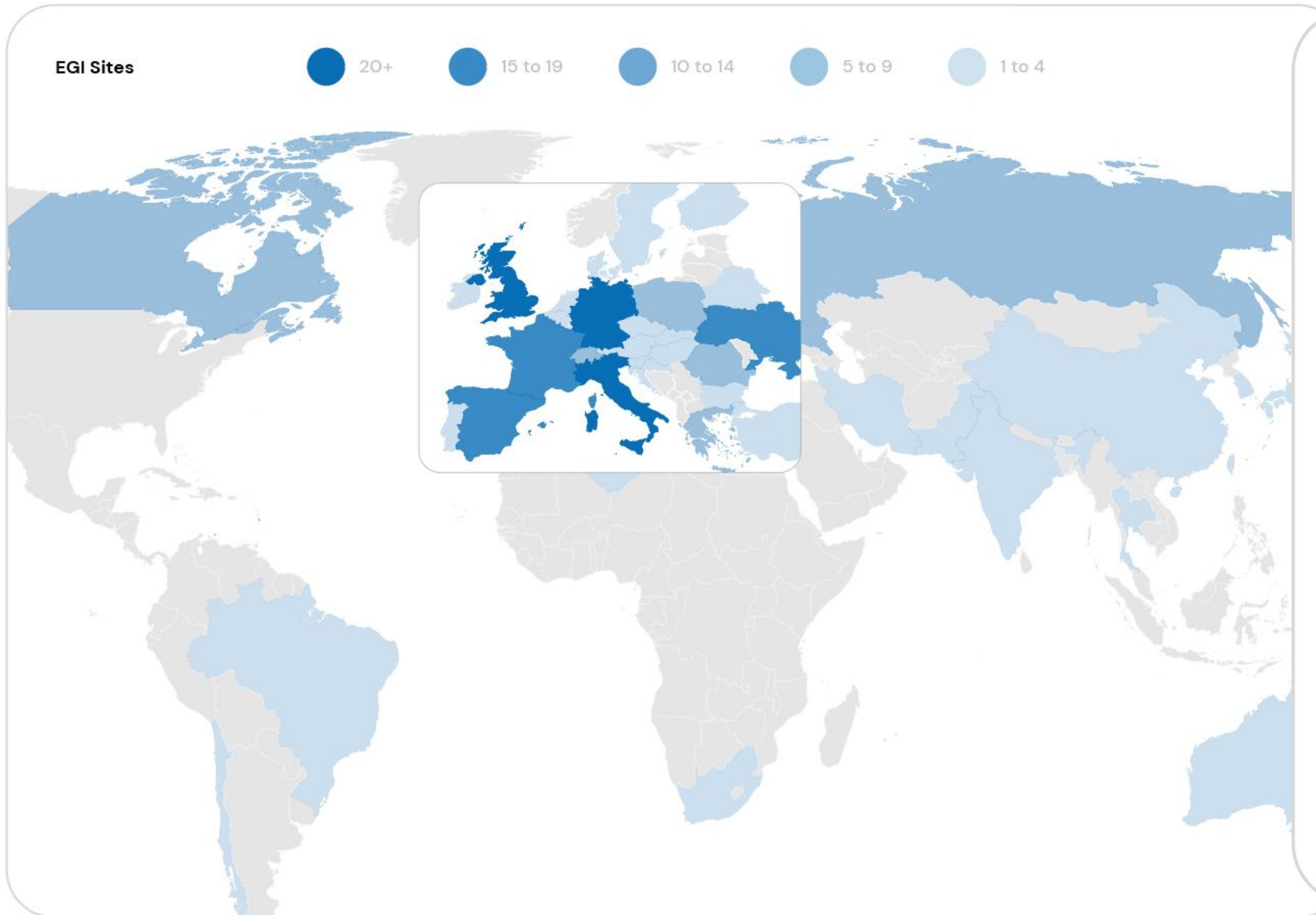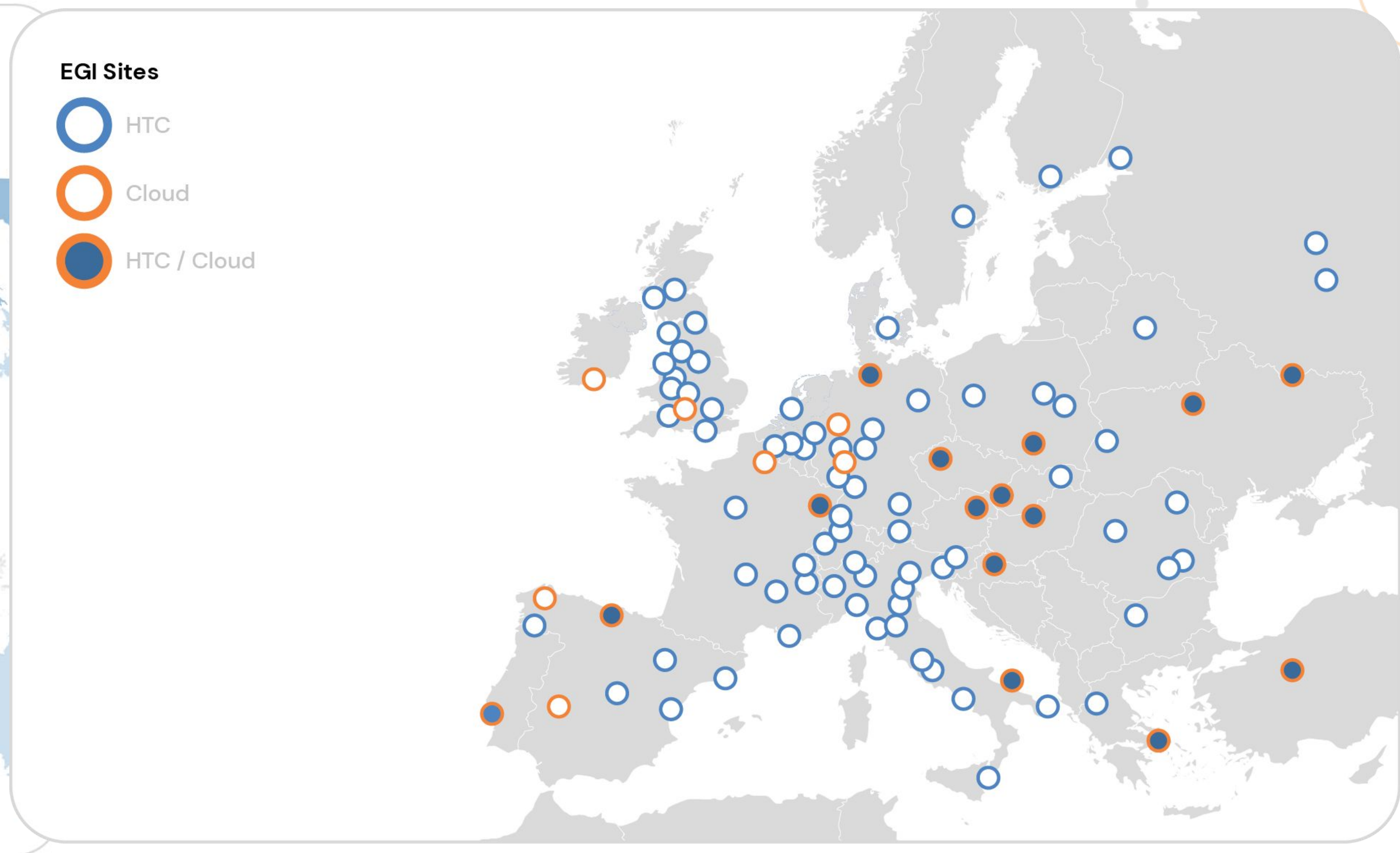
In the context of the SPECTRUM Project, High-Energy Physics and Radio-Astronomy representatives are developing a technical blueprint and strategic research agenda for a compute & data continuum

www.spectrumproject.eu

EGI Federation in 2030

Global Open Research Commons - Essential Elements

Data Continuum

Data → Information → Insight
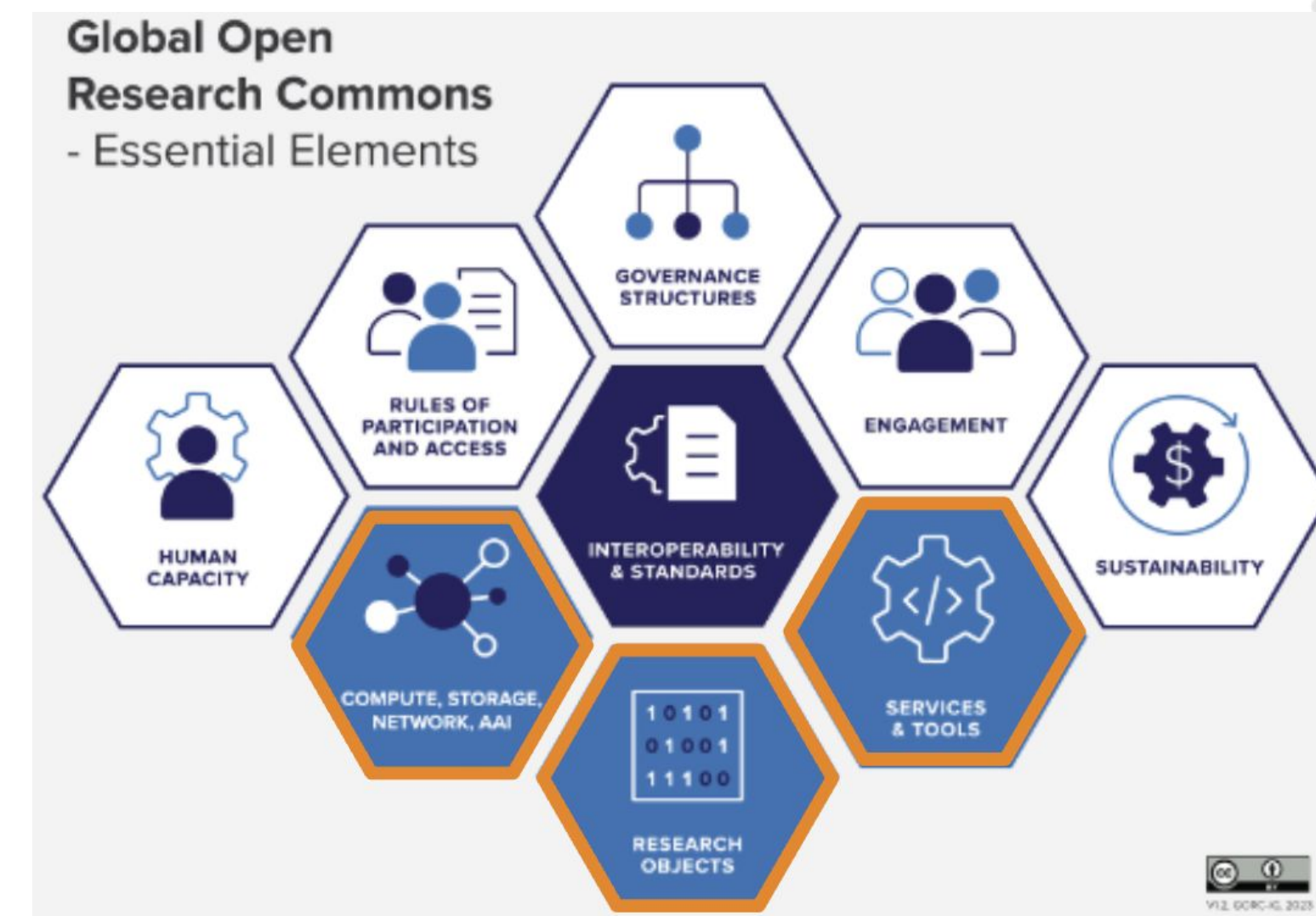
Orchestration & Interoperability

On-device → Edge → Cloud → HTC → HPC → Quantum

Compute Continuum
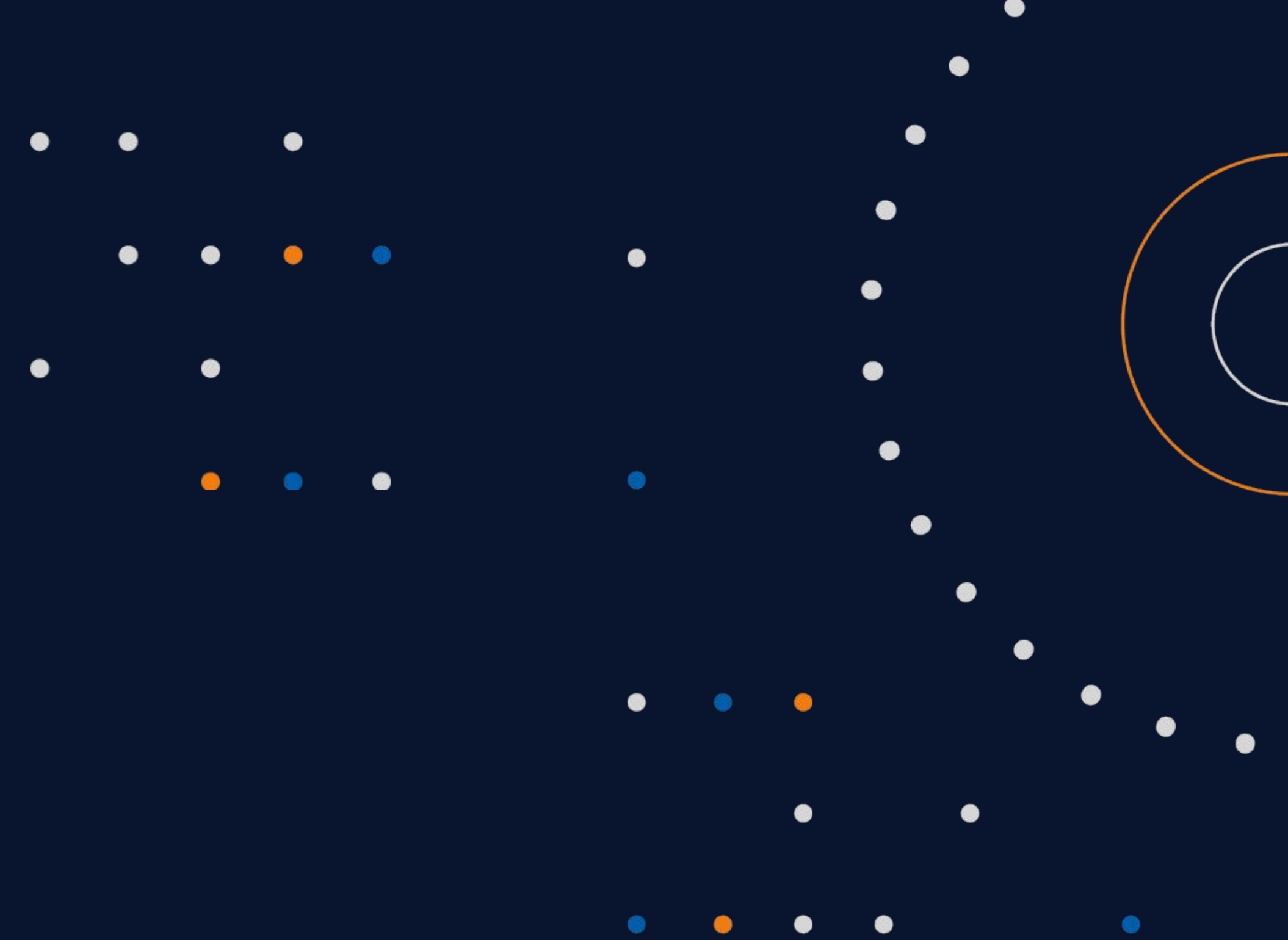
2/The crisis of accessibility to research data

RI-SCALE

1/

Increasing size and complexity of research data holdings

2/

Limited data analytics capacity, integrating technical knowledge, compute environments, data staging, data analysis and AI techniques

3/

Lack of business models for the data economy when applied to research

Examples of ESFRI RI use cases (credits: RI-SCALE consortium)

Climate modelling

CMIP6 data repository at DKRZ, 20PB in total

1. High-resolution **downscaling** of climate scenarios and **risk trend analysis** in agriculture
2. Smart **detection of anomalies** in climate data usage
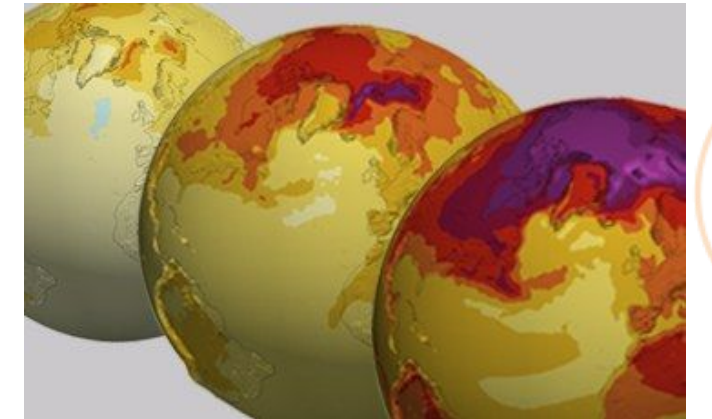
Near-Earth geospace studies

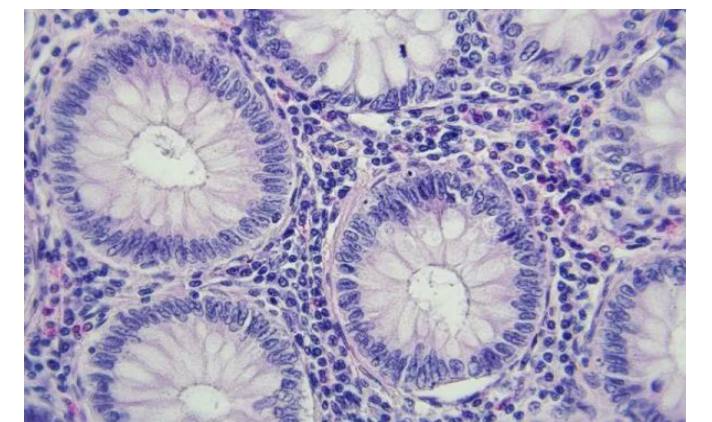*40 years of EISCAT and new EISCAT3D data, few PBs/year*

1. **Intelligent Scheduling** of Radar Observations and Experiments
2. Space **Debris and Anomaly** Detection
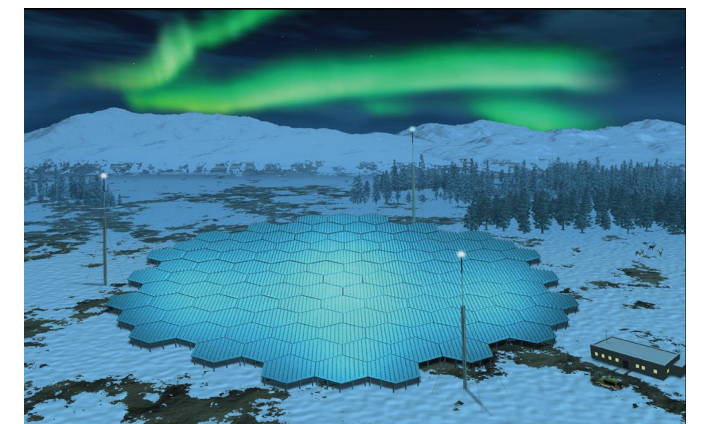
Cancer treatment

*Cancer & digital pathology data at 3 BBMRI sites, 3 PB in total*

1. Colorectal **cancer prediction** with explainable AI
2. **Synthetic data** for computational pathology

Imaging in life sciences

BioImage Archive and EMPIAR DBs at EMBL-EBI, 4.5 PB in total

1. **Foundational models** for heterogeneous biological image data
2. **Generative AI-Powered assistant** for data discovery and analysis
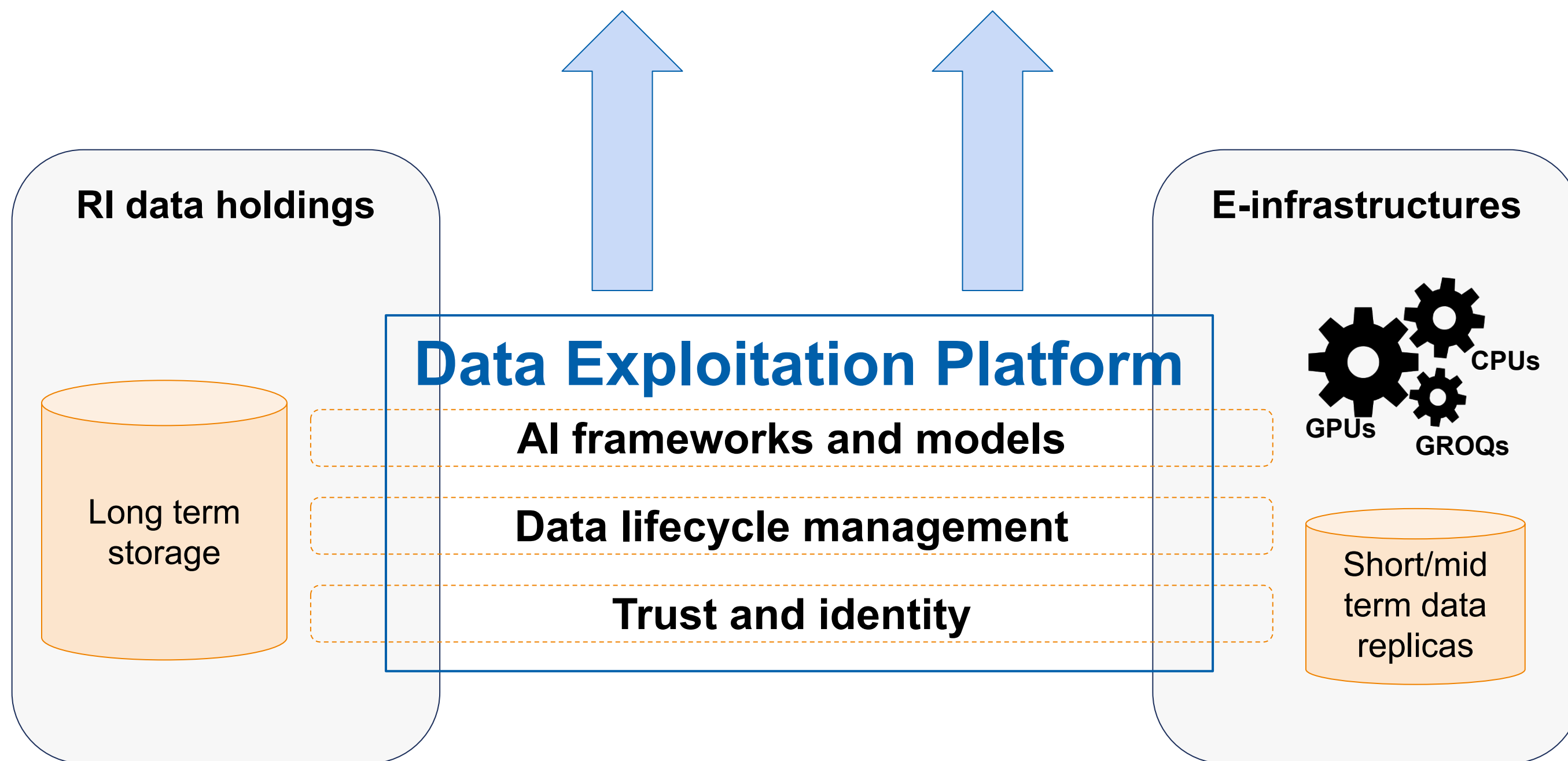
RI-SCALE: technology innovation for scalable RI data exploitation
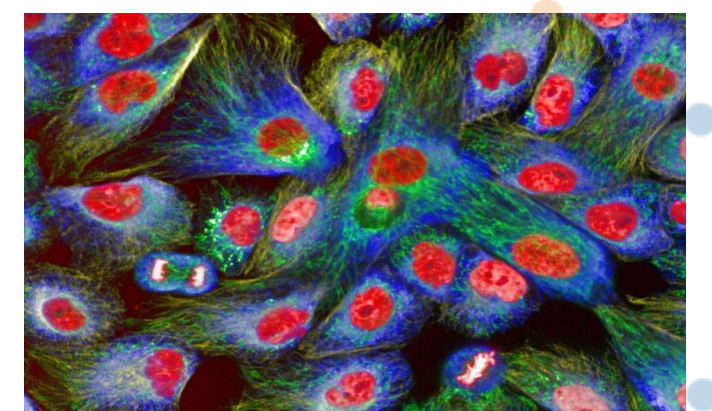
Value for data managers
Data enrichment, error detection, etc.

Value for scientist
Data mining, analysis, AI-powered helpdesk, etc.

RI data holdings

Long term storage

Data Exploitation Platform

AI frameworks and models

Data lifecycle management

Trust and identity

E-infrastructures

CPUs
GPUs   GROQs

Short/mid term data replicas

- **Goal: scalable data exploitation**
- **Policies for data replication scaling out data to the EGI Federation**
- **AI frameworks for data valorisation**
- **Credit based data usage accounting**
- **Interoperability validation with data spaces**
  - Destination Earth, Copernicus Data Space, EUCAIM cancer image federation
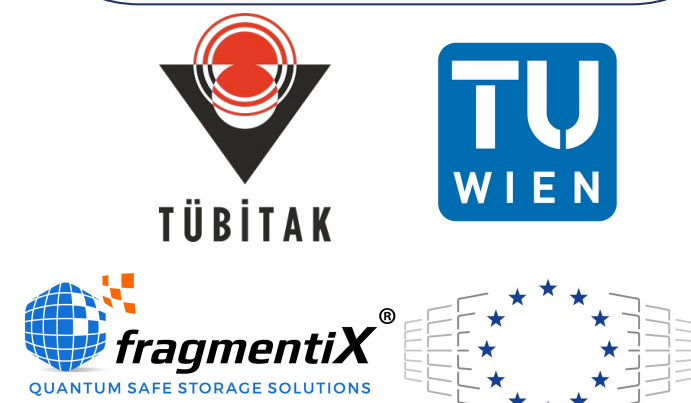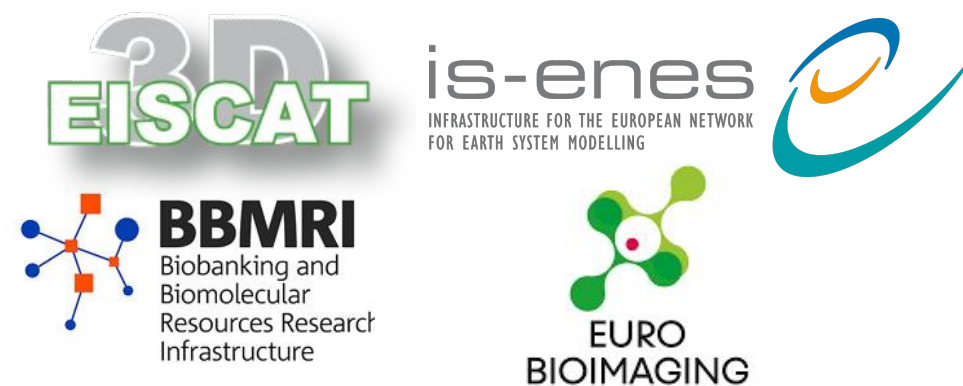
interTwin

interTwin develops a Digital Twin Engine to support the development of Digital Twins for various science domains
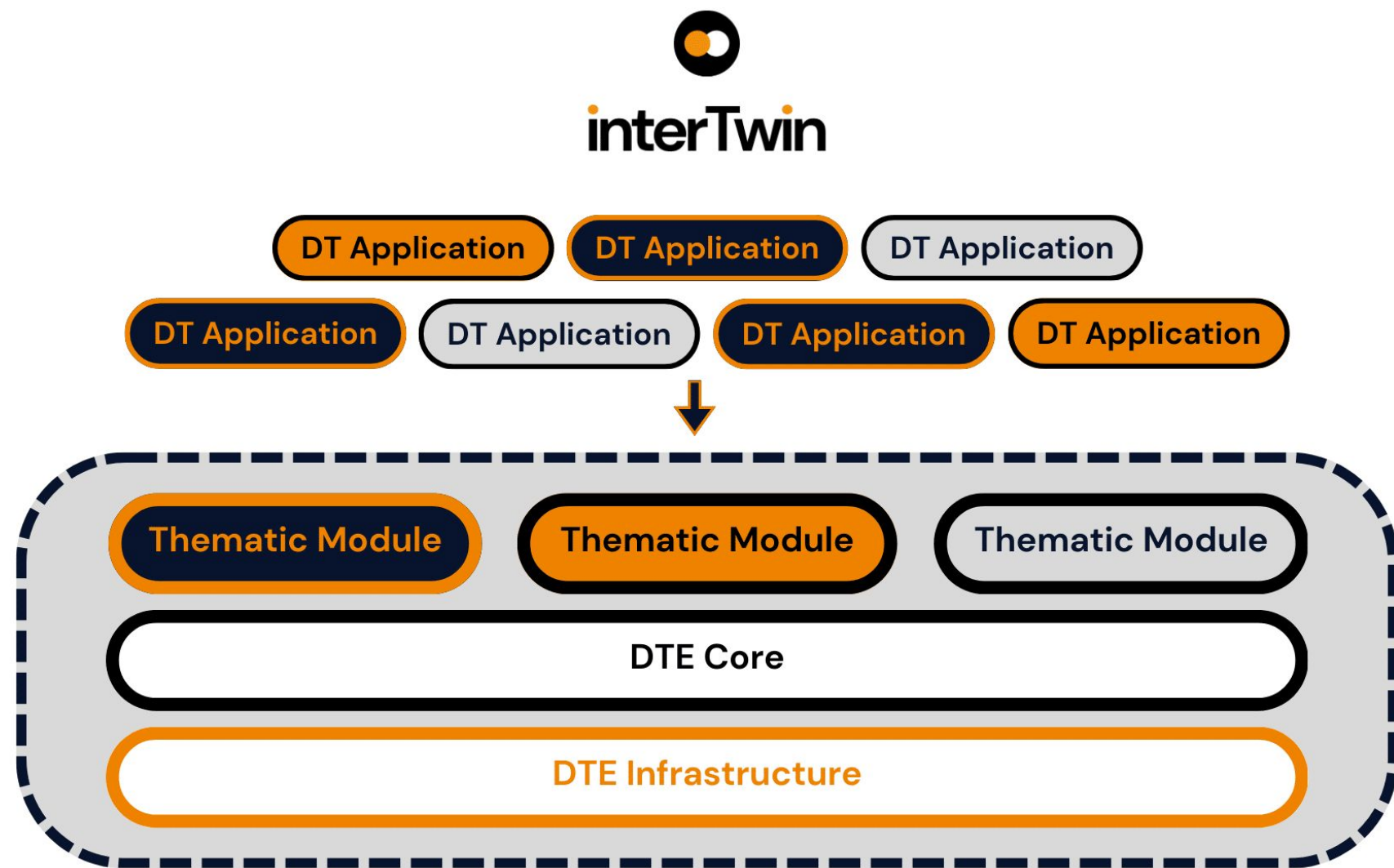
The interTwin Digital Twin Engine (DTE)

Toolkit for AI workflow and method lifecycle management

DT Quality Framework

DTE Infrastructure implementing computing continuum and Data Federation

WildFire Hazard Map Generation
CMCC, CNRS, Univ. of Trento

Early warning for Extreme events
Deltares, EURAC, TU Wien

Extreme events impacts
CERFACS, EURAC, Deltares

Tropical Cyclone Detection
CMCC, CNRS, Univ. of Trento

Radio Astronomy
Max Planck Society

High Energy Physics
CERN, CNRS

Lattice QCD
CSIC, CNRS, ETHZ

Gravitational Wave - Astrophysics
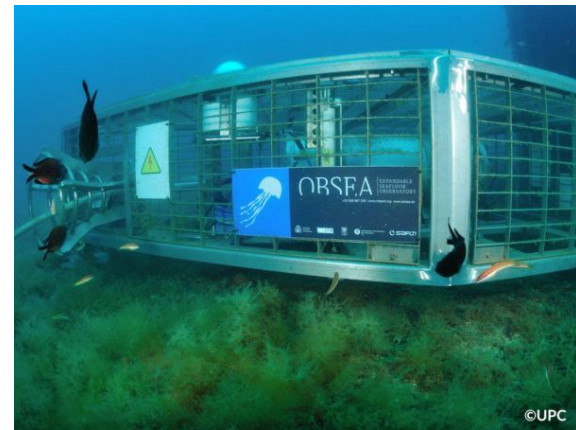INFN

21
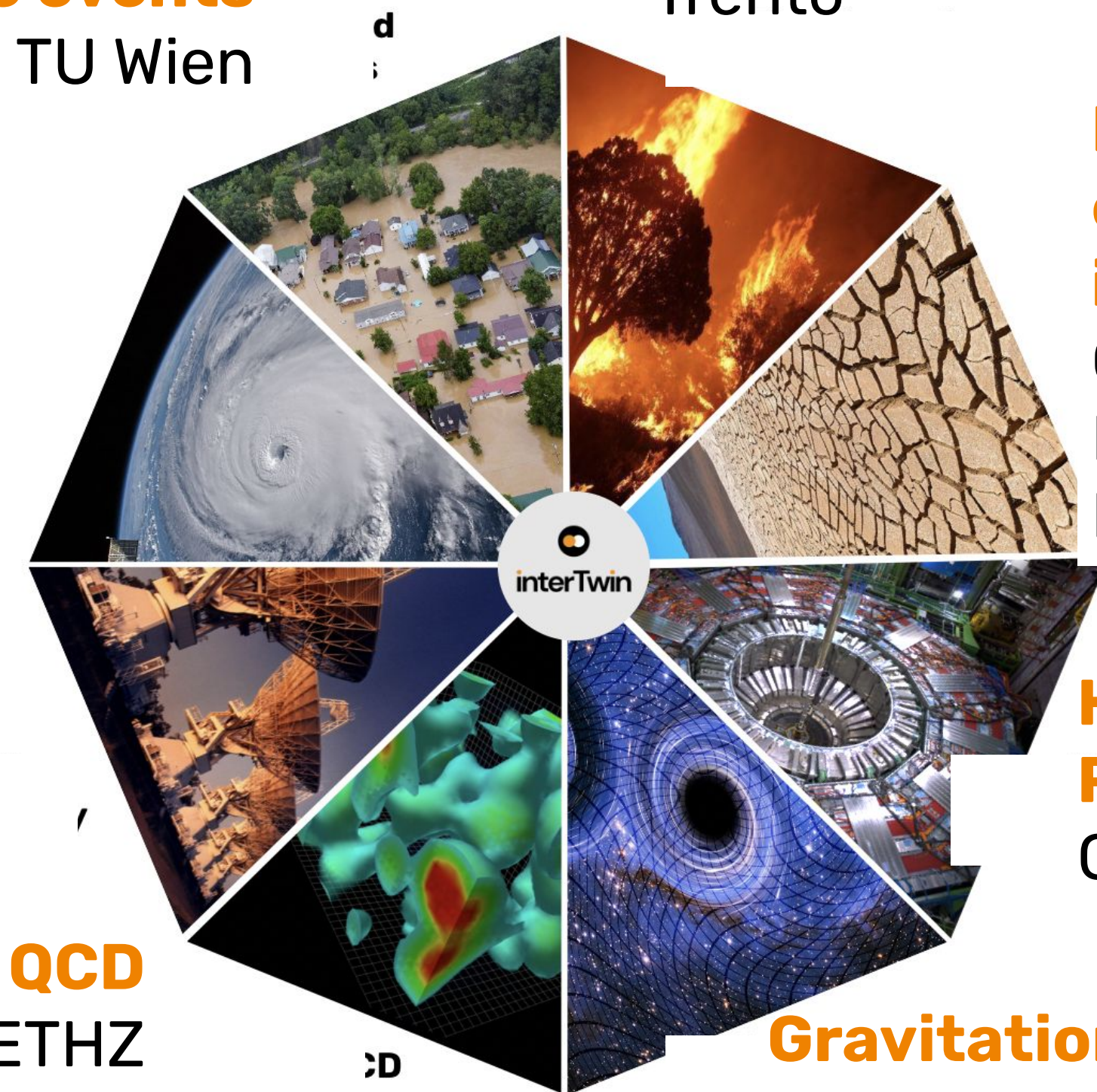
3/The need of AI democratisation

Image data tsunami in aquatic sciences

iMagine

In Europe, we spent circa **1.4 Billion Euro a year** in marine data acquisition (in-situ and remote sensing data)
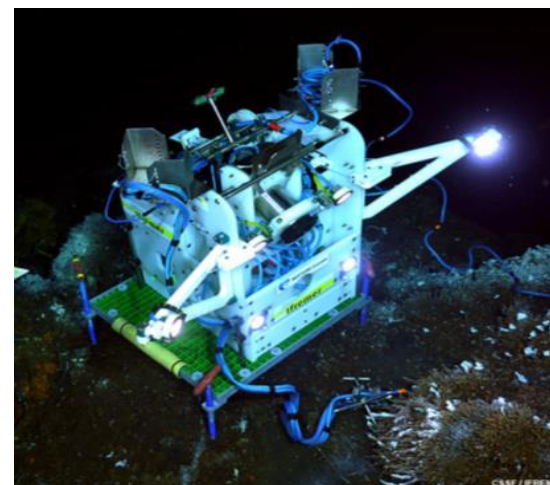
Underwater cameras

Microscopes

CCTV cameras

Seafloor observatories

Satellites

Drones

ACQUISITION

EMODnet
European Marine Observation and Data Network

SeaDataNet

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Copernicus Marine Service

... Euro Argo

MANAGEMENT

AI/ML

From DATA to KNOWLEDGE

iMagine

VALORISATION
Fish monitoring and classification
Marine litter detection
Oil spill prediction, ...

Approach

Thematic Services

AI/ML Model development, training and delivery
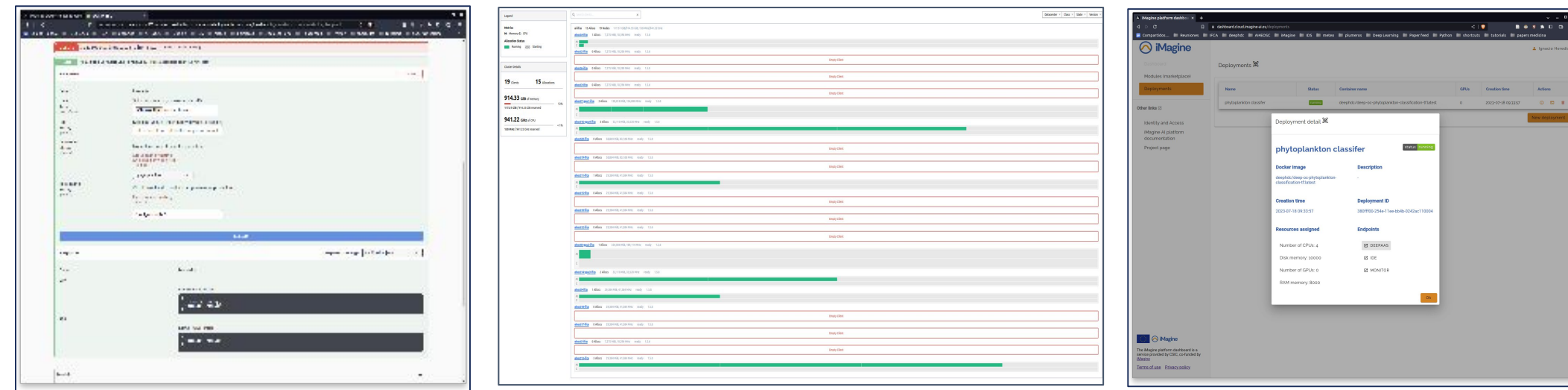
GPU intensive **CPU intensive**
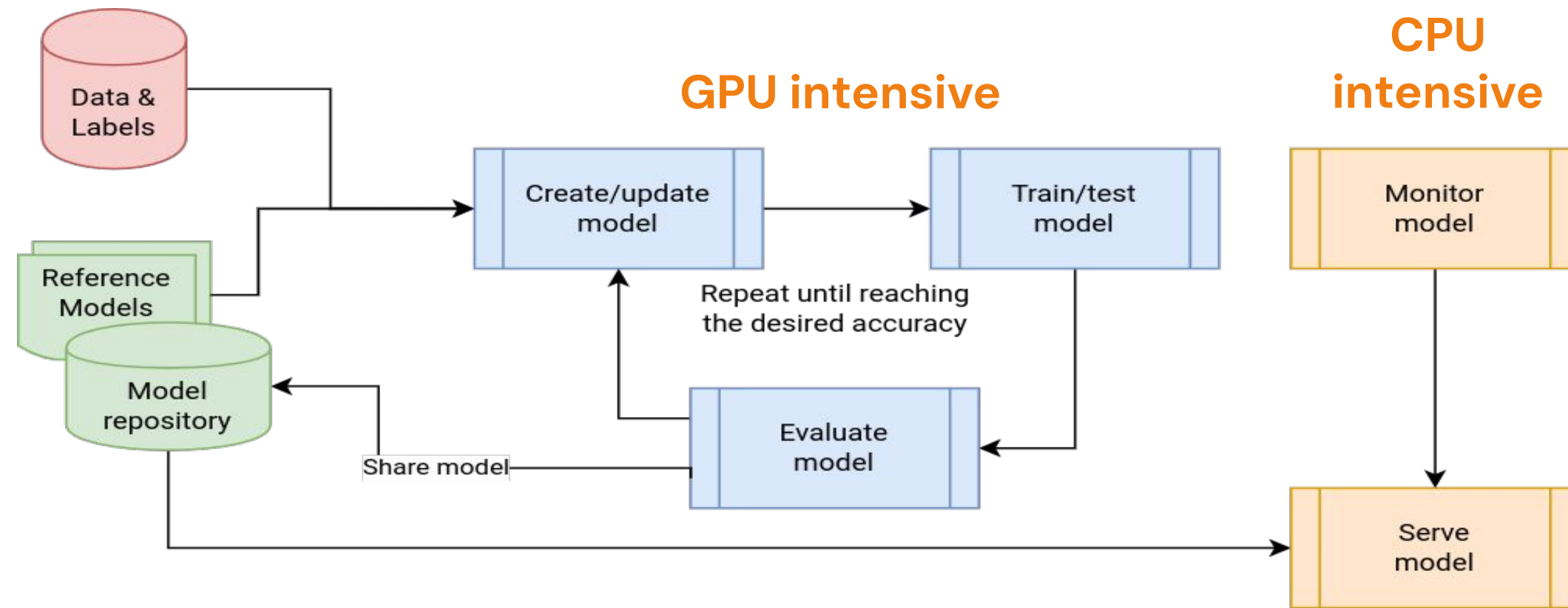
Thematic Services Competence Centre

8 internal use cases **&**
3 external use cases

Benefitting **10+** Research Infrastructures and linking to

Platform Service

iMagine AI Platform
Generic, scalable platform for developing and sharing AI/ML applications.

Infrastructure

EGI Federated Cloud
- **1500** TB-months
- **132,000** GPU-hours
- **6,000,000** CPU-hours

OpenStack clouds with GPUs, CPUs, Storage in Spain, Portugal, Turkey & Ireland.

iMagine

Powered by EGI

The iMagine Open Call for AI-powered image analysis in aquatic sciences is NOW OPEN!

We offer:
- 10-month collaborative projects

Support for
- AI model training
- Large-scale image analysis
- Use of the iMagine AI platform to develop and train AI models
- Access to cloud resources (GPUs, CPUs, storage) to store images and to scale up analysis workflows
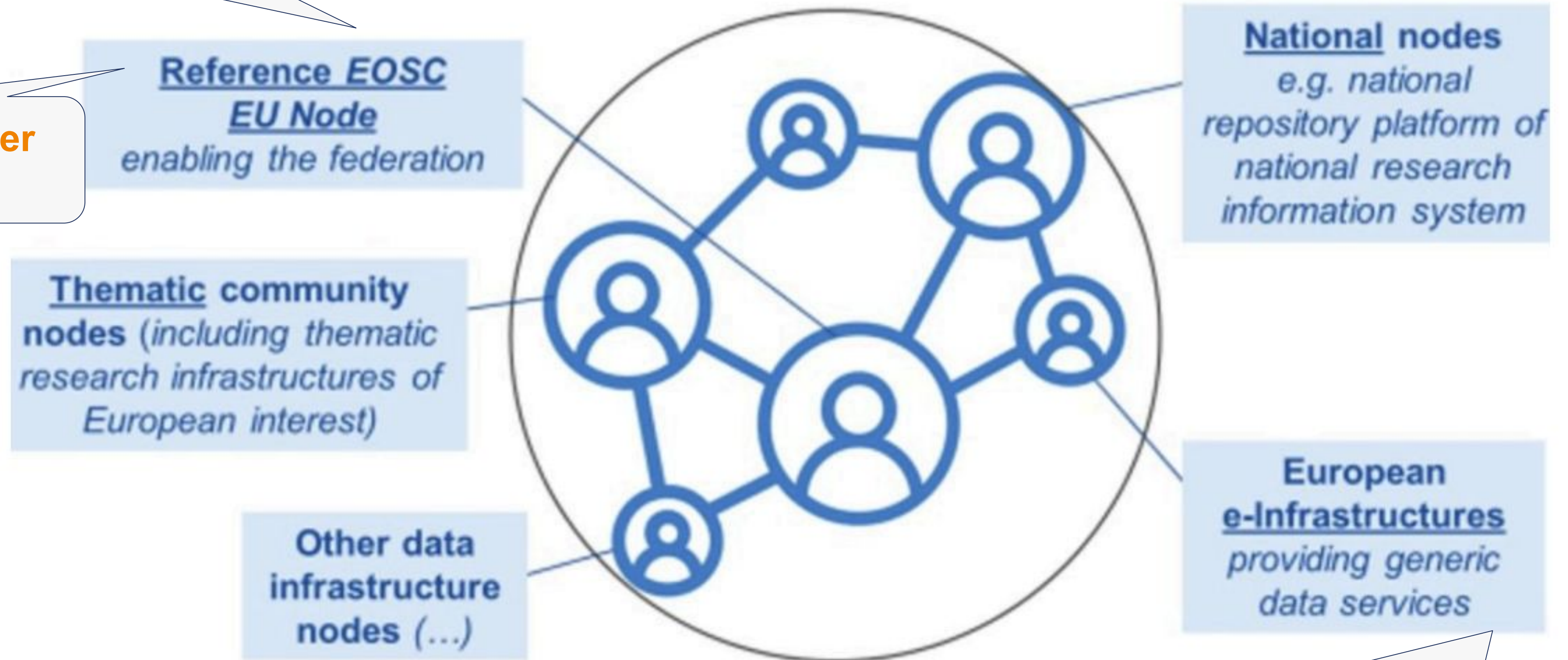
EGI Role
in the European Open Science Cloud

EGI Federation contribution to EOSC

1/ **Provider of federating services for the EOSC EU Node**

2/ **Contributor to Jupyter Notebooks**

Reference *EOSC EU Node* enabling the federation

Thematic community nodes *(including thematic research infrastructures of European interest)*

Other data infrastructure nodes *(...)*

National nodes *e.g. national repository platform of national research information system*

European e-Infrastructures *providing generic data services*

3/ **Candidate EGI Federation Node (B2C, B2B)** Scalable compute-storage, Generic platforms, Thematic data and services

# EGI Contribution to the EOSC Federation

Discussion paper
V. 1.2 07-05-2024

Abstract

Compute, storage, research objects, services and analytics tools are integral to the Open Science Commons and the realisation of the EOSC vision. This paper describes how EGI Federation envisages their provision as a collaborative effort of Research Infrastructures and e-Infrastructures, to jointly provide compute and co-located data services for international, grand challenge applications under the EOSC Node architecture.

This discussion paper reflects the EGI Federation's current thinking on collaborative service provision within the EOSC framework. We believe a strong, collaborative approach is essential for the success of EOSC. We welcome your feedback and insights, which will inform the further development of this position.

+31 (0)20 89 32 007
contact@egi.eu
www.egi.eu

EOSC Data Commons
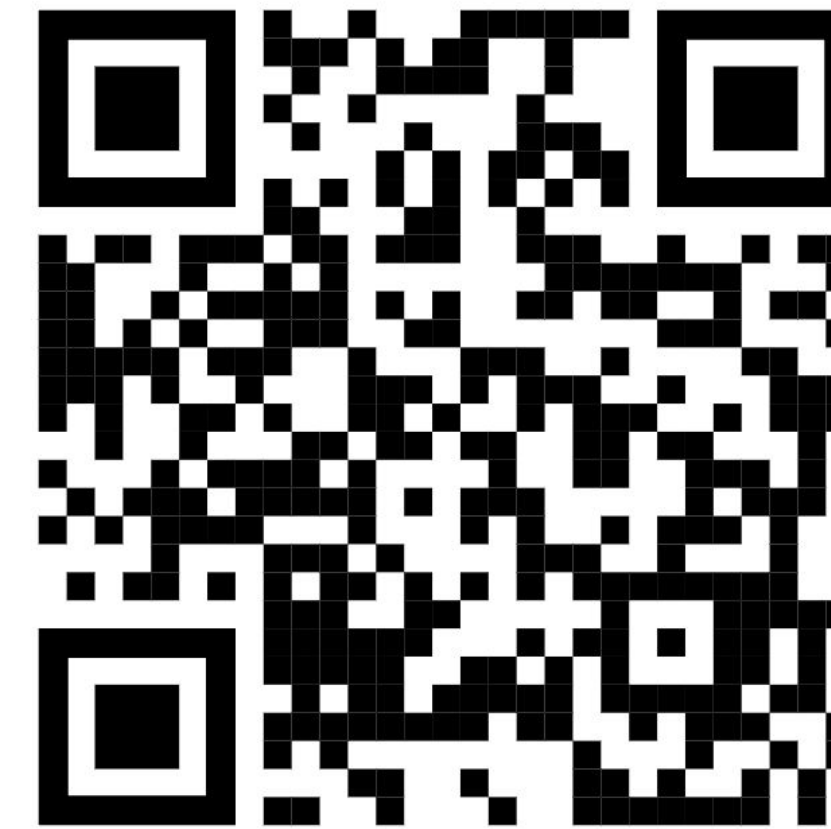
Services for inter- and cross-disciplinary data discovery, access, sharing and reuse in the EOSC Federation

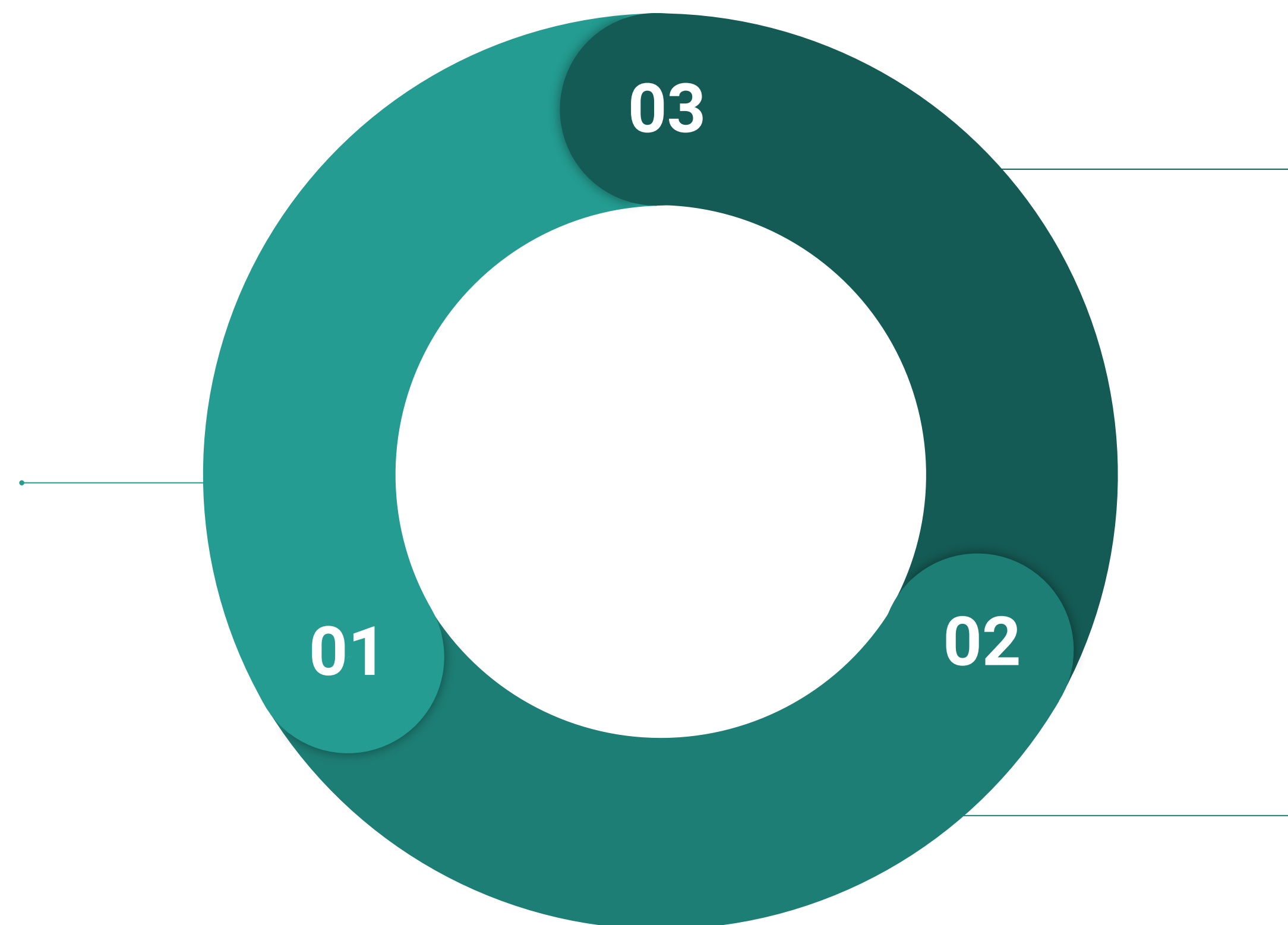**EGI Federation and Repository Owners and Data Providers**

5 Multidisciplinary Repositories

7 Thematic Repositories

New EOSC Core Services

Next generation of Core Services for the next release of the EOSC EU Node and as reference implementation for other nodes

Node-based technical architecture

EOSC Federation Technical Architecture taking into account the introduction of the concept of Node

Sandbox and integration suite

Testing, validation and integration of EOSC nodes with data spaces

01

02

03

- **EGI Federation participants and partners were forerunners in establishing distributed data processing and analysis for data-intensive science**
- **Many technical, policy and governance challenges remain ahead of us**
- **Collaboration remains key!**

Be part of the future EGI Federation strategy 2025-2029!

2029

https://go.egi.eu/2029

Thanks!

Coordinated Projects

Participated Projects