



# The power of digital e-infrastructures: EOSC-Hub

Jorge Gomes (LIP / INCD)



eosc-hub.eu



@EOSC\_eu





## Project Overview

*20 digital research infrastructures,  
EGI, EUDAT CDI and INDIGO-DataCloud  
jointly offering  
services, software and data  
for advanced, data-driven research &  
innovation*

**EOSC-hub: Integrating and managing services for the European Open  
Science Cloud**

**Grant Agreement ID 777536**

**Tot budget: €33,287,542**

**EU Budget**

**€ 30 000 000**

**100 Partners, 76 beneficiaries**

**3830 PMs, 106 FTEs**

**+150 staff involved**

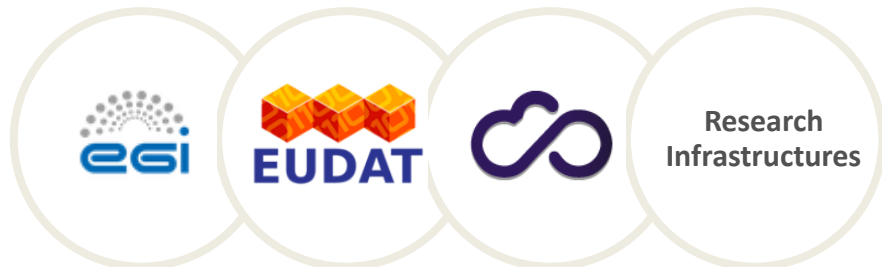
**Coordinator  
STICHTING EGI**

**Consortium  
100 partners  
53 countries**

**36 Months Jan 2018 – Dec 2020**



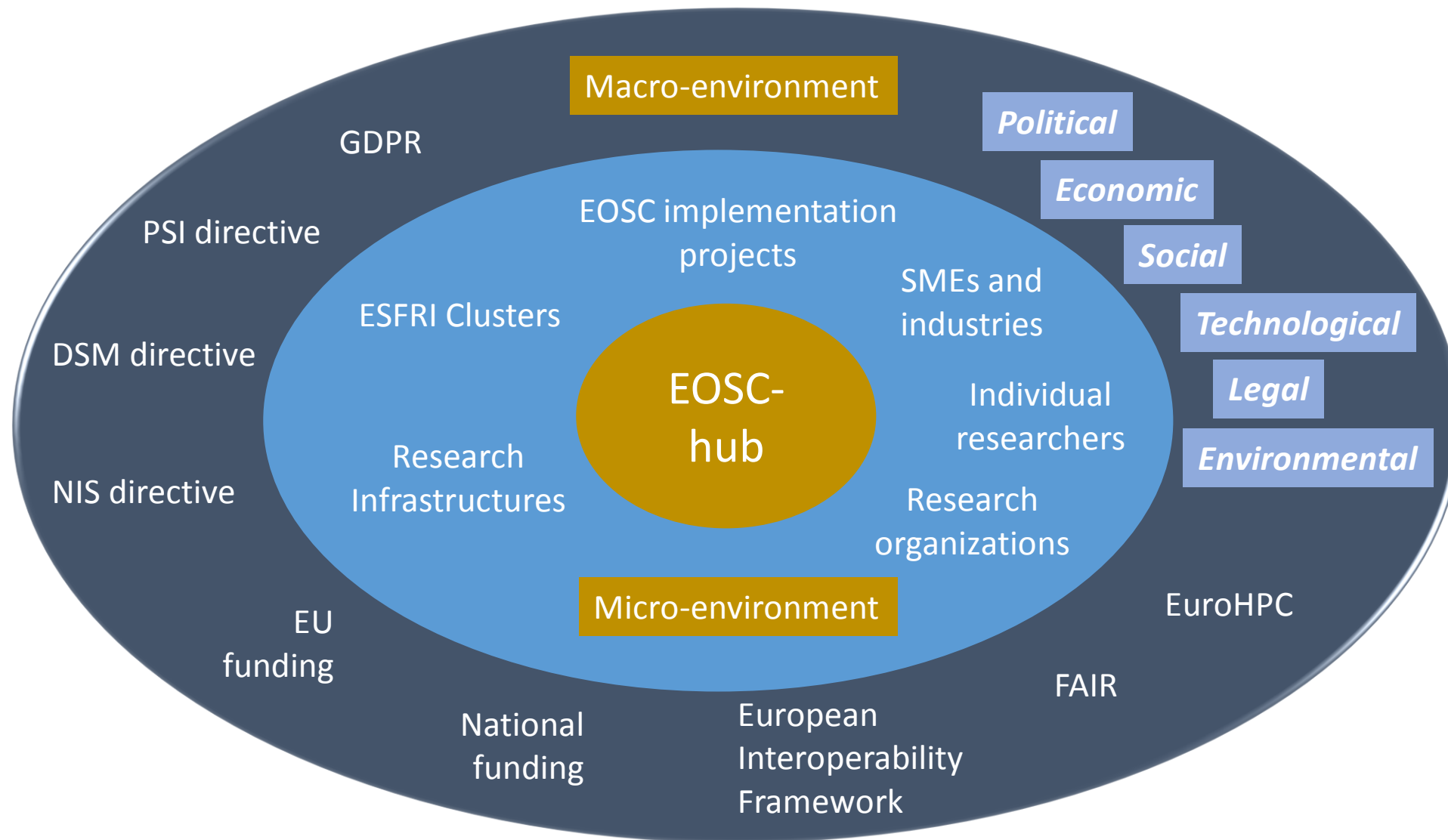
# **EOSC-hub**



## Mission

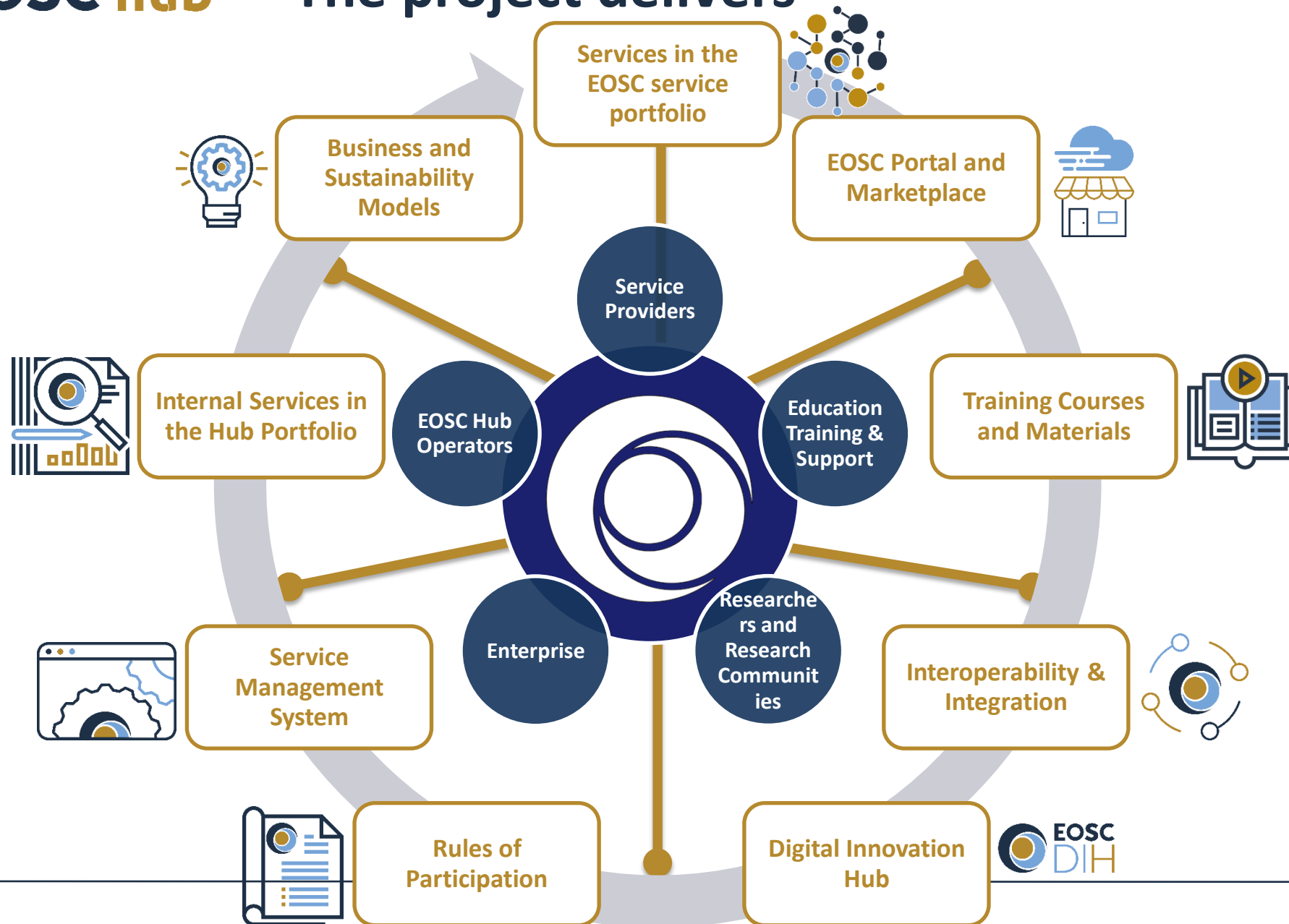
*The EOSC-hub project mobilises providers of pan-European relevance offering services, software and data for advanced data-driven research and innovation.*

*These resources are offered via the Hub – the integration and management system of the European Open Science Cloud, acting as a European-level entry point for all stakeholders.*

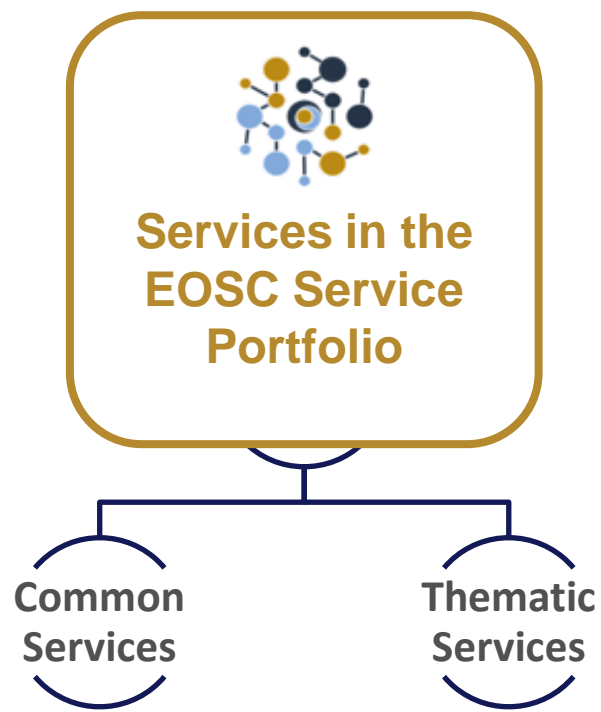


- Publish, discover, access services and resources for all scientific disciplines
- Open to national, regional, pan-European providers, and supports different exploitation models (e.g. free at point of use, commercial)
- Provide thematic services integrated with European compute/data platforms for data exploitation
- Single sign on, integrated access and order management
- Services to share and discover research artefacts (publications, datasets, software, workflows etc.), research artefacts data sources (publication repositories, publishers, data archives, software archives, etc.)

# The project deliverables

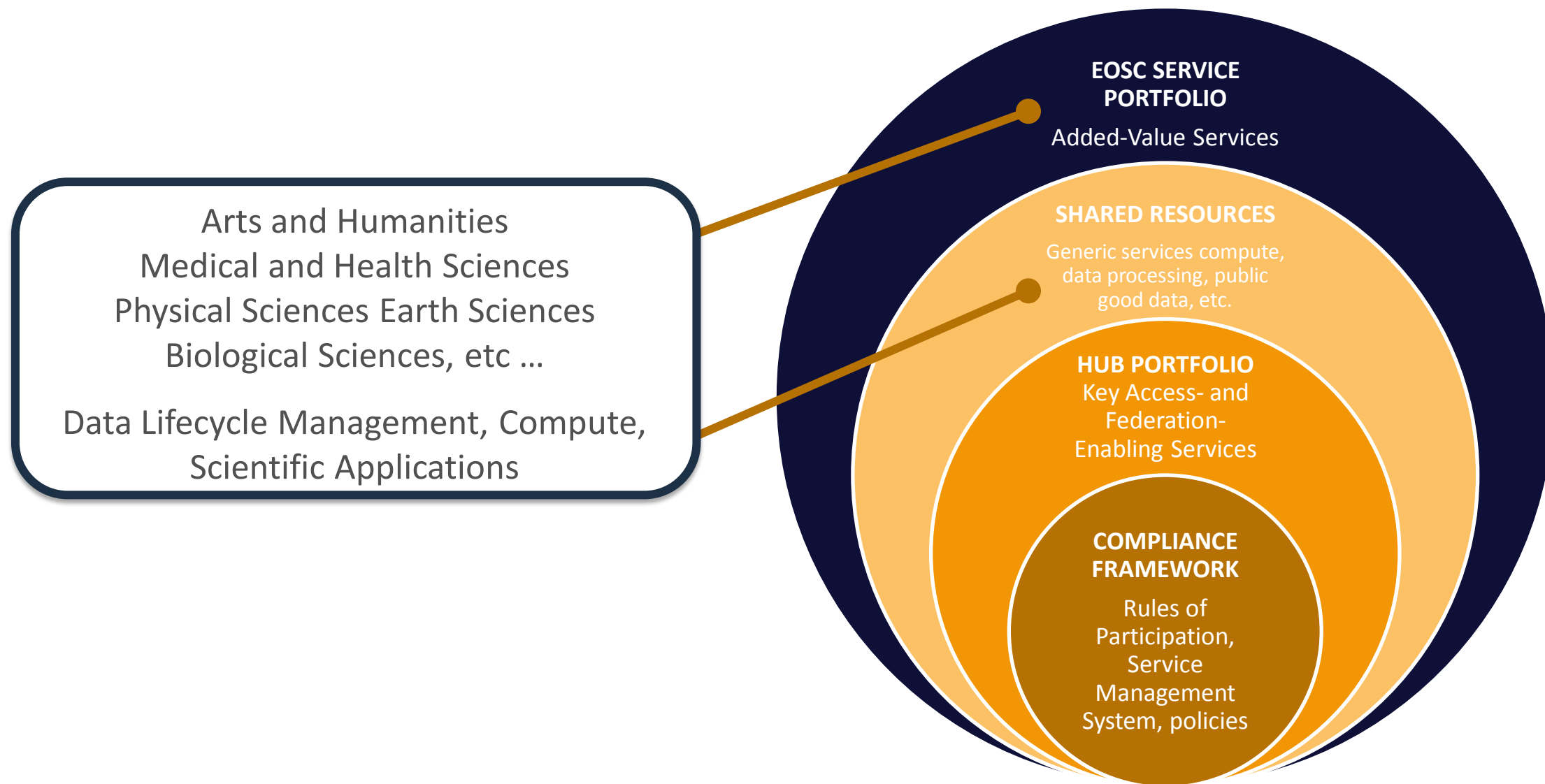


*Simplify access to a broad portfolio of **products, resources and services** provided by the major pan-European and international organizations through an **open and integrated service catalogue***

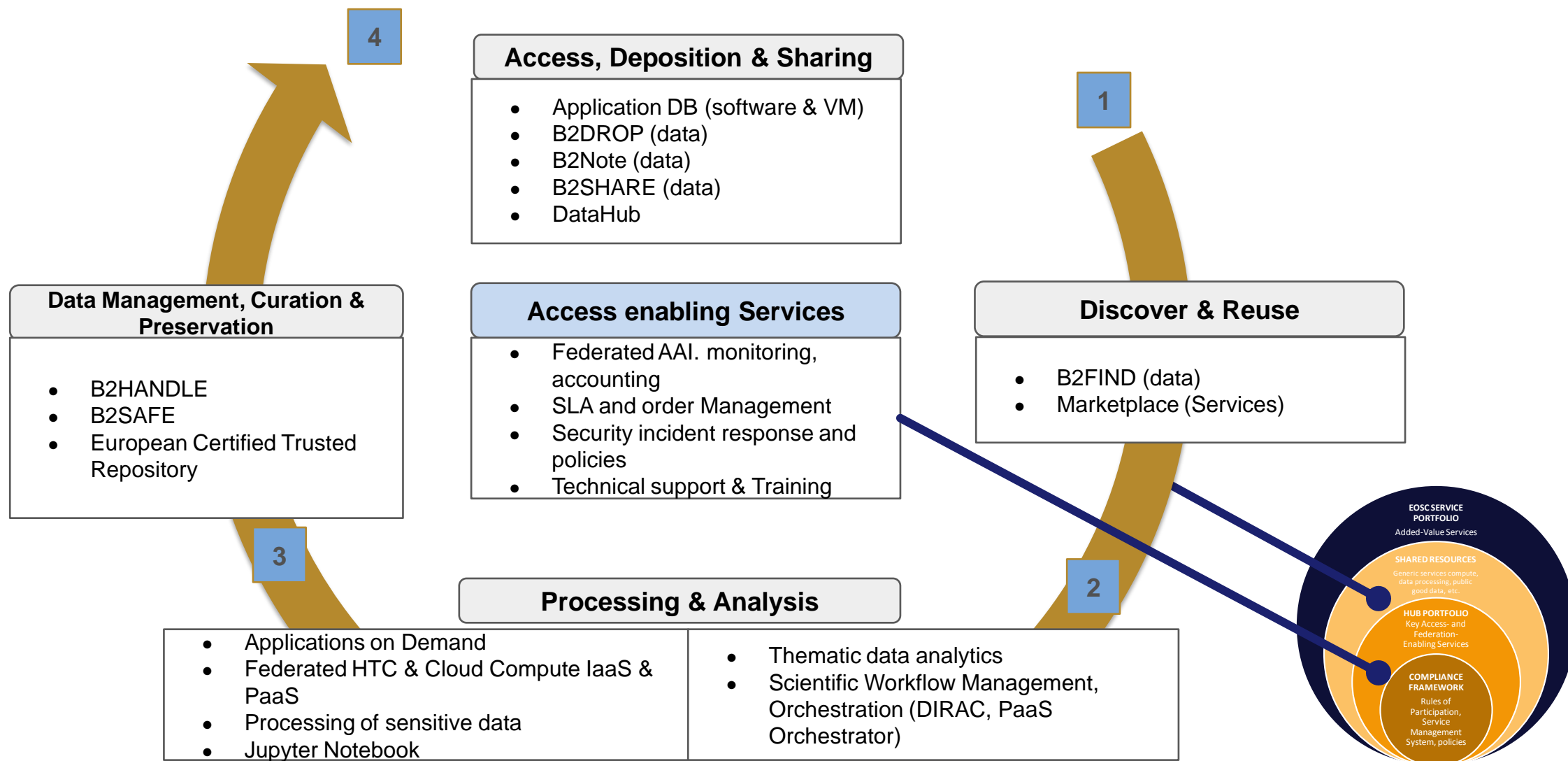


Support a federated service ecosystem:

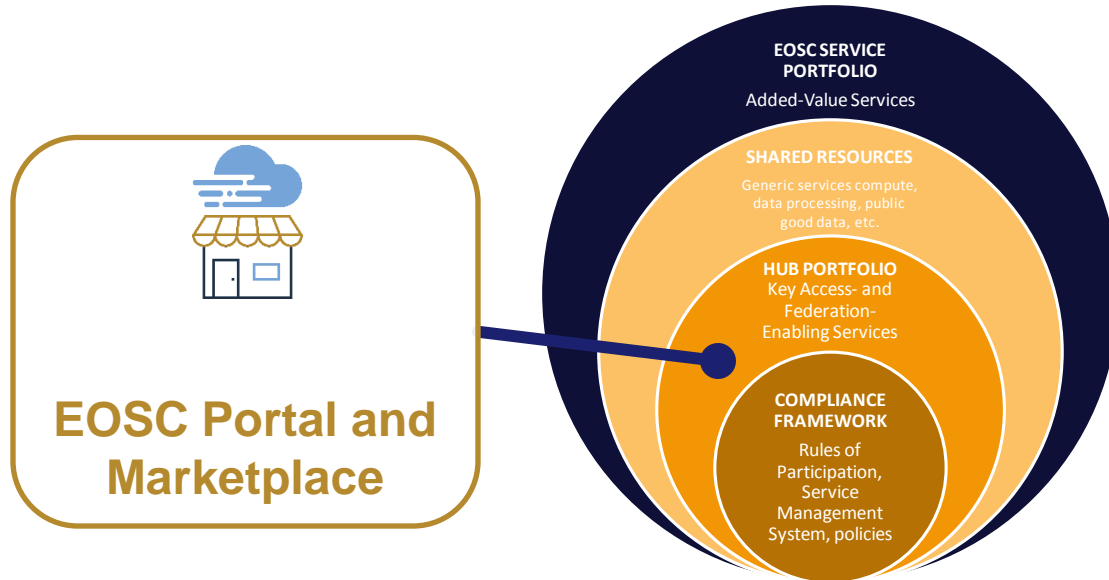
- service discovery
- access to market







*Expand the access to services to **all user groups** including researchers, high-education, business organizations and enlarge the user base*



## Service discovery and access to market

- Entry point to access the services
- Researchers can compare solutions and reuse their credentials and knowledge
- Higher service visibility and reach out
- Service providers will obtain increased interest in their services

## EOSC (hub) Service Portfolio

### Thematic Services

- Scientific applications and data
- Addressing a specific research community or multidisciplinary
- From ESFRIs, national projects, international initiatives

### Common Services

- Provide generic functionalities that can apply to multiple domains
- From technology providers, research projects, national groups
- Includes services also seen as components in **Hub Portfolio**

CLARIN



### Virtual Language Observatory

A facet browser for fast navigation and searching in huge amounts of metadata.

Provided by: [CLARIN ERIC](#)

Research area: [Humanities, Social Sciences](#)

Dedicated for: [Researchers](#)

☆☆☆☆☆ (0.0 /5) 0 reviews

ABOUT

REVIEWS (0)

A facet browser for fast navigation and searching in large amounts of metadata. This portal enables the discovery of language data and tools, provided by over 40 CLARIN centres, other language resource providers and Europeana.

The VLO also provides access to the [Virtual Collection Registry](#) metadata and can be used as a starting point to process language data with the [Language Resource Switchboard](#).

Sensitive data



### CSC ePouta

Secure and cost-effective cloud computing for processing sensitive data

Provided by: [CSC](#)

Research area: [Interdisciplinary, Social Sciences, Arts, Philosophy, ethics and religion, Electrical, electronic and information engineering](#)

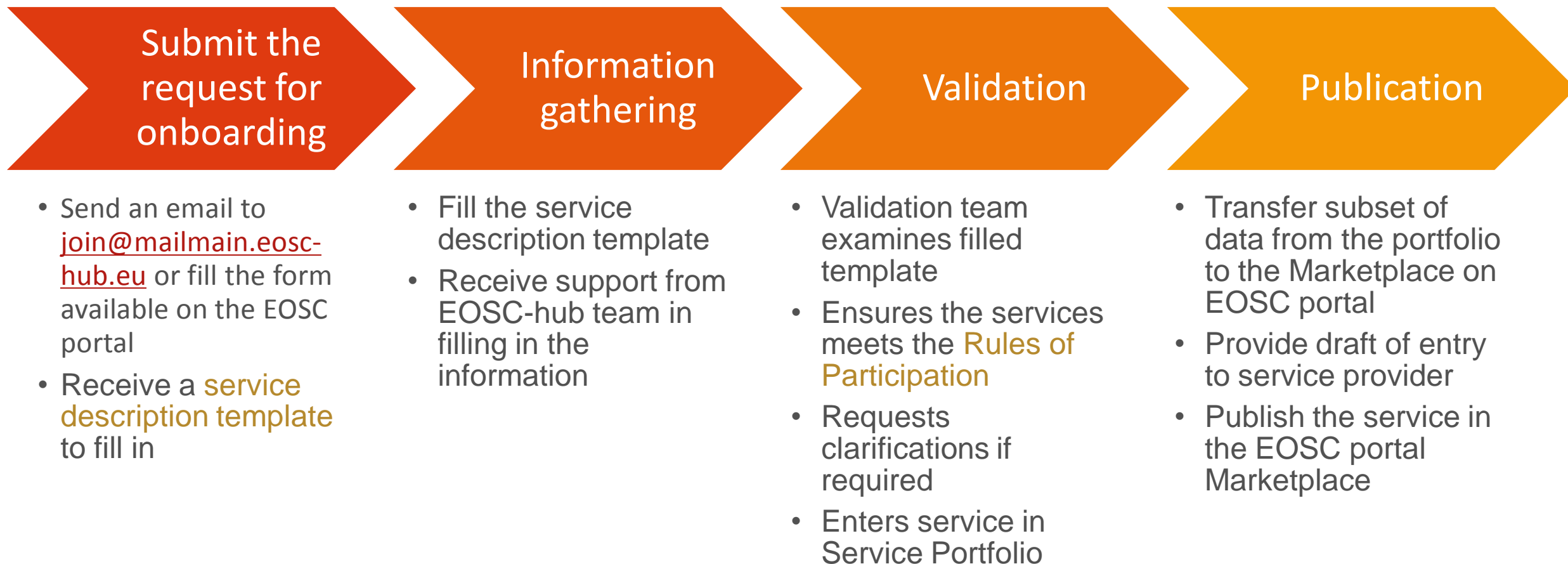
Dedicated for: [Researchers, Research organisations, Research group, Providers](#)

☆☆☆☆☆ (0.0 /5) 0 reviews

ABOUT

REVIEWS (0)

This service provides a infrastructure as a-service for running analysis on sensitive data. The ePouta Virtual Private Cloud service allows customers to provision virtual machines and storage resources directly to their own internal networks. It provides an easy to use admin web interface and a programmable API for managing virtual machines, networks and storage. CSC ePouta meets elevated information security level regulations and is targeted for sensitive data processing



- **Virtual Language Observatory:** uniform search and discovery functionality for language resources and tools
- **Virtual Collection Registry:** a registry of collections pointing to digital objects, enabling easy citation
- **Language Resource Switchboard:** find a matching tool for a given (language) dataset

LRS - VA Metrics
Visits: 528/month
Harvested metadata records: 909'388
VCR - VA Metrics
# Collections: 7
VLO - VA Metrics
# Tools: 70

Service	Integration with Hub
Virtual Language Observatory (VLO)	EGI Cloud Compute, B2FIND, B2SHARE
Virtual Collection Registry	B2FIND, B2SHARE
Language Resource Switchboard	B2DROP, B2SHARE

Virtual Language Observatory Search Help

VLO / Faceted search

Search

Showing all 1646828 records Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language Collection Resource type Modality Format Keyword Availability Search options

**EXMARaLDA Demo corpus**  
(Part of Hamburger Zentrum für Sprachkorpora (HZSK))  
A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system.; HIAT (simplified); HIAT; free comment; suprasegmental information; accentuation/stress; English translation; Standard German translation; German translation; English translation; code-switch

**The Hamburg MapTask Corpus (HAMATAC)**  
(Part of Hamburger Zentrum für Sprachkorpora (HZSK))  
Audio and two-video recordings of map tasks with adult L2 users of German and one L1 speaker. The speakers' L1 and their L2 proficiencies vary. The maps used for the tasks are available.; orthographic transcription/simplified HIAT; Fine-grained part of speech tagging using TreeTagger and the STTS tagset.; superordinate...

**Czech news subcorpus from 2012 (ces\_news\_2012\_1M)**  
(Part of Leipzig Corpora Collection)  
Czech news subcorpus based on material from 2012 (1,000,000 sentences)

**Southern Sotho Web text subcorpus (South Africa) from 2018 (sot-za\_web\_2018\_10K)**  
(Part of Leipzig Corpora Collection)  
Southern Sotho Web text subcorpus (South Africa) based on material from 2018 (10,000 sentences)

**Main use cases:**

Publication of collections of research data from the Arts & Humanities with integrated meta-data handling and search

- **DARIAH Science Gateway:** provides a set of generic and customized services and tools
  - **Semantic Search Engine (SSE):** allows users to search in Open Access Document Repositories and Data Repositories and to discover new correlations about document and data
  - **Parallel Semantic Search Engine (PSSE):** enabling users to semantically correlate contents in geographically distributed digital repositories
  - **Simple Cloud Access:** workflow creation and execution
  - **DBO@Cloud:** Cloud repository of Bavarian dialects
- **Invenio-based repository in the cloud:**
  - enables users to easily create, deploy and configure their own Invenio-based repository instance in the cloud
- **DARIAH repository:** storing and searching objects in research projects

Service	Integration with Hub
DARIAH Science Gateway	EGI Cloud Compute EOSC-hub AAI Accounting Monitoring
Invenio repo	INDIGO IAM INDIGO Orchestrator EGI Cloud Compute
DARIAH repo	EOSC Federation Services

DARIAH - VA Metrics
Number of supported applications: 5
Country reached: 10

## Domain specific training to data providers and data scientists

- Focused on the preparation, and delivery of training contents and events to target the data providers and data scientists within structured scientific communities, linked to the Competence Centres and Thematic Services

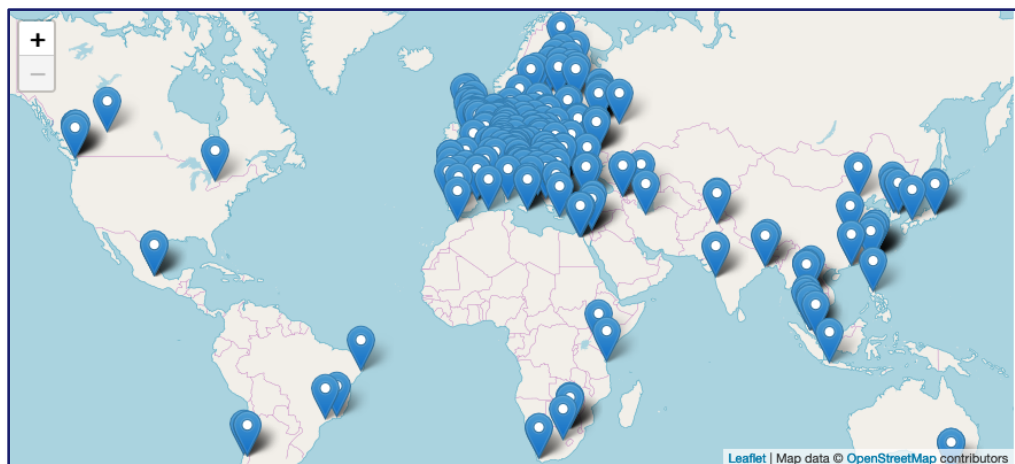


... and

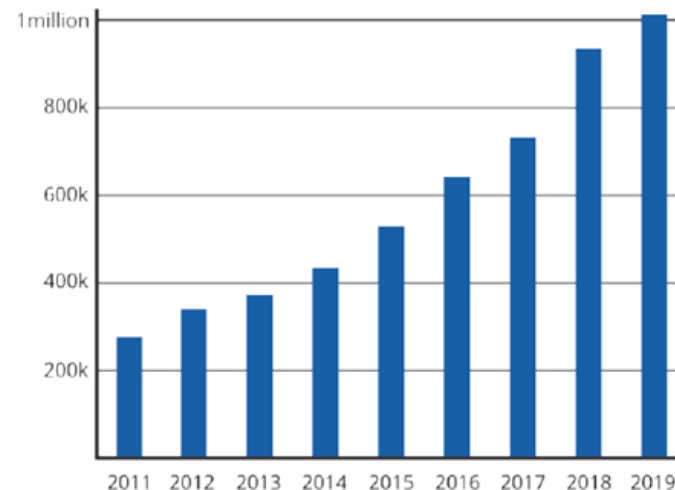




- Enabling exploitation of publicly funded resources at pan-European level



EGI: more than 250 federated resource centres in 47 countries



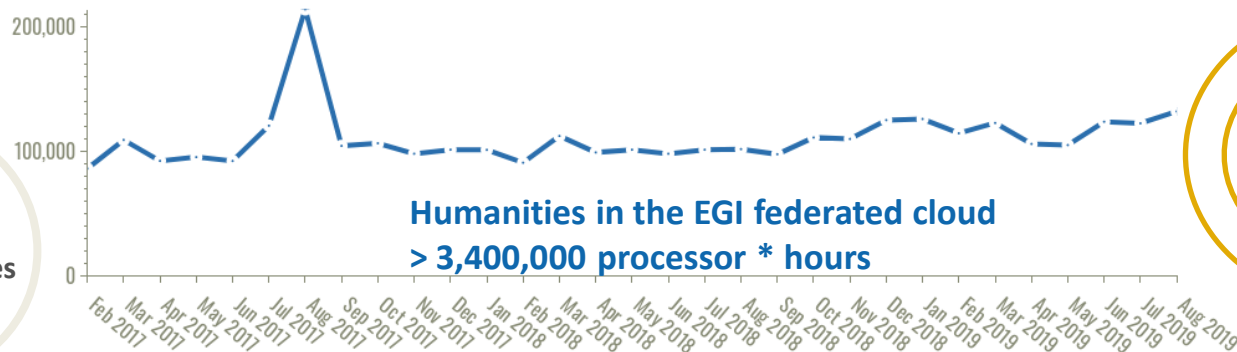
5.0 Billion  
CPU core wall  
time / year

> 1 Million  
computing  
cores in 2019

> 740 PB disk  
& tape

2,915 service  
end-points

Elapsed time \* Number of Processors (hours) by Scientific Discipline and Month





## Early adopters

- *Pilot*
- *Use*
- *Co-develop*



## Users

- *Access range of services*
- *Support*
- *Improve with feedback*
- *Use*

[early-adopter@mailman.eosc-hub.eu](mailto:early-adopter@mailman.eosc-hub.eu)

[contact@mailman.eosc-hub.eu](mailto:contact@mailman.eosc-hub.eu)

## Competence Centres

- *Planning & design*
- *Requirement collection*
- *Setup proof-of-concepts*
- *Service co-creation*
- *Training*

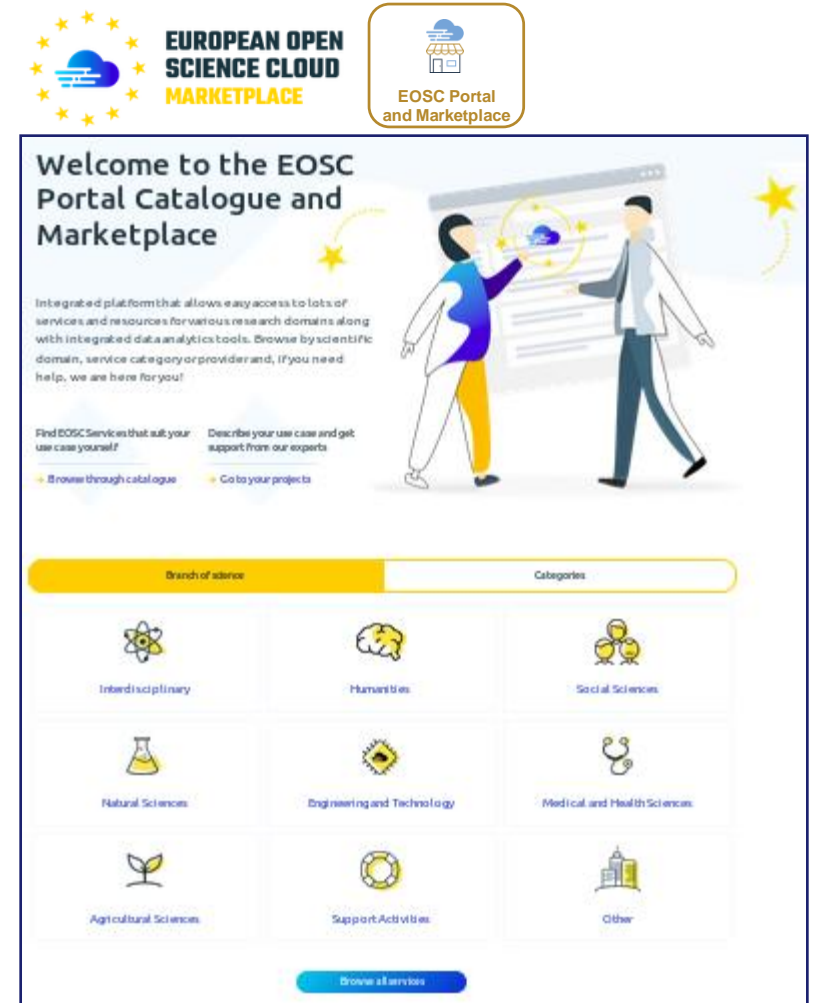


## Join as provider

- *Deliver open services*
- *Production quality*
- *Increase capacity*
- *Benefit from federation*
- *Reach wider audience*

[contact@mailman.eosc-hub.eu](mailto:contact@mailman.eosc-hub.eu)

<https://eosc-portal.eu/join-provider>



<https://marketplace.eosc-portal.eu/>

**Thank you  
for your attention!**

---

*Questions?*

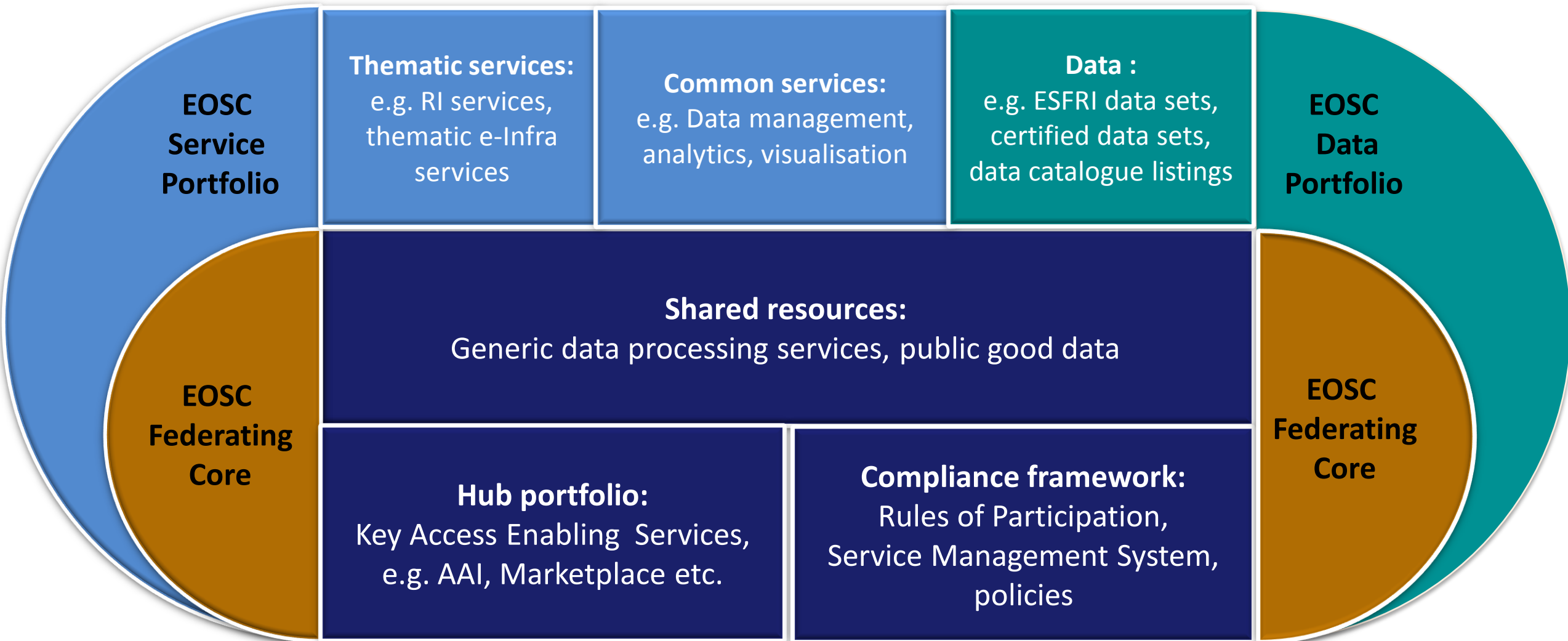


**EOSC-hub**

 [eosc-hub.eu](https://eosc-hub.eu)  [@EOSC\\_eu](https://twitter.com/EOSC_eu)

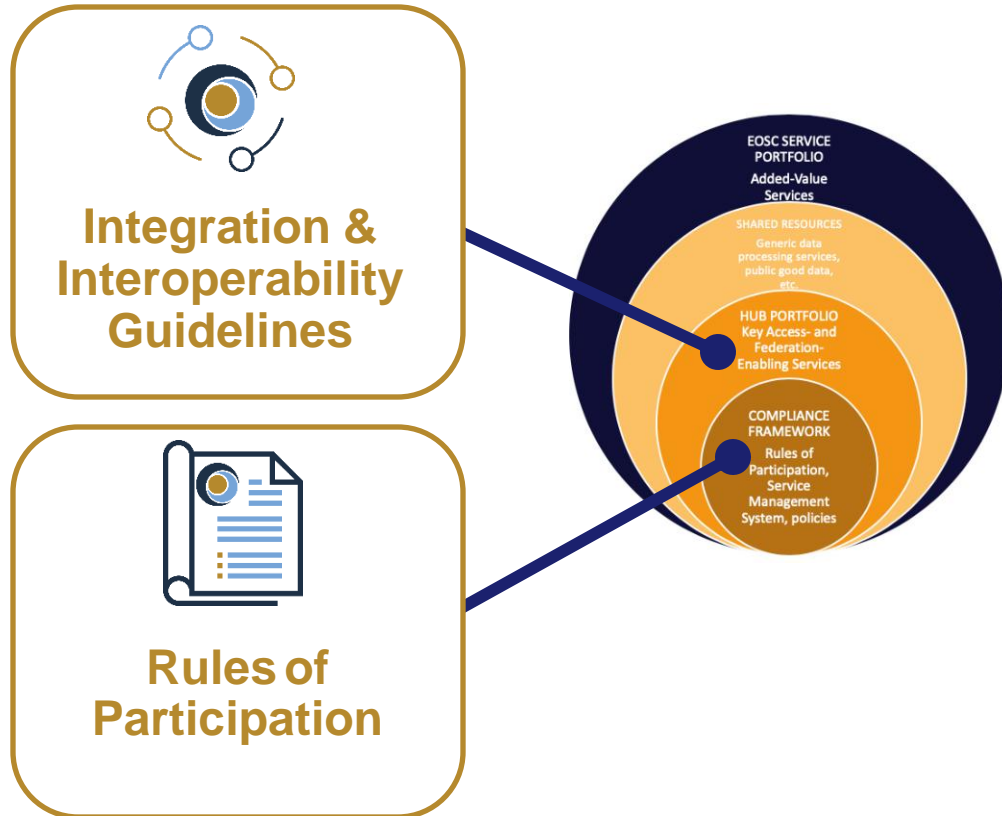


This material by Parties of the EOSC-hub Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



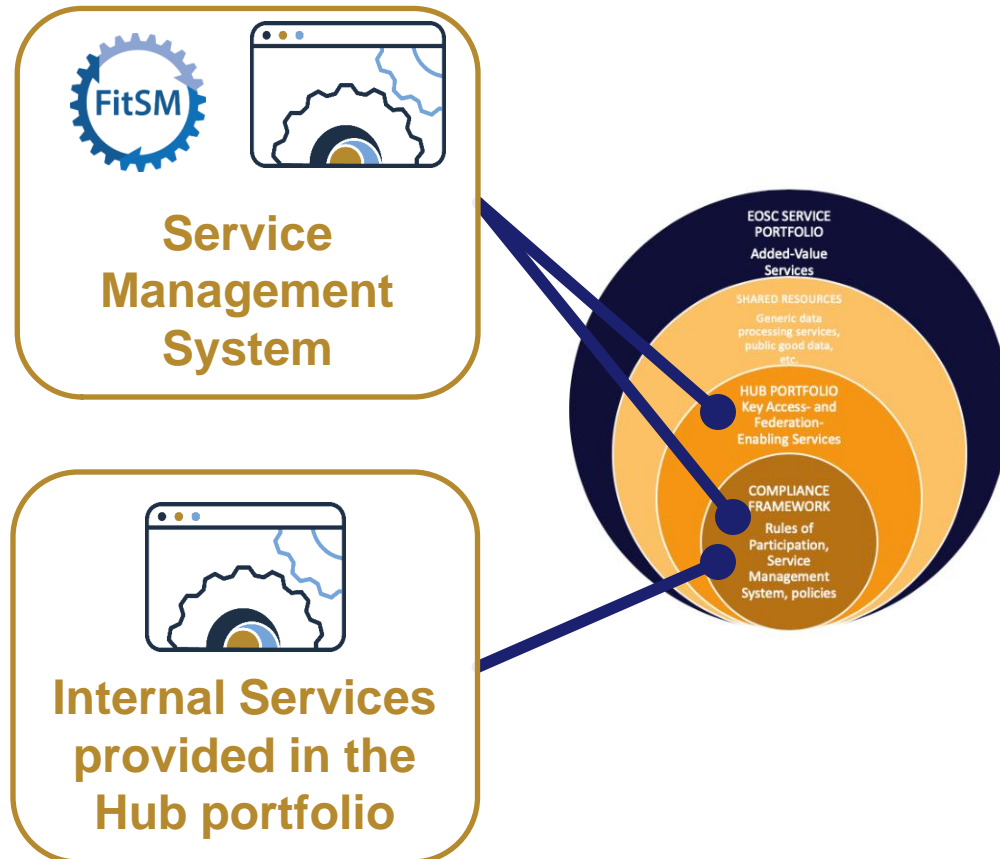


*Remove fragmentation of service provisioning and access to digital services in Europe and beyond, through the **technical integration** and **adoption of standards** for **interoperability** of compute, storage, data and software platforms*

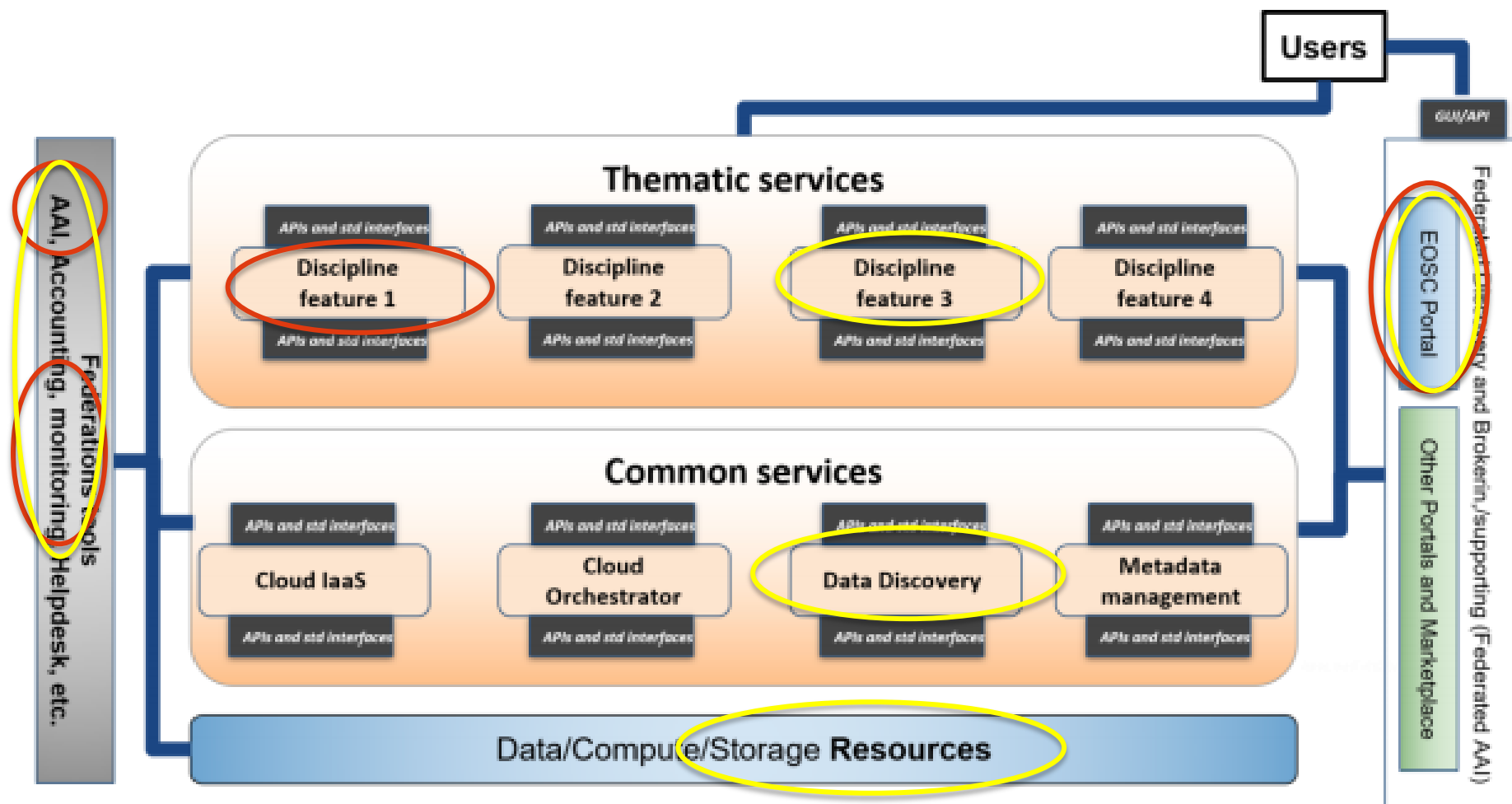


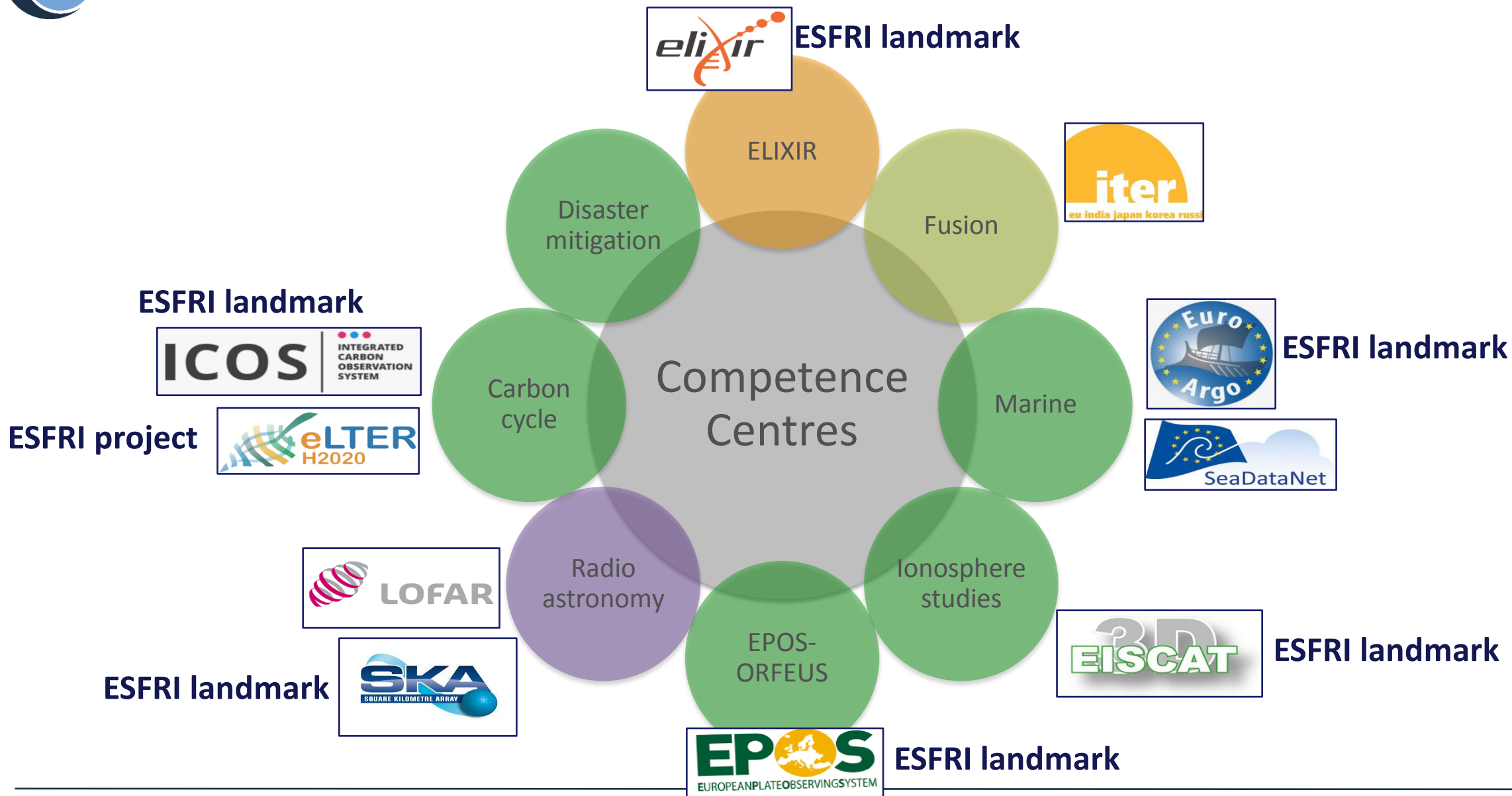
- Technical guidelines to minimise the technical design and development effort needed to integrated services
- Harmonized coherent policies and recommendations for new services
- Enable service composition

*Consolidate e-Infrastructures by expanding **capacity** and **capabilities** and improving service **quality***



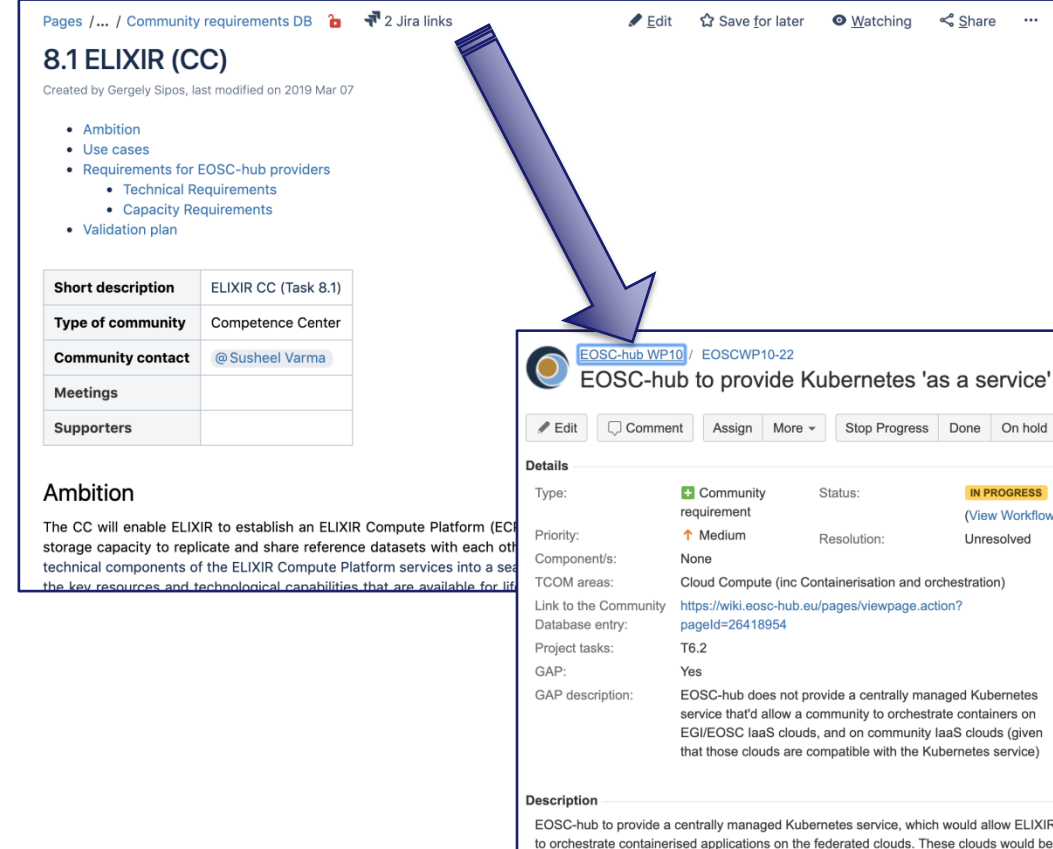
- Ensure robust service delivery
  - Validation of providers
  - Incident Management
  - SLA negotiation
  - Customer Relationship Management
  - Capacity Management
- Tools to integrated services
  - Federated AAI
  - Order Management
  - Monitoring and Accounting
  - Helpdesk etc.





- Who
  - Research community experts
  - Service providers (common services)
  - Technology providers
- What
  - System planning & design
  - Requirement collection
  - Setup proof-of-concepts
  - Dissemination and training
  - Service co-creation & bringing new services into the EOSC catalogue

## Community requirements Database & Jira



Pages / ... / Community requirements DB 2 Jira links Edit Save for later Watching Share

### 8.1 ELIXIR (CC)

Created by Gergely Sipos, last modified on 2019 Mar 07

- Ambition
- Use cases
- Requirements for EOSC-hub providers
  - Technical Requirements
  - Capacity Requirements
- Validation plan

Short description	ELIXIR CC (Task 8.1)
Type of community	Competence Center
Community contact	@Susheel Varma
Meetings	
Supporters	

**Ambition**

The CC will enable ELIXIR to establish an ELIXIR Compute Platform (ECP) with storage capacity to replicate and share reference datasets with each other. The technical components of the ELIXIR Compute Platform services into a set of the key resources and technological capabilities that are available for life science research.

**Details**

Type: Community requirement Status: **IN PROGRESS** (View Workflow)

Priority: Medium Resolution: Unresolved

Component/s: None

TCOM areas: Cloud Compute (inc Containerisation and orchestration)

Link to the Community Database entry: <https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=26418954>

Project tasks: T6.2

GAP: Yes

GAP description: EOSC-hub does not provide a centrally managed Kubernetes service that'd allow a community to orchestrate containers on EGI/EOSC IaaS clouds, and on community IaaS clouds (given that those clouds are compatible with the Kubernetes service)

**Description**

EOSC-hub to provide a centrally managed Kubernetes service, which would allow ELIXIR to orchestrate containerised applications on the federated clouds. These clouds would be