www.egi.eu
@EGI_eInfra

Accessing EGI cloud resources

Tiziana Ferrari, EGI Foundation

PaNOSC Annual Meeting, 2019

The work of the EGI Foundation
is partly funded by the European Commission under H2020 Framework Programme

EGI is a federation of > 250 computing and data centres spread across Europe and the rest of the world.

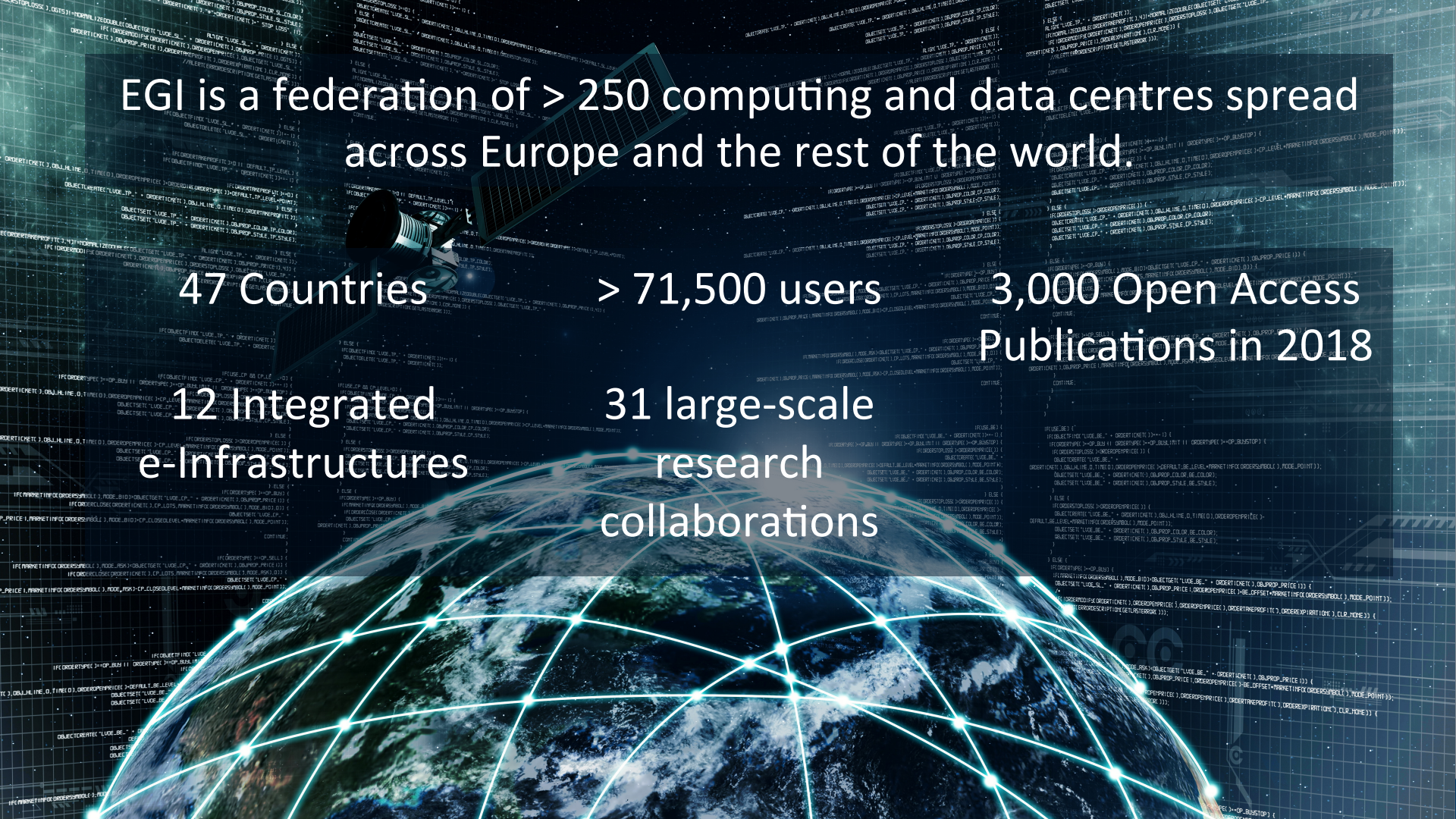
47 Countries

> 71,500 users

3,000 Open Access Publications in 2018

12 Integrated e-Infrastructures

31 large-scale research collaborations



The EGI Federation (Sep 2019)

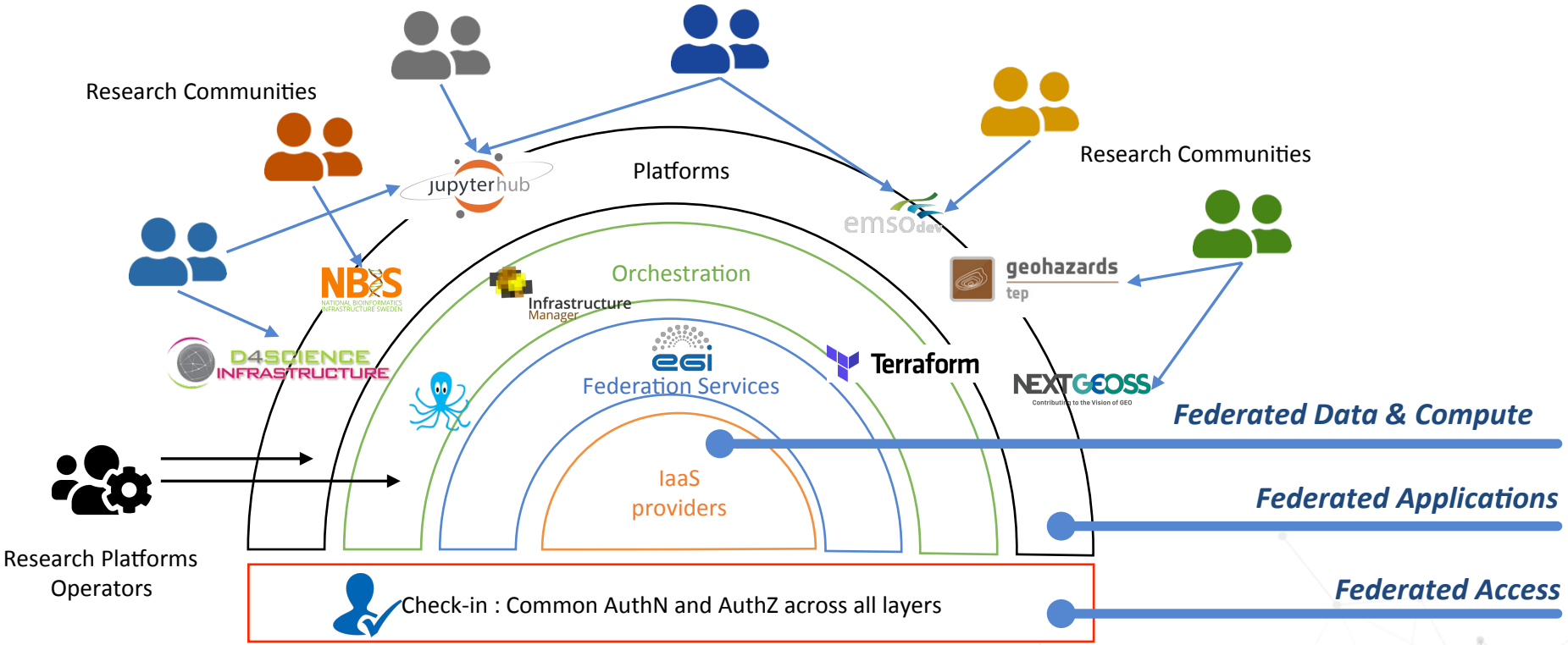
5.0 Billion CPU core wall time / year

> 1 Million computing cores in 2019

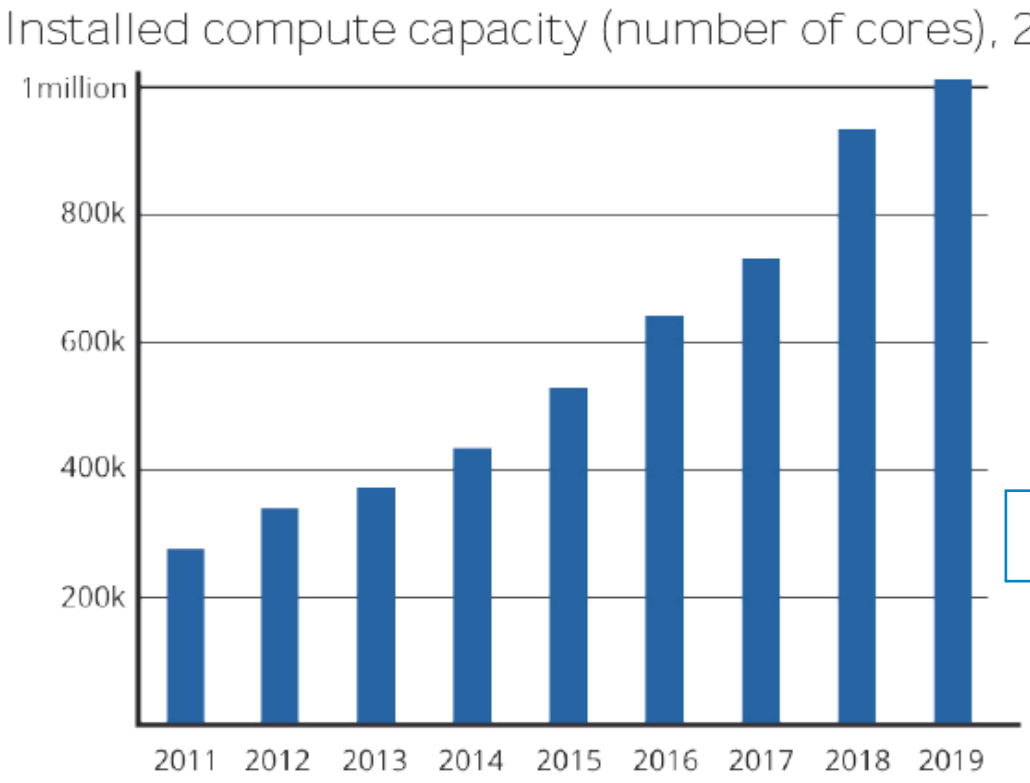
> 740 PB disk & tape

2,915 service end-points

Data-centric Scientific Computing



Federated Compute Capacity 2011-2019



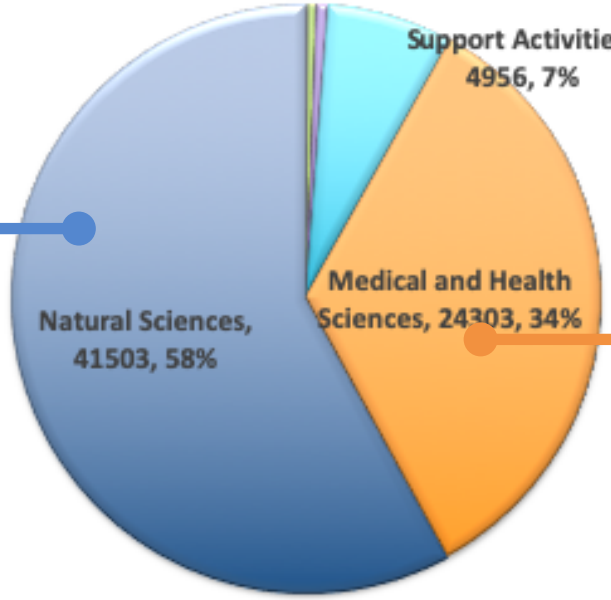
The number of cores installed on the data centres of the EGI Federation has consistently increased since the current measurement system was implemented in 2011, during the EGI-InSPIRE project.

The federation crossed the 1,000,000 cores milestone earlier in 2019.

EGI User Statistics (Sep 2019, +16% in 2018)

Physical Sciences > 41500 users
LHC, BELLE, ICECUBE,
LOFAR, VIRGO, CTA, XENON,
LSST, SKA, KM3NET, ..

Of which
Biological Sciences > 750 users
LifeWatch
Earth Science > 16700
EISCAT, EMSO
Chemical sciences > 550



Medical & Health Sciences > 24300 users

Bioinformatics
Neuroimaging
Structural Biology
Hematology

> 20 Cloud providers

100%IT

RECAS BARI

IFIN-HH

LIP

INFN Sezione di Catania

IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

i3M Instituto de Instrumentación para Imagen Molecular

metacentrum cesnet

grnet

CESGA Centro de Supercomputación de Galicia

IFCA Instituto de Física de Cantabria

GWDG

BEgrid The Belgian Grid for Research

CYFRONET

Institute of Informatics Slovak Academy of Sciences

INFN Sezione di Padova

TÜBİTAK ULAKBİM

JÜLICH FORSCHUNGSZENTRUM

Fraunhofer SCAI

Université de Lille

New!

eodc

CloudFerro

Finalising integration!

Scientific Disciplines/Cloud

Scientific Discipline Cloud — Elapsed time * Number of Processors (hours) by Scientific Discipline and Year

Scientific Discipline	2015	2016	2017	2018	Total	Percent
Other	0	0	0	2,771	2,771	0%
Medical and Health Sciences	1,305	1,295,248	1,190,800	2,206,120	4,693,473	5.34%
Engineering and Technology	4,059	863,467	2,256,929	2,277,126	5,401,581	6.15%
Humanities	2,632,434	3,201,538	1,313,777	1,248,065	8,395,814	9.56%
Support Activities	3,409,107	8,470,420	11,173,822	9,212,083	32,265,431	36.73%
Natural Sciences	9,669,539	7,403,863	7,977,551	12,025,093	37,076,047	42.21%
Total	15,716,444	21,234,536	23,912,880	26,971,257	87,835,117	
Percent	17.89%	24.18%	27.22%	30.71%		

1 - 6 of 6 results ‹ 1 › Number of rows per page 30

Humanities: **CLARIN** (2018), **DARIAH** (+28%, 2017),
Environment: **EMSO** (2018), **LifeWatch** (+45%, 2015)
BIOISI (+45%, 2016), **BIOMED** (+575%, 2016), **CHIPSTER** (+256%, 2016), **GEOHAZARDS** (+19%, 2016), **OpenCOAST** (2018), **EXTRAS** (2018),

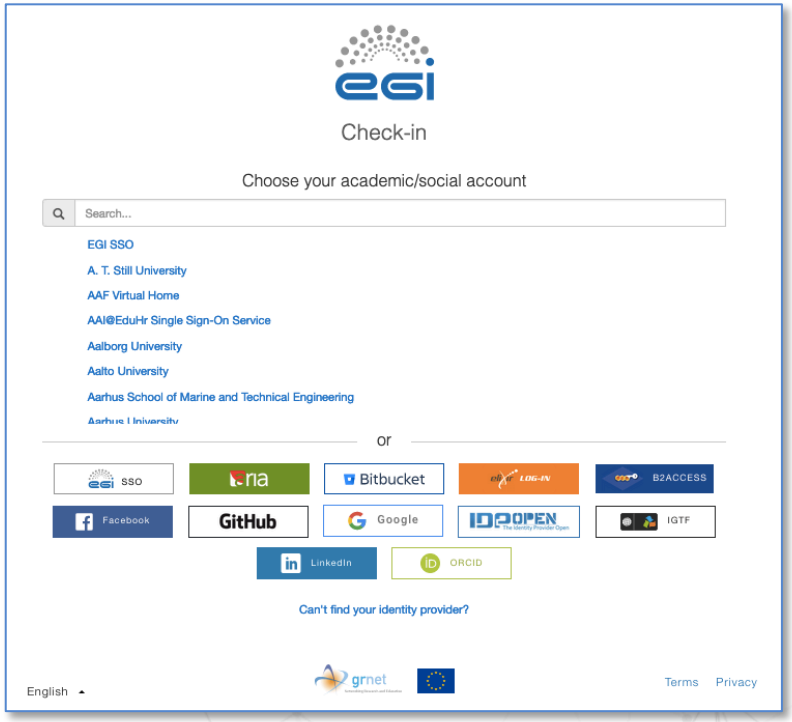
685
cloud virtual appliances
requested through AppDB in 2018

Access and capacity allocation

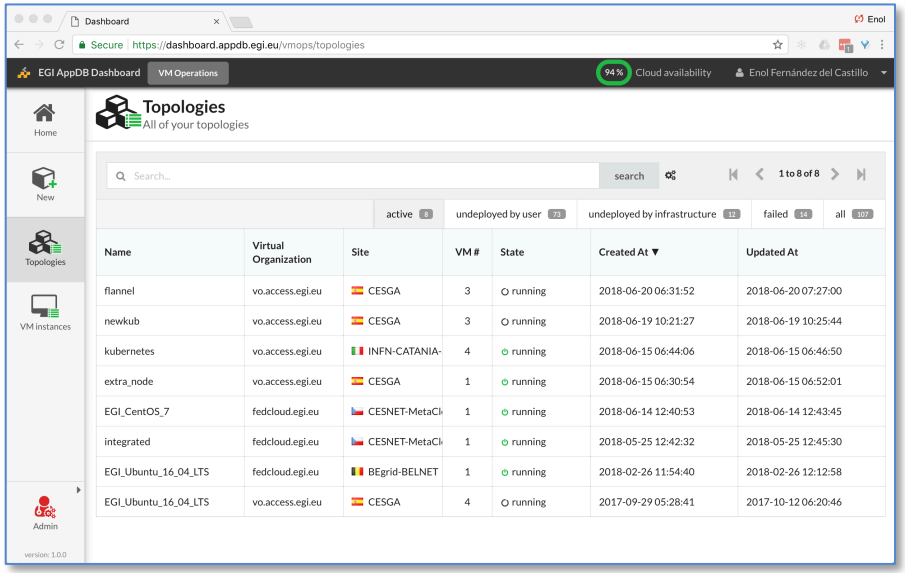
VO	Community VOs
Membership requirements	Access via Check-in, EGI AAI Proxy
AppDB VMOps	Depends on VO
CLI/API access	✅
Membership duration	1 year renewable membership
Resource limits	Specified in the SLA: opportunistic, pledged, time based
Access Policy supported	National sponsored access or pay for use
Available providers	As per SLA

New: Check-in & EGI Cloud

- **Check-in** provides**:**
  - **Single Sign-On** through eduGAIN, social media and other institutional or community-managed identity providers
  - Harmonised **authorisation information**, aggregated from multiple sources
  - **OpenID Connect** allowing web and non-web access to services
- **Check-in** cloud access:
  - **Native support at all the EGI Cloud layers** (IaaS providers, IaaS Orchestration, AppDB VMOps) and at EGI services/platforms running on top: Notebooks, Containers
  - FedCloud client https://aai.egi.eu/fedcloud for easily getting individual tokens for CLI/API access

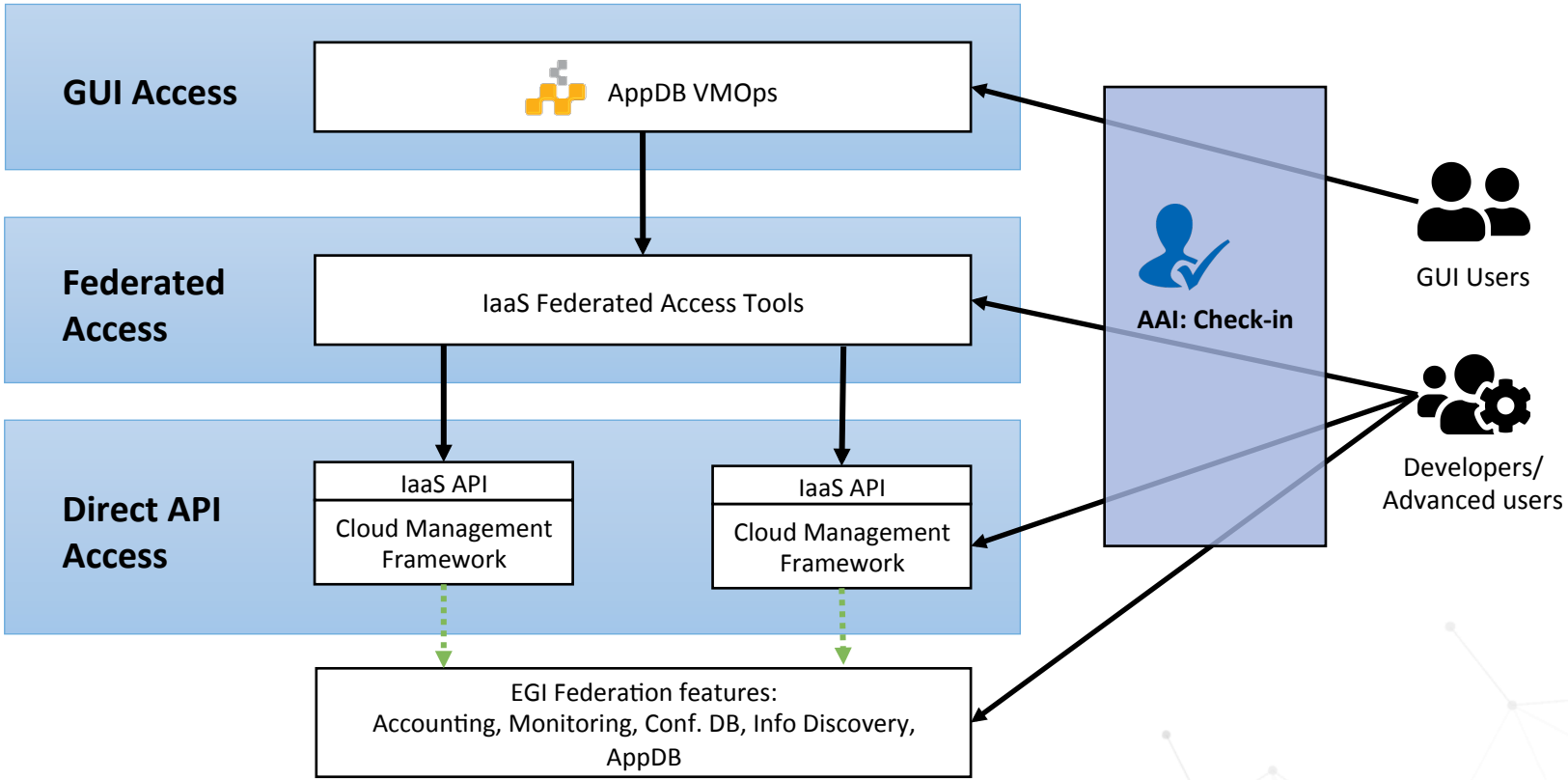


AppDB VMOps - User Friendly GUI

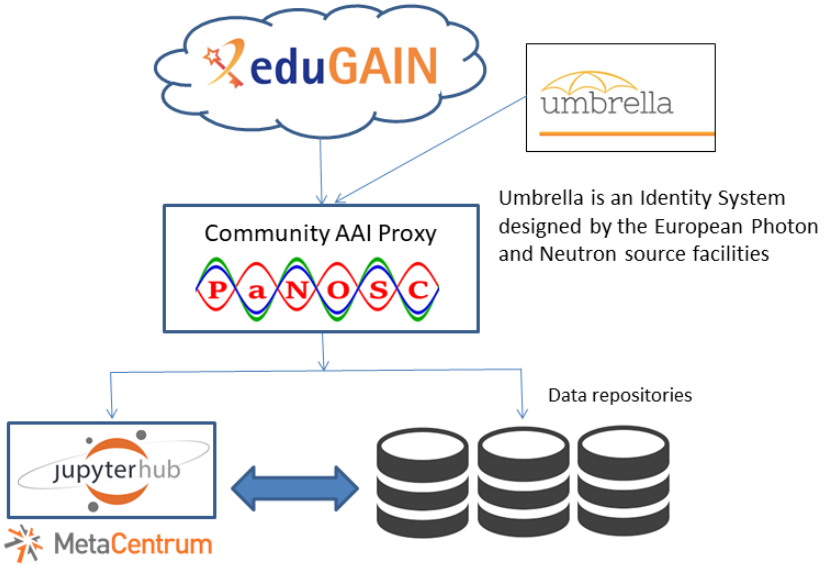


- Single Web dashboard to manage VMs in the federation
  - Point-and-click solution to create new VMs
- Integrated with:
  - Check-in, discovery, VM catalogue, monitoring
- Powered by Infrastructure Manager

Interfaces and Check-in



Open software and solutions



- ESRF (France)
- ILL (France)
- ESS (Sweden)
- ELI-DC (Belgium)
- XFEL (Germany)
- CERIC-ERIC (Italy)
- EGI Foundation (Netherlands)

dCache ONEDATA FTS EGI

Notebooks
Federated Cloud

PaNOSC / Resource overview

CESNET is responsible for supporting the increasing PaNOSC computing needs by providing and technically supporting a cloud compute infrastructure offering Cloud compute capacity according to the following:

- PY1: 1000 Hours of special server (64 cores/1 TB Ram)
- PY2: 2000 Hours of special server (64 cores/1 TB Ram)
- PY3: 4000 Hours of special server (64 cores/1 TB Ram)
- PY4: 10000 Hours of special server (64 cores/1 TB Ram)

The compute infrastructure will be complemented by a data archiving facility offering 300 TB of archival storage capacity.

DESY is responsible for supporting the increasing PaNOSC computing needs by providing and technically supporting a cloud compute infrastructure offering Cloud compute capacity according to the following:

- PY1: 20000 Hours of standard server (16 cores/128 GB Ram)
- PY2: 40000 Hours of standard server (16 cores/128 GB Ram)
- PY3: 100000 Hours of standard server (16 cores/128 GB Ram)
- PY4: 200000 Hours of standard server (16 cores/128 GB Ram)

The compute infrastructure will be complemented by a data archiving facility offering 200 TB of storage capacity and an online data facility offering TB.

STFC will be also responsible of configuring, operating and supporting a data storage facility based on the B2SAFE service offering 4 PB space. The data infrastructure will be operated in two phases:

- PY1: setting up of the infrastructure and service enabling by PaNOSC applications (Effort: 8 PM)
- PY2-3-4: operations and support (effort: 3 PM/year)

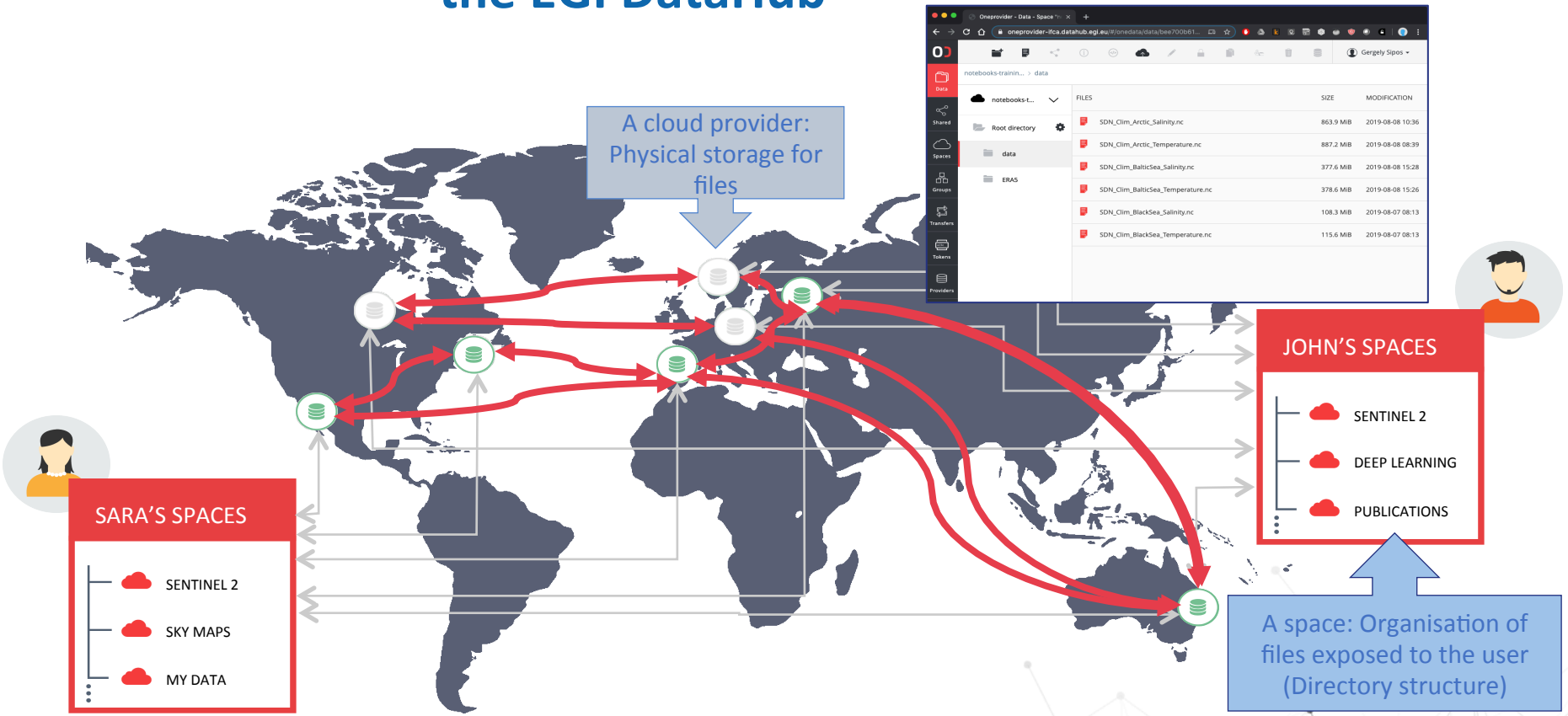
EGI and the European Open Science Cloud

Data, applications and computing infrastructure as integrated solution

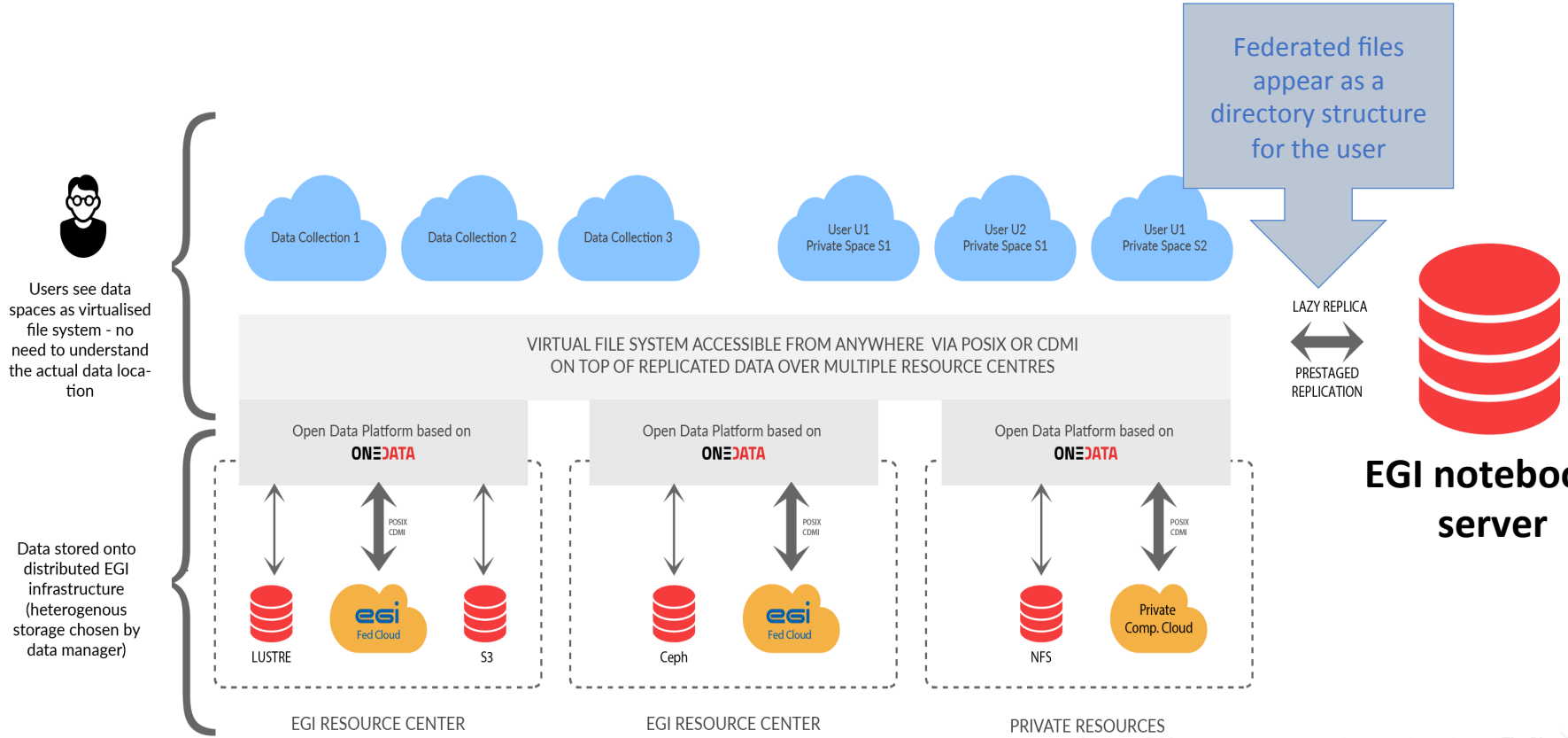
Towards a data-centric compute infrastructure

- Growing demand of Cloud access to HTC, GPU and HPC resources
  - Machine learning adopted by many applications running in the EGI Federation
  - Hybrid provisioning is key
- End-users engaged at the SaaS level
  - Easy to use solutions bring together
    - Federated AAI
    - Federated data discovery and management
    - Federated data analytics capabilities
- Data-centric approach to Cloud
  - Zero download access to large research data holdings e.g Copernicus

Federating data in a multi-cloud with the EGI DataHub

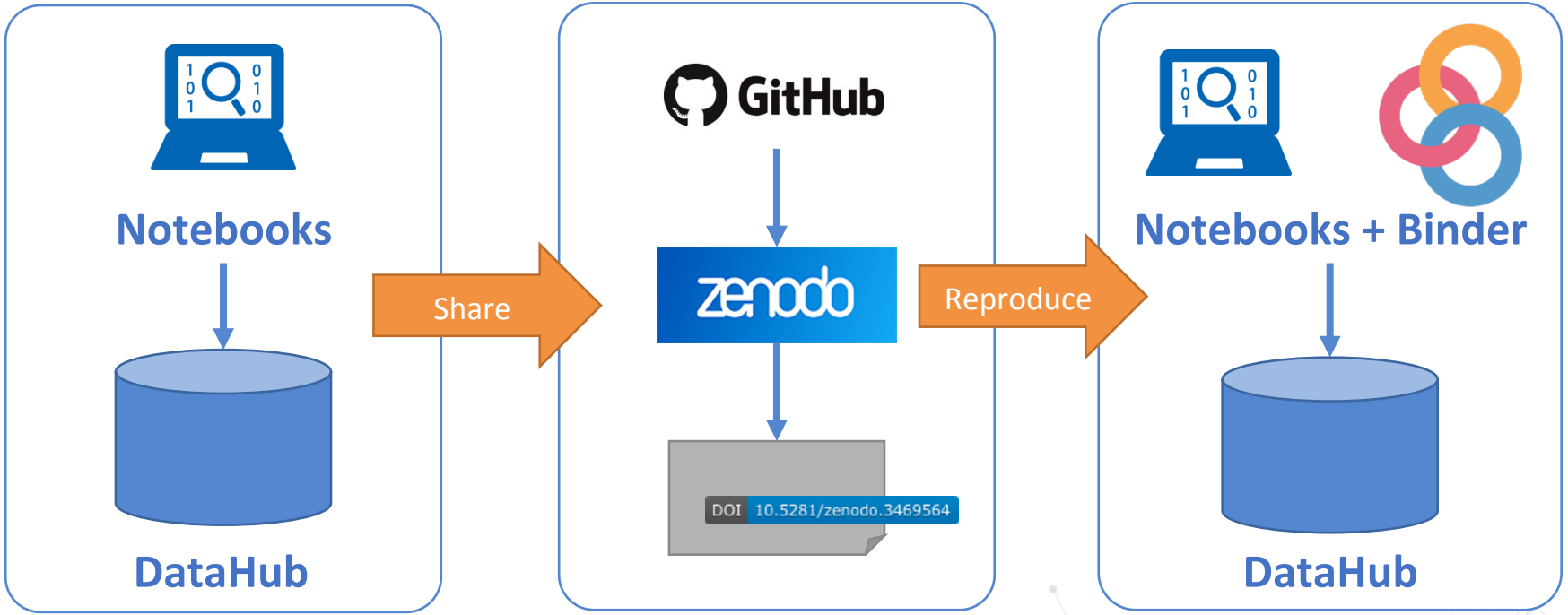


Example: EGI Notebooks



Open science with EGI and OpenAIRE Services

Reproducible and discoverable analysis



EGI towards EOSC

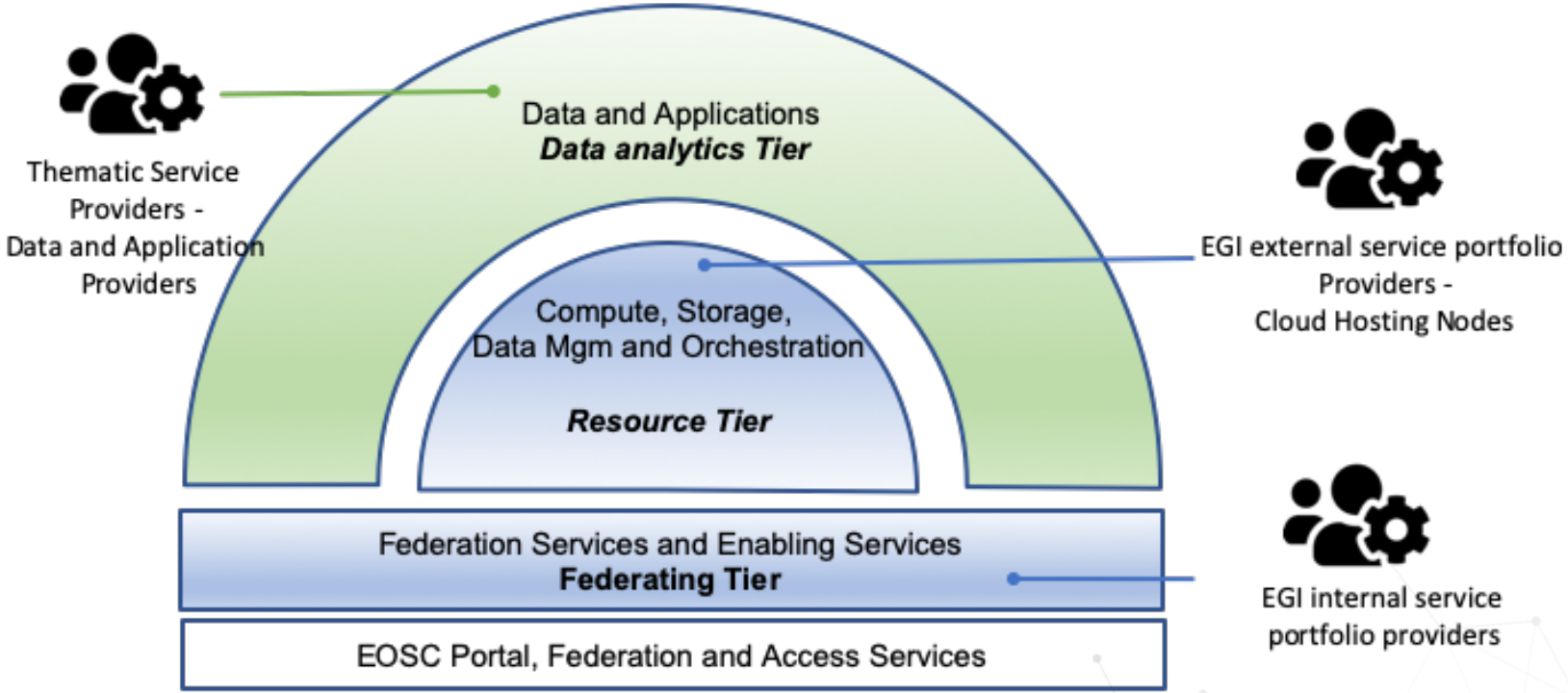
Vision

Computing, storage, data, software and a rich portfolio of research community-specific services for data analytics will be federated and made available for on-demand access in the European Open Science Cloud through the participation of EGI members and partners.

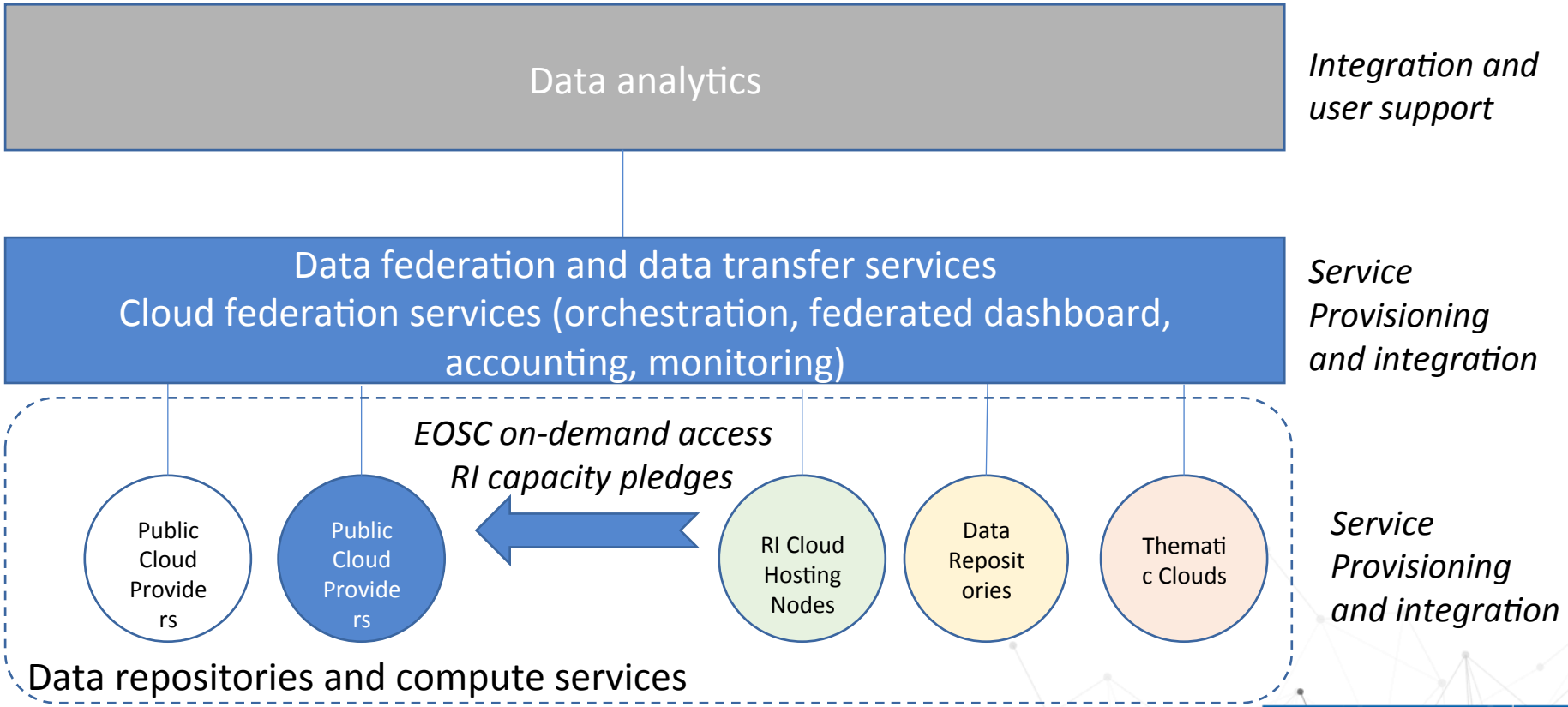
EGI Mission

*To implement the **data processing pillar** of the European Open Science Cloud, by **extending the on-demand capacity and capabilities** of the EGI external service catalogue,*
- leveraging the EGI federation services and the funding, development and operational efforts of EGI participants and partners.

EGI tiered architecture in EOSC

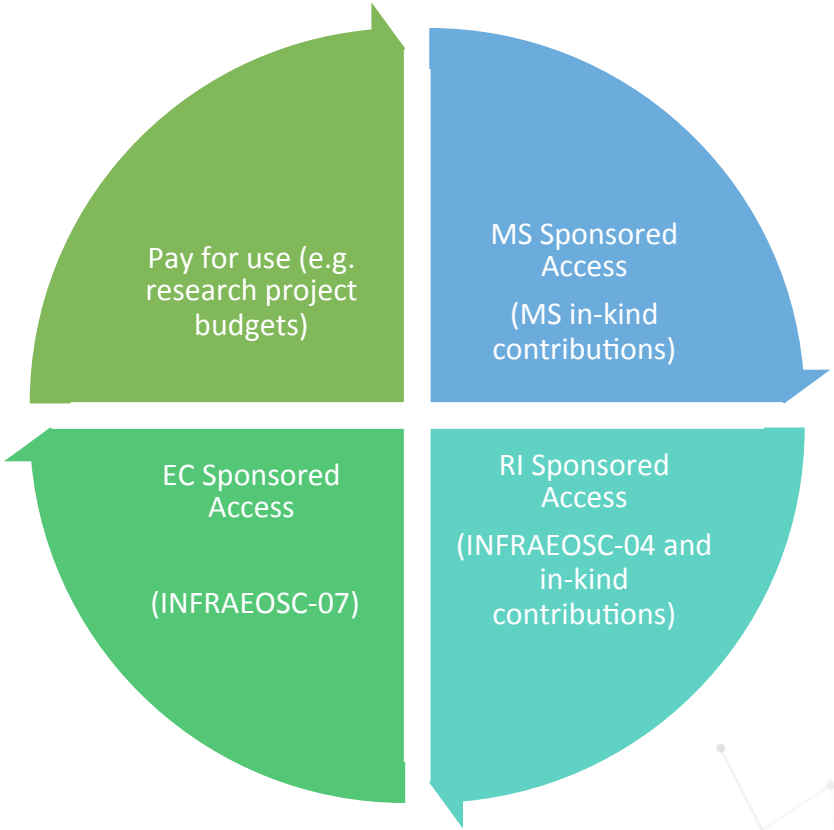


Resource Tier



Resource Tier: possible funding model

Co-programmed partnership



H2020 INFRAEOSC-07 A1

(a1) Distributed and cloud computing resources enabling researchers and other users to process and analyse data in a distributed computing environment. The services should include, but not be limited to, running virtual machines on demand with complete control over computing services, executing compute and data-intensive workloads, analysing large datasets and executing parallel computing tasks, utilizing large amount of processing capacity over long periods of time, sharing resources and enabling collaborative research.

Scope of the call

- Proposals should build on the competences of pan-European e-Infrastructures of diverse domains to ensure multidisciplinary research and synergies with national and regional programmes. The progressive federation of the services and resources under the awarded proposals, together with the progressive connection of ESFRI research infrastructures (INFRAEOSC-04) and thematic clouds developed under other parts of the Horizon 2020 programme, should allow the EOSC Portal to provide a catalogue that increasingly meets the researchers' needs covering the full research life cycle.
- Proposals should foresee dedicated activities for cooperation with the other selected projects and earmark appropriate resources.

Conclusions

- Scientific computing infrastructures are a necessary instrument of scientific excellence and open science in Europe
- EGI: from HTC infrastructure, to hybrid HTC & HPC infrastructure supporting cloud IaaS, PaaS, SaaS access
  - Federated identity provisioning and access
  - Federated data management towards exabyte scale data processing
- European Open Science Cloud
  - Supporting the entire research data lifecycle from production to zero download data exploitation

EGI: Advanced Computing for Research

www.egi.eu
@EGI_eInfra

Thank you
for your attention.

Questions?

The work of the EGI Foundation
is partly funded by the European Commission under H2020 Framework Programme