EGI towards the European Open Science Cloud

Tiziana Ferrari

Technical Director/EGI Foundation

EGI Conference 2017, 9-12 May 2017

EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142

EGI Services

EGI Service Catalogue
Marketplace
Current and new services
Success stories

EGI Service Catalogue

Online catalogue

https://www.egi.eu/services/

Printed publication

http://go.egi.eu/ServiceCatalogueBrochure

External service catalogue for researchers

https://www.egi.eu/services/

Compute

Cloud Compute

Run virtual machines on demand with complete control over computing resources

Cloud Container Compute BETA

Run Docker containers in a lightweight virtualised environment

High-Throughput Compute

Execute thousands of computational tasks to analyse large datasets

Storage and Data

Online Storage

Store, share and access your files and their metadata on a global scale

Archive Storage

Back-up your data for the long term and future use in a secure environment

Data Transfer

Transfer large sets of data from one place to another

Training

FitSM training

Learn how to manage IT services with a pragmatic and lightweight standard

Training infrastructure

Dedicated computing and storage for training and education

Internal service catalogue for EGI Federation members

https://www.egi.eu/internal-services/

Accounting
Track and report the usage of your services

Operations Coordination and Support
Coordinate activities to ensure seamless operations

Community Coordination
A joint approach to user engagement

Communications
Share your successes at a larger scale

Project Management and Planning
A joint approach to planning and management

Helpdesk
Your point of contact to ask for support at EGI

Service Monitoring
Monitor the performance of IT services

Validated Software and Repository
Benefit from a repository of high-quality software validated for the EGI infrastructure

Operational Tools
Integrate resources and operations in a federated ecosystem

Security Coordination
Enhance local security for a safer global infrastructure

ITSM Coordination
Ensures professional service management for EGI IT services

Technical Coordination
Progress and innovation through collaboration

Strategy and Policy Development
One federation, one vision, one strategy

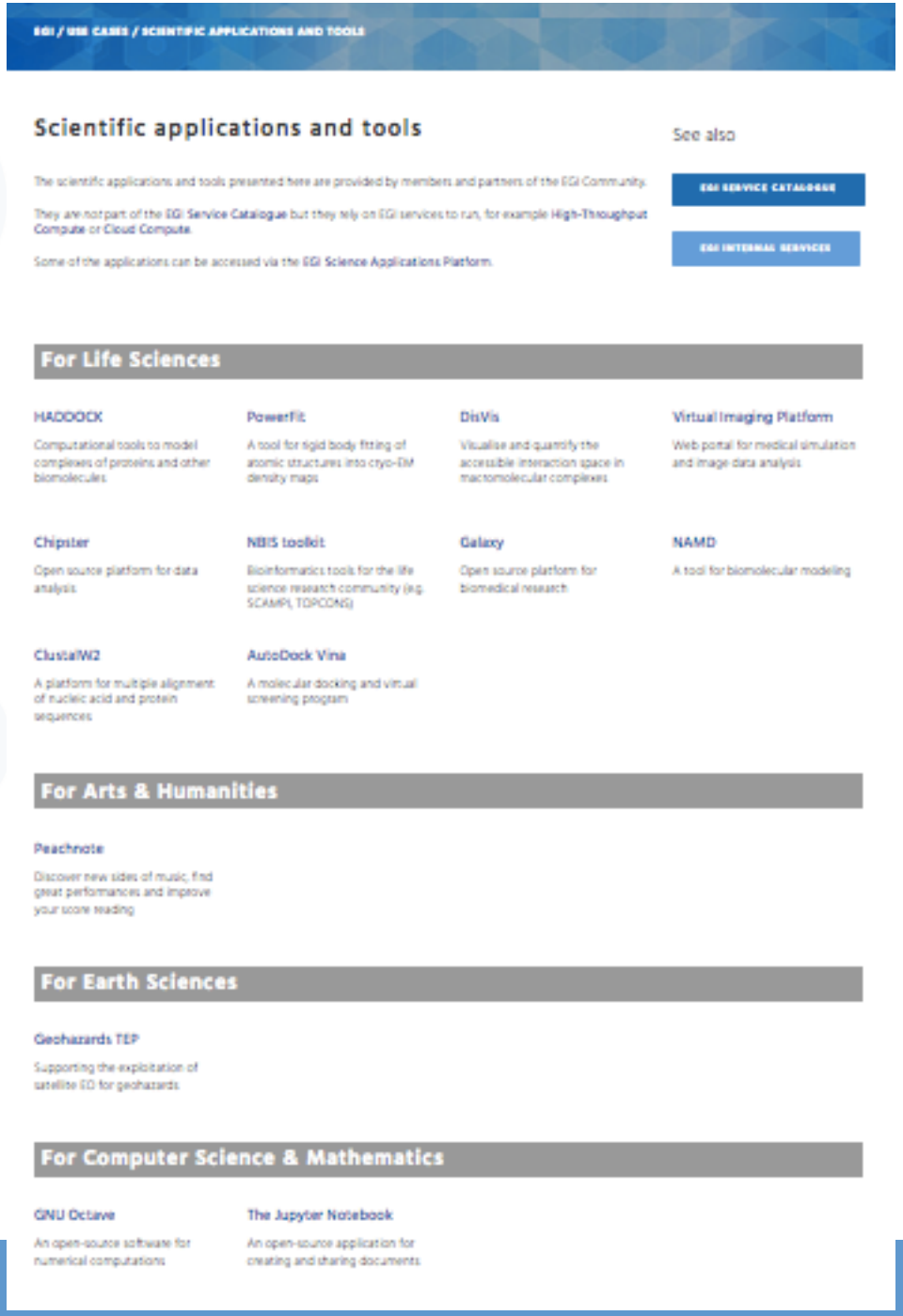
Configuration Database
Manage the configuration information of federated e-infrastructure assets and their functional relations

Collaboration Tools
IT tools for better coordination

EGI-Engage

Thematic services from EGI partners

https://www.egi.eu/use-cases/scientific-applications-tools/



EGI / USE CASES / SCIENTIFIC APPLICATIONS AND TOOLS

Scientific applications and tools

See also

The scientific applications and tools presented here are provided by members and partners of the EGI Community.

They are not part of the EGI Service Catalogue but they rely on EGI services to run, for example High-Throughput Compute or Cloud Compute.

Some of the applications can be accessed via the EGI Science Applications Platform.

EGI SERVICE CATALOGUE

EGI INTERNAL SERVICES

For Life Sciences

HADDOCK
Computational tools to model complexes of proteins and other biomolecules

PowerFit
A tool for rigid body fitting of atomic structures into cryo-EM density maps

DisVis
Visualise and quantify the accessible interaction space in macromolecular complexes

Virtual Imaging Platform
Web portal for medical simulation and image data analysis

Chipster
Open source platform for data analysis

NBIS toolkit
Bioinformatics tools for the life science research community (e.g. SCAMPI, TOPCONS)

Galaxy
Open source platform for biomedical research

NAMD
A tool for biomolecular modeling

ClustalW2
A platform for multiple alignment of nucleic acid and protein sequences

AutoDock Vina
A molecular docking and virtual screening program

For Arts & Humanities

Peachnote
Discover new sides of music, find great performances and improve your score reading

For Earth Sciences

Geohazards TEP
Supporting the exploitation of satellite EO for geohazards

For Computer Science & Mathematics

GNU Octave
An open-source software for numerical computations

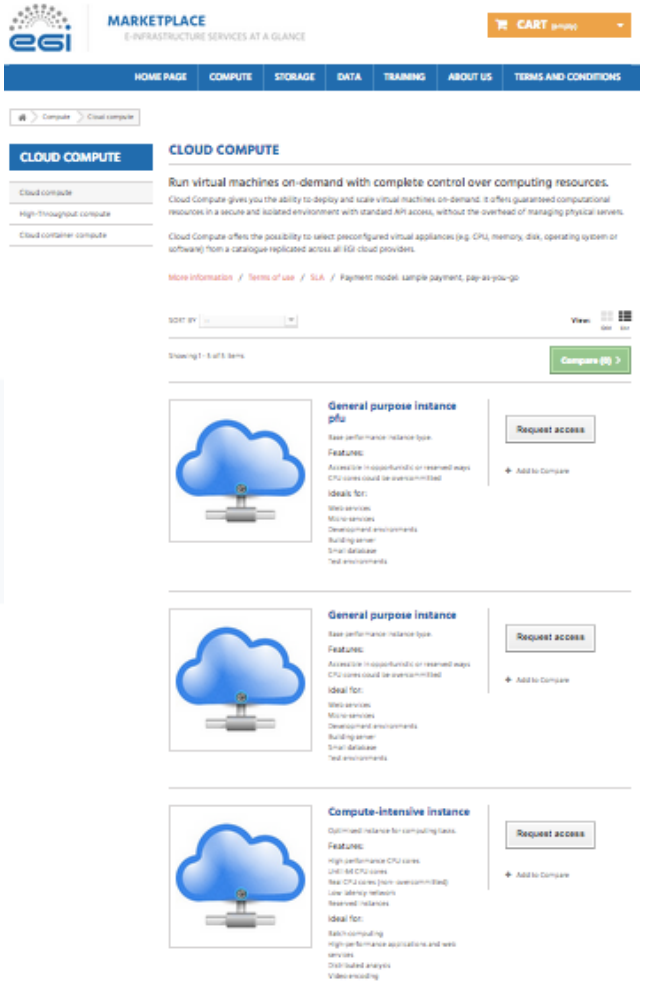
The Jupyter Notebook
An open-source application for creating and sharing documents

Service discoverability and easier access

http://marketplace.egi.eu/

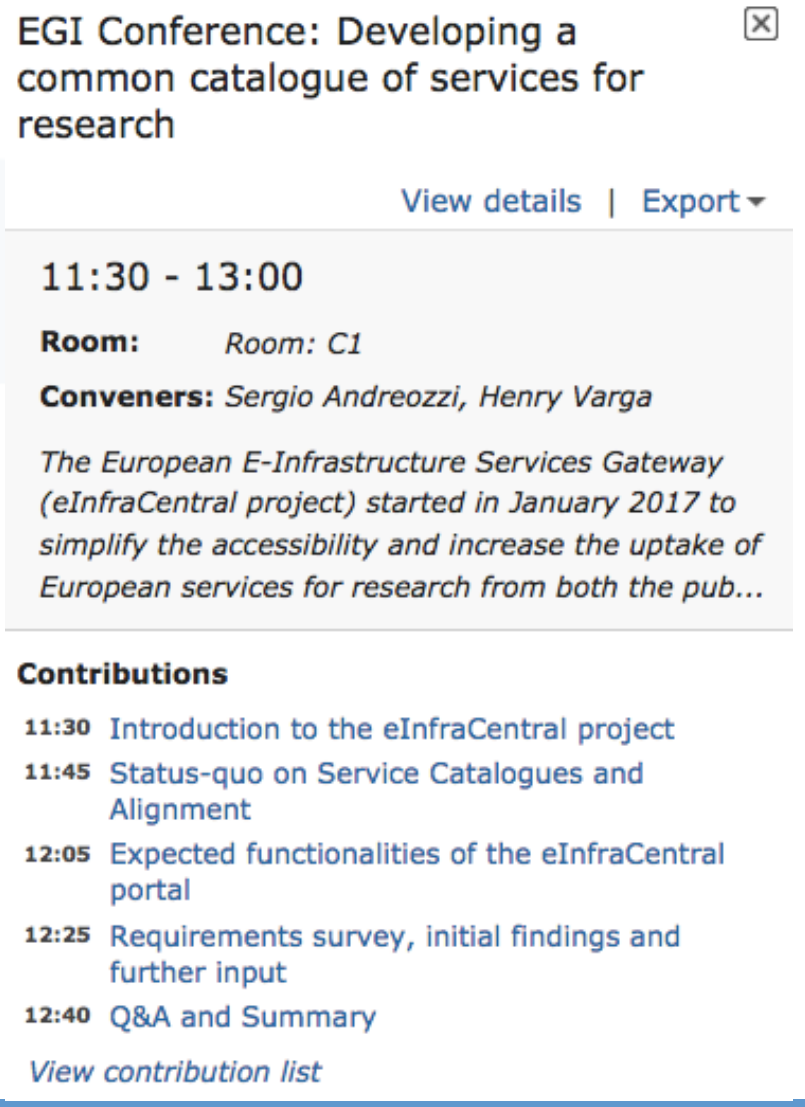
- Discovery and access to local, regional, national services that are open for access, use and reuse
- Visibility of data and thematic services from EGI partners
- More inter-disciplinary research
- Programmatic access to multiple services
  - Compute, storage, orchestrators, data, thematic services
- Different access policies, including pay for use

→ for EGI members, partners, integrated e-Infrastructures and collaborating organizations



Towards a joint catalogue: eInfraCentral

- Wed 11 May, 11:30 AM
- Participate to survey by 31 May!
  - https://www.surveymonkey.com/r/e-infra_service_catalogue



EGI Conference: Developing a common catalogue of services for research

View details | Export

11:30 - 13:00

Room: *Room: C1*

Conveners: *Sergio Andreozzi, Henry Varga*

The European E-Infrastructure Services Gateway (eInfraCentral project) started in January 2017 to simplify the accessibility and increase the uptake of European services for research from both the pub...

Contributions

11:30 Introduction to the eInfraCentral project
11:45 Status-quo on Service Catalogues and Alignment
12:05 Expected functionalities of the eInfraCentral portal
12:25 Requirements survey, initial findings and further input
12:40 Q&A and Summary

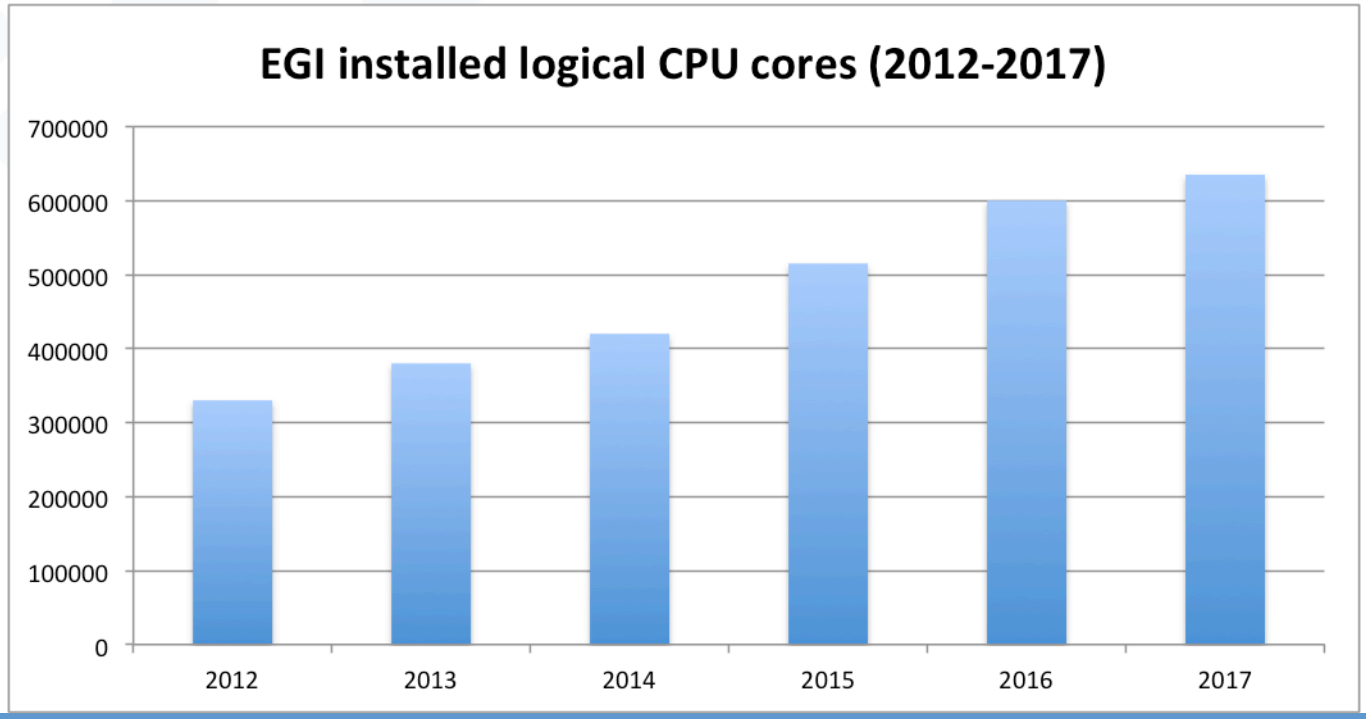
View contribution list

EGI Services

High Throughput Compute

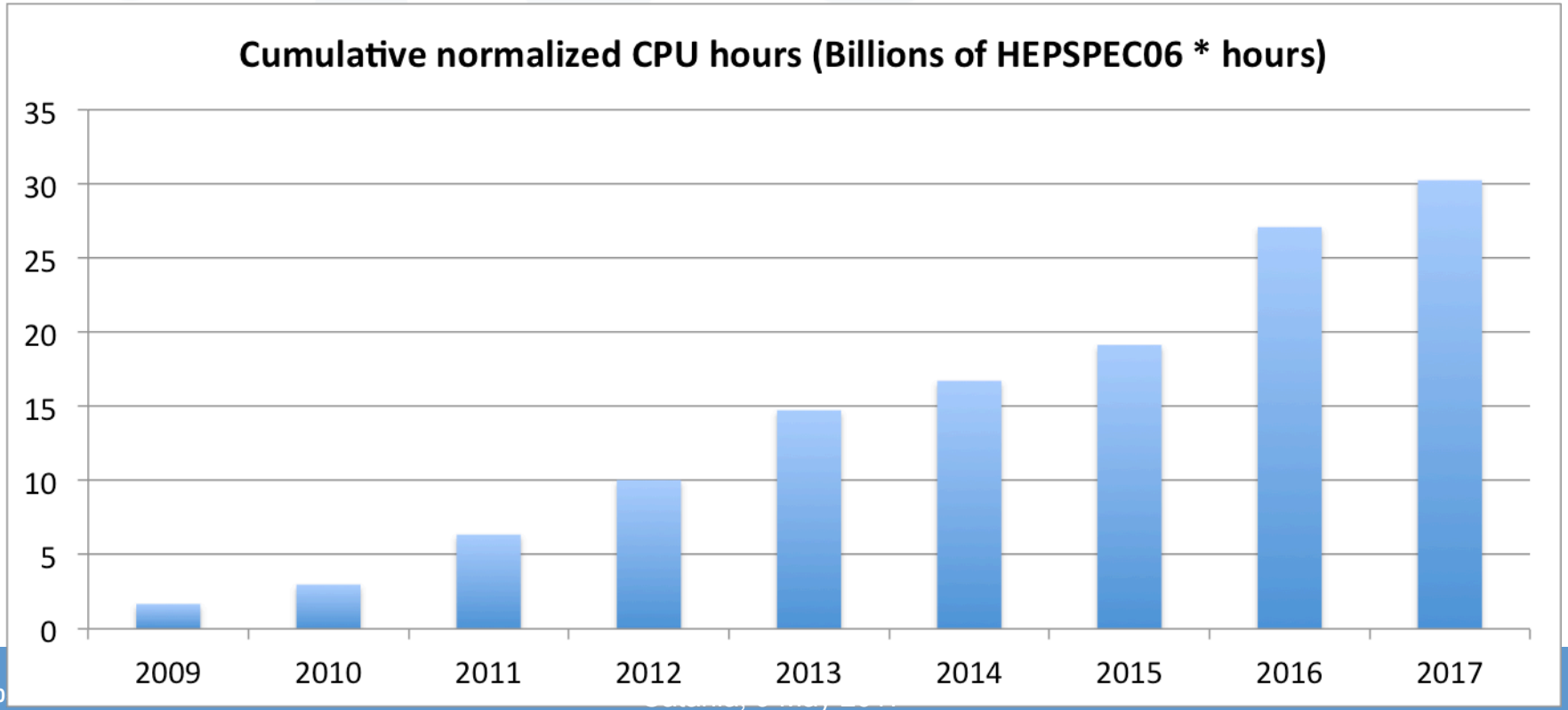
High Throughput data analysis platform (HTC)

	Logical cores	Computing power (HEP_SPEC06)	Online storage (PB)	Nearline storage (PB)
March 2017	635,000	6,680,000	268	272
Increase from March 2016	+ 5%	+ 15%	+ 3%	+ 13%



HTC capacity consumption

Metric	In the last 12 months	From the previous 12 months
Normalized elapsed time	30.2 Billions wall clock time hours* HEP_SPEC06	+10.4 %
Number of jobs	592 Million	-2.7%
Num. multicore jobs	53 Millions	+32%



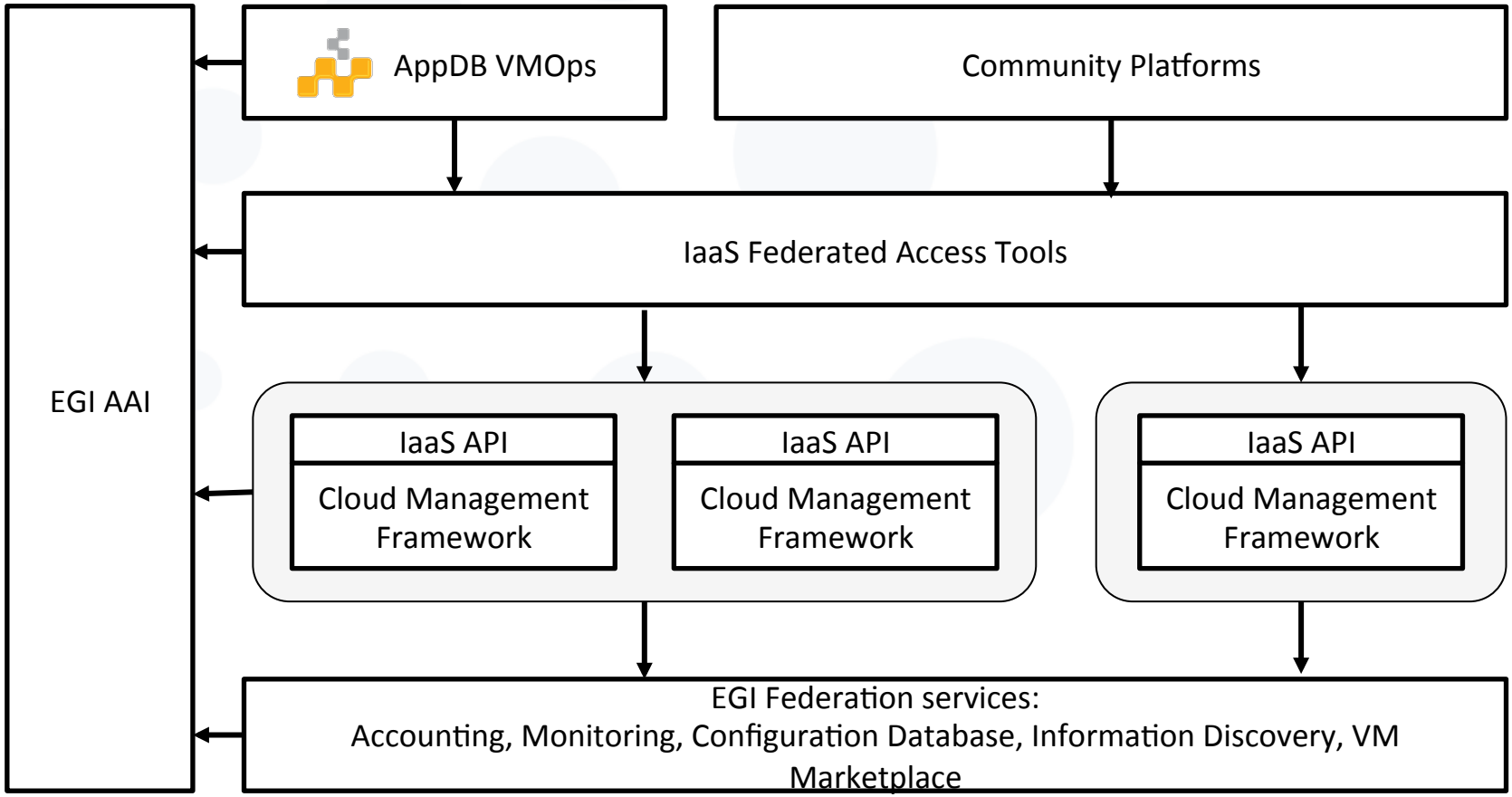
EGI Services

Cloud Compute

Cloud Compute

- The EGI Federated Cloud is a multi-national cloud system that integrates community, private and/or public clouds into a scalable computing platform for research.
- The Federation pools IaaS, PaaS and SaaS services from a heterogeneous set of cloud providers using a single authentication and authorization framework that allows the portability of workloads across multiple providers and enable bringing computing to data.

EGI Federated Cloud: architecture evolution

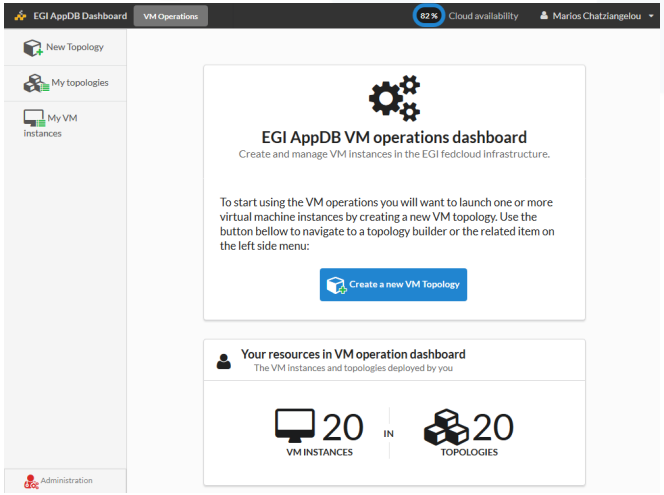


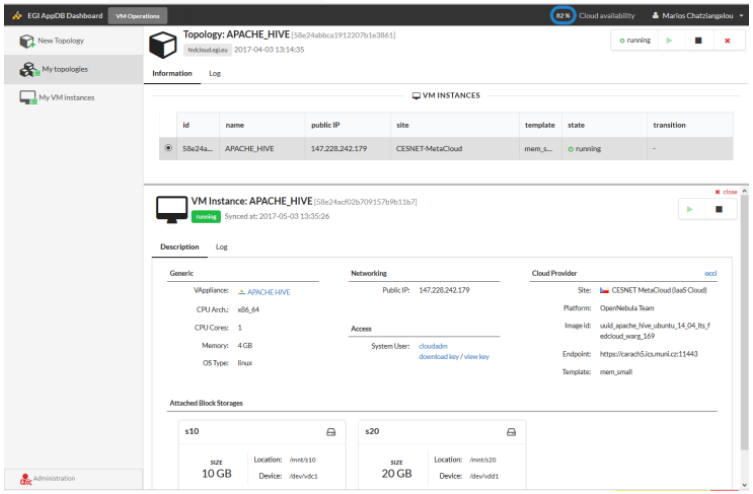
OCCI and OpenStack Realms

AppDB VMOps Dashboard

https://dashboard.appdb.egi.eu/vmops

- Unified resource management on the Federated Cloud
- Complete EGI CheckIn integration
- Personalized, wizard-like, topology builder:
  - Manage multiple-VM topologies with associated storage and contextualization
  - Fine grained control of VMs within topologies

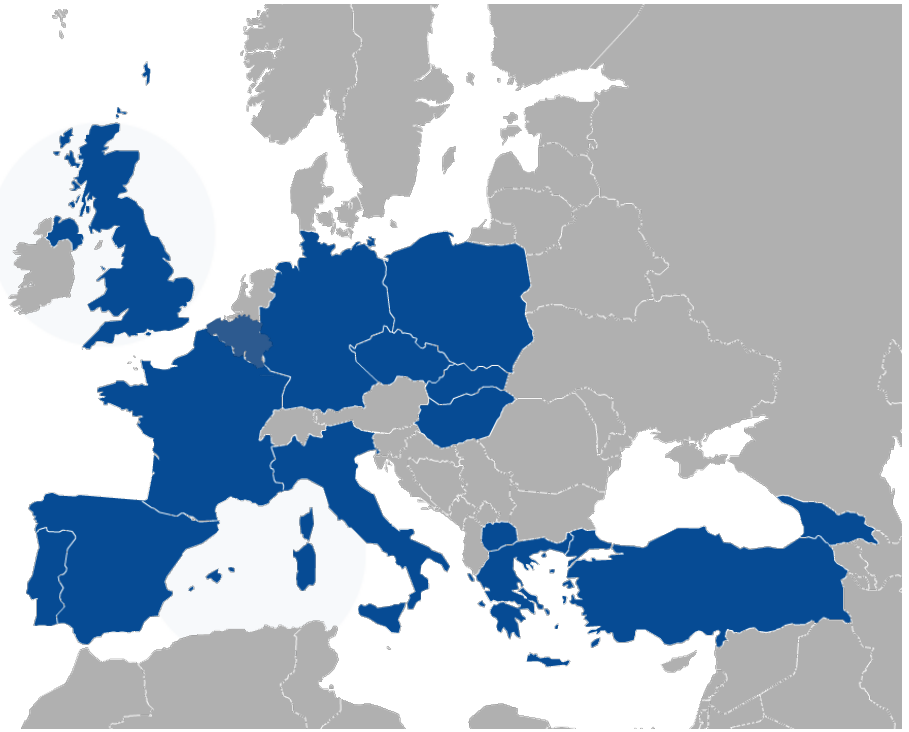




EGI Federated Cloud participants

As of	Number of resource centres
2016	24 (+2 integrations in progress)

Cloud management system	Number of resource centres
OpenStack	17
Open Nebula	6
Synnefo	1



Federated CPU cores	Increase from 2016
6600	+10%

Cloud Roadmap

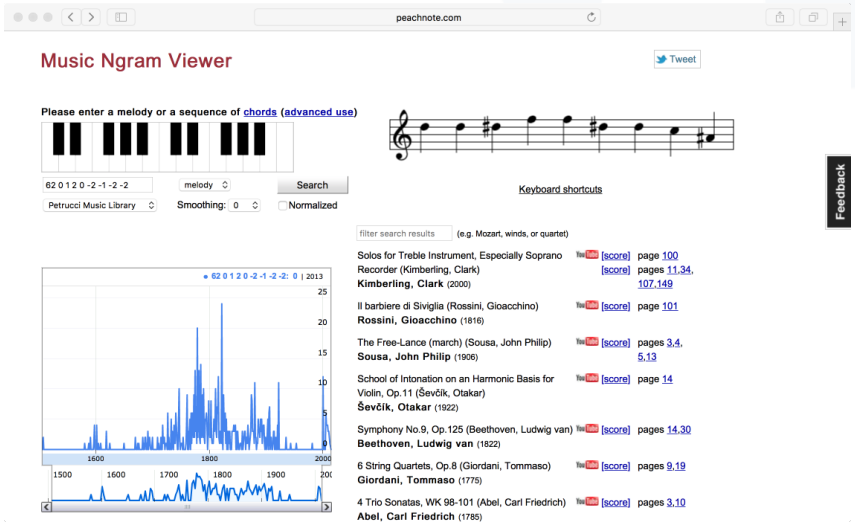
- TCB-Cloud established in Q3 2016 to define/maintained the Technical Roadmap and Architecture of the Federated Cloud

- Highlights:
  - Improve reliability of federation with enhancement of monitoring and integration components (AAI, information discovery, accounting, VM marketplace)
  - Promote IaaS Federated Access Tools Layer as interoperation mechanism
  - Expand the portfolio: PaaS/SaaS

Peachnote

http://www.peachnote.com/

- Music score search engine for users to discover performances and improve their score reading from a library of more than 380K scores of over 115K works
  - +7000 registered users
  - On 2016: +64M page views, +23M score views



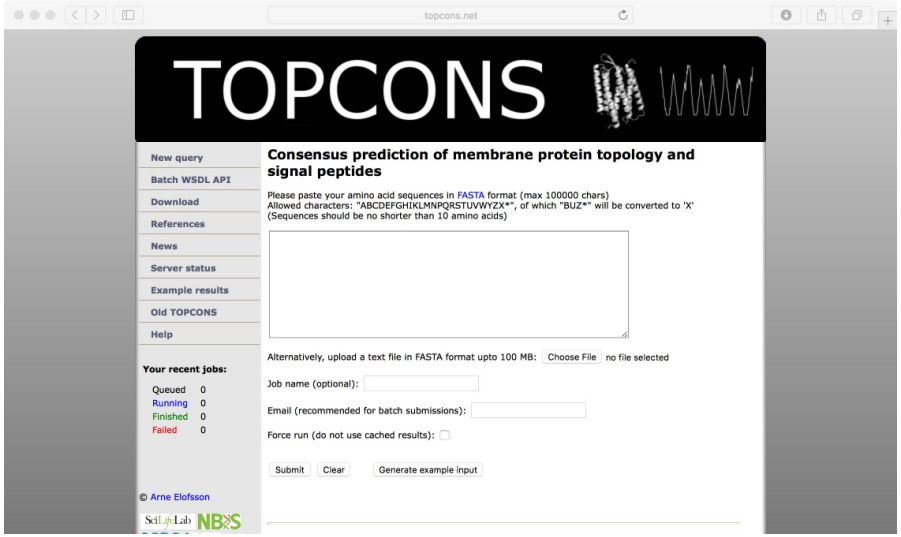
EGI provides:

- 104 CPU cores
- 162 GB RAM
- 8TiB Storage

NBIS/BILS

http://nbis.se

- NBIS is a distributed bioinformatics infrastructure supporting life sciences in Sweden
- 8 web-accessible thematic services running with EGI Federated Cloud resources
  - ~8K users from 73 countries



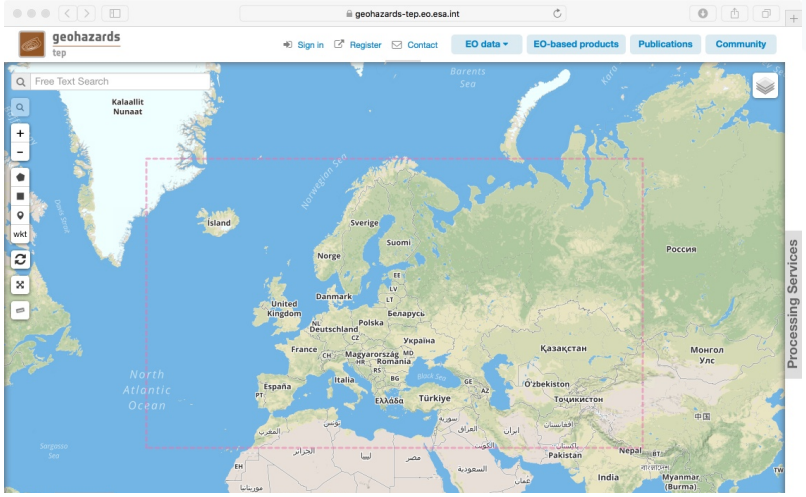
EGI provides:

- 172 CPU cores
- 400 GB RAM
- 9TiB Storage

GeoHazards Thematic Platform

https://geohazards-tep.eo.esa.int/geobrowser

- Earth Observation Data processing services for geohazards analysis and monitoring based on the exploitation of satellite data
- User communities
  - CEOS Seismic Hazards Pilot, a three-year demonstration project of the Committee on Earth Observation Satellites (CEOS)
  - EPOS (European Plate Observing System)
  - +500 registered users, ~50% resources on EGI FedCloud



EGI provides:

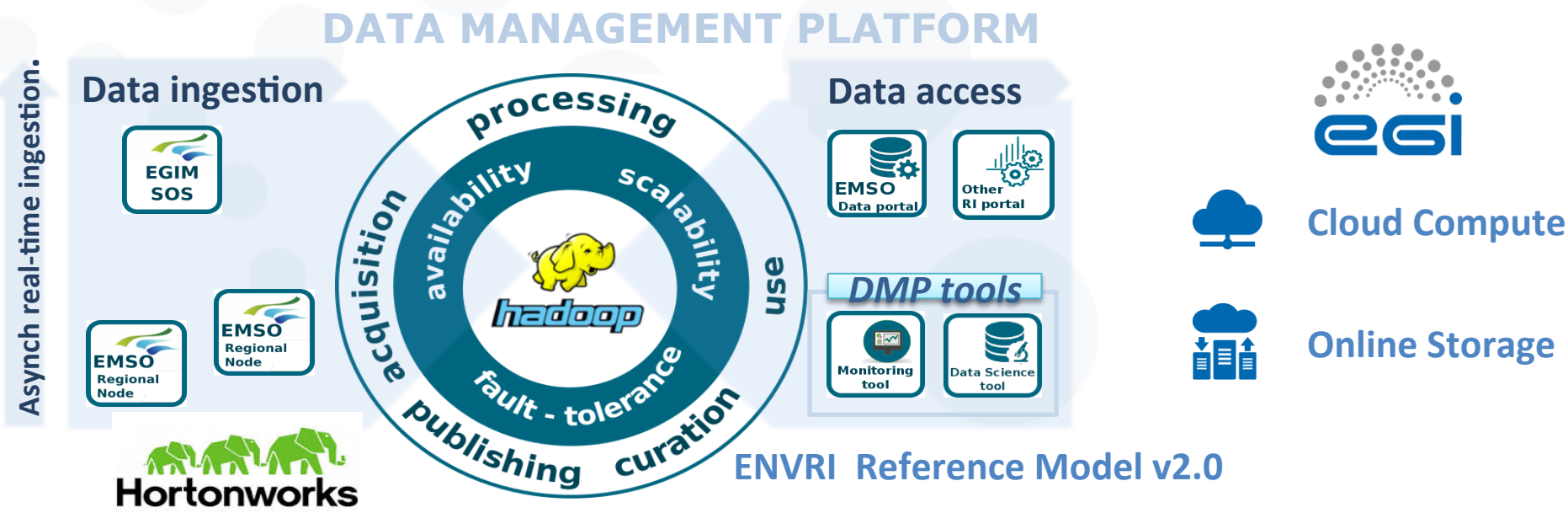
- 360 CPU cores
- 800 GB RAM
- 10TiB Storage

EMSODEV Data Management Platform

EGI-Engage

European Multidisciplinary Seafloor and water column Observatory (*EMSO*)

Data MGM Platform: Acquisition, processing publishing and curation of all the data collected by the deep see observatories



EGI-EMSODEV SLA (vo.emsodev.eu):

- 4 Resource Providers: RECAS-BARI, INFN-Padova, CESGA, LIP
- Cloud Compute: **~ 300 cores ~ 600 GB RAM** & Online Storage: **9 TB**
- Current deployment: **~10 VMs (8 CPUs + 16GB RAM + 40GB HD), 5TB**

Open Research Cloud Declaration

- Public discussion towards the implementation of an international research cloud federation.
  - Aim to produce a high level set of principles on the value for federating a distributed open cloud for researchers and scientists internationally.
- Participants: cloud providers for and by research, academic and scientific institutions
  - Arizona Open Compute; OpenNebula; NSF-Cloud (JetStream, CloudLab, Chameleon); EGI Federated Cloud; CERN; MassOpenCloud; Nectar
  - Hosted by the OpenStack Foundation and by MIT, 11-12 May 2017

EGI Services

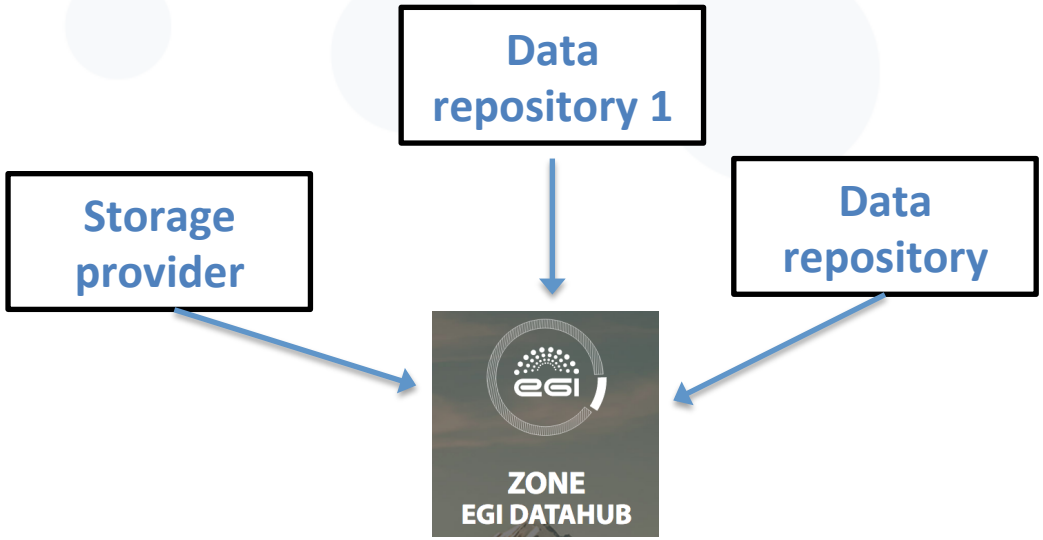
New Services

Advancing EGI Data Management

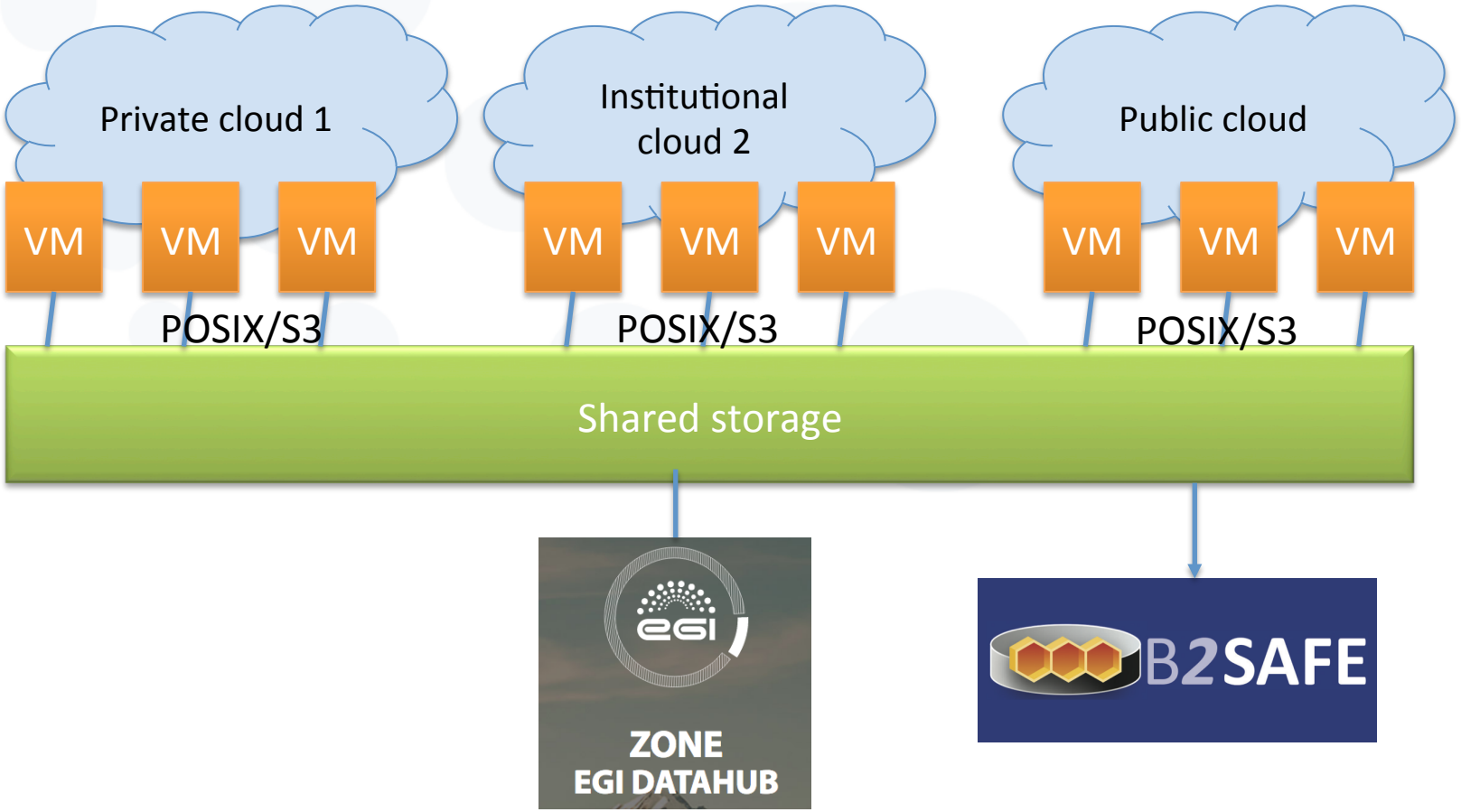
- Increased accessibility of third party research data to users (Data as a Service) for downstream analysis
  - Bridge preserved data and computing
  - Federate distributed data repositories
  - Mint DOIs, make data discoverable
  - Federated Identity Provisionin

EGI DataHub (alpha)

- New DaaS offering based on Onedata Onezone
- Use case 1. Cross-domain federation of existing storage providers
- Use case 2. Distributed platform for managing replicas of publicly available data collections to be made available on EGI Infrastructure



Data intensive computing across multiple clouds



EGI Applications on Demand service

https://access.egi.eu/start

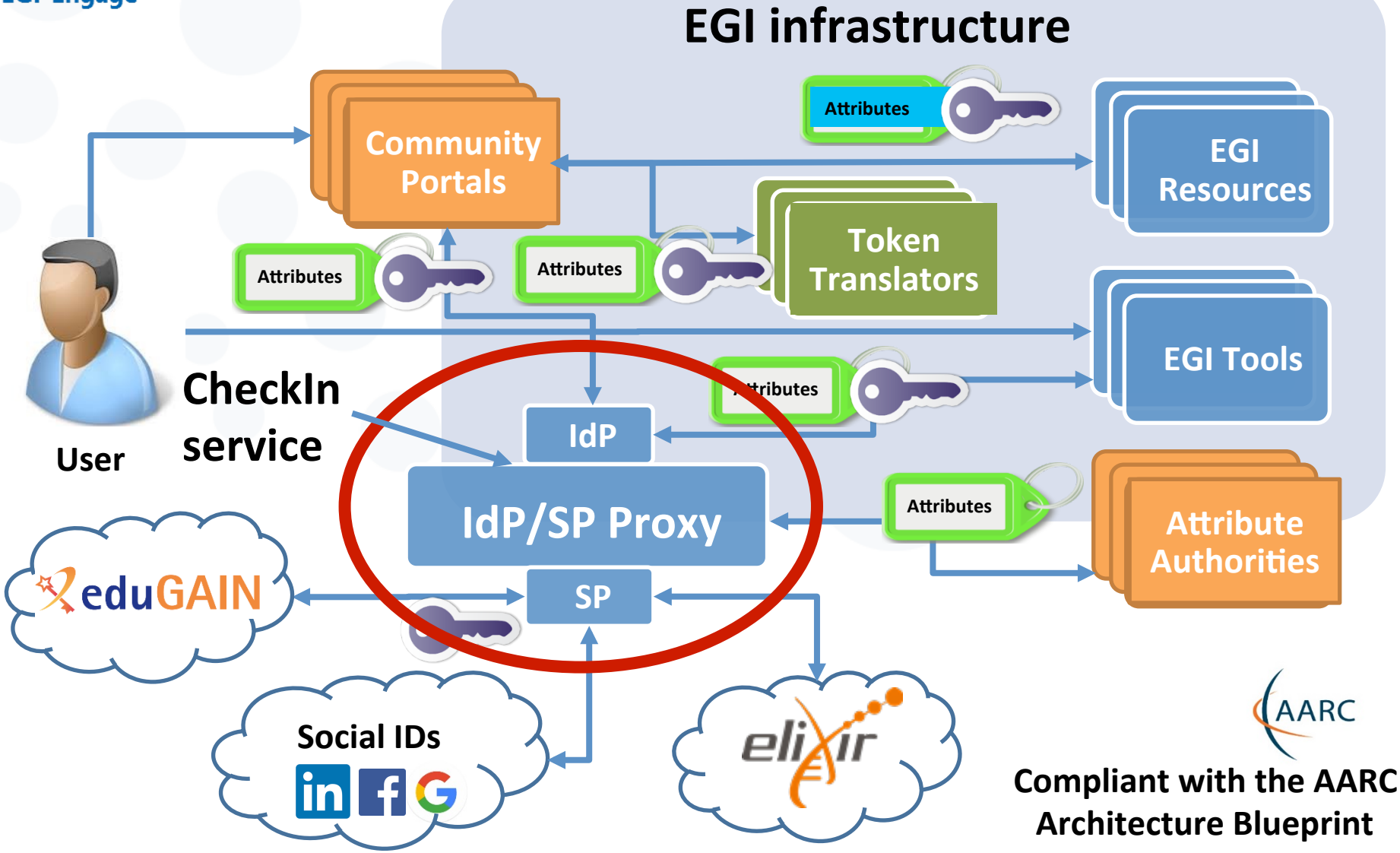
- Allow researchers, without dedicated access to computational and storage resources, to run custom scientific applications or applications/tools available from a provided library through a quota of compute and storage resources.
  - Researchers can access scientific applications underpinned by a compute and storage infrastructure, to carry out compute and data-intensive analysis.
  - Developers can integrate custom applications into the service and share them as with other researchers
  - Providers of compute and storage services can offer reach application library on local clusters and clouds to serve national user communities

Ported applications and tools

Core requirements for the AAI

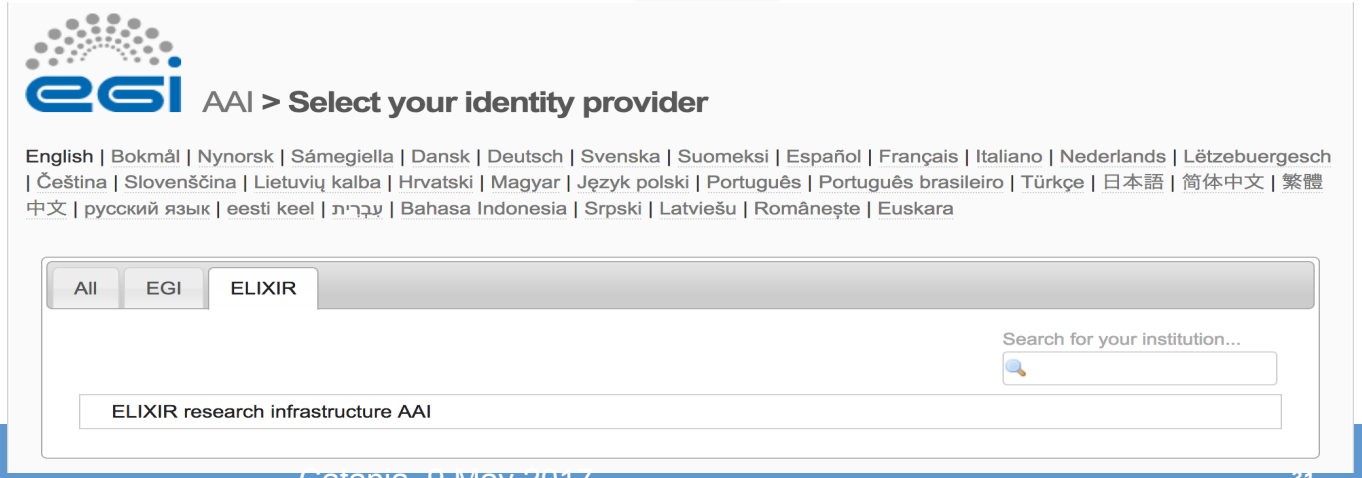
- **Credentials/Tokens**
  - Users access EGI services with credentials released by his/her home organisation (eduGAIN support)
  - Take into account the so-called *homeless users*
  - Level Of Assurance (LoA) for each credential type
- **Open architecture**
  - Support the most common technologies to manage federated identities: SAML, OpenID Connect, X.509, etc.
  - Support several attributes sources
  - Easily extensible and interoperable with other infrastructures
- **Hide the complexity to the Service providers**
  - Token Translator Services (TTSs)
    - Conversion of credentials into tokens recognised by the service

The new EGI AAI architecture



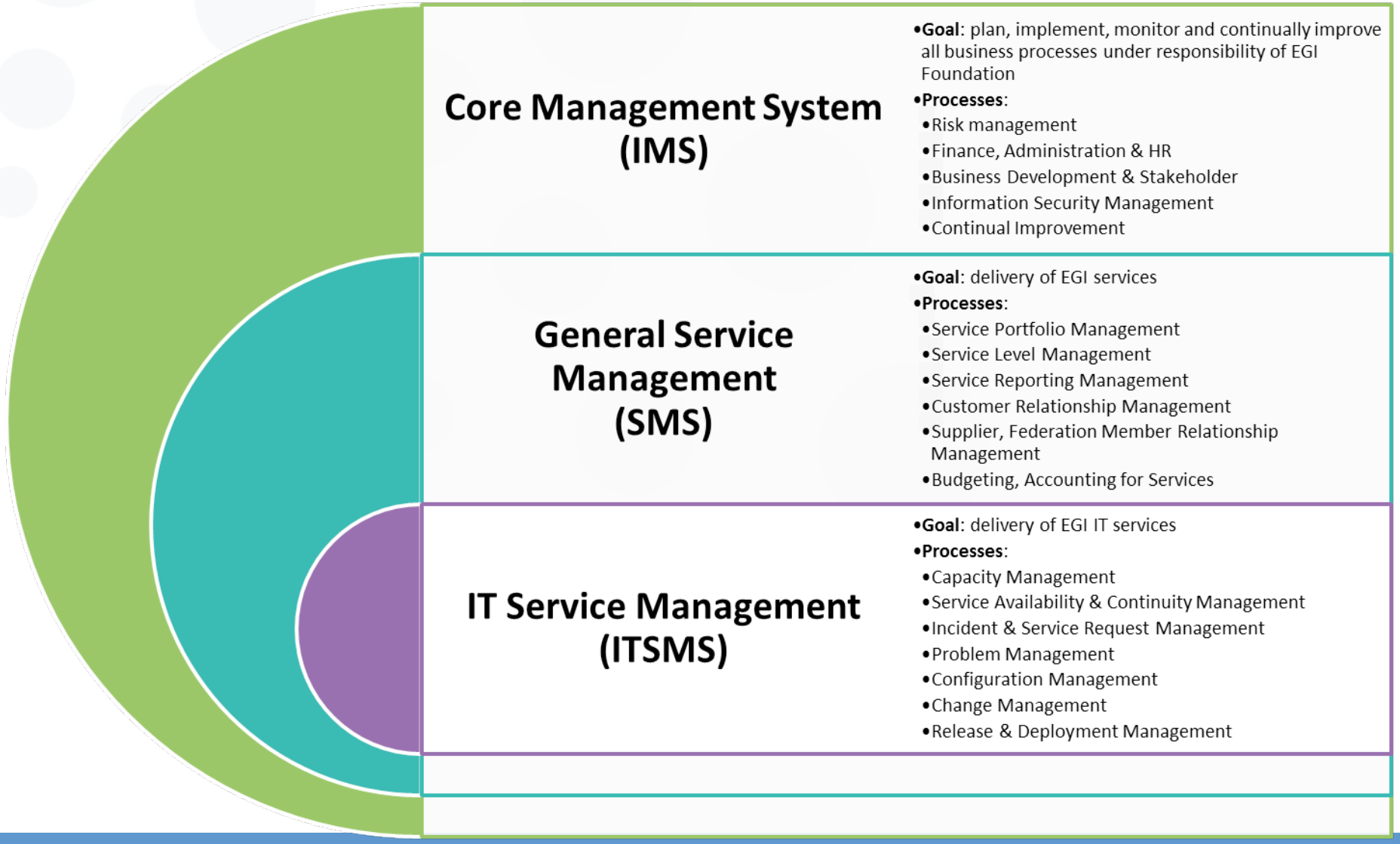
User community success story: ELIXIR

- CheckIn integrated with the ELIXIR AAI IdP Proxy
- ELIXIR users allowed to access:
  - GOCDB: site admins can register and manage infrastructure resources (cloud and storage)
  - APPDB: application developers can register Virtual Appliances for publishing on the ELIXIR cloud sites and sharing with the life science community.



Integrated Management System

Integrated Management System



ISM Processes

IT Service Management	•**Goal**: delivery of EGI IT services •**Processes**: •Capacity Management •Service Availability & Continuity Management •Incident & Service Request Management
General Service Management (SMS)	•**Goal**: delivery of EGI services •**Processes**: •Service Portfolio Management •Service Level Management •Service Reporting Management •Customer Relationship Management •Supplier, Federation Member Relationship
Core Management System (IMS)	•**Goal**: plan, implement, monitor and continually improve all business processes under responsibility of EGI Foundation •**Processes**: •Risk management •Finance, Administration & HR •Business Development & Stakeholder •Information Security Management •Continual Improvement

NA1 Activities & Achievements

ISO certifications

ISO 9001

ZERTIFIKAT ■ CERTIFICATE ■ 認證證書 ■ СЕРТИФИКАТ ■ CERTIFICADO ■ CERTIFICAT

TÜV SÜD Management Service

CERTIFICATE

The Certification Body
of TÜV SÜD Management Service GmbH
certifies that

EGI Foundation / EGI.eu
Science Park 140
1098 XG Amsterdam
Netherlands

has established and applies
a Quality Management System for

**The management system implemented to
plan, implement, monitor and continually improve
all business processes under responsibility and to deliver
all services covered by the service catalogue of EGI Foundation.**

An audit was performed, Report No. **707067543**.
Proof has been furnished that the requirements
according to

ISO 9001:2015

are fulfilled.

The certificate is valid from **2017-03-07** until **2020-03-06**.

Certificate Registration No.: **12 100 53643 TMS**.

Product Compliance Management
Munich, 2017-03-08

TÜV SÜD Management Service GmbH • Zertifizierungsstelle • Ridlerstraße 65 • 80339 München • Germany
www.tuev-sued.de/certificate-validity-check

ISO 20000

ZERTIFIKAT ■ CERTIFICATE ■ 認證證書 ■ СЕРТИФИКАТ ■ CERTIFICADO ■ CERTIFICAT

TÜV SÜD Management Service

CERTIFICATE

The Certification Body
of TÜV SÜD Management Service GmbH
certifies that

EGI Foundation / EGI.eu
Science Park 140
1098 XG Amsterdam
Netherlands

has established and applies
a Management System for IT-Services.

**The management system implemented to
plan, implement, monitor and continually improve
all business processes under responsibility and to deliver
all IT services covered by the service catalogue of EGI Foundation.**

An audit was performed, Report No. **707067543**.
Proof has been furnished that the requirements
according to

ISO/IEC 20000-1:2011

are fulfilled.

The certificate is valid from **2017-03-07** until **2020-03-06**.

Certificate Registration No.: **12 410 53644 TMS**.

Product Compliance Management
Munich, 2017-03-08

TÜV SÜD Management Service GmbH • Zertifizierungsstelle • Ridlerstraße 65 • 80339 München • Germany
www.tuev-sued.de/certificate-validity-check

- EGI IMS based on FitSM
- Lightweight standard for IT service management. With FitSM Training you will learn the fundamentals of IT service management and how to implement FitSM in your organisation through a combination of lessons and examples.
  - Check training opportunities in June!
    - https://www.egi.eu/services/fitsm-training/calendar/

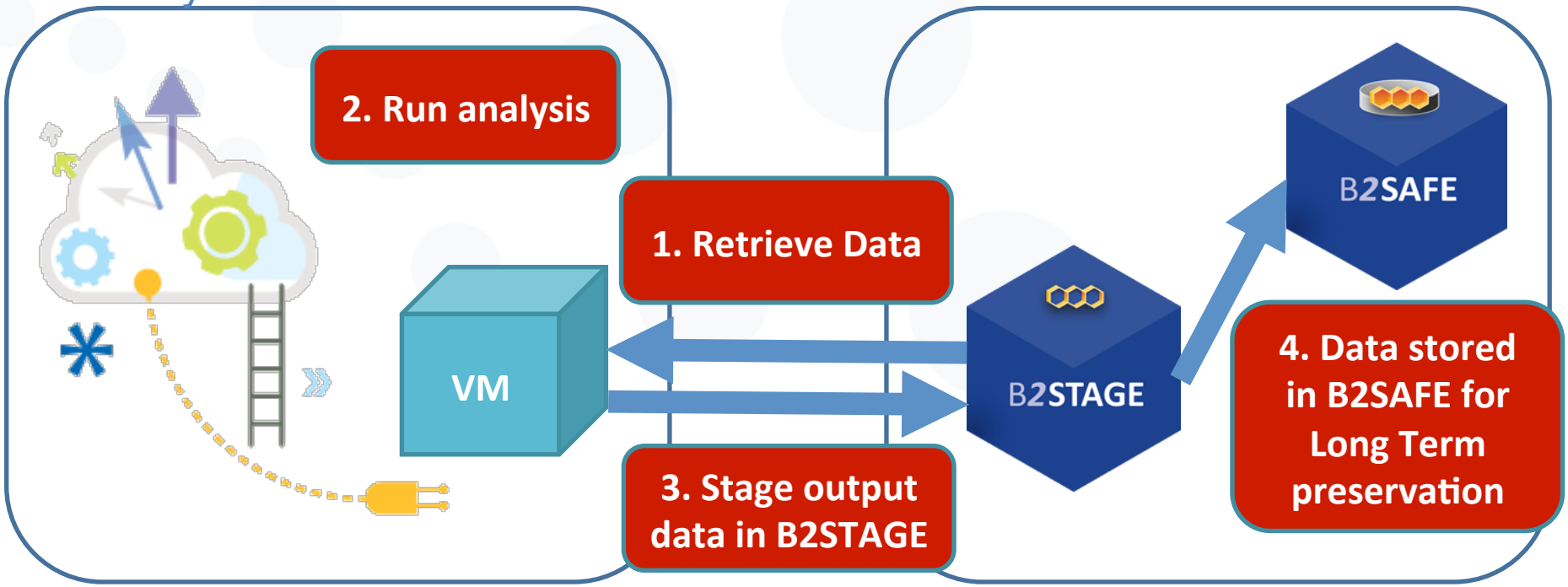
E-Infrastructure integration

EUDAT
D4SCIENCE

EGI-EUDAT Integration Pilot:
Processing in EGI, long-term storage in EUDAT

(1) Access to EGI and EUDAT services with a single user identity

(2) Data Staging between EGI Federated Cloud and EUDAT services



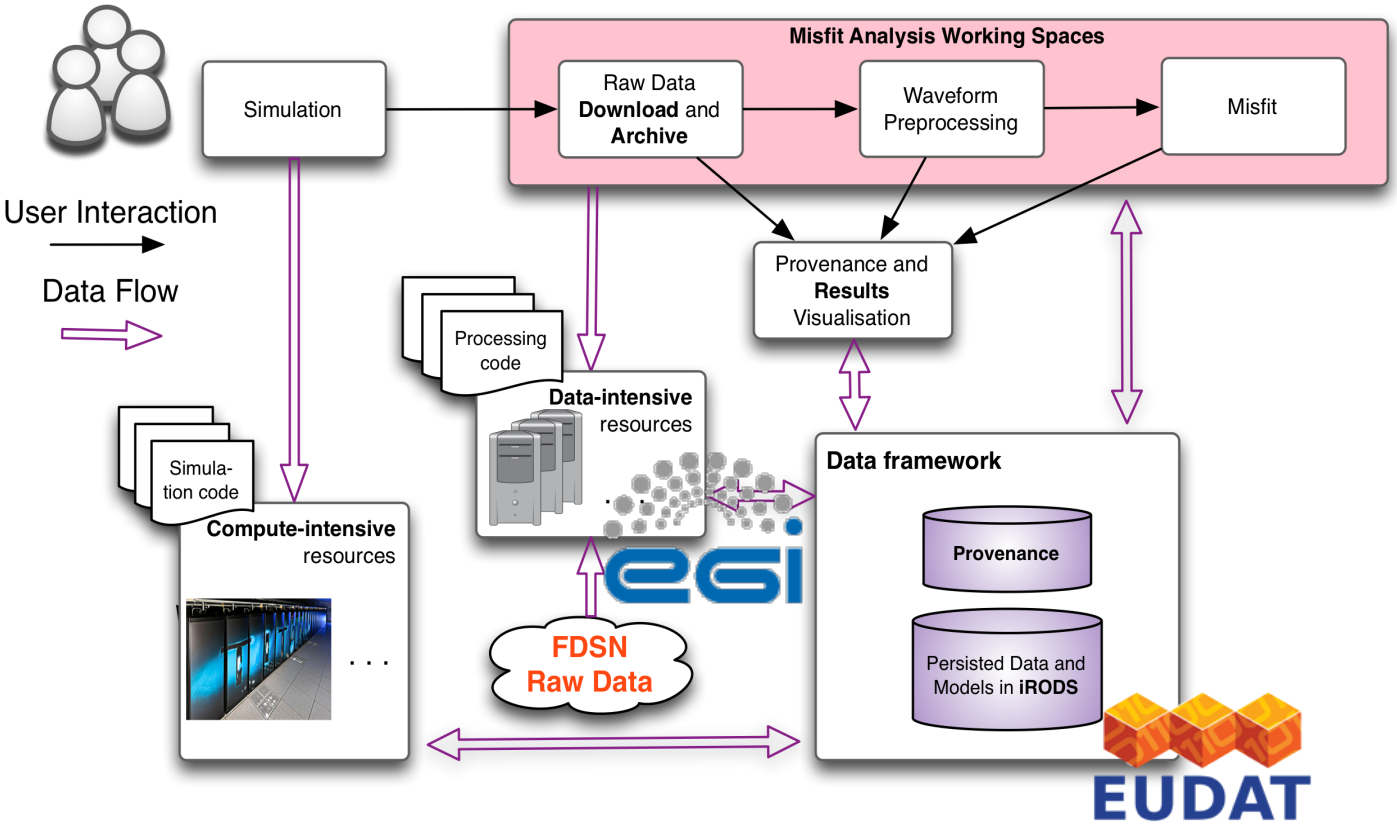
Federated Cloud

EPOS – Computational Seismology TCS Misfit Use Case

Processing and comparison of data resulting from the simulation of seismic wave propagation and real measurements recorded by seismographs following a real earthquake

Cloud Compute

Online Storage



ICOS Carbon Portal Footprint tool

EGI-Engage

ICOS Carbon Portal

Mission: enable research to understand the greenhouse gas budgets and perturbations in Europe/adjacent regions

Cloud Compute

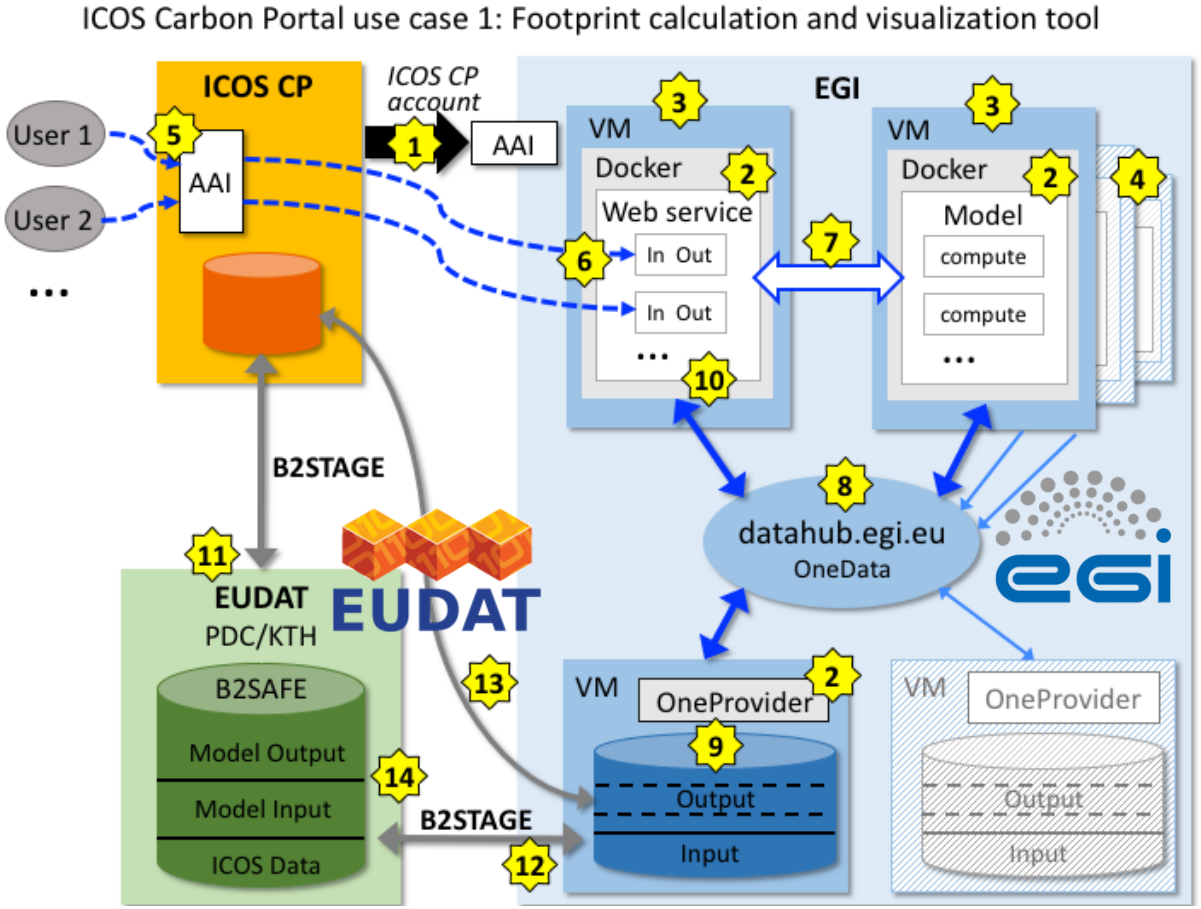
Online Storage

Data Hub

EUDAT

B2STAGE

B2SAFE



EGI Federated Cloud integration with D4SCIENCE

Credits: P. Pagano/CNR

- D4Science is a Data and a Computational e-Infrastructure
  - Implements the notion of e-Infrastructure as-a-Service
  - Used by several Projects: BlueBRIDGE, PARTHENOS, SoBigData, ENVRIPlus, OpenAIRE Connect
  - Hosts 91 VREs serving the Marine, Social Mining, Environmental, Cultural Heritage, ICT, Agriculture, and Digital Library Communities

- Completed integration with EGI Cloud Compute

Towards EOSC

Vision
Architecture
EGI Contribution

Defining the EOSC

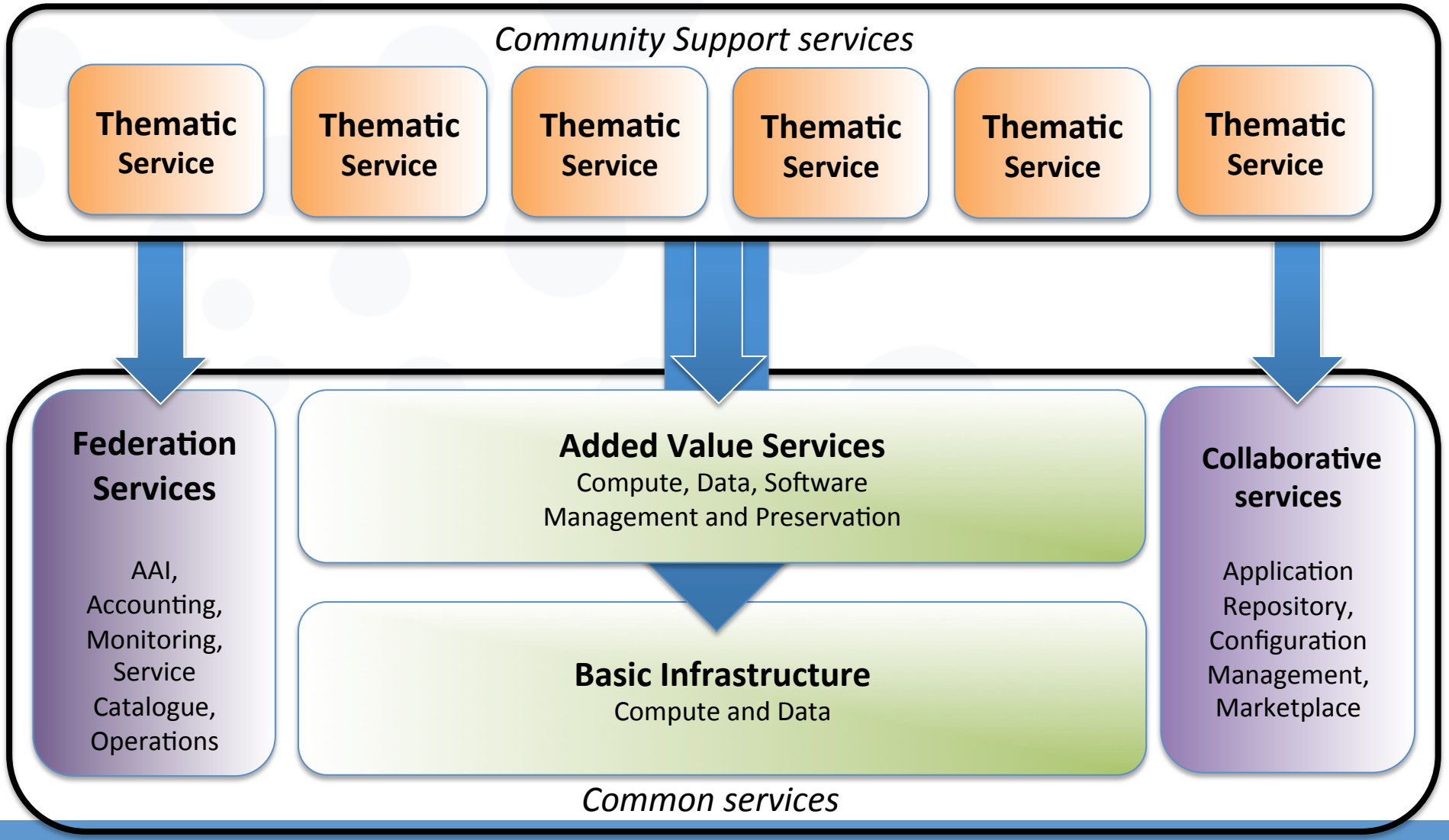
"The European Open Science Cloud (EOSC) is a vision for a federated, globally accessible, multidisciplinary environment where researchers, innovators, companies and citizens can publish, find, use and reuse each other's data, tools, publications and other outputs for research, innovation and educational purposes."

Credits: Open Science Policy Platform EOSC wg

Obstacles

- Fragmented efforts in service provisioning at national and European level
- Lack of basic AAI interoperability
- Disconnected compute/data infrastructures
- Heterogeneous funding models
- Different access policies
- No easy mechanism to procure services from e-infrastructures
- Dependency on short-term project-based funding
- …

EOSC Services



EOSC Architecture Components

1. Service pillars

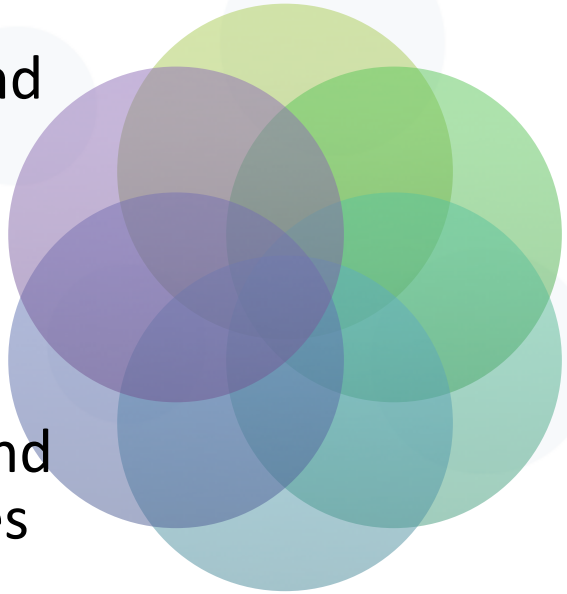
2. Federation services

3. Collaborative services

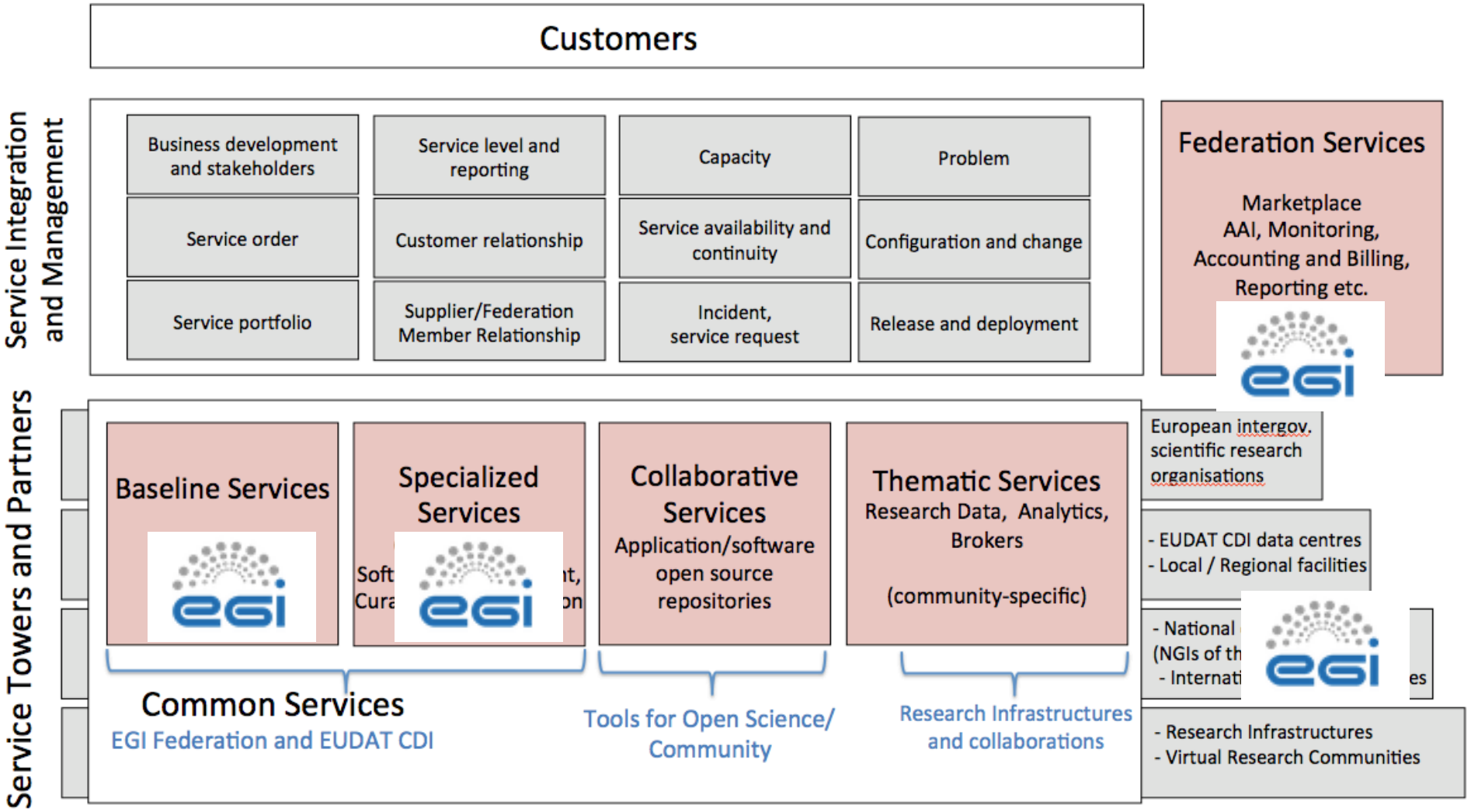
4. Policies (service providers and users) and standards

5. Service integration and management processes

6. Service integration and management tools



Architecture and EGI contribution



EOSC-hub project proposal, EINFRA-12

Mission

EOSC-hub mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-infrastructures 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 single entry point for all stakeholders.

Thank you for your attention.

Questions?

www.egi.eu