

# Towards an Open Science Commons

**Yannick LEGRÉ**

EGI.eu Managing Director



Slides credit: EGI Federation members

[www.egi.eu](http://www.egi.eu)

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



- EGI today
- Achievements
- Big shifts
- Towards 2020 and an Open Science Commons



EGI today

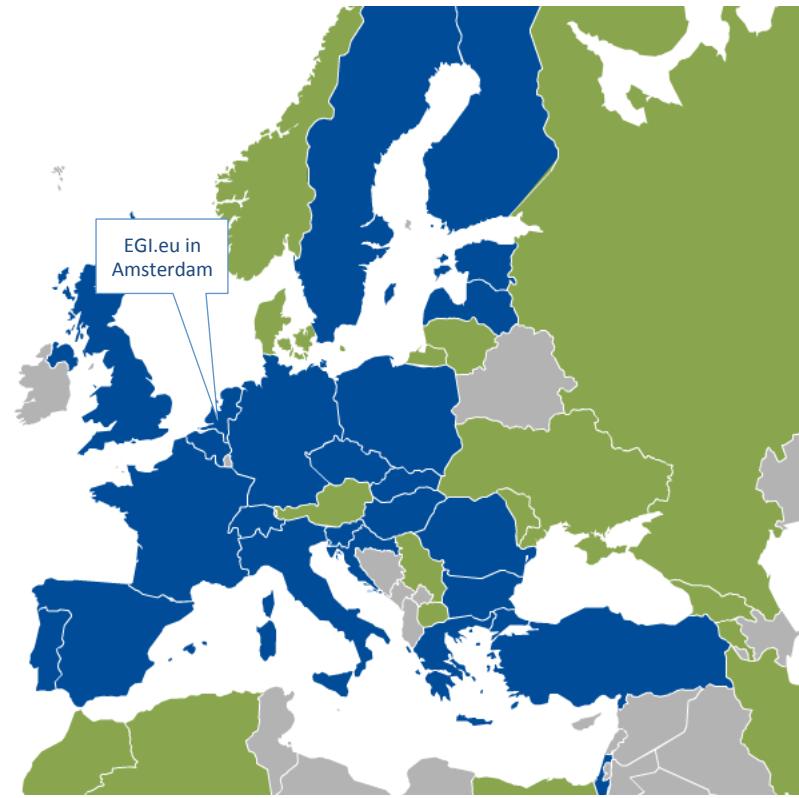
- Governance, the power of federating

Researchers from **all disciplines** have **easy, integrated and open access** to the advanced digital capabilities, resources and expertise needed to collaborate and to carry out compute/data intensive science and innovation

Create and deliver **open solutions**  
for science and research  
infrastructures  
by **federating** digital capabilities,  
resources and expertise  
across communities and national  
boundaries

# EGI.eu and its participants - 2015

- 26 participants: 24 NGIs and 2 EIROs (CERN, EMBL-EBI)
  - Opening membership to research communities
  - Opening membership to non-European countries
- Affiliation programme
  - Lower barriers of entry to widening countries



## Participants

CERN, EMBL-EBI, Belgium, Bulgaria, Croatia, Czech Republic, Estonia, Finland, France, Germany, Greece, Hungary, Israel, Italy, FYR of Macedonia, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Switzerland, Sweden, Turkey, UK

## Under discussion

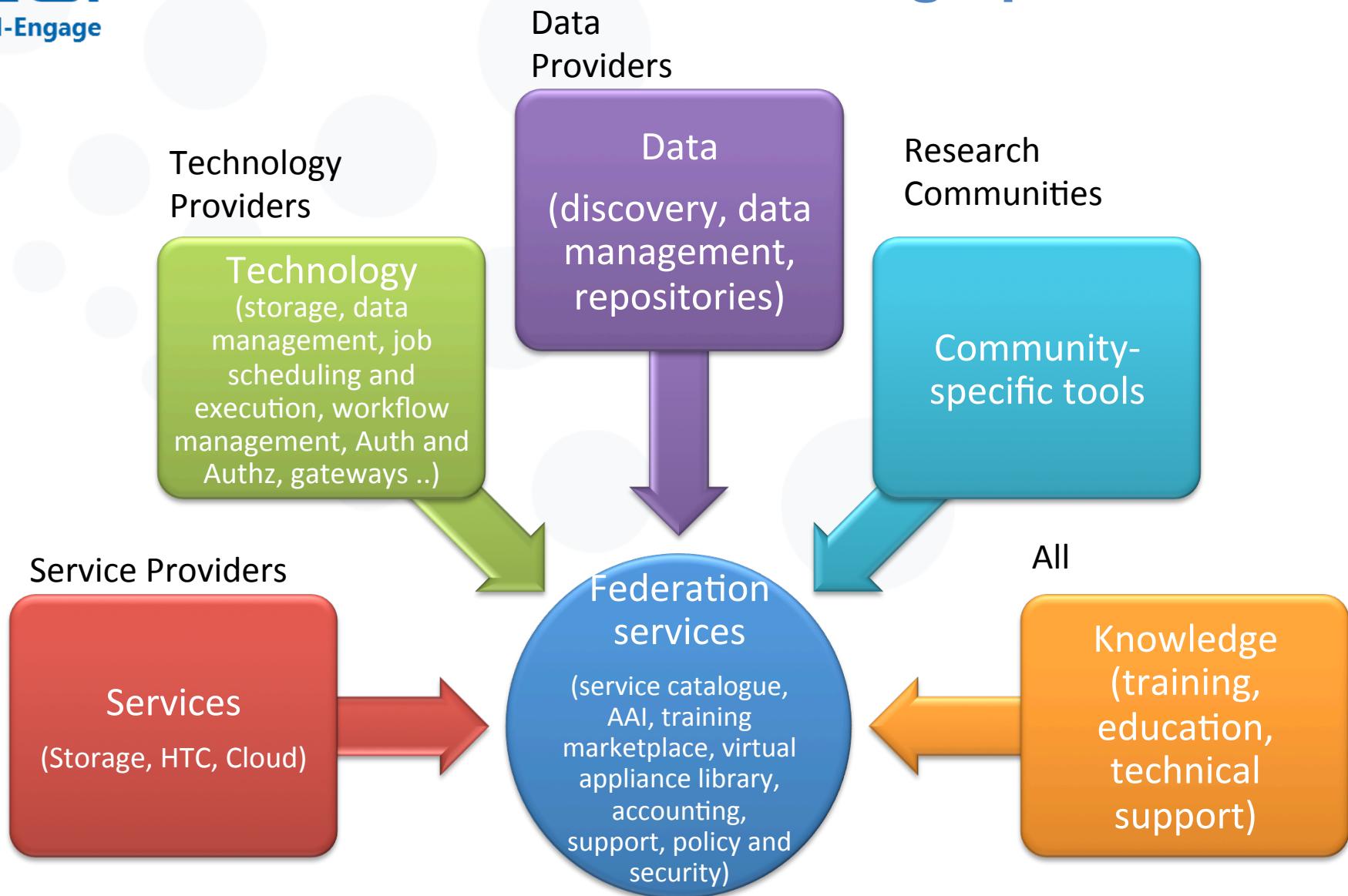
Armenia, Austria, Belarus, Denmark, Moldova, Norway, Russia, Ukraine

- High-Throughput Data analysis
- Federated Cloud
- Federated Open Data Processing
- Federated Operations
- Community driven Innovation and Support
- Policy Advice

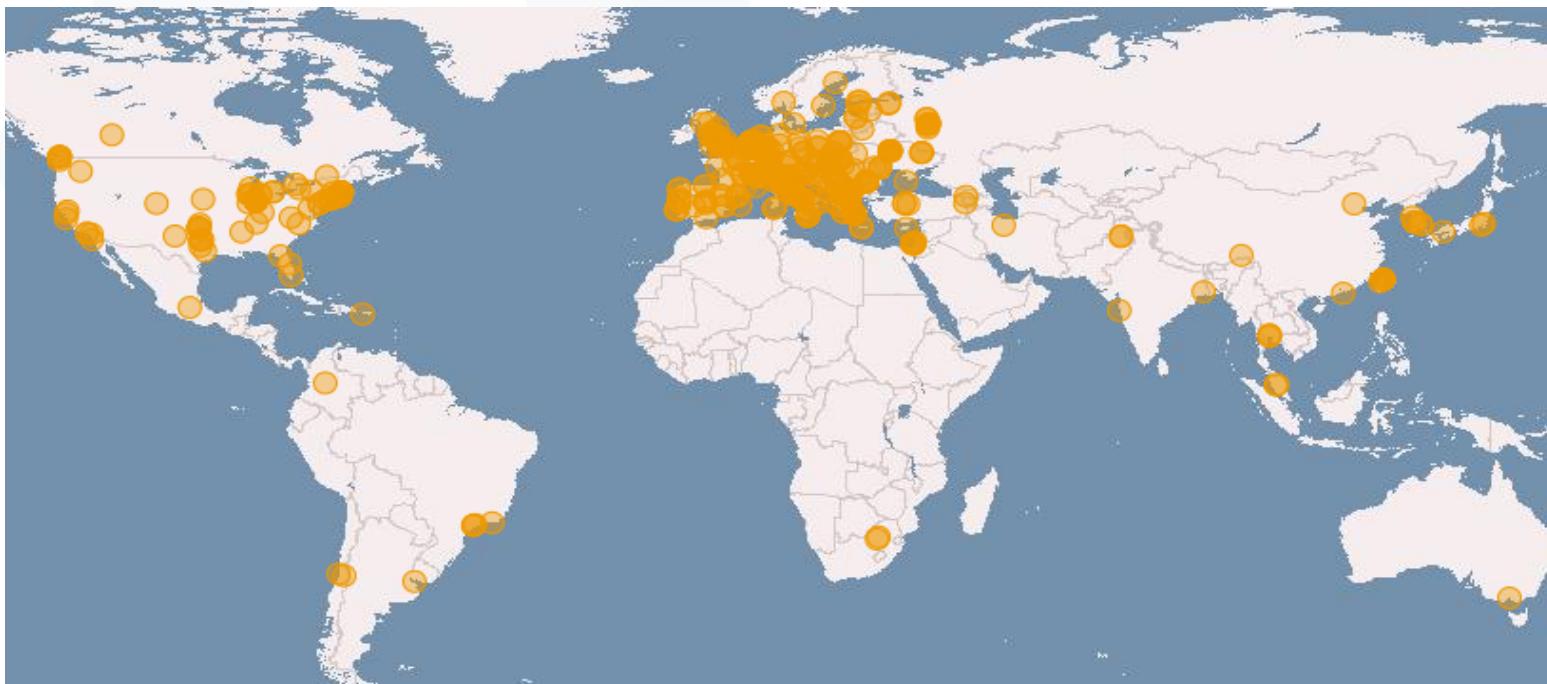
# Science is inherently distributed

- Discoverability of services and knowledge
- Portability
  - data, applications, software
- Sharing and openness
- Common access policies, security
- One accounting infrastructure
- One support infrastructure
- Single sign on
- Federated service management
- Aggregation of demand and offer

# Federating open science

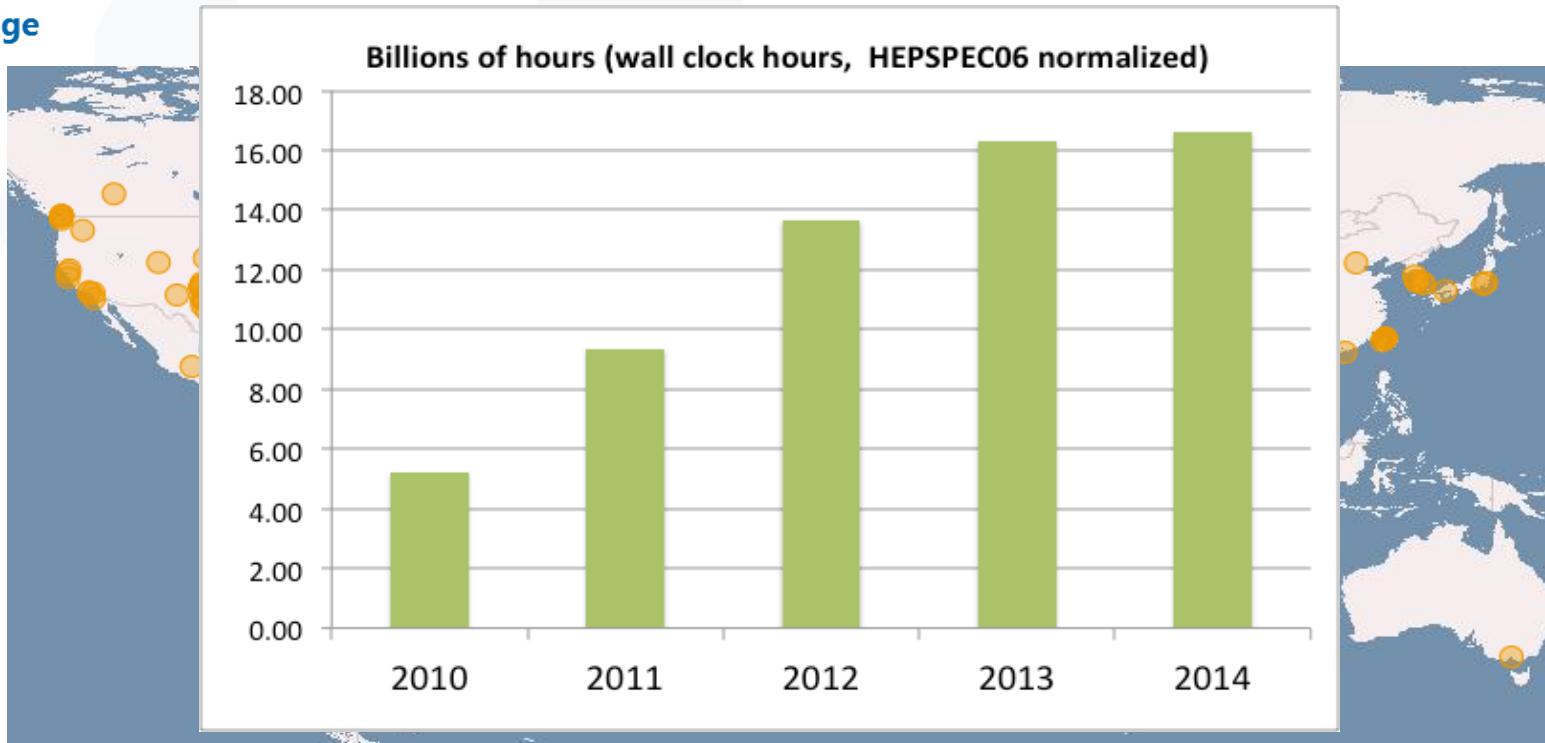


# Federating e-infrastructures and data 1/2



- Distributed, federated storage, HTC and cloud facilities
- Virtual Research Environments
- > 200 registered user research projects
- ~350 resource centres in 54 countries
- 550,000 logical CPU cores
- >290 PB disk, 180 PB tape
- > 99.6% reliability

# Federating e-infrastructures and data 1/2



- **Distributed, federated** storage, HTC and cloud facilities
- Virtual Research Environments
- > 200 registered user research projects
- ~350 resource centres in 54 countries
- 550,000 logical CPU cores
- >290 PB disk, 180 PB tape
- > 99.6% reliability

# Federating e-infrastructures and data 2/2

EGI Council members  
Integrated Infrastructures  
Peer Infrastructures



- More than 6,000 jobs/year to OSG
- More than 68,000 jobs/year in IDGF
- 840 M CPU hours/year in Asia Pacific

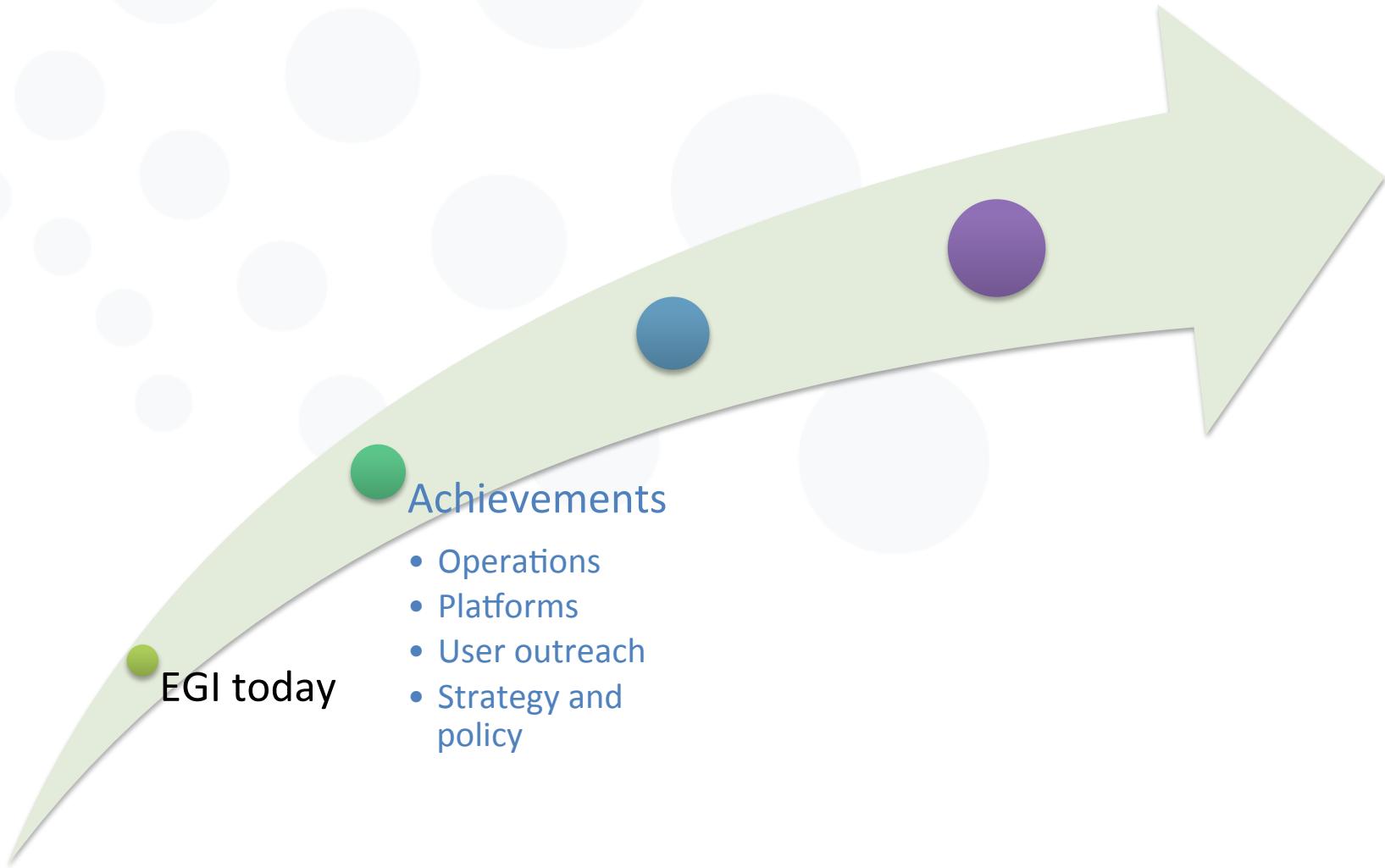


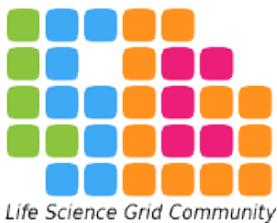
Canadian Astronomy Data Centre



E-CEO

E-Collaboration  
For Earth  
Observation



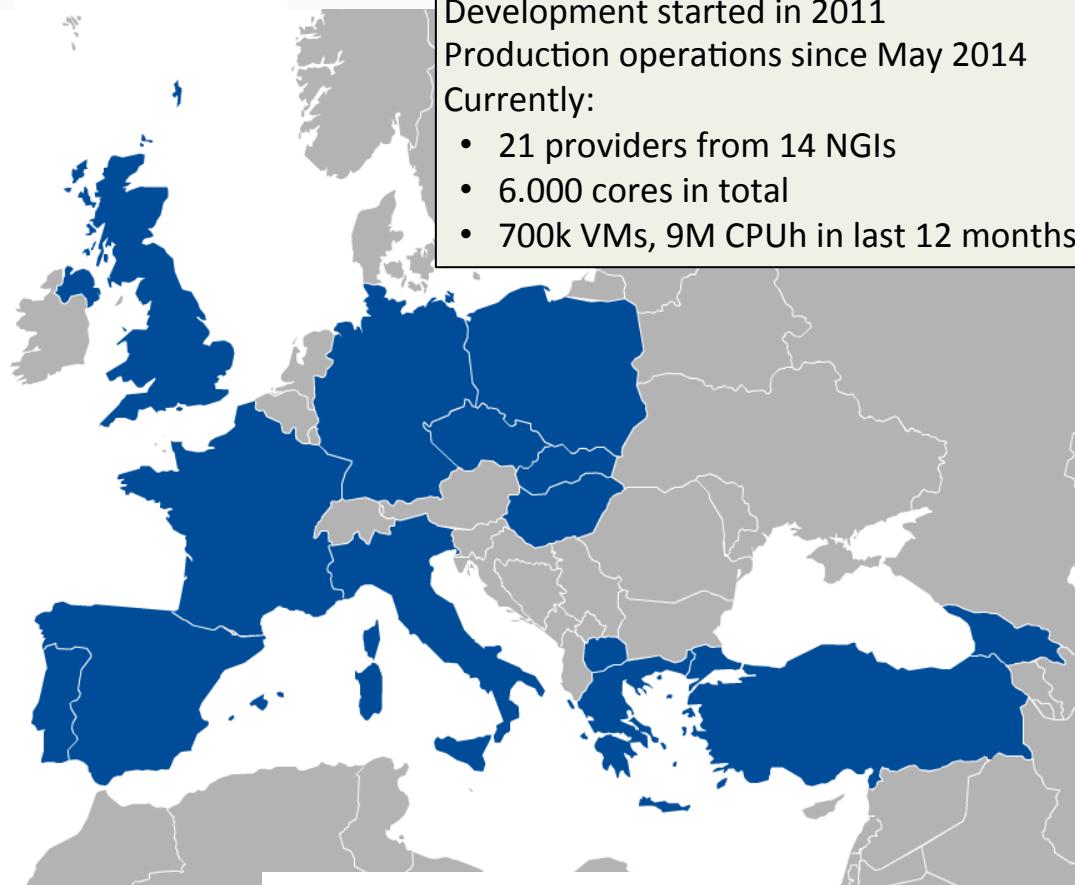


# Get infrastructure services

Resource allocation for national and international resources

- e-GRANT
  - Pooling of distributed infrastructure resources (HTC and cloud)
  - Matchmaking demand  $\leftrightarrow$  offer
  - Allocation
  - SLA negotiation (user community  $\leftrightarrow$  EGI.eu)
- Monitoring of service level targets

# EGI Federated cloud

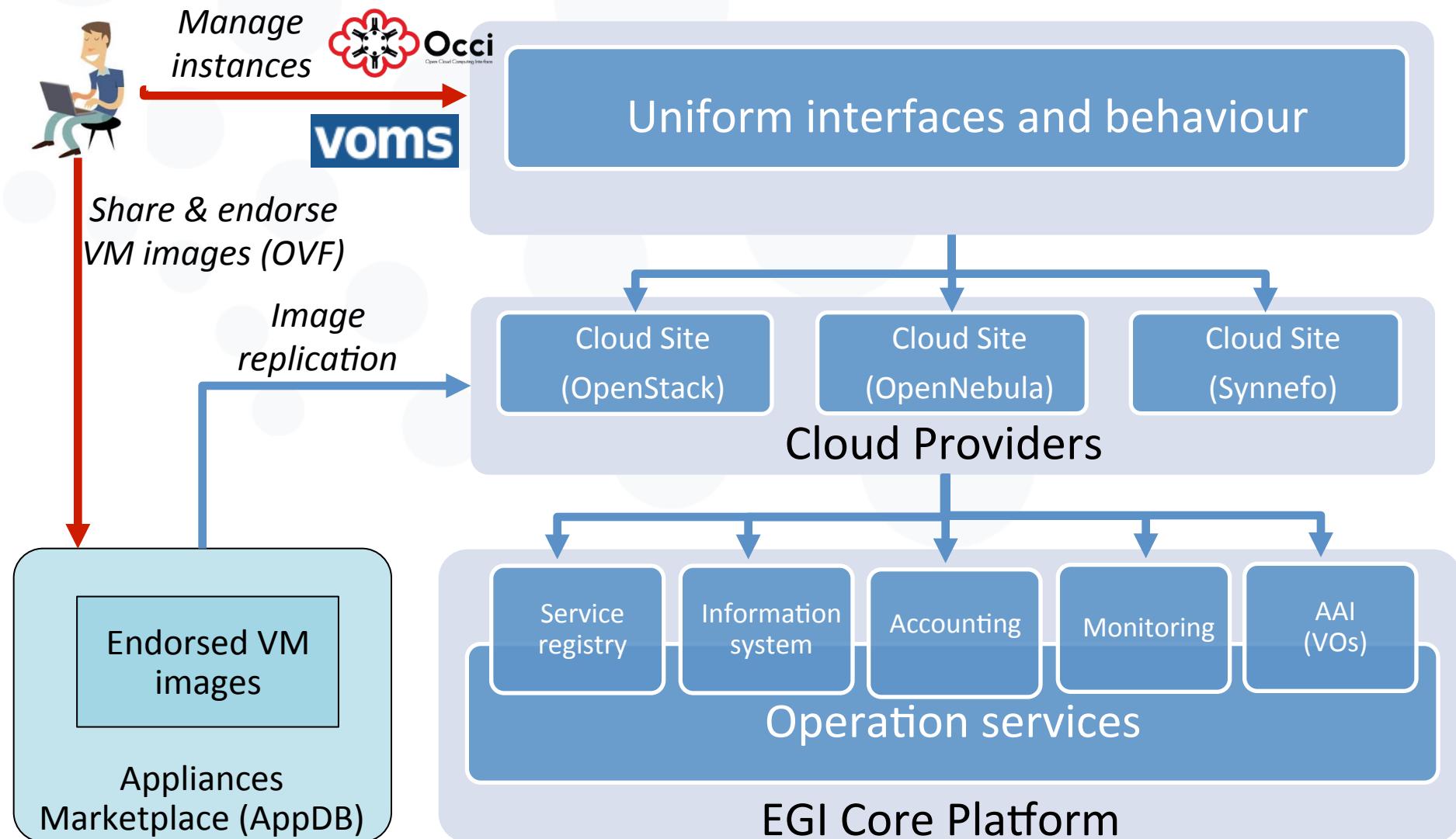


Development started in 2011  
Production operations since May 2014  
Currently:

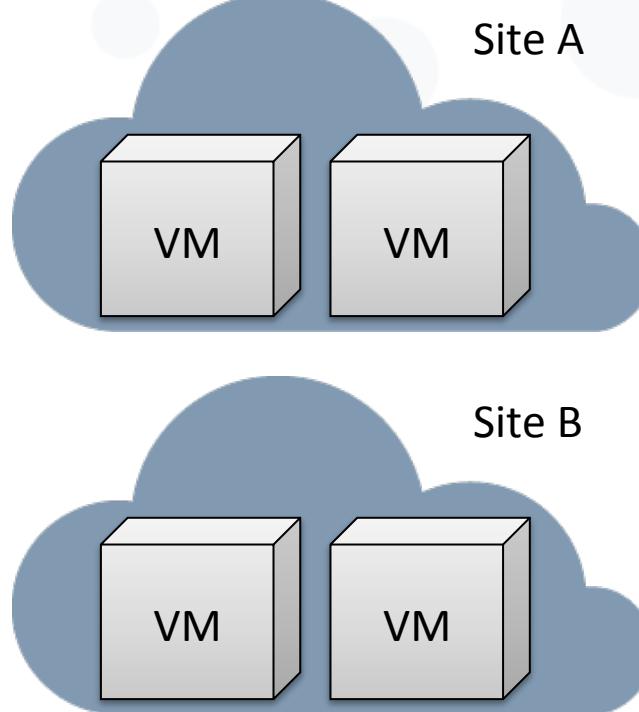
- 21 providers from 14 NGIs
- 6.000 cores in total
- 700k VMs, 9M CPUh in last 12 months



# Architecture of the EGI Federated Cloud



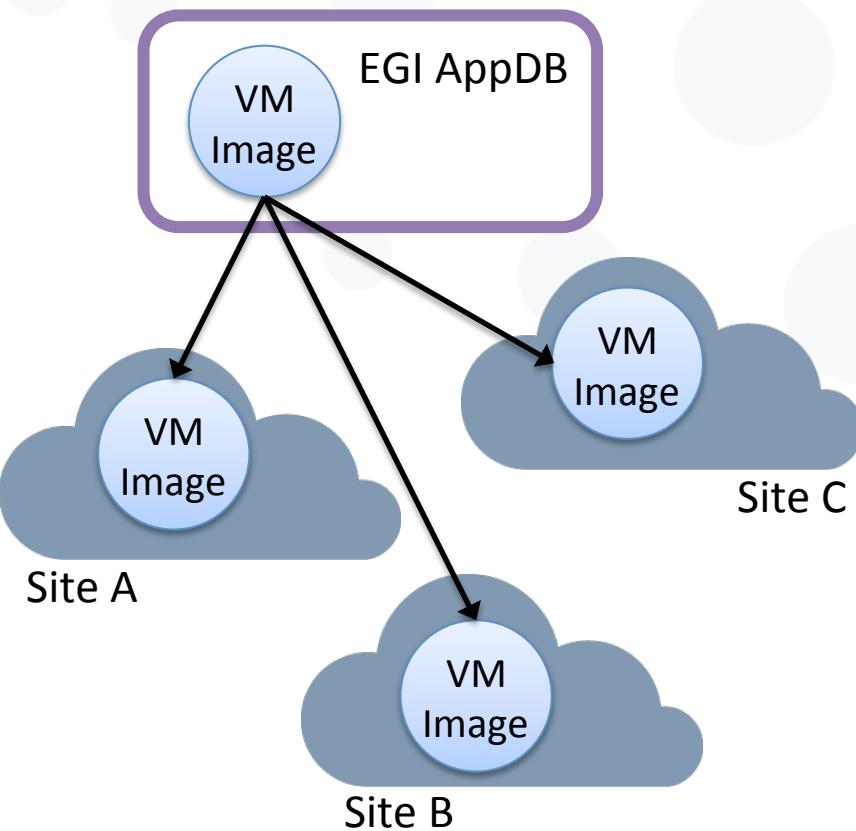
## On demand compute to run any kind of workloads on virtual machines



- Easy provisioning**
  - OCCI API across the whole infrastructure
  - VMs start immediately
  - Ruby and Java clients
- Customize**
  - Select your OS
  - *root* access
  - Contextualization
- Scale to your needs**
  - Select VM size (cores, RAM)
  - Create and destroy VMs as needed

# VM Image Management

## Automatic and secure distribution of endorsed VM images for Virtual Organisations



Web based

- Easy creation from AppDB
- Re-use and extend images

EGI endorsed images

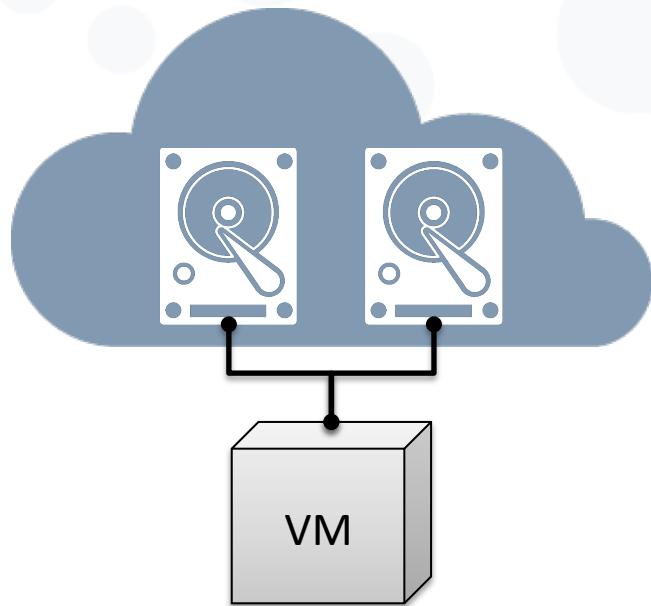
- Basic OS ready to use and contextualize
- Available on every site

VO-level control

- VO endorse images
- Automatically distributed to sites supporting the VO

# Block Storage

## Persistent Block Level Storage to attach to VMs



Simple usage

- Manage with [OCCI](#) and use as any other block device from VMs (i.e. POSIX)
- Snapshotable

High Performance

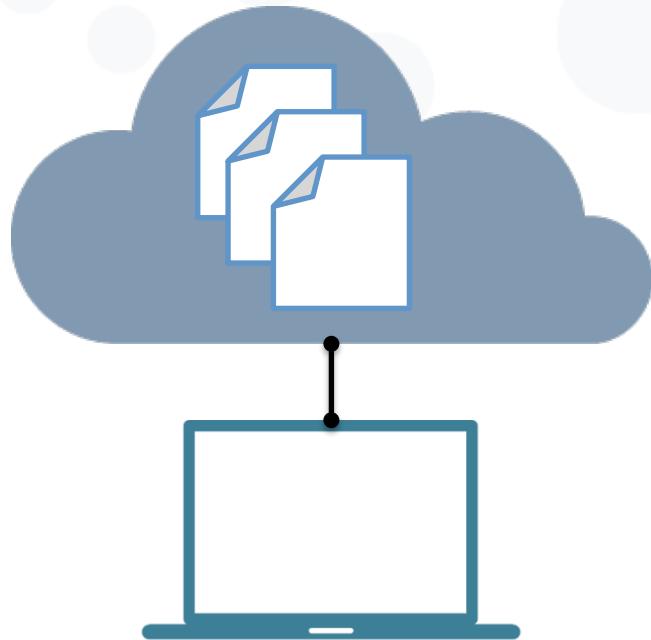
- Consistent and low-latency performance
- SSDs (in some sites)

Scale to your needs

- From GB to TB
- Create and attach to VMs on demand

# Object Storage

**Data storage infrastructure for storing and retrieving data from anywhere at any time**



## API Access

- CDMI REST API for managing and accessing data

## Sharing

- Define ACLs on each object, share publicly your data

## Scale to your needs

- Store as much data as needed
- Get accounted only for the space used

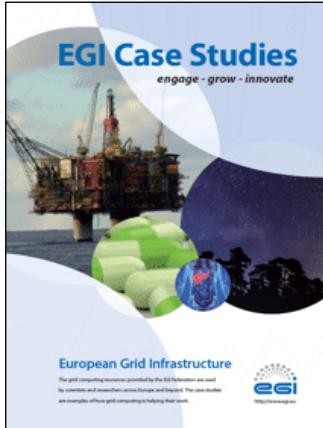
## 12 months of Federated Cloud activities

- 26 communities
  - Biological sciences
  - Physical sciences
  - Earth sciences
- 59 use cases currently supported, 5 from commercial organisations
- 700,000 VMs instantiated

# Strategy, Policy, Business Development

- New EGI strategy for 2020 in consultation with the EGI.eu Executive Board and the EGI Council
- The Open Science Commons
- Pay-for-Use pilot
  - 30 providers across 12 countries publishing pricing information (~10 ready/able to sell)
  - Emerging business models
  - Tools adapted (GOCDDB, AppDB, e-GRANT), including GUI
  - Final Report

# Impact



## EGI case studies

3,600 service endpoints, 47 UMD releases, 38,000 users

### Increasing use of new disciplines

- 220 research projects, 76 new
- Astronomy and astroparticle Physics, Structural biology, Hydrology and climate, Medical and Health Sciences

### Better services for the long tail

- 46% of the new users)

### Support to Research Infrastructures

- BBMRI, CTA
- Testing: EISCAT-3D, ELIXIR, ELI-NP, LifeWatch, LOFAR, KM3NeT

2,400+ Peer-reviewed papers, 620 new registered applications



## Compendium of RI requirements



EGI today

Achievements

The big shifts  
through EGI-  
Engage and  
sister projects

- AARC
- INDIGO Datacloud
- EDISON
- Technical support (ENVRI Plus, ELITRANS)

# The big shifts

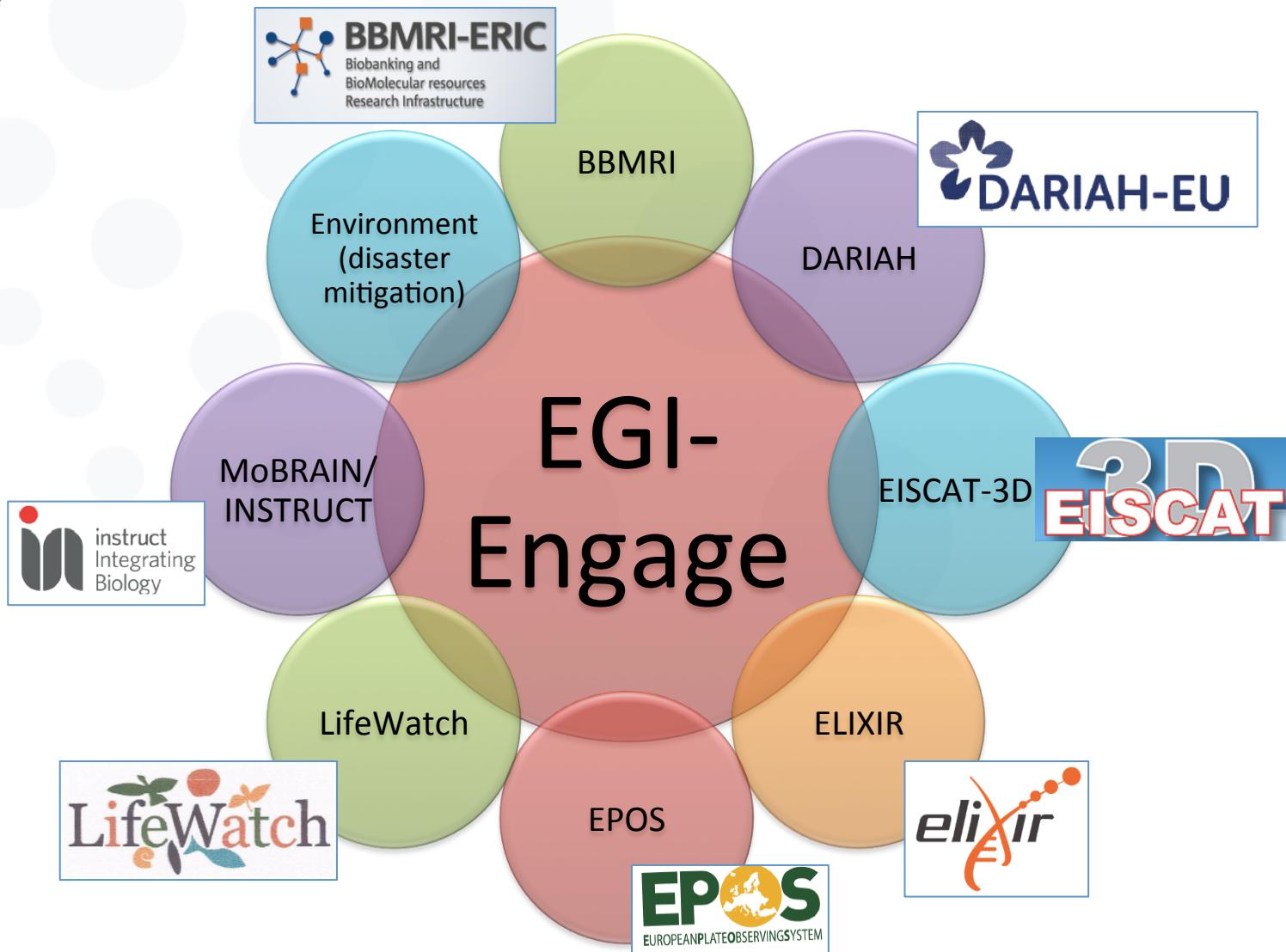
New governance to community engagement

The Distributed Competence Centre

# Distributed Competence Centre (DCC)

- Promote reuse of solutions of common interest across research communities
- Evolve the EGI technical services with community requirements and provide a test environment with NGIs/EIROs → co-development
- Promote the integration of community services
  - Scientific applications
  - Joint training programme
  - Technical user support

# EGI-Engage support to the DCC



# Outcomes

Community clouds

GPU federations

Advanced AAI

Data products

Science gateways

Active Repositories

Scalable applications

Training resources

User support

Open Science

MoBRAIN/  
INSTRUCT

Environment  
(disaster  
mitigation)

BBMRI

DARIAH

EISCAT-  
3D

ELIXIR

LifeWatch

EPOS

EGI community

Federated Cloud

High  
Throughput  
Computing

Security, access  
control

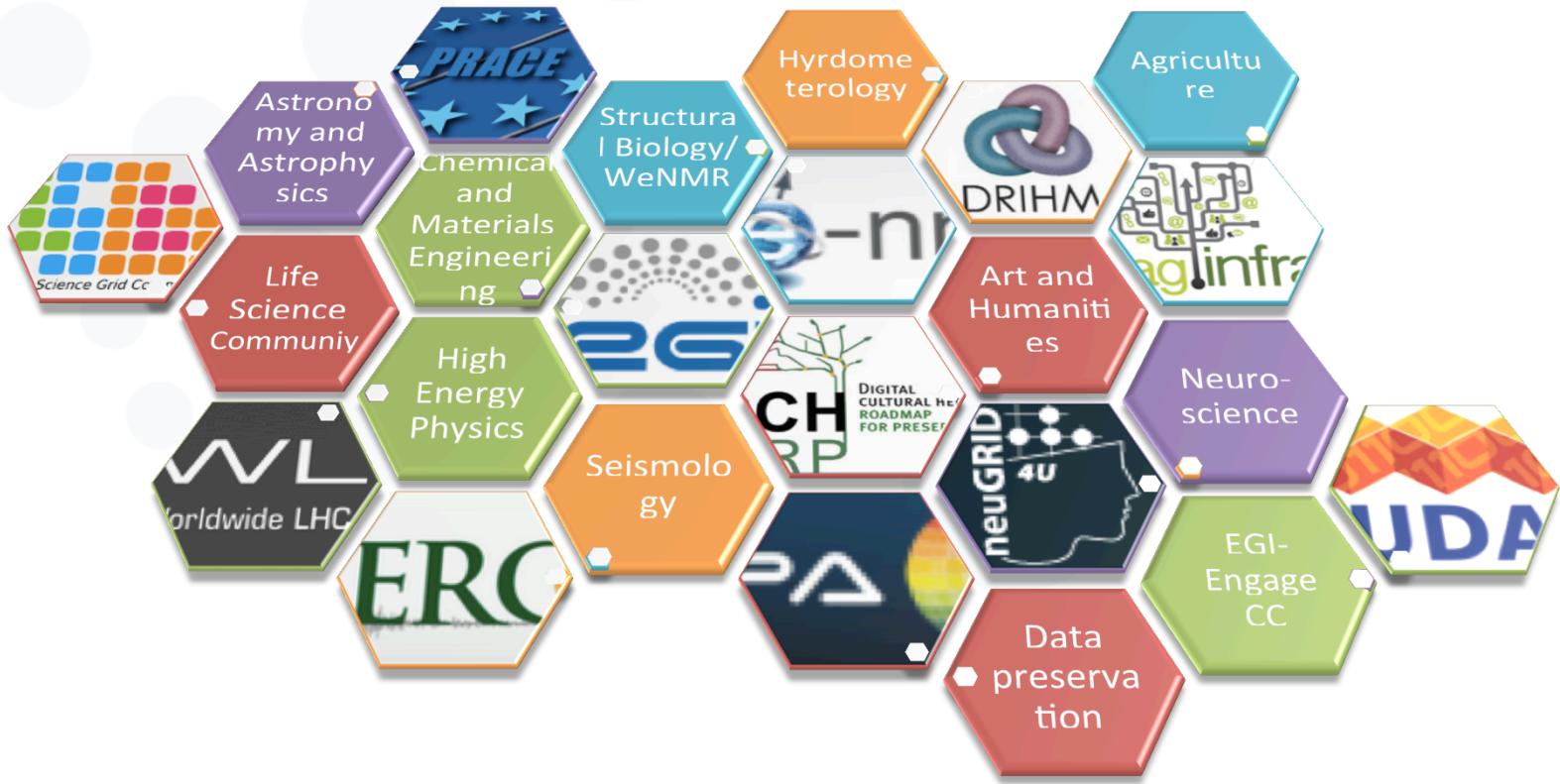
Data  
management  
and federation

Gateway  
systems

# Actors – present and future



# Federate Knowledge in Europe



# The big shifts

New governance to community engagement

The Distributed Competence  
Centre

Better services for the long tail

Centrally provided services for  
reduced access barriers

# Services for the long tail of science

- Move towards a “zero (technical) barrier” e-infrastructure
  - Services dedicated to individual users or very small collaborations:
    - No certificate, no VO, full EGI experience
- User facing features
  - Log in using their federated identity
  - Provide the additional information not available in the IdP
  - Discover (marketplace) and submit a request for resources
- EGI/NGIs facing features:
  - Assign UIDs to users of the long tail of science platform
  - Approve user request
  - Monitor usage of resources

# The big shifts

New governance to community engagement

The Distributed Competence Centre

Better services for the long tail

Centrally provided services for reduced barriers

New AAI

Service proxy/  
virtual IdP

Token translation  
IdP for homeless users

# Advancing AAI

- EGI users are directly/indirectly using x509 credentials to access the production services
- Objective: allow users to use their existing institutional credentials by
  - Replicating the current architecture to manage user communities in the other authentication technologies already used by the users
  - Integrating other federated identities into EGI services
- Testing and deployment of AAI services, and requirements analysis in close collaboration with the CCs and the other communities
  - Catch all IdP service (EGI sso), online CA, attribute authorities to manage users without X.509 ceritificate
  - Service proxy/Virtual IdP: technical service AND support to help communities to integrate easily their IdP with EGI. Integrating new IdP and attribute authorities in a one-step.
- Collaboration with AARC project

# The big shifts

New governance to community engagement

The Distributed Competence Centre

Better services for the long tail

Centrally provided services for reduced barriers

New AAI

Service proxy/virtual IdP, token translation, IdP for homeless users

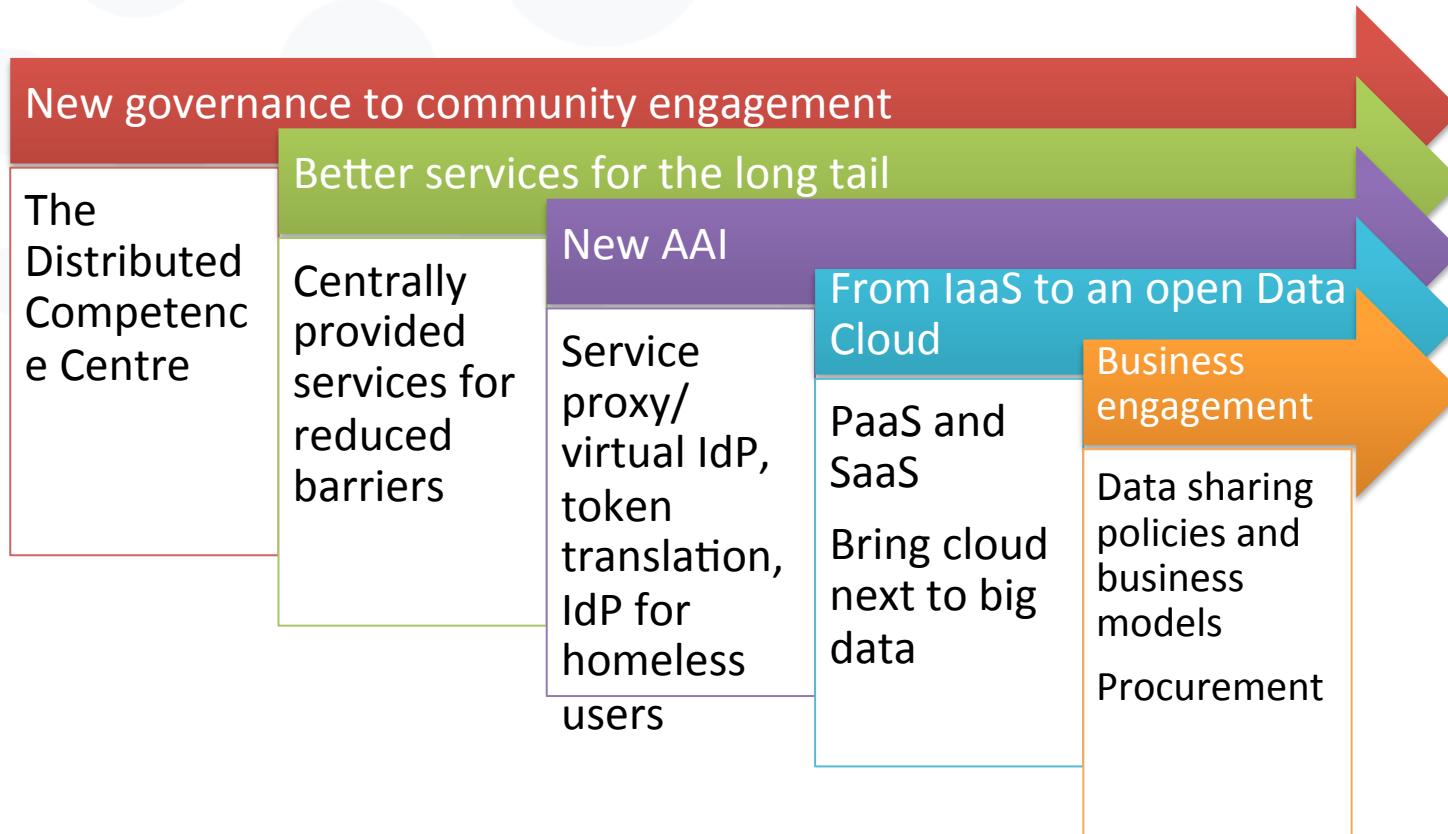
From IaaS to an open Data Cloud

PaaS and SaaS  
Bring cloud next to big data

# Federated Cloud + Open Data: Open Data Cloud

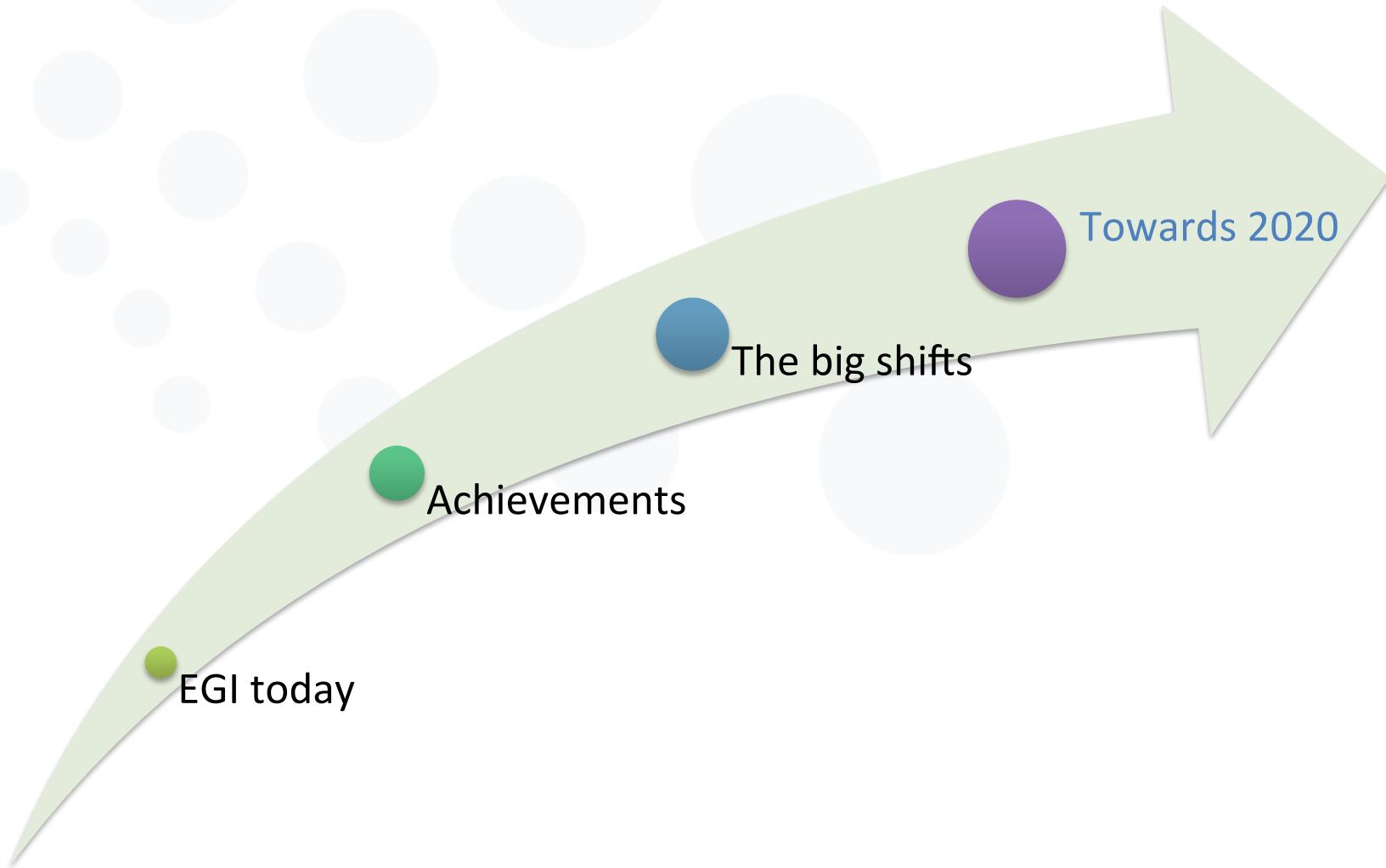
- Objective: scalable access to open research data for discovery, access and use
- Remove policy and technical barriers
  - Bring cloud service next to distributed data repositories
  - Replicate open research data of research/commercial relevance
  - Discovery, accounting
  - Provide PaaS and SaaS and evolve the federation services
  - Virtual appliance library of community tools and data for
    - Repeatability of science, training and education ([EDISON](#))
- Collaboration with [EUDAT](#) and [INDICO-Datacloud](#)
- Multiple stakeholders involved

# The big shifts



# Policy and business

- Pay-for-use and cross-border procurement
- Facilitate collaboration with SMEs (focus on consumer side) via a model to be adopted and adapted for a wider number of NGIs/Resource Centres
  - Use cases from agriculture, fishery and marine sciences, biodiversity, earth science
- Explore with SMEs opportunities and threats around the Open Data and co-develop business models for their exploitation
  - Market analysis and user requirements
  - Data Sharing Policies and Legal Aspects



# Digital ERA – State of play 2015

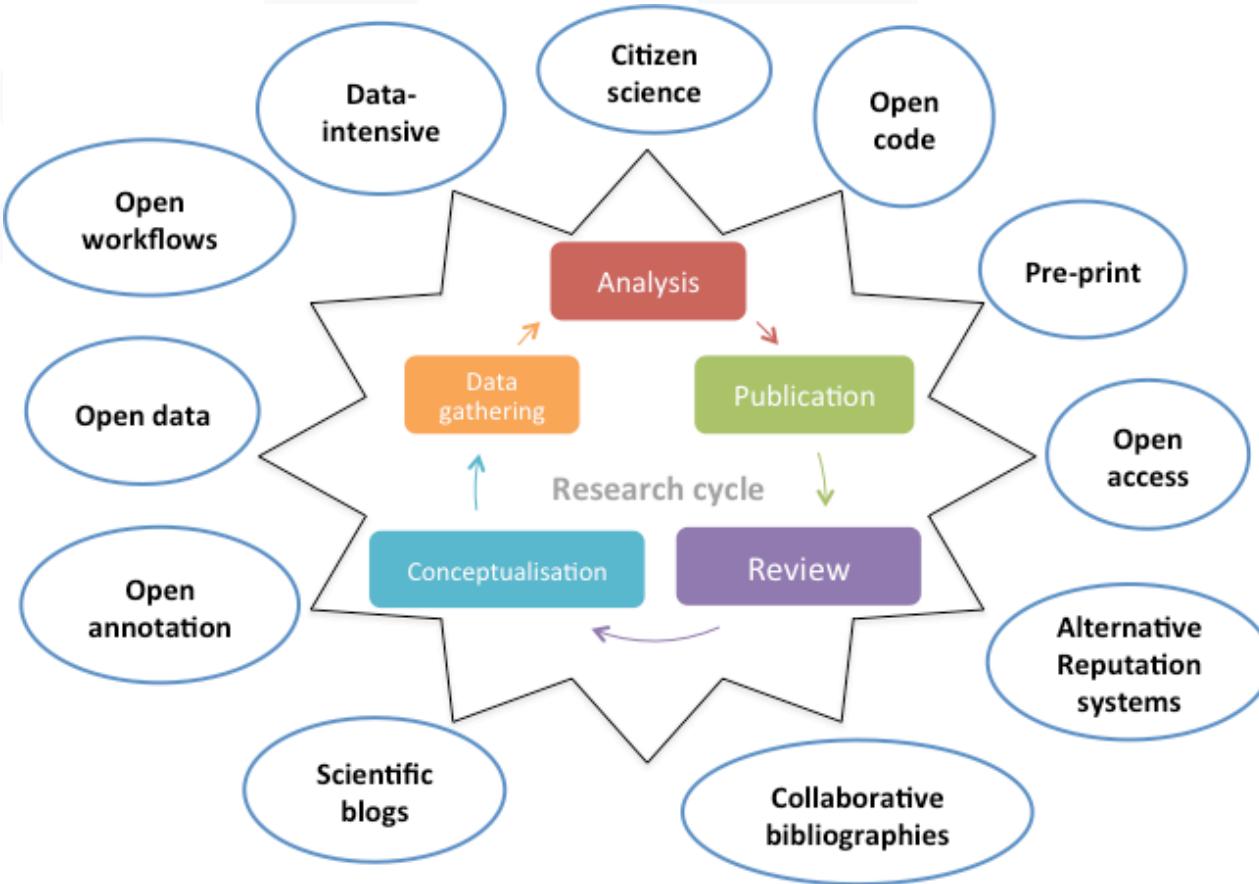
- Incomplete national roadmaps for Research and e-Infrastructures
  - E-Infrastructures and RIs should be components of the same research system
- e-Infrastructure Commons not fully achieved yet
  - Lack of e-Infrastructure capacity for multidisciplinary research and the long tail of science
  - Different access policies for user groups in each access
  - Incomplete technical interoperability, different access policies
  - The “Commons” governance principle not widely adopted
  - Non organized landscape of multiple service providers and research communities, lack of cross-border procurement/funding scheme that allows coordinated resource management across Europe (except for GEANT)
- Lack of one ‘backbone’ of European ICT capabilities

# Developing an Open Science Commons

- Defining Open Science
- Defining the Open Science Commons
- Developing an Open Science Commons

# Open Science

**Opening of the creation and dissemination of scholarly knowledge** towards a multitude of stakeholders, from professional researchers to citizens



Source: <http://goo.gl/uO9MK5>

Greater social value



# Defining Open Science: 5 Schools of Thoughts

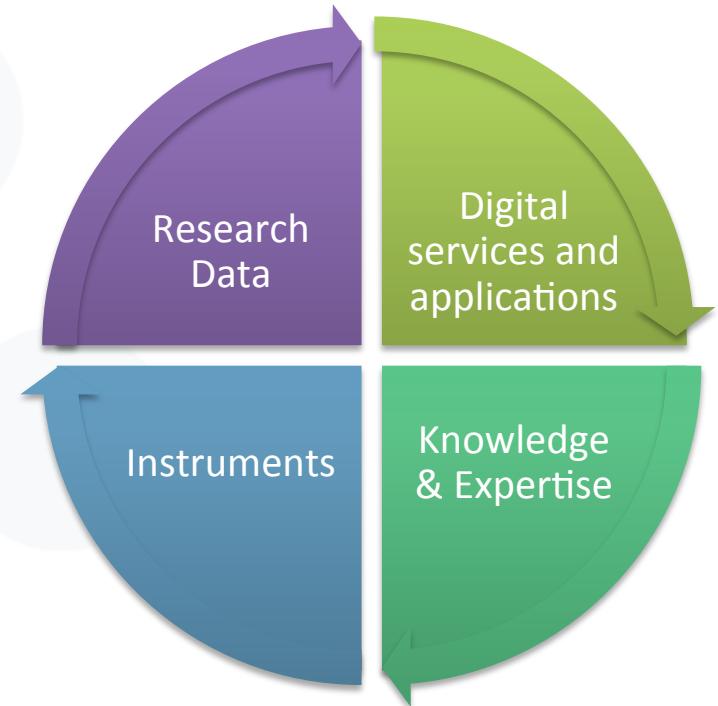
School of thought	Involved groups	Central assumption	Central Aim	Tools & Methods
Democratic	Scientists, politicians, citizens	The access to knowledge is unequally distributed	<b>Making knowledge freely available for everyone</b>	Open access, intellectual property rights, Open data, Open code
Public	Scientists & citizens	Science needs to be made accessible to the public	<b>Making science accessible for citizens</b>	Citizen Science, Science PR, Science Blogging
Infrastructure	Scientists & platform providers	Efficient research depends on the available tools, applications <b>and shared infrastructures</b>	<b>Creating openly available platforms, tools and services for scientists</b>	Collaboration platforms, tools, <b>computing platforms</b>
Pragmatic	Scientists	Knowledge creation could be more efficient if scientists collaborated	<b>Opening up the process of knowledge creation</b>	Wisdom of the crowds, network effects, Open Data, Open Code
Measurement	Scientists & politicians	Scientific contributions today need alternative impact measurements	<b>Developing an alternative metric system for scientific impact</b>	Altmetrics, peer review, citation, impact factor

In red: S. Andreozzi additions

Source: [Opening Science book, 2013](#)

# Open Science: a Complex Resource System

- Shared resources
  - Integrated, easy and fair access
- Engaged communities
  - Participating in the process
  - Culture of sharing
  - Collaborating in the management and stewardship
- Governance
  - Rules to access
  - Rules to resolve conflicts
  - Rules to balance quality vs. openness
- Financial support
  - For long-term availability



# A common endeavor (EU perspective)

EMBL



**Orfeus**



Research data



Instruments



3D  
**EISCAT**

BBMRI-ERIC  
Biobanking and BioMolecular resources Research Infrastructure



**GÉANT**



**EUDAT**

**OpenAIRE**

Digital services and applications

Knowledge & Expertise

Innovation Centres

Centres of Excellence



European Institute of Innovation & Technology

**netherlands eScience center**



Institutionalised community governance of the **production** and/or **sharing** of a particular type of resource (from natural to intellectual)



[GÉANT: European Communications Commons](#)

[Constructing Genome Commons](#)

Wikipedia

...

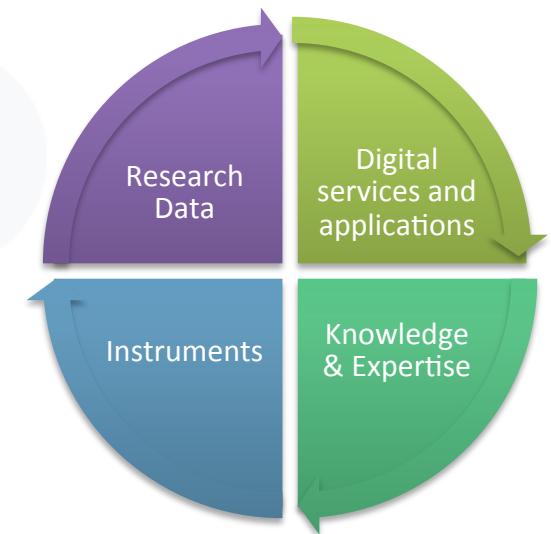
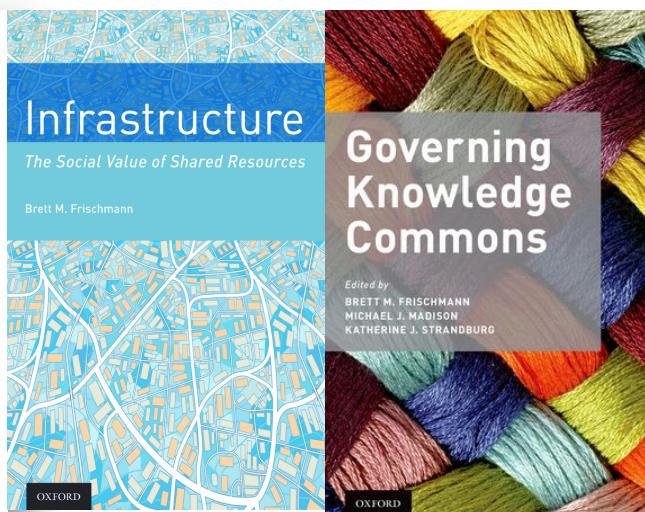
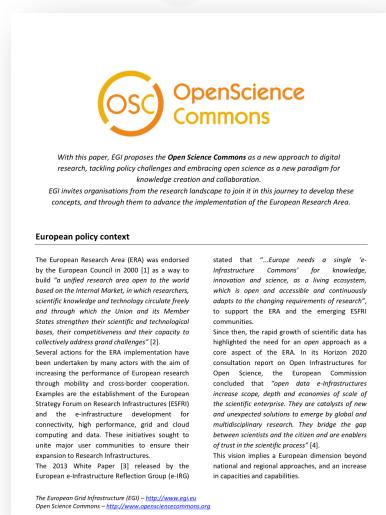
[e-Infrastructure Commons](#)

Linux

Internet

# Open Science Commons: Definition

- A set of interrelated resource systems governed as commons that support the open creation and dissemination of scholarly knowledge
- An area of study in the commons theory applied to open science



website: [www.opensciencecommons.org](http://www.opensciencecommons.org) - paper: <http://go.egi.eu/osc>

*Researchers from all disciplines have easy, integrated and open access to the advanced digital services, scientific instruments, data, knowledge and expertise they need to collaborate and achieve excellence in science, research and innovation.*

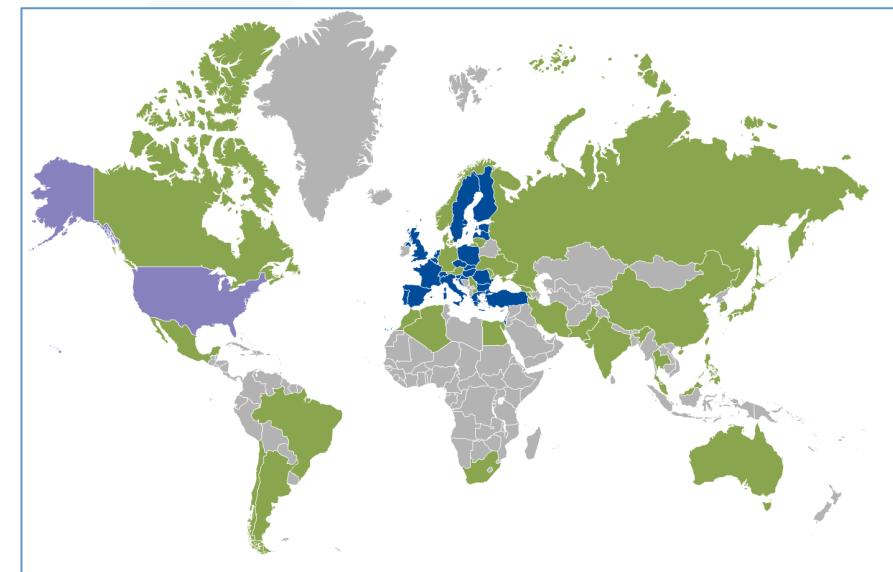
*They feel engaged in governing, managing and preserving these resources for everyone's benefit, with the support of all stakeholders.*

# Key aspects in Open Science

Principles of the Commons	What it means to the Open Science Commons
<b>Shared community resources</b>	Research data, scientific instruments, digital services, software, scientific publications, educational and training, expertise
<b>Community-based rules and procedures with incentives for sharing and responsible use</b>	<a href="#">Access modes</a> are <a href="#">well defined</a> and non-discriminatory (e.g. see charter for open access to RIs)
<b>Governance: the community is part</b>	<a href="#">Governance model</a> with <a href="#">multiple stakeholders</a> , it should include the users of the resources
<b>Long-term, persistent care of resources</b>	<a href="#">Long-term support</a> of funding agencies for stability and sustainability Community committed to manage, preserve

# Commoning in EGI

- **Types of shared resources**
  - Large-scale computing/storage/cloud IaaS-PaaS-SaaS/data services
  - Applications, tools, science gateways
  - Knowledge, expertise, training
- **Rules**
  - Various types of access modes
  - e.g., policy-based, excellence-driven, membership-based
    - Not yet fully harmonized across Member States
- **Governance**
  - Mainly service providers at the moment
  - Evolving to include research infrastructures
  - Advisory board for user communities
- **Funding**
  - National funding agencies, EC, service providers, user communities



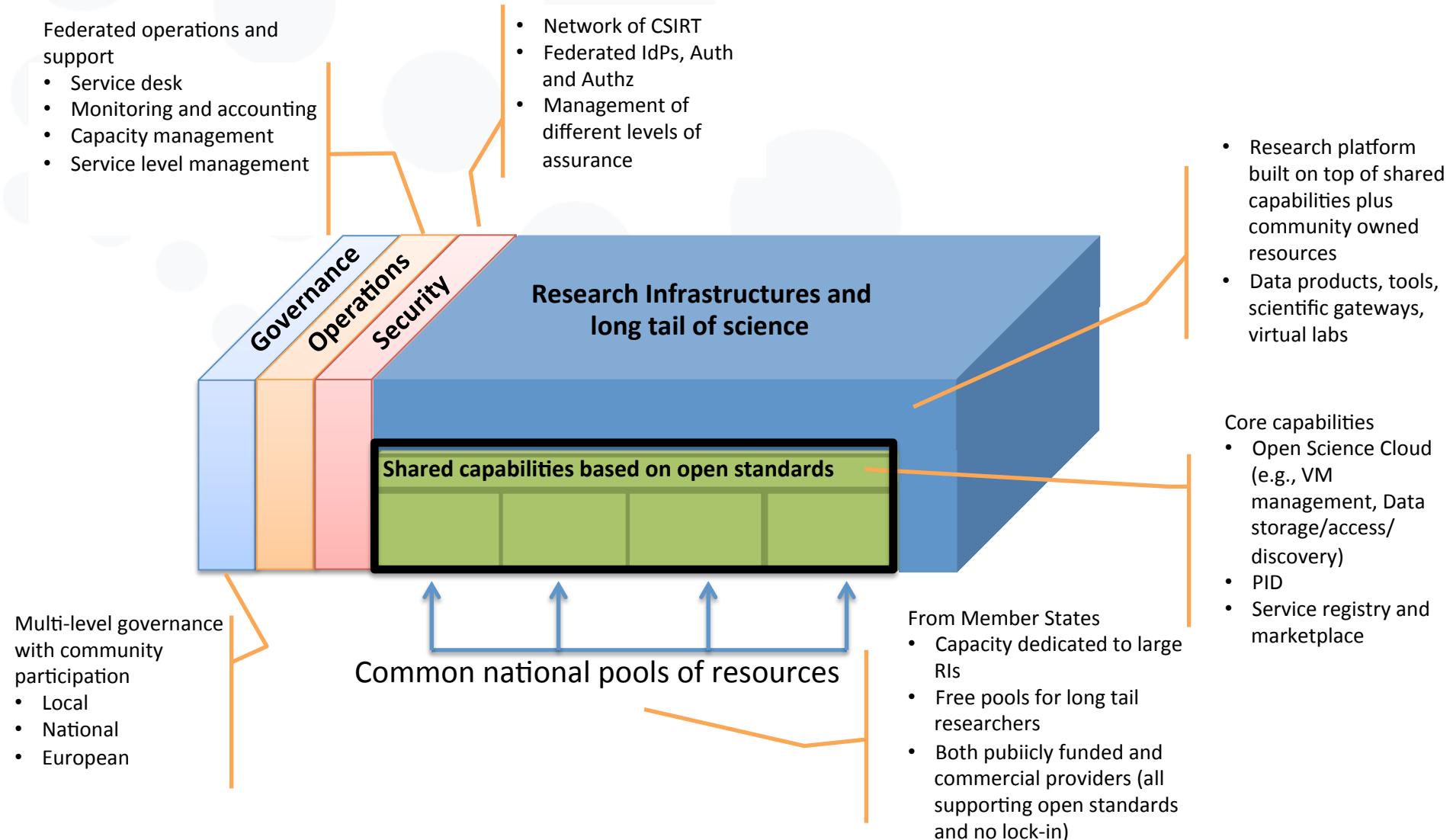
# Developing an OSC (1): Governance structure and funding models

- Analyse governance structure of existing infrastructure/knowledge resource systems in open science
  - Identify best practices and patterns for commoning
  - Develop guidelines
- Define a multi-level governance – European and national – bringing together the different stakeholders including communities
- Identify funding models for sustainability and capacity building

## Developing an OSC (2): Open Science Backbone

- The set of **standardised generic ICT capabilities** across countries and communities offering
  - Compute/Data intensive capabilities
  - Easy discovery, access, use and reuse of open data
  - Shared capacity for RIs and long tail
- Supporting **open standards** (both API and data formats)
- Complemented by community-specific and community-managed services
  - Research Infrastructures

# Developing an OSC (2): Shared Open Science Infrastructure Backbone



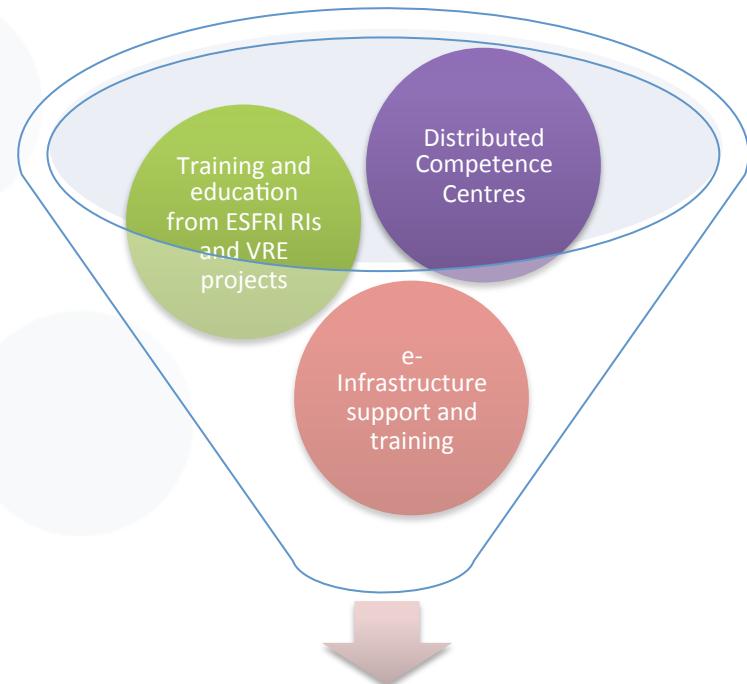
## Developing an OSC (3): Research data

- Stimulate the creation of public repositories of open research data
- Stimulate a culture of sharing and the right incentives to contribute and maintain
- Address legal and policy issues
  - Prefer minimal IPR or non-exclusive licensing
  - Many initiatives exist (e.g. RDA, CODATA)

# Developing an OSC (4): European open knowledge hubs

A coordinated network of competence centers

- Offering knowledge and expertise for scientific software, applications, tools
  - Knowledge and expertise from a network of European training and education centres
  - Scientific software is open, documented, discoverable, supported
  - Support to access different capabilities (HTC, HPC, cloud, open research data, tools, applications, software...)



Open Knowledge Hub

# How EC could contribute

- Fund a **design study** to develop guidelines and best practices on “commoning” resources (infrastructure, data, people) for science
- Coordinate with Member States to **include e-Infrastructures into national roadmaps** for contributing to an open science infrastructure backbone
- Stimulate the development of **digital capabilities** that are based on **open standards** and **certified**
  - Avoid proprietary enclosures

# How Member States could contribute

- Prioritise investment to further develop and operate e-Infrastructures (as already done for ESFRI)
  - Contributing to the open science infrastructure backbone
  - Includes capacity building
- Stimulate sharing of infrastructure and knowledge resources
- Develop various access policies for different research segments including the long tail
  - Excellence drive, policy based, market based, ...

# How Research and e-Infrastructures could contribute

- Open by design
  - Open API, open formats, open licenses, open tools, ...
- Strengthen collaboration via joint strategies, events, common roadmaps and working groups
- Design a governance according to the commons principles
- Ensure an healthy development of shared capabilities (e.g., open innovation processes)

# How can EGI contribute?

Federate digital capabilities, resources and expertise



Operate services across the federated infrastructure



Co-create and integrate open and user-driven services  
and solutions



Be a trusted adviser on data and compute intensive  
science

## Future steps

- Broaden the dialogue on OSC to more stakeholders
- Establish a framework of discussion that encourages participation and contribution
- Identify/analyse other key building blocks that need development

# Conclusion

- Open Science
  - Needs a complex resource system of shared infrastructure and knowledge resources
  - Needs a ‘whole’ approach
- Open Science Commons
  - Developing a commons-oriented approach when designing systems
  - Ensuring maximization of return from public investment
  - Initial elements for Europe
    - An open science infrastructure backbone for network, computing and data
    - Stimulating research data as commons
    - Networks of knowledge hubs for skills development, innovation and expert support



<https://www.opensciencecommons.org/getinvolved/>

# Thank you for your attention.

**Questions?**



[www.egi.eu](http://www.egi.eu)



<http://cf2015.egi.eu>

This work by Parties of the EGI-Engage Consortium is licensed under a  
[Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

