European Grid Infrastructure

Services of big data processing in a secure federated environment

Tiziana Ferrari/EGI.eu

Brussels, 08-05-2014

e-infrastructure

Outline

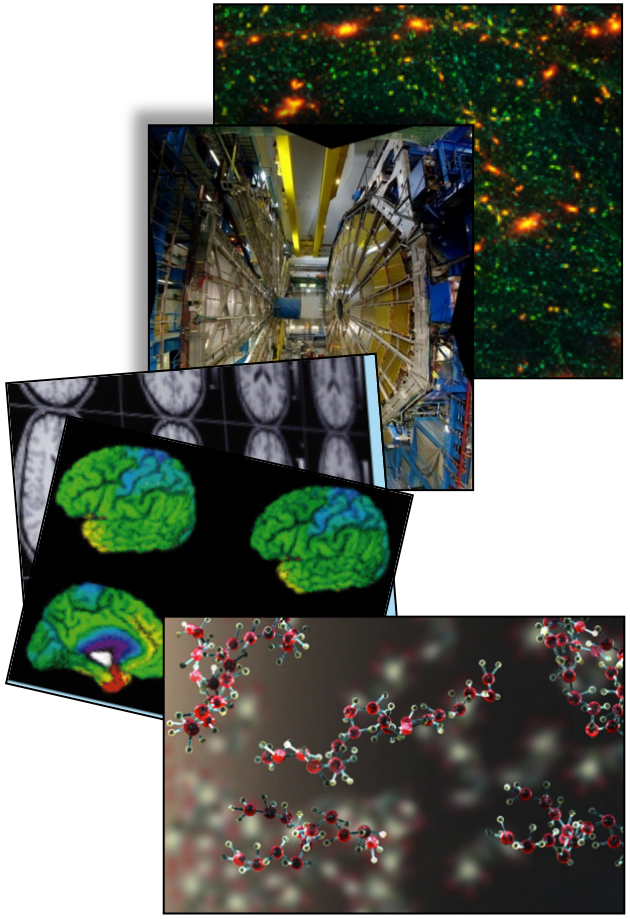
- Current and new services
- Vision
- Sustainability of e-Infrastructures and RIs

EGI Mission

MISSION: To support international researcher collaborations from all disciplines with the reliable and innovative ICT services they need to accelerate excellent science

- Natural and physical sciences
- Medical and health sciences
- Engineering and technology
- …

EC **EGI-InSPIRE** project (2010-2014)



http://www.egi.eu/case-studies/

Principles

- Uniform access to heterogeneous data and compute services
  - Grid and Cloud platforms
- Federation of services from
  - Publicly funded infrastructures
  - Institutional infrastructures
  - Commercial providers (partnership with Helix Nebula)

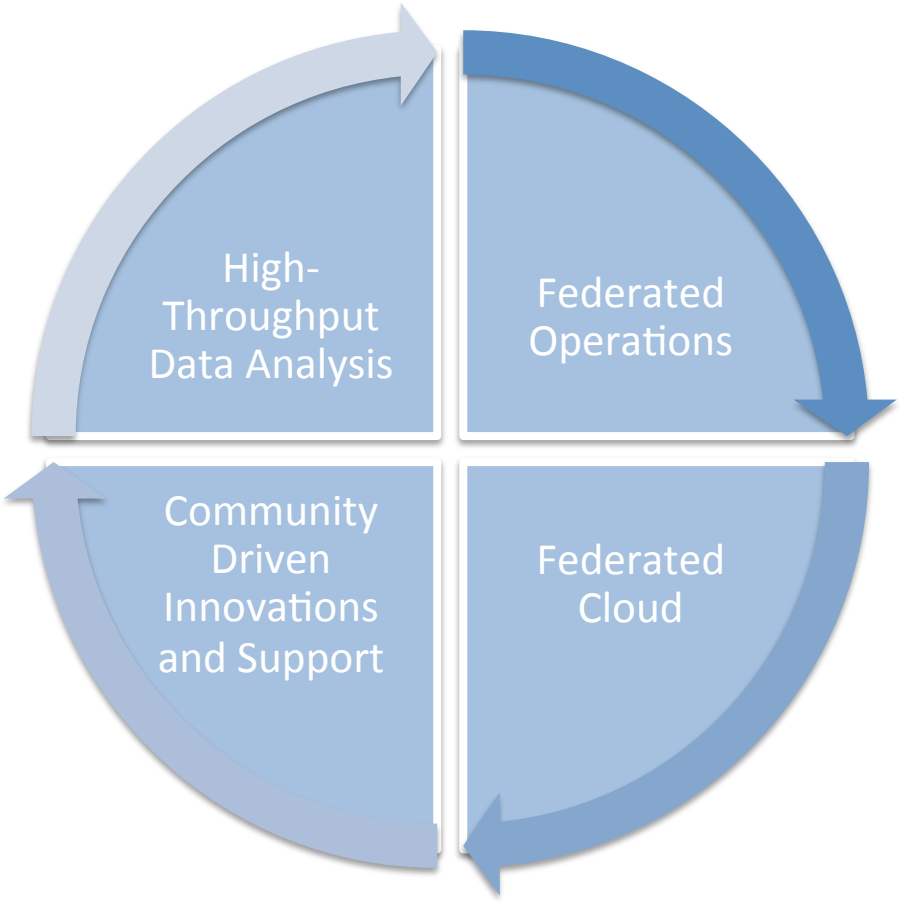
  → free at point of delivery/pay per use

EGI Services for long tail and big science

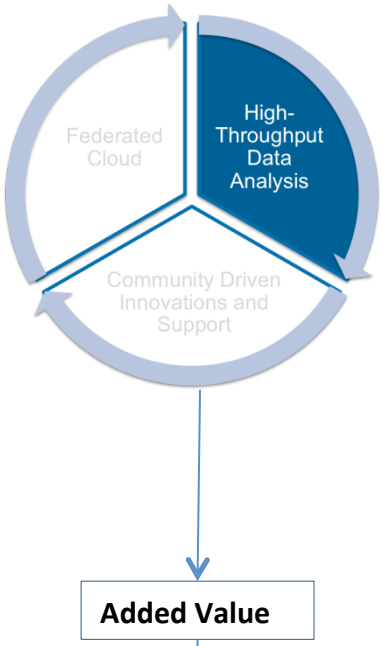
Individual Researchers & Teams

Research Communities & Institutions

Resource Centres & Service providers



Services for High-Throughput Data Analysis



For management and analysis of large datasets and execution of thousands of computational tasks

European federation of publicly-funded clusters

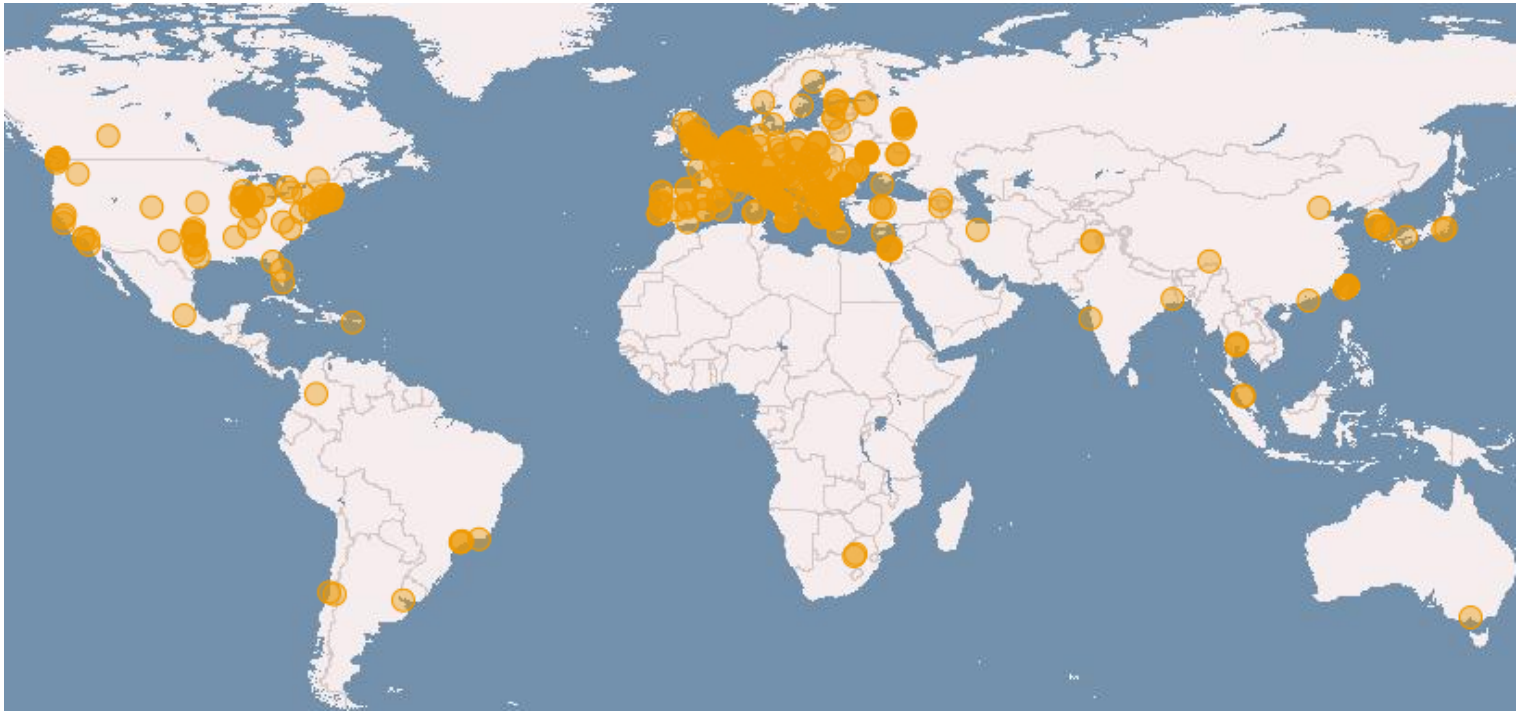
- **Grid Compute**: computational jobs
- **Grid Storage**: store/access/retrieve files
- **Data Management**: e.g. metadata catalogue, file transfer service

Based on **open standards** and **open source software**

Integrate heterogeneous infrastructure/technologies

- **Uniform access** to distributed computing capabilities to run large-scale computational jobs processing big data and preventing single vendor lock-in
- Possibility to **federate your own resources**
- Facilitate **collaboration** across communities and borders by **sharing compute and data**

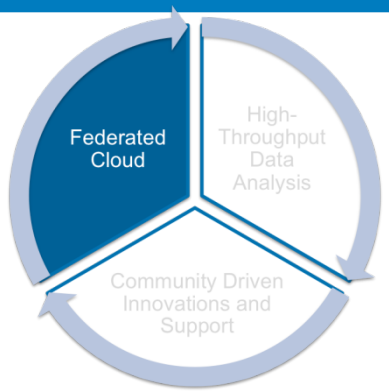
European Grid Infrastructure



Distributed and federated data and computing facilities
Grid and Cloud compute platform
340 data centres in 34 National Grid Initiatives/EIROs
435,000 logical CPU cores

10 years of support to science
> 200 research projects
190 PB disk, 180 PB tape
1.6 M job/day
> 99.6% reliability

Federated Cloud



Added Value

Infrastructure to deploy on-demand IT services for managing and processing your research data?

European federation of publicly-funded community clouds
- **Cloud Compute**: deploy/manage virtual machines (VM)
- **Cloud Storage**: store/access/retrieve digital objects
- **Image Repository**: store/retrieve certified VM images

Based on **open standards** and **open source software**

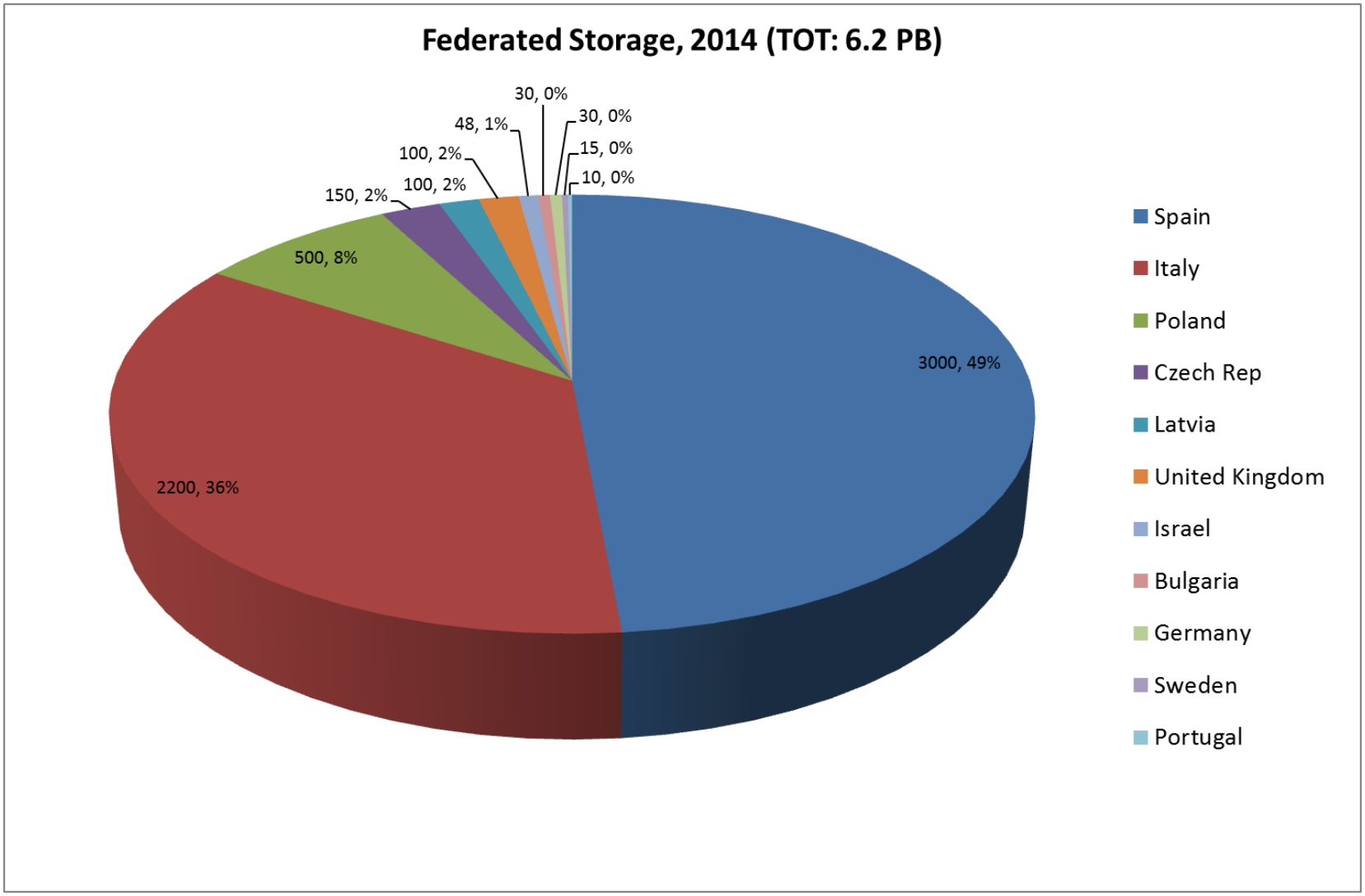
Integrate heterogeneous cloud technologies
- OpenStack, OpenNebula, CloudStack,…

Integrate with commercial providers
- Within EGI or through the Helix Nebula Marketplace

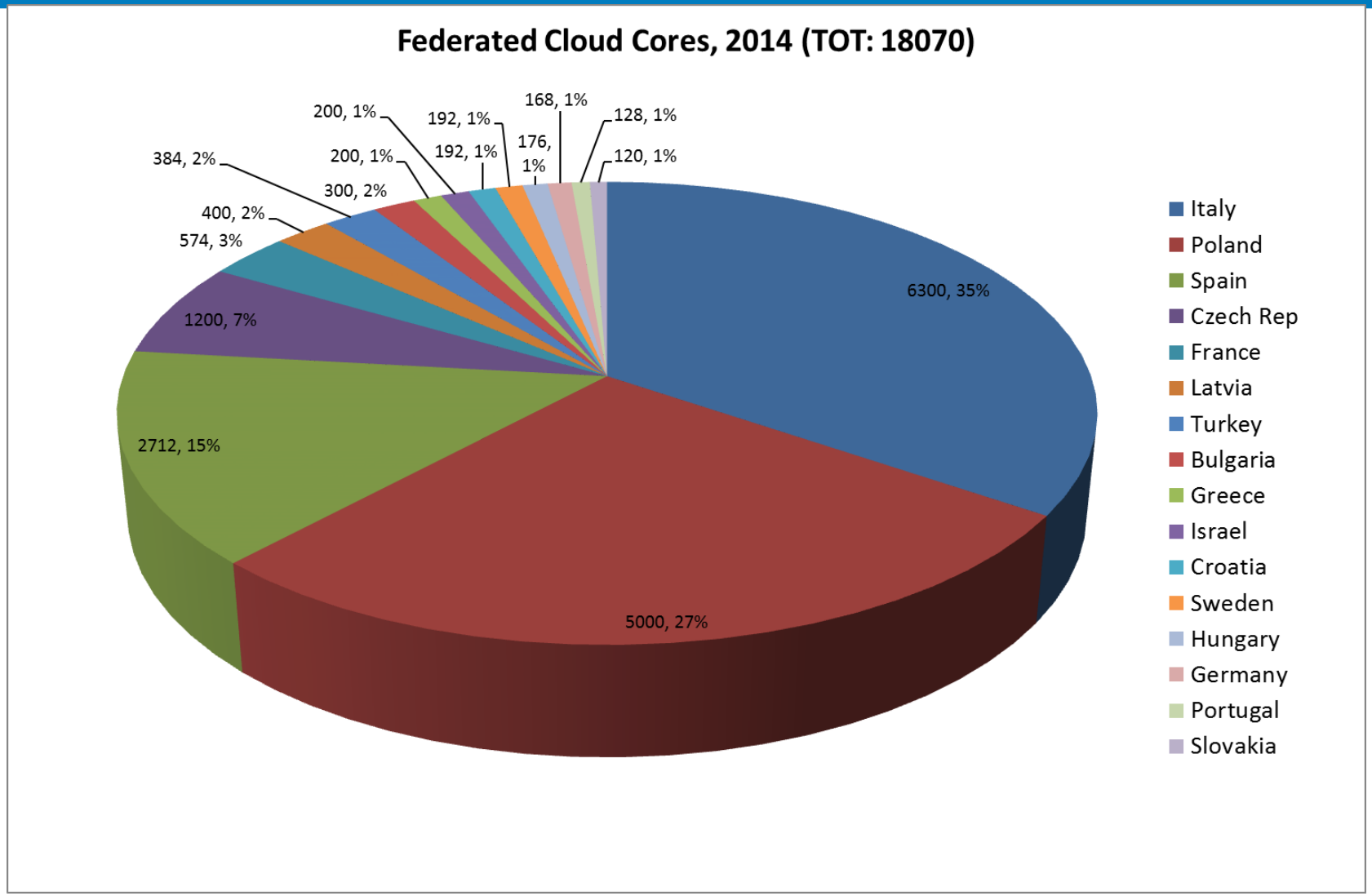
- Uniform access, **no lock-in** and **on-demand scale out** capabilities
- Easy deployment of **own/customised services**
- Possibility to **federate** existing institutional clouds
- **Efficiency** by **co-locating big data + cloud computing**

Launch into production: May 2014

Federated Cloud/Cloud Storage



Federated Cloud/Compute IaaS



Services for federated operations

- Activities and tools for the operations of distributed services
  - Central operations tools (message brokers, operations dashboards, VO management, service and security monitoring, service registry)
  - Federated accounting (distributed repositories and portal)
  - Technical support and incident management
  - Security operations coordination, policy development, software vulnerability
  - Software distribution, verification, validation

Innovation and Support

- Call for participation of RIs to Competence Centres → MAY 2014
  - Aim: support the integration of RI virtual research environments with EGI compute/data services, user co-development of new service features
  - from design to pre-production and deployment
  - Participated by users, technology providers, service providers
  - Get in touch with us: egi-engage (at) mailman.egi.eu

Long-term vision for RIs

- One European HTC and cloud infrastructure for RIs
  - technically integrated with EUDAT and PRACE and world-wide
  - complemented by commodity services from commercial providers
- Distributed network of competence centres

Discussion

- Sustainability of e-Infrastructures and RIs
  - ensure the long-term availability
  - operations and innovation
- European coordinated action for procurement of services
  - Economies of scale in sharing costs of operations and innovation