Introduction to EGI

Diego Scardaci

diego.scardaci@egi.eu

Technical Outreach Expert

(EGI User Community Support Team)

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

EGI-Engage

Outline

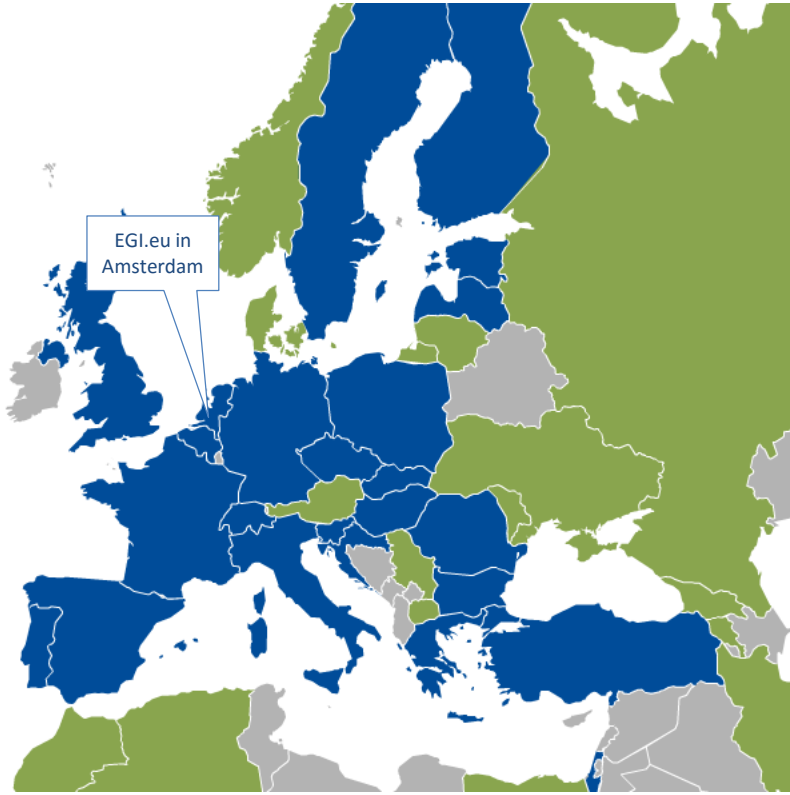
- Introduction to EGI
- EGI Communities
- EGI Technology
- Conclusions

EGI-Engage

Introduction to EGI

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



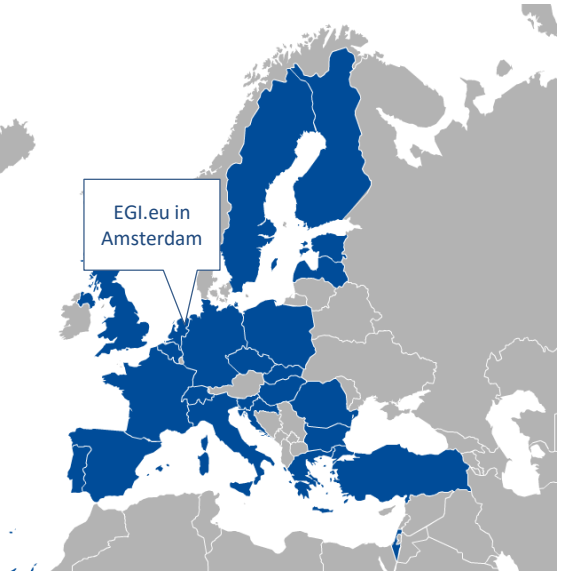
Participants

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

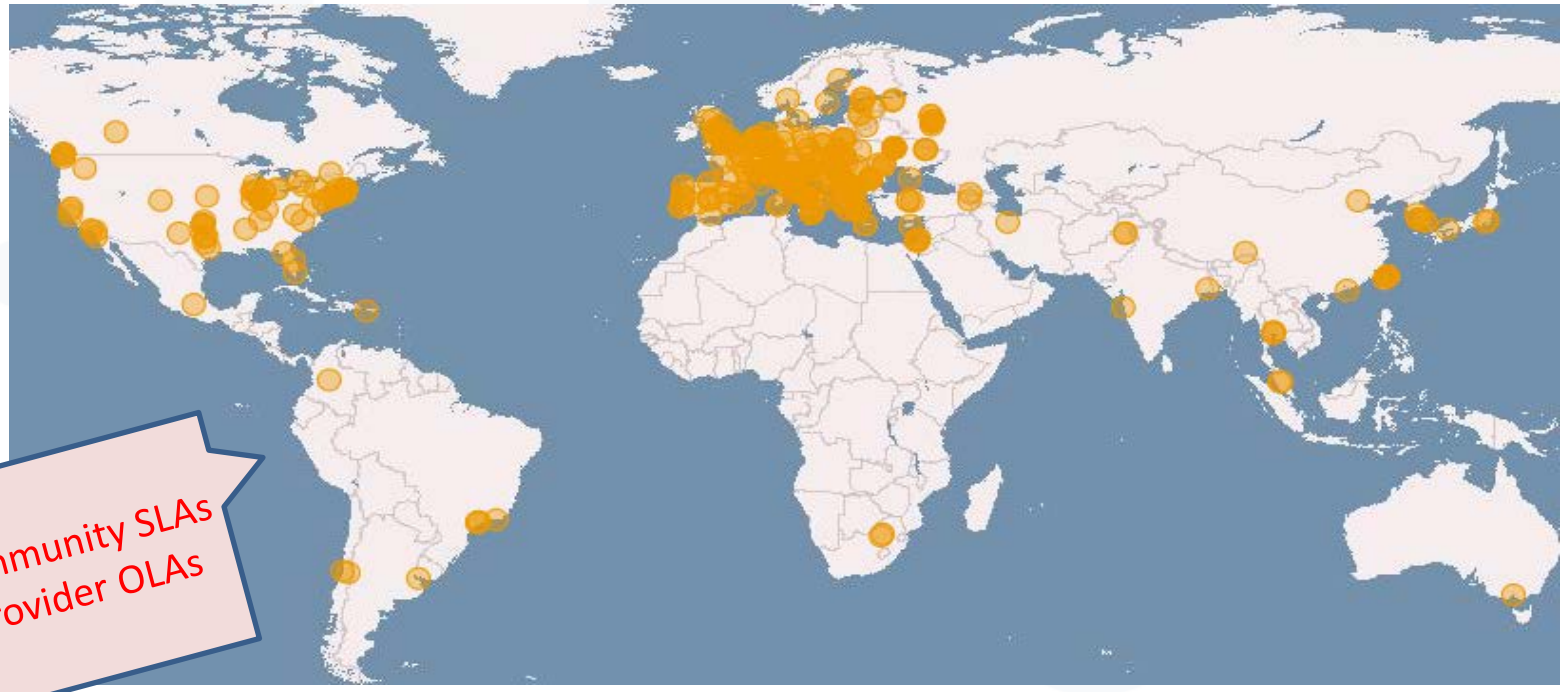
Under discussion

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

- **EGI = Infrastructure**
  - Federation of 340 Resource Centres across 54 countries
  - Provides distributed computing and storage resources to accelerate data-intensive research
- **EGI.eu = Coordination Body**
  - Coordinator of the EGI federation
  - Non-profit foundation based in Amsterdam (~20 staff)
  - 25 participants (e.g. NGIs, EIROs) form governing body (EGI Council)
- **EGI-Engage = EC-funded project**
  - H2020 project started in March 2015, for 30 months
  - Accelerate the implementation of the Open Science Commons
- **Other projects with EGI.eu membership**
  - INDIGO-DataCloud (from May 2015), AARC (from May 2015), EDISON (from September 2015)
  - BioMedBridges, FedSM, CloudWATCH, Civic Epistemologies
- **Partner projects**
  - Partnership formalised with an MoU
  - E.g. technology provider; User community; Resource provider; etc.



Enabling Global Infrastructures



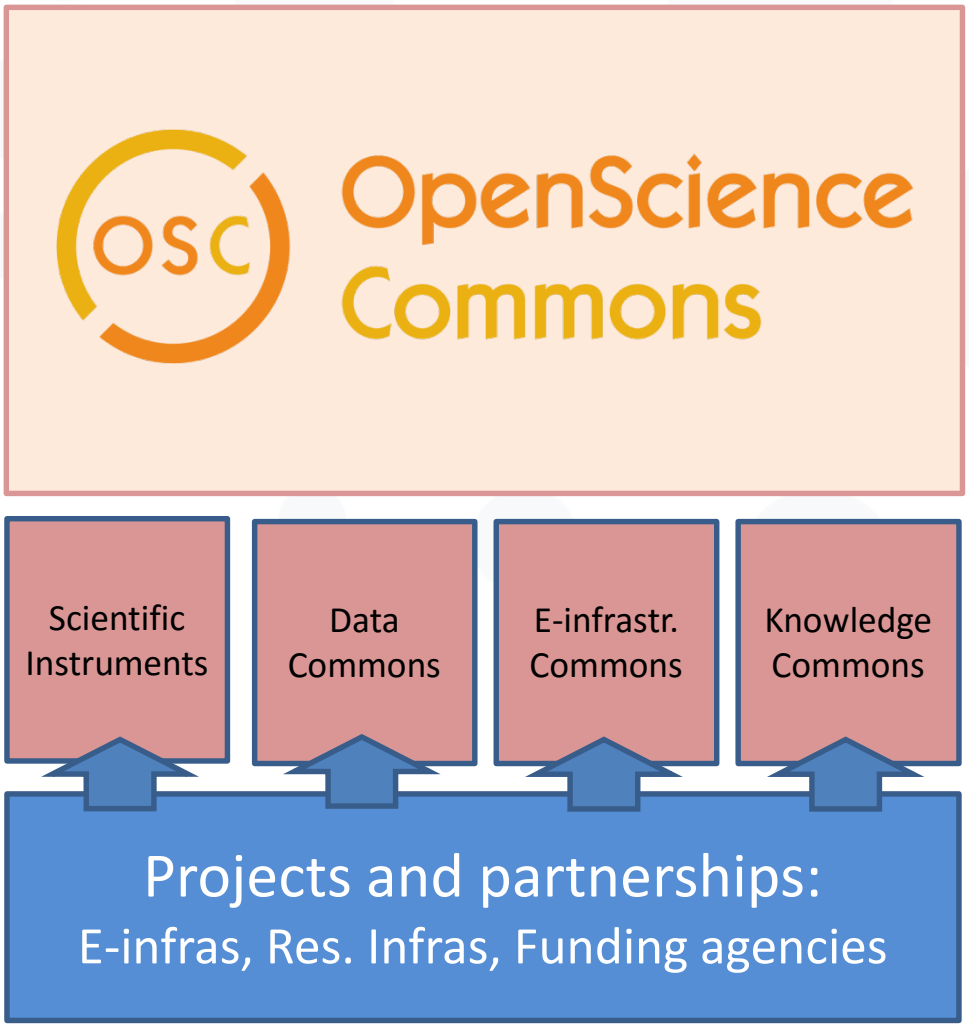
Community SLAs
Provider OLAs

- Distributed, federated storage and compute facilities
- Compute platforms (Grid, Cloud)
- Virtual Research Environments
- > 200 user research projects

Total capacity (grid + cloud):

- 340 resource centres in 54 countries
- 620,000 logical CPU cores
- 270 PB disk, 220 PB tape

Implementing the Open Science Commons



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 to achieve excellence in science, research and innovation.

www.opensciencecommons.org

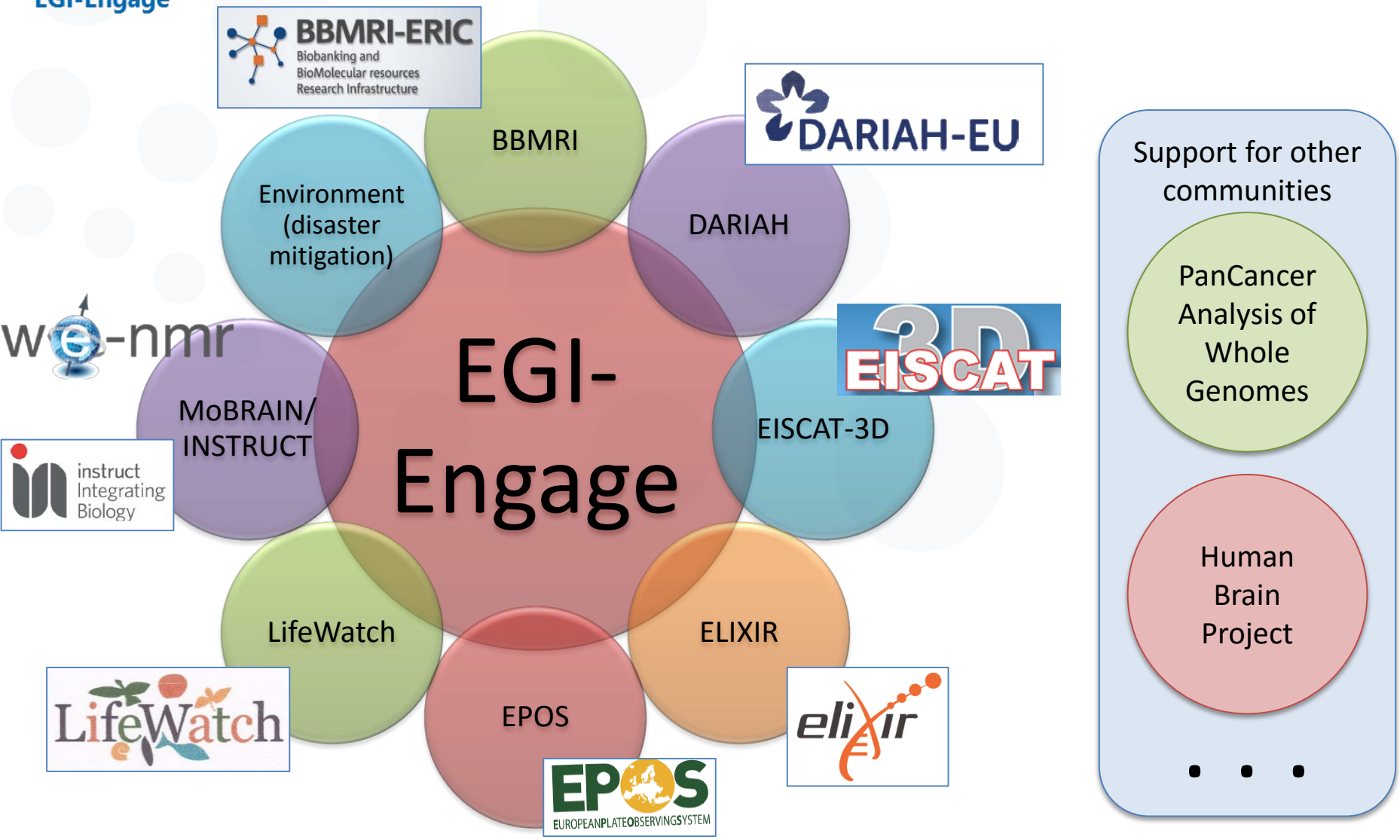
EGI-Engage H2020 project:

- 30 months, Start: 1/March/2015
- 42 partners
- 8 m Euro EC contribution

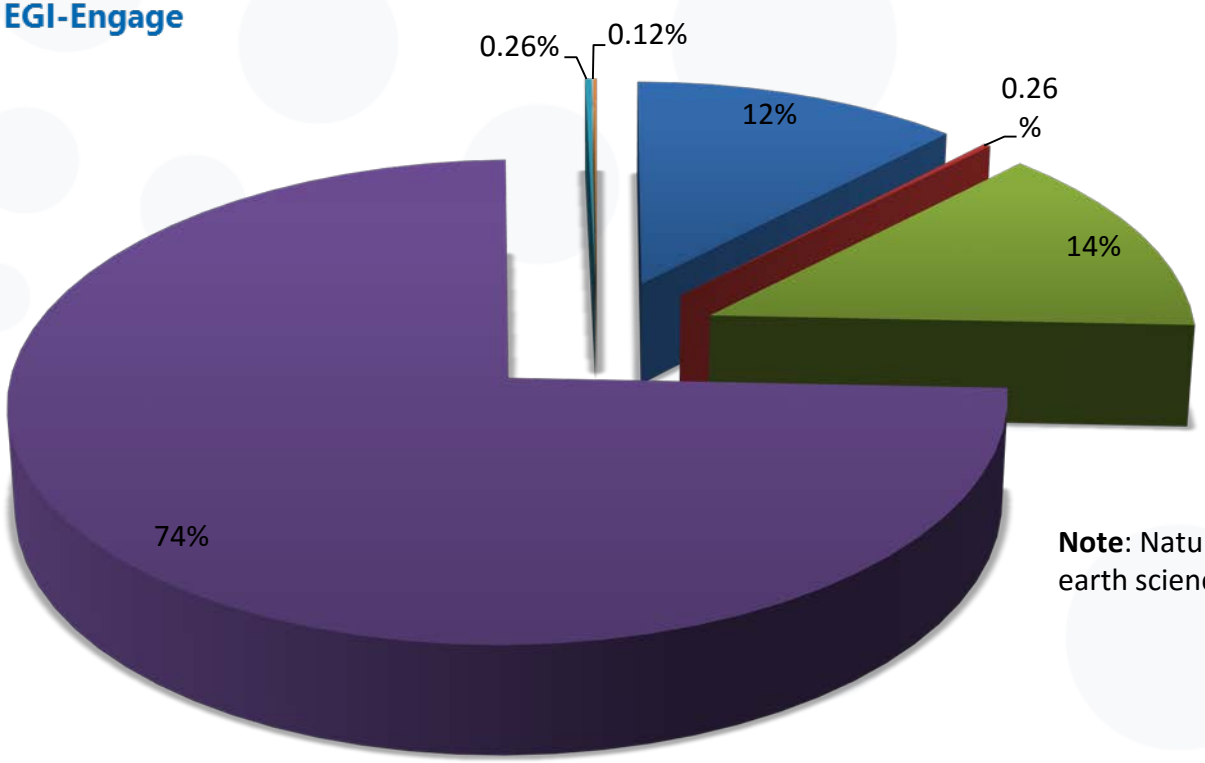
EGI-Engage

EGI Communities

EGI-Engage Competence Centers



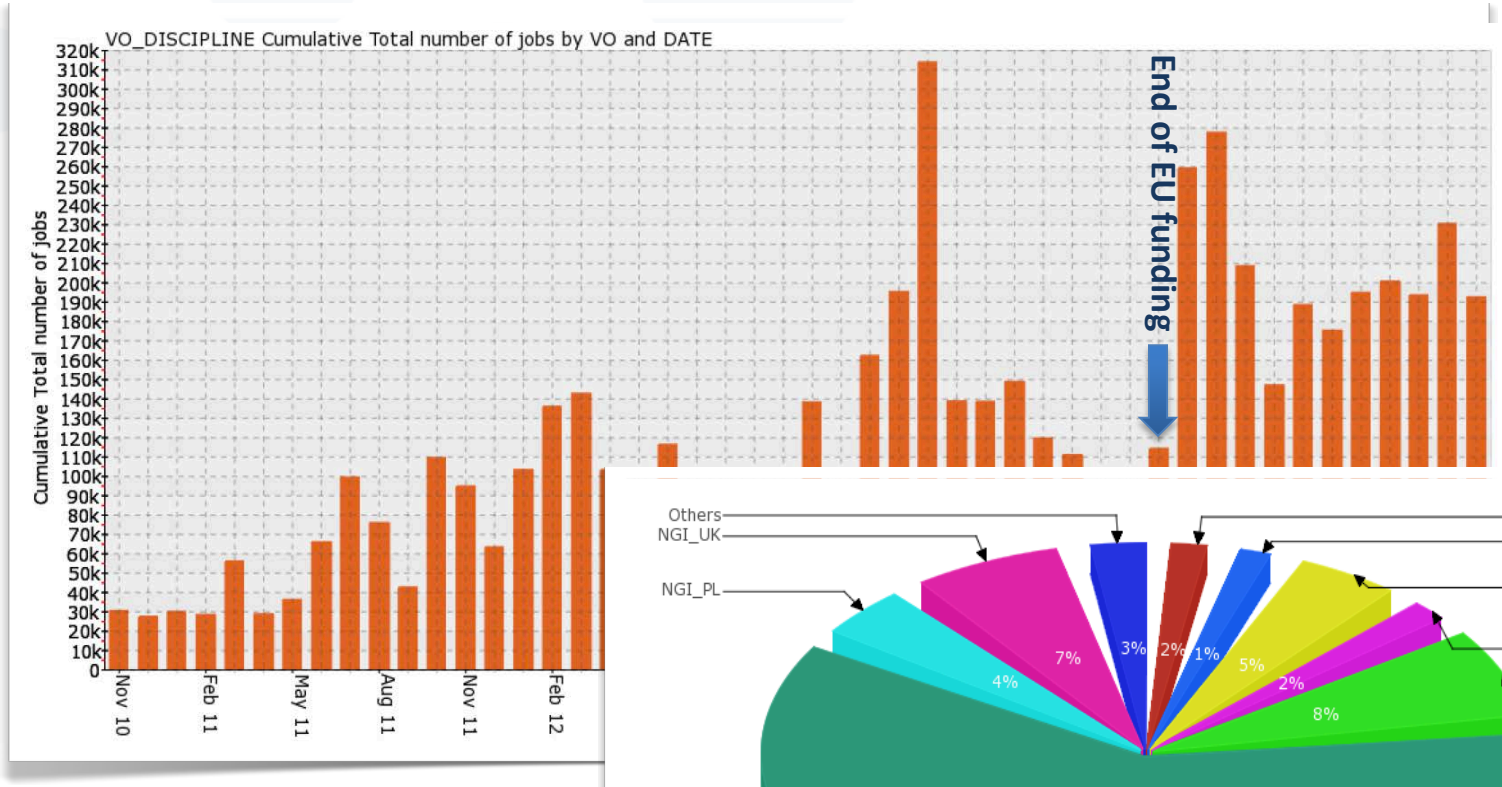
EGI User Communities



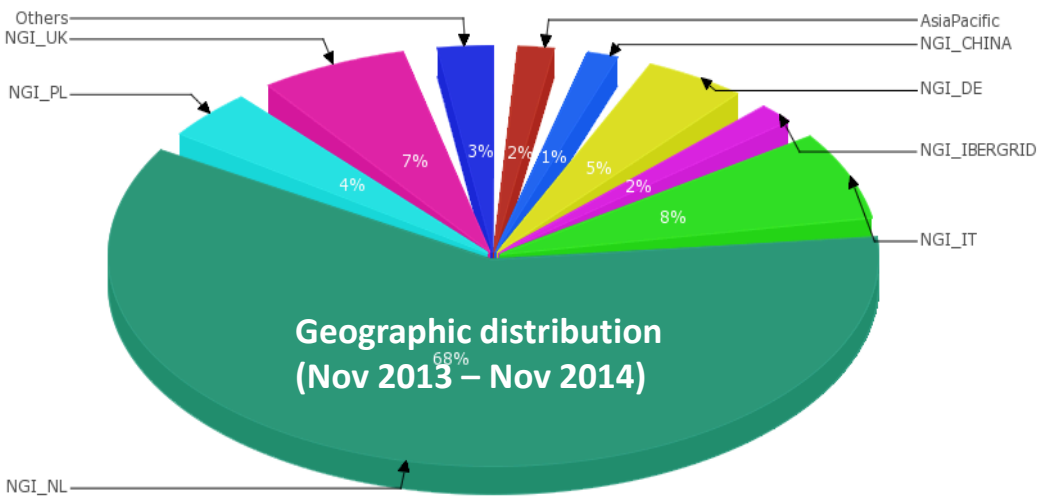
Note: Natural Sciences include physics, chemistry, earth science and some biology, etc

CPU time usage of EGI Scientific User Communities from the 03 - 08 2015

EGI Usage statistics

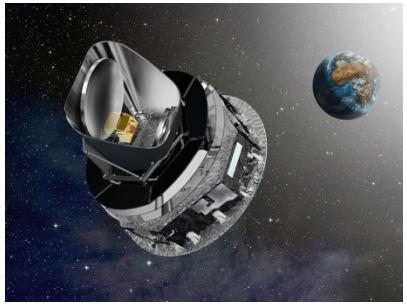


Excluding OSG resources

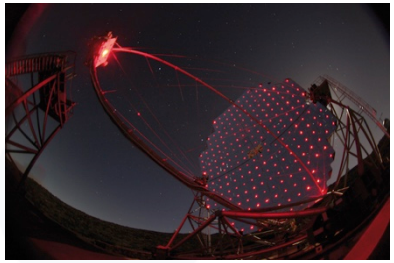


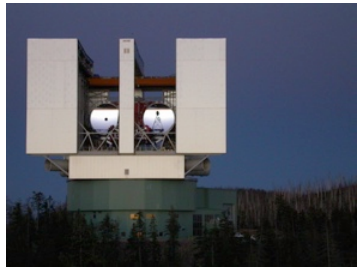
Enabled via EGI federated operation

A&A Heavy User Community



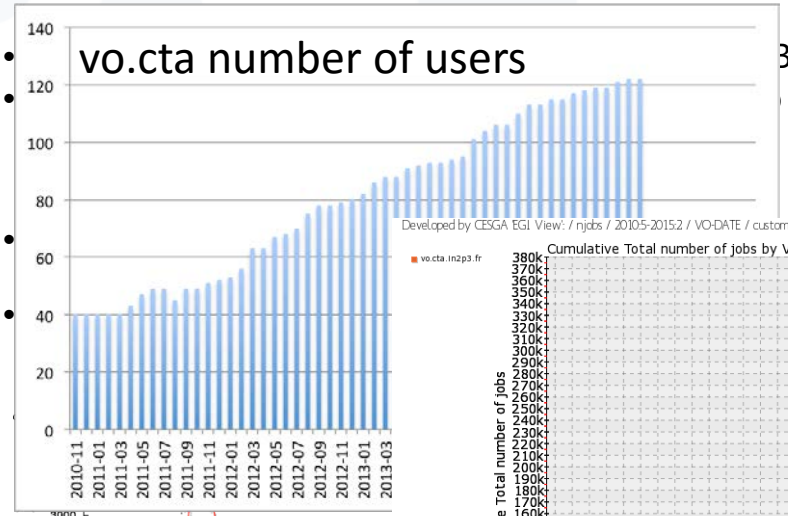
- Research communities:
  - Radio Astronomy, Gamma Ray Astronomy, Helio-physics, Stellar Astrophysics, cosmology….
- The A&A Virtual Organizations
  - 10 with
  - about 500 users
  - about 80000 cores
  - about 100 TB storage
  - More in other VOs

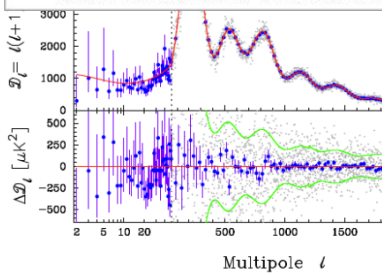


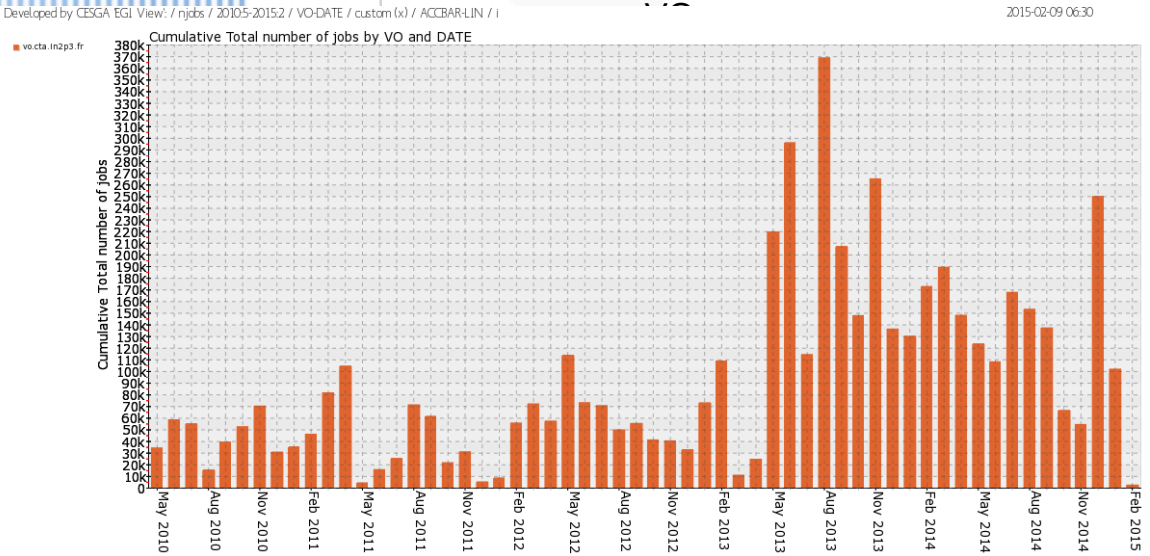


- A light weight coordination
- The A&A **Heavy User Community**
- Identify commonalities and common technical solutions.

Success Stories







- **CTA:**
  - open observatory to a wide astrophysics community
  - explore our Universe in depth in Very High Energy

ires large HTC

LOFAR

- LOFAR telescope:
  - array of simple omni-directional antennas
  - make radio pictures of the sky
  - Wide Area Sensor Network - sensors for geophysical research and studies in precision agriculture
- 7000 antennas arranged in clusters
- Observational data:
  - rates up to 60 Gbps (650 TB per day)
  - once processed, amount of data to be kept for a longer period is significantly reduced
- Pathfinder for SKA

LOFAR use cases in the EGI Federated Cloud

- LOFAR Long Term Archive (LTA)
  - Distributed information system created to store and process the large LOFAR data volumes
  - LOFAR data in the EGI FedCloud
- Evaluation of an innovative calibration pipeline

CANFAR

- Cloud ecosystem for data intensive astronomy
  - National facility for open access
  - Telescope collections:
    - Multiple missions, facilities and wavelengths
    - Pointed and survey observations
    - 12 telescopes
    - 6 advanced data collections
  - Services
    - Archive services & Data curation
    - Community projects
    - Operating primarily on Compute Canada resources
- More than 7000 users and 1000 TB handled in the last year

CANFAR requirements

Combine data from CANFAR and European Astronomy centers:

- unique AAI oriented cloud ecosystem based on CANFAR approach
- IVOA standards
- Open datasets for scientific collaborations and projects

International Virtual Observatory Alliance (IVOA)

- Standardization of data and metadata
- Standardization of data exchange method
- Use of a service registry
- provide competing and co-operating data services between data services

CANFAR use cases in the EGI Federated Cloud

- Two telescopes with the same area of the sky
  - Canada-France-Hawaii telescope (CFHT) data is available at CADC (Canada)
  - Large Binocular Telescope (LBT) is archived at IA2 (Italy)
- Combining data for source detection and cross-matching
- Involves:
  - Authentication
  - Data discovery and programmatic access
  - Data-location-aware virtual clusters
  - Storing results in a project data space
  - Collaborative analysis with science team

EGI-Engage

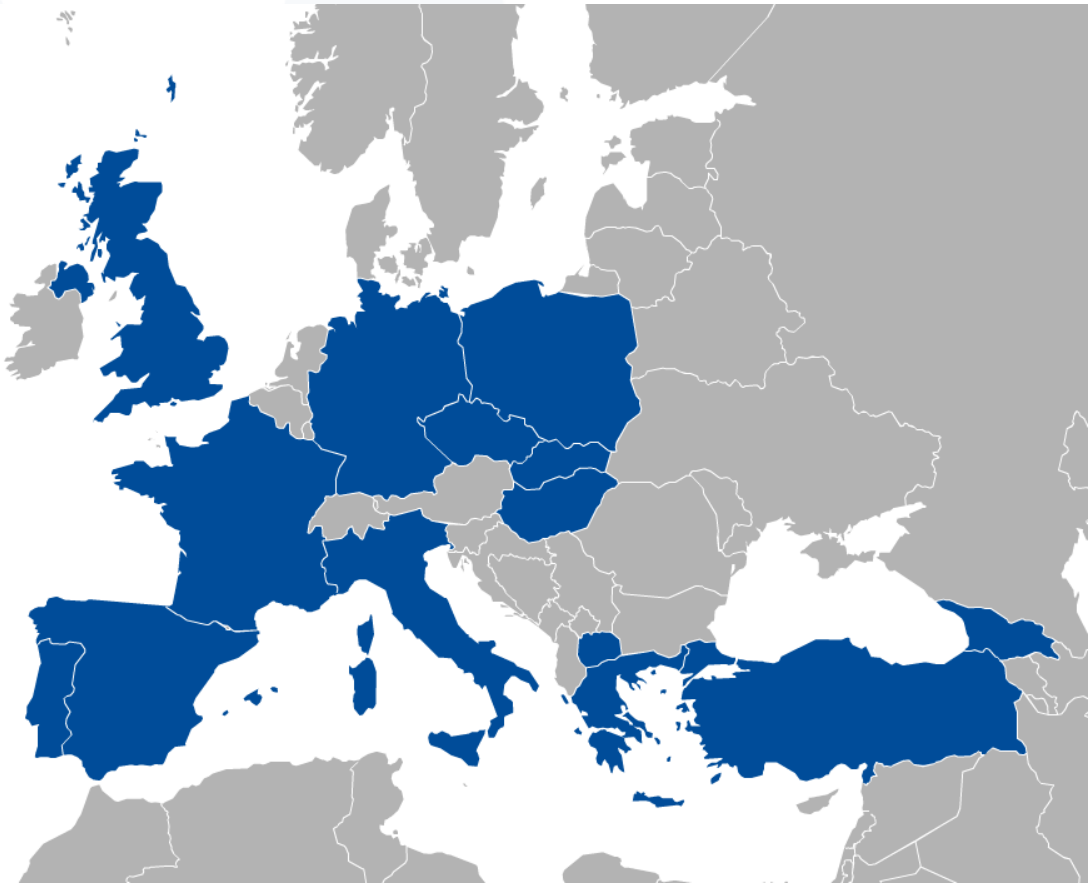
EGI Technology

Platform diversification in EGI

- 2005-2014:
  - Grid platform (gLite, ARC, Unicore, QCG, DG)
- 2014-2020:
  - Federated Cloud platform
  - Long-tail of science platform
  - Open Data platform
  - (GPGPU platform)
  - Container (Docker)
  - Community platforms

EGI Federated cloud

EGI-Engage



IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

INFN Istituto Nazionale di Fisica Nucleare

CYFRONET

IFCA Instituto de Física de Cantabria

LIP

MTA SZTAKI

TÜBİTAK ULAKBİM

Grid y Computación de Altas Prestaciones GRyCAP

bifi

100%IT

SLOVAK ACADEMY OF SCIENCES

grnet Networking Research and Education

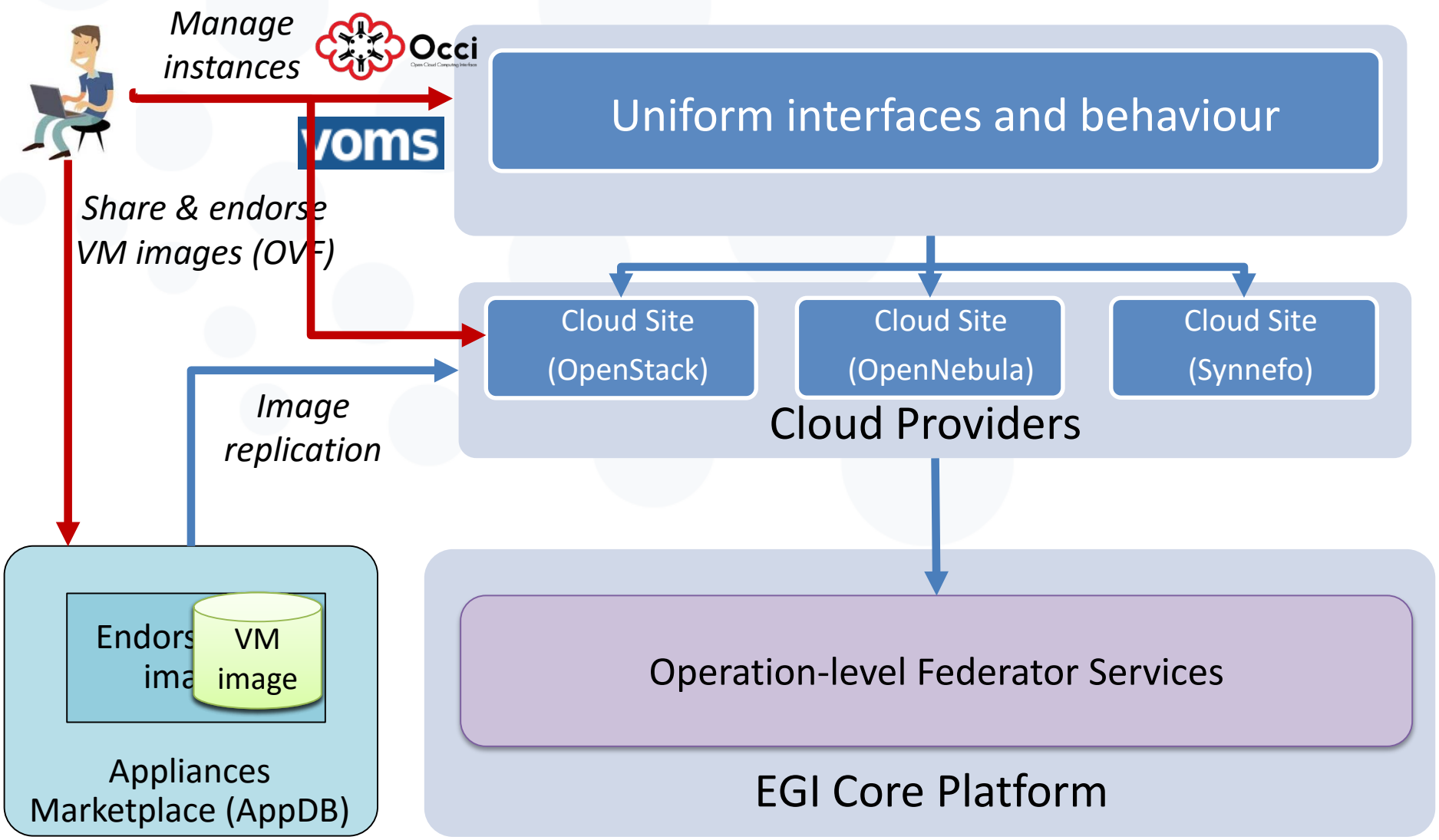
CENTRO EXTREMEÑO DE TECNOLOGÍAS AVANZADAS CETA Ciemat

JÜLICH FORSCHUNGSZENTRUM

MetaCentrum

Ss. Cyril and Methodius University in Skopje FACULTY OF COMPUTER SCIENCE AND ENGINEERING

Architecture of the EGI Federated Cloud

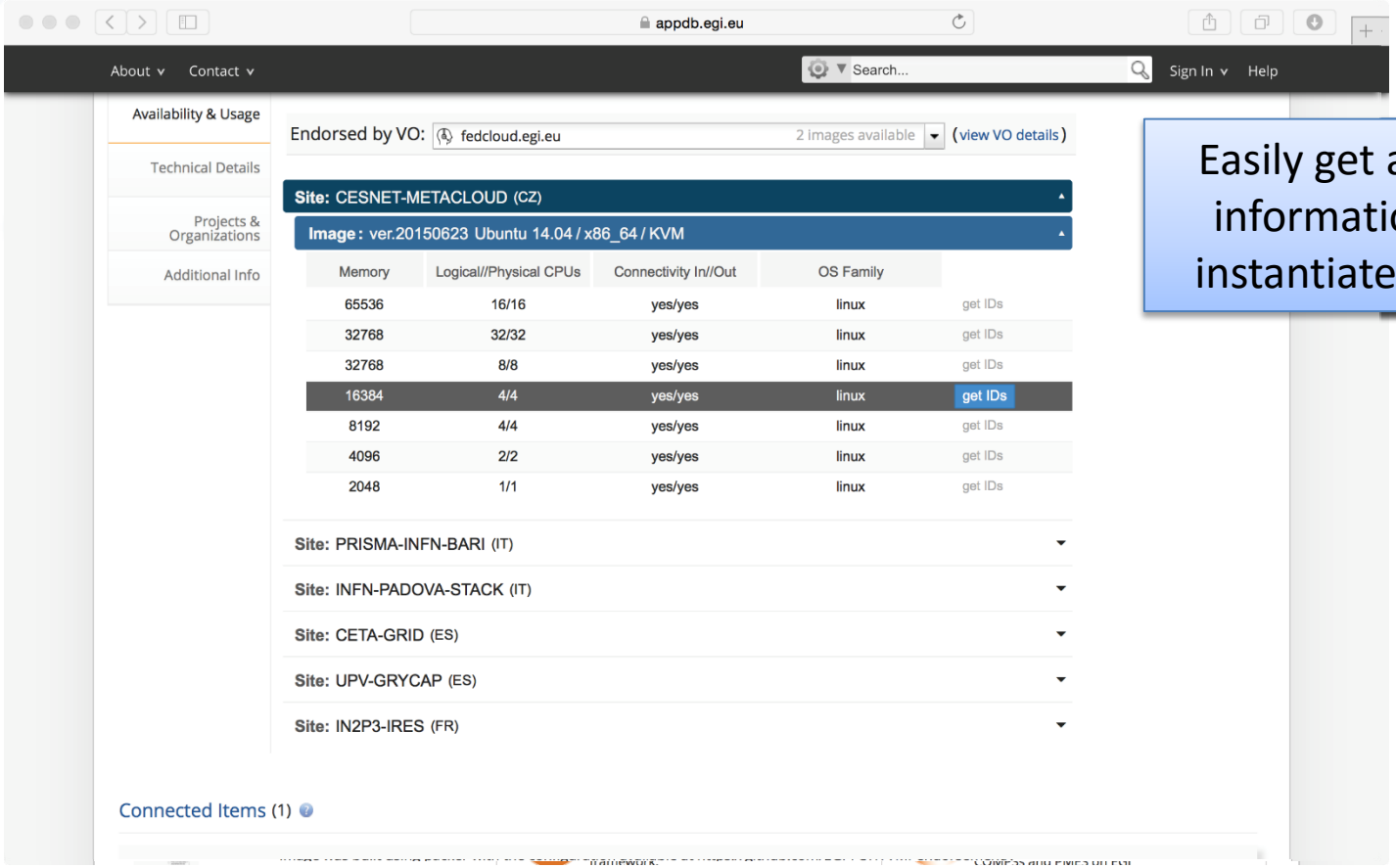


Typical Usage Models

- **Service Hosting**
  - Long-running services (e.g. web, database or application servers)
- **Compute and data intensive workloads**
  - Batch and interactive computing with scalable and customized environments
- **Datasets repository**
  - Store and manage large datasets for your applications
- **Disposable and testing environments**
  - Hosting for demos, trainings, tests with minimal overhead

VM Image Catalogue

Library of Virtual Appliances (bundle of VM images) for use on a cloud or personal download



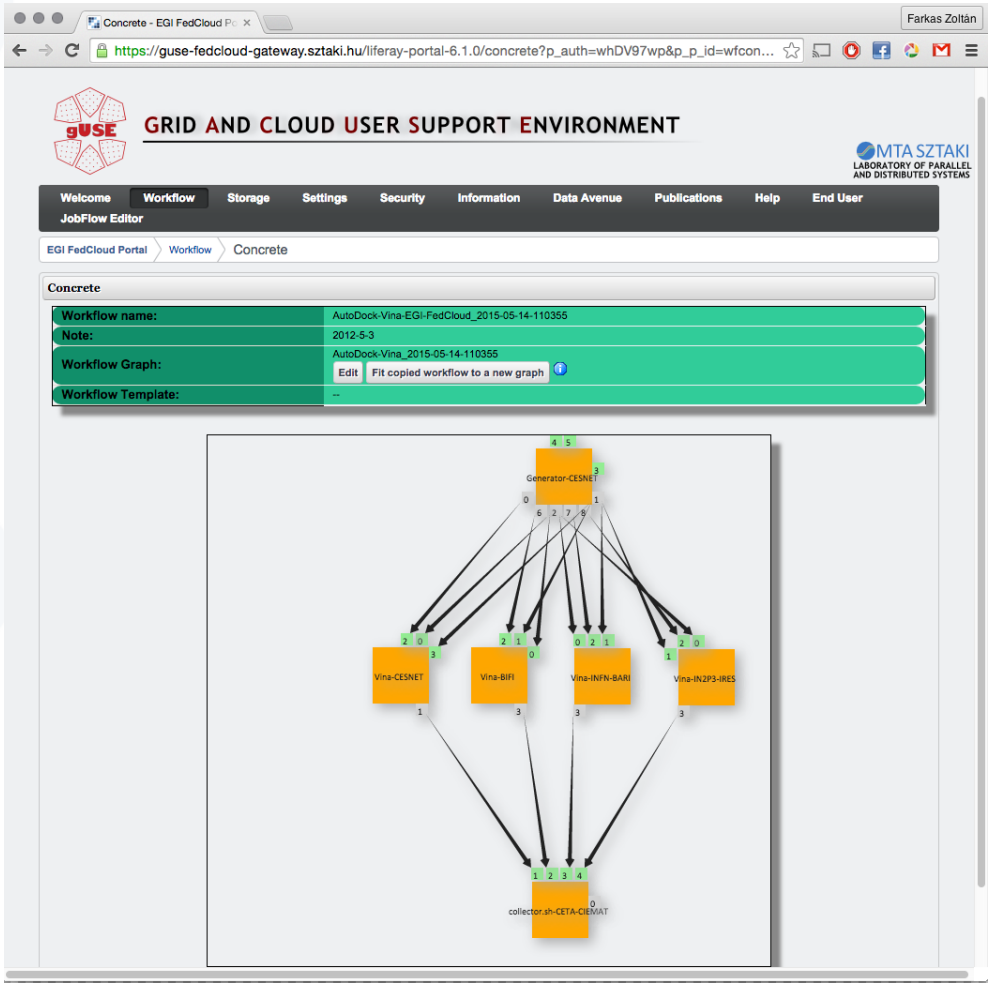
Easily get all the information to instantiate a VM

VM Images

- EGI provides a set of base images:
  - Basic OS, well-configured, **secure**, and up-to-date
  - Documented process for creation, configuration and publishing
  - **Automatically built** using packer (can be integrated on a CI system)
- Community images:
  - Every community can have its own image set
  - Curated by community managers
  - **Automatic distribution** to supporting sites

Frameworks for building VRCs

- IaaS support with OCCI
  - CLI, Ruby, Java SDK
- AppDB
  - Web GUI + RESTFUL APIs
- Several tools extend the Iaas capabilities of the EGI cloud (PaaS/SaaS):
  - External contributions (→ support for other clouds too
  - Manage workflows of VMs and full VM lifecycle support
  - E.g.: CSGF, VMDirac, WS-PGRADE, COMSs, SlipStream



A compute-intensive workflow in WS-PGRADE on the EGI cloud
http://sourceforge.net/projects/guse/

EGI Support & Training

- Guide users from prototyping to production
  - Identify adequate solutions and technical experts
  - Create VMs and containers
  - Deploy resources and services
  - Integrate components
- Training
  - F2F / webinars
  - Modules (slides/videos):
    - IaaS with rOCCI now available
    - Docker on EGI Federated Cloud
    - IaaS for developers (using Java or Ruby)
  - Training cloud infrastructure for hand-on activities

EGI Federated Cloud Evolution

- EGI Federated Cloud keeps evolving:
  - Docker Support (2015 Q4)
  - Enable certificate-less access for users and system administrators (2015 Q4)
  - Create VM snapshots, resize VMs, migrate VMs with OCCI (2016)
  - Management of VMs from AppDB (2016-17)
- Community specific service developments (PaaS/SaaS):
  - for Competence Centers (ELIXIR, BBMRI, MoBRAIN, …)
  - for HumanBranProject, Marine and Fisheries, CANFAR, …

Integration with other e-Infrastructures

- Integration with EUDAT
  - Access to EUDAT service (B2STAGE, B2DROP, B2FIND and B2SAFE) from the federated cloud
- Integration with other cloud federations:
  - CANFAR, FogBow, HARNESS, NeCTAR, CERN, etc.
  - Technology exchange; Interoperability; User support and training

Conclusions

- EGI is a federation of national entities that operates an e-Infrastructure for e-Science
  - 340 resource centres in 54 countries
  - 620,000 logical CPU cores
  - 270 PB disk, 220 PB tape
- EGI is already heavily used by astronomers and astrophysics
  - Plank ESA satellite mission, CTA, LOFAR
  - New use case: CANFAR
- The EGI Federated Cloud is offering new to exploit the EGI infrastructure
  - Service Hosting, Interactive computing, etc.

Thank you for your attention.

Questions?

http://cf2015.egi.eu

www.egi.eu