

# FedCloud in EGI-ENGAGE

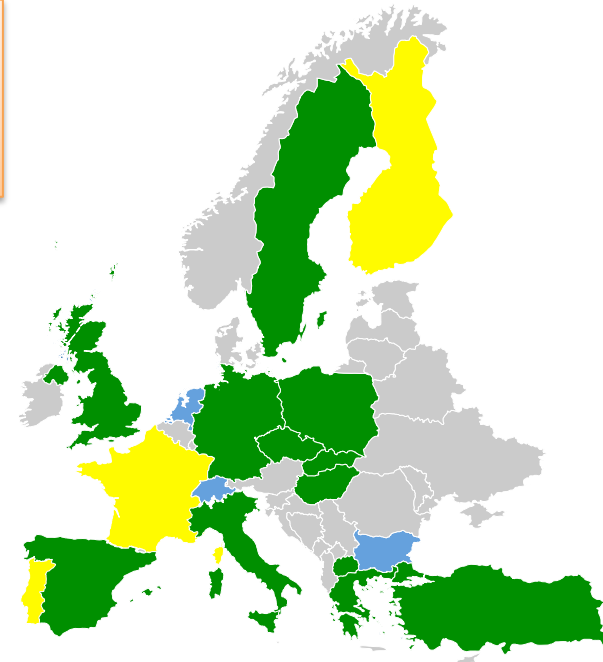
Enol Fernández

Reunión JRU Española en EGI-ENGAGE

The EGI Federated Cloud is federation of institutional private Clouds, offering Cloud Infrastructure as a Service to scientists in Europe and worldwide.

EGI Federated Cloud is based on:

- **Standards and validation:** federation is based on common **Open-Standards** – OCCI, CDMI, OVF, GLUE, etc...
- **Heterogeneous implementation:** no mandate on the cloud technology, fostering the adoption of the standard interfaces and services.



\* Not shown on map

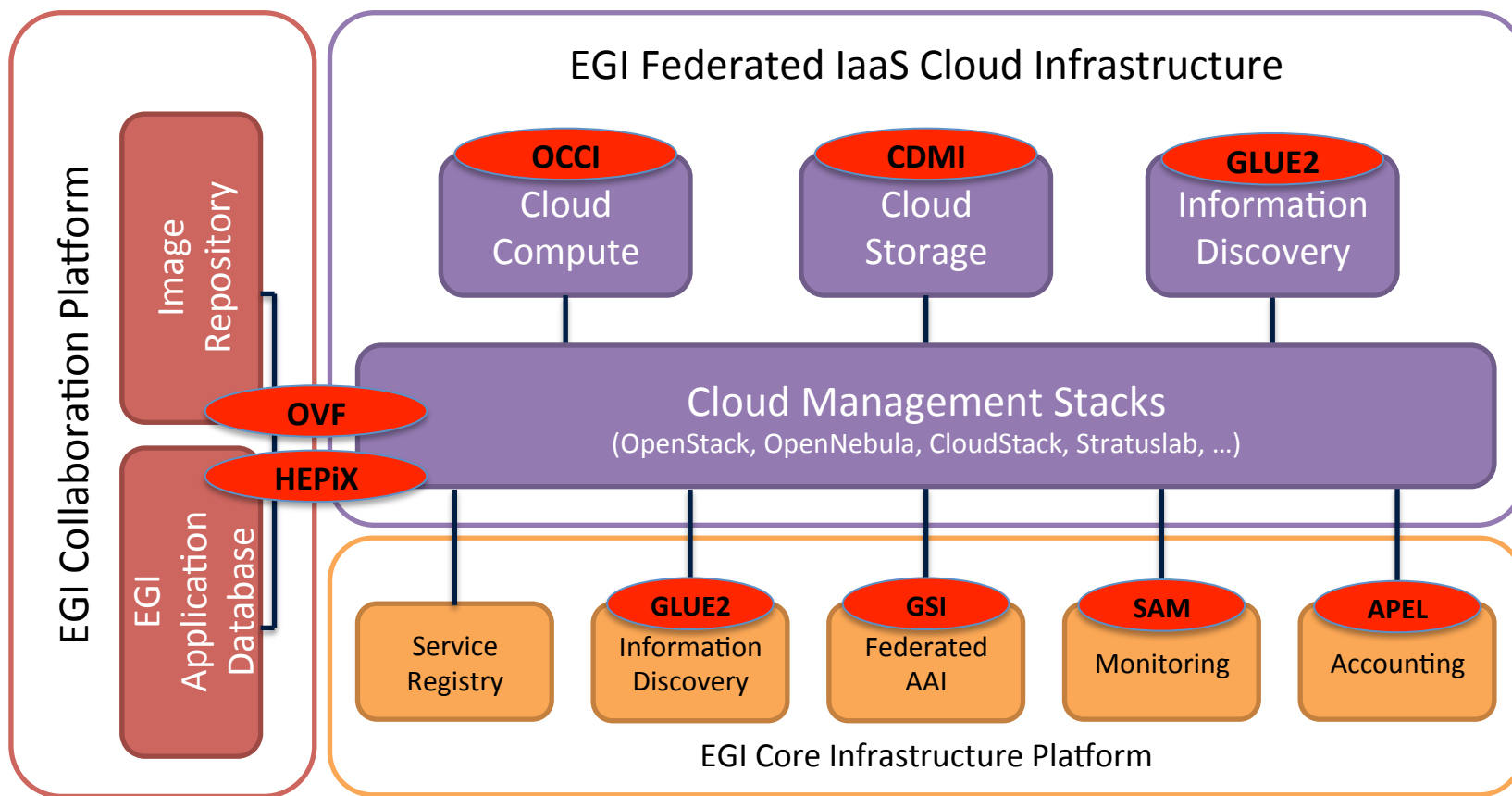
Supported use cases:



- 12 countries provide certified resources
  - Czech Republic, Germany, Greece, Hungary, Italy, Finland, Poland, Slovakia, Spain, Sweden, Turkey, United Kingdom
- 19 resource providers
- 4 countries currently integrating
  - France, Finland, Portugal, South Korea\*
- 5 countries interested
  - Bulgaria, Denmark, Israel\*, The Netherlands, Switzerland
- Worldwide interest
  - South Africa\* (SAGrid)
  - United States\* (NIST, NSF Centres)
  - Australia\* (NeCTAR)
  - Microsoft

**January 2015**

- Document describing EGI FedCloud:
  - Federation Model
  - Cloud specific interfaces
  - Integration with EGI core services
  - Adding sites to FedCloud
- EGI-InSPIRE MS-520 (v1), MS-521 (v2)
- v3 on preparation



**End users:**

Access services running in VMs



Create/destroy/access VMs

**Cloud sites**

**Site:**

Supports a research group

Download VM images

Download VM image and use 'locally'

EGI Applications Database

Cloud Marketplace

Publish image

Include image in a research group set

**Technical developer:**  
Prepares VM Images



**Research group Manager:**  
Endorses suitable VM images



1. VM Management
2. Data Management
3. Information Systems
4. Accounting
5. Monitoring
6. Notification → without any relevant activity
7. Federated AAI
8. VM Image Management
9. Brokering
10. Contextualisation → concluded

- OCCI
  - OGF Standard, implemented by all FedCloud RP
  - Extended to support contextualization
- EGI-ENGAGE (see Alvaro's talk)
  - OCCI v1.2
  - “flavours” extension
  - Vertical scaling, snapshots
- Still missing:
  - Network management → new network scenario in FedCloud



- Open FedCloud to new APIs and services
  - provide new features not covered by OCCI/CDMI
  - maintaining the integration with core services
- New PROCEDURE being defined
  - PROC19: <https://wiki.egi.eu/wiki/PROC19>
  - Integrate a new cloud management stack or grid middleware in EGI
  - Allow to assure a certain level of quality

- Current OpenStack integration in FedCloud
  - keystone-voms → supports X509+VOMS
  - caso → provides accounting information
  - vmcatcher/atrope → subscribe to VM image lists
  - cloud-bdii-provider → provides OCCI identifiers but easily extensible to native ids
- Main points missing:
  - monitoring
  - gocdb service types, SU, UMD, e-Grant, security

- CDMI
  - SNIA/ISO Standard to access object-storage
- Current status:
  - CDMI implementation for Okeanos
  - Unmaintained support in OpenStack
  - dCache pilot implementation
- EGI-ENGAGE:
  - Ongoing effort to support OpenStack CDMI
  - OneData
  - INDIGO-Datacloud

- FedCloud resources are discoverable through EGI's top-BDII
  - GlueSchema v2 representation of resources
  - Extensions proposed to OGF and currently under discussion
- cloud-bdii-provider
  - OpenNebula, OpenStack support
  - Synnefo integration undergoing
- EGI-ENGAGE:
  - Use OCCl for information discovery?

- **Cloud Usage Record**
  - New fields agreed on F2F
- **Accounting information:**
  - New extractors being deployed now
  - Still need to clarify things in the repo
- **EGI-ENGAGE:**
  - Accounting for cloud storage
  - Improvement of the portal views

- All services in EGI are monitored by SAM
- Cloud related probes:
  - CloudBDII (not really testing anything relevant now!)
  - OCCI (submission of VMs)
  - APEL (accounting records up to date)
- Under development
  - CDMI
  - VM Image management
- EGI-ENGAGE
  - Integration with “standard” SAM instances?
  - New monitoring infrastructure SAM/ARGO

- Current EGI AAI based on X.509 + VOMS
  - FedCloud also using X.509+VOMS
  - Known usability issues
- PerUser SubProxies
  - Pilot implementation available in OS (tested by BIFI) and ONE
- EGI-ENGAGE
  - transition to new mechanisms in > 1 year time frame
  - should we allow other methods in the meantime?

- AppDB manages VM image metadata
  - VM images stored somewhere else (e.g. EGI Virtual Appliance repository) in OVA format
- HEPiX lists to distribute images on RP
  - Sites subscribe lists in AppDB and download endorsed images
- EGI-ENGAGE:
  - Alternative implementations to vmcatcher
  - TaskForce for maintaining FedCloud images
  - Security audits as a service



- **Infrastructure & Application Brokers**
  - Catania Science Gateway Framework – SaaS with ID federations
  - SlipStream – PaaS for automating deployments; Helix Nebula
  - COMPSs – programming framework for auto-parallelisation
  - VMDIRAC – abstraction on top of HPC/HTC/cloud
  - WS-PGRADE – workflow development and enactment
  - Vcycle - a VM lifecycle manager that implements the vacuum model on IaaS Cloud services
  - CompatibleOne - All in one proprietary solution for brokering, accounting, user management, monitoring, etc..
- **EGI-ENGAGE**
  - e-Grant
  - INDIGO-Datacloud will bring new PaaS developments
  - Probably others

- There is a github project for EGI FedCloud developments
- Currently the authoritative source for (most) integration software
  - No clear rules on how to contribute
- **EGI-ENGAGE**
  - Transition to UMD

Requirements	Description
<b>Contextualisation tool(s) for Windows OS</b>	offer a contextualization solution for Windows OS virtual machines too
<b>Extend the EGI Federated Cloud OCCl client to get information about the cloud site configuration</b>	a user should be able to query a EGI Federated Cloud site to know its policy about IP management, ports, etc. The OCCl client should allow users to create new security rules (e.g. open port 22) during for their virtual machines
<b>Offer a VM with the EGI Federated Cloud CLI pre-installed</b>	to speed-up the use case preparatory work
<b>Reliable service management</b>	Well-defined procedures, rules and policies to access the EGI Federated Cloud resources (for free, for pay, how much to pay and with which rules)

Requirements	Description
<b>Reliable storage solutions</b>	offer production quality block and object storage solutions
<b>Automatic scalability to exploit cloud elasticity</b>	Offer tools allowing users to exploit horizontal (start/stop new virtual machines when needed) and vertical cloud scalability (change characteristics of already started virtual machines, e.g. increase the number of cores, the amount of memory, etc.).
<b>Custom DNSs</b>	Users should be able to create custom DNS configurations
<b>Java API for OCCl</b>	useful to easily integrate user applications to the EGI Federated Cloud
<b>Big Data Platform</b>	official solution should be offered by the EGI Federated Cloud to copy and replicate the big data sets used by the use cases

- EGI-ENGAGE includes a network of eight Competence Centres
- Joint effort from NGIs, user communities, technology and service providers
- Collect requirements, integrate community-specific applications into state-of-the-art services, foster interoperability across e-Infrastructures, and evolve services

- Pan-European RI (17 countries) to build a sustainable infrastructure for biological information
- CC will select representative use case workflows to develop demonstrators targeting FedCloud
- CC will also create best practices materials

- Biobanking a BioMolecular resources RI.
- FedCloud related objectives:
  - Secure data storage infrastructure → AAI
  - Data Sharing → respecting protection level required by owners of data and samples
  - Biobank Workflow support → Big-data processing with BioBankCloud (based on Hadoop)

- Research from Molecule to Brain
- Builds on the experience of WeNMR and NeuGrid4You
- FedCloud related objectives:
  - Bring computing to the data → managing data, running jobs with Scipion & visualization
  - Facilitate the operation of web services and portals → move to DIRAC4EGI
  - Offer custom virtual machines to research for specific scenarios
  - Enable GPU processing in the EGI infrastructure → GPUs in FedCloud



- CC to ease research from the areas of the Arts and Humanities
- FedCloud related objectives:
  - DARIAH eScience Gateway on EGI → based on WS-PGRADE/gUSE
  - Storing and Accessing DARIAH contents on EGI (SADE) → based on gLibrary
  - Multi-Source Distributed Real-Time Search and Information Retrieval → big data stream and batch processing techniques.

- LifeWatch CC will capture and address the requirements of Biodiversity and Ecosystems research communities.
- FedCloud related objectives:
  - Big data & Ecological Observatories → tools for managing and processing the data from observatories & simulations, R support
  - Workflows & Virtual Labs → web gateways, COMPSs
  - Citizen Science → receive data by citizens, analyze it, open data

- Next generation incoherent scatter radar system
- FedCloud related objectives:
  - Develop the EISCAT portal → access to the data archives
  - Basic reanalysis in the portal → filtering data before download

- Design of the use of grid and cloud for the integrated solid Earth Sciences research as part of the European Plate Observing System
- FedCloud related objectives:
  - Identify and validate authentication and authorisation services
  - Test cloud resources and usage models
    - compute/data resources for applications
    - hosting services
  - Science gateway → no specific platform yet

- Customised IT services to support the climate and disaster mitigation researchers
- FedCloud related objectives:
  - web portals for tsunami wave propagation simulations and for WRF-based weather simulation
  - Application of the simulation portals for scientific scenario in disaster mitigation

- AAI
  - not just for the IaaS, also at the application/portal/gateway level
- Big Data
  - not clear what big data is for each CC
  - both **storage** and processing will be needed
- VM image management
- Hosting portals/gateways/applications
  - scaling in/out
  - orchestration?
- Training
  - documentation

## Discussion

