GARR

The Italian Academic & Research Network

GARR Cloud Strategy

GARR strategy towards the provisioning of Cloud Services

Mario.Reale@garr.it
ELCIRA WP5

Outline

- GARR perspective on the Cloud
- Cloud Requirements
- **GARRbox** Pilot Cloud Storage:
  - Current status of the pilot Cloud Storage service (GARRbox)
  - Future developments
- **DOGS** Data on Grids Services portal
  - Requirements for Grid Data Management
  - Architecture
  - Status
  - Wrap Up
- Conclusions

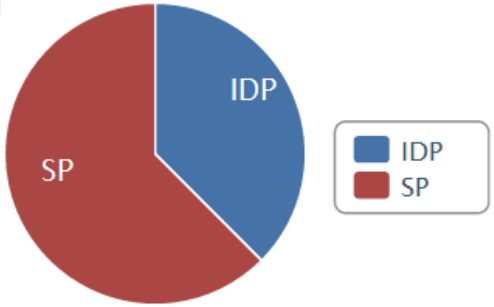
GARR Perspective on Cloud

- GARR aims at supporting its user community in providing Cloud services
- Facilitating the integration of the provided Cloud services with AAI based on the National Identity Federation of trust (IDEM)
- Currently acquiring direct hands-on experience through a Cloud pilot service devoted to **Cloud Storage** provisioning
- GARR does not primarily aim at becoming a Cloud Computing services provider
- Basic role is to be key enabler for inter-federated, trusted, reliable services provided by its member partners and available for the community of GARR users

IDEM – the Italian Identity Federation

www.garr.it

- IDEM is the National Italian Identity Federation for Research
- The objective of IDEM is to create and support a framework to manage shared access to online resources
  - Through Single Sign On (SSO)
- IDEM is open to federate systems based on the SAML2 standards
  - interoperable with Shibboleth
- IDEM is currently including
  - 42 Identity Providers
  - 70 Service Providers



- IDEM enables Italian researchers to connect to inter-federation EduGAIN and EduROAM services

GARR Requirements for Cloud

- Secure authentication mechanism
  - When possible, based on Identity Federations
- Data integrity
- Ubiquitous access
- Dynamic Management of resources
  - Possibility to dynamically manage file access rights
  - Delegation of Management of subset of resources and Multi-tenancy
- Data encryption
  - For data and metadata

GARRbox

Fabio Farina, Fulvio Gaelazzi, Andrea Biancini (INFN Mib), Simon Vocella, Cristiano Valli, Mario Reale

GARRbox: a Cloud Storage pilot

- GARR started an internal pilot project on Cloud computing (GARRbox) in Q3 2011 – currently still ongoing
- Main Goal is to acquire hands-on know how on Cloud Storage provisioning in view of setting up a service for the BioMedical Community
- In particular, experience has been gained through the pilot on:
  - AAI based on the national identity federation (IDEM)
  - Performances and main features of distributed file systems
  - Providing different interfaces to users for accessing their data
  - Business continuity / Disaster recovery

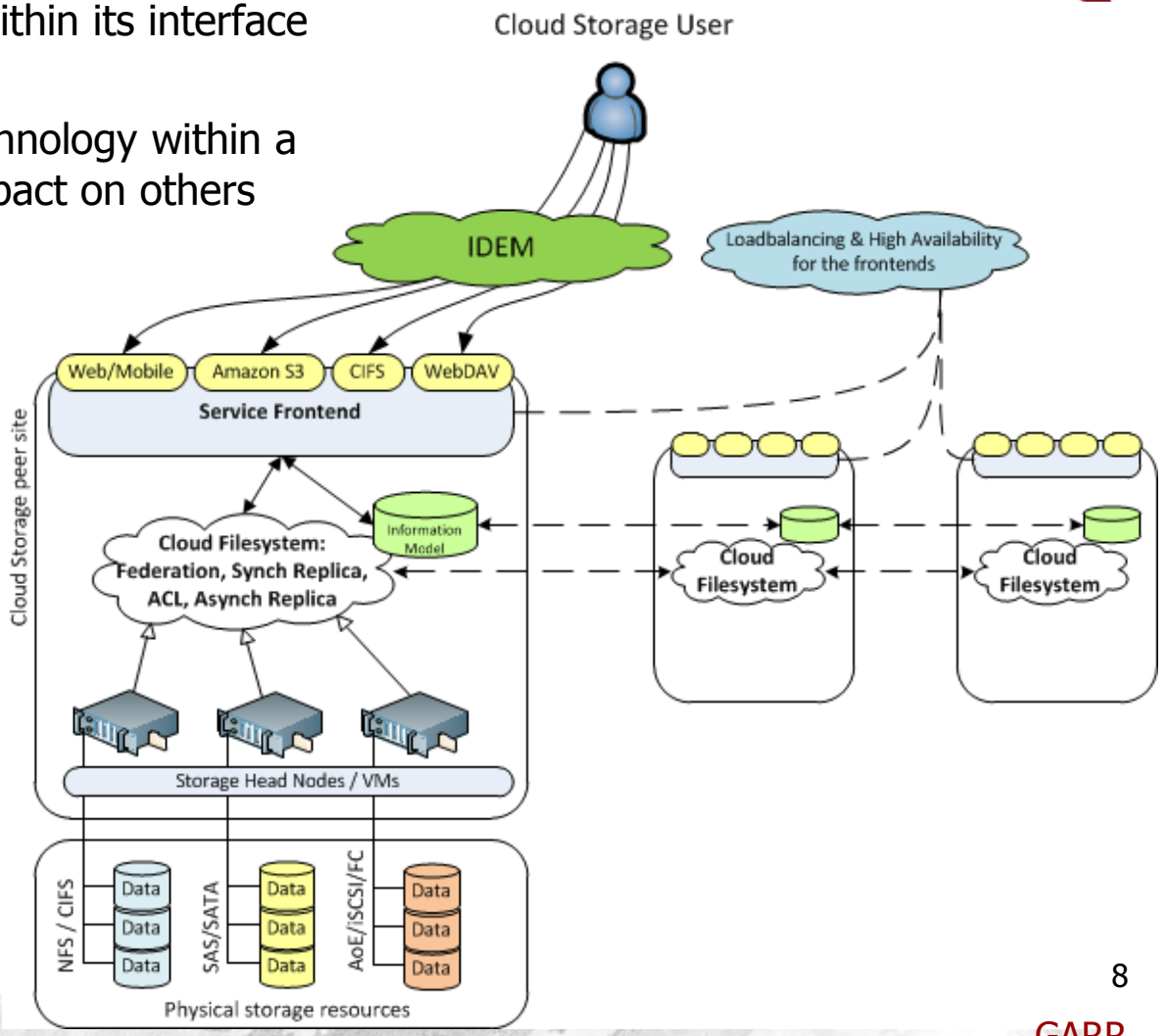
GARRbox architecture

- Layered architecture
- Each layer self-contained within its interface boundaries
- Protocols and underlying technology within a layer replaceable without impact on others
- Resilience / High Availability

Presentation layer

Abstraction & Aggregation Layer

Physical Resources



Technological choices - as of today

- Distributed File System : GlusterFS

- Web Front-End: Ajaxplorer

- S3 Front-End: DragonDisk

- SSO SAML2 Implementation: Shibboleth

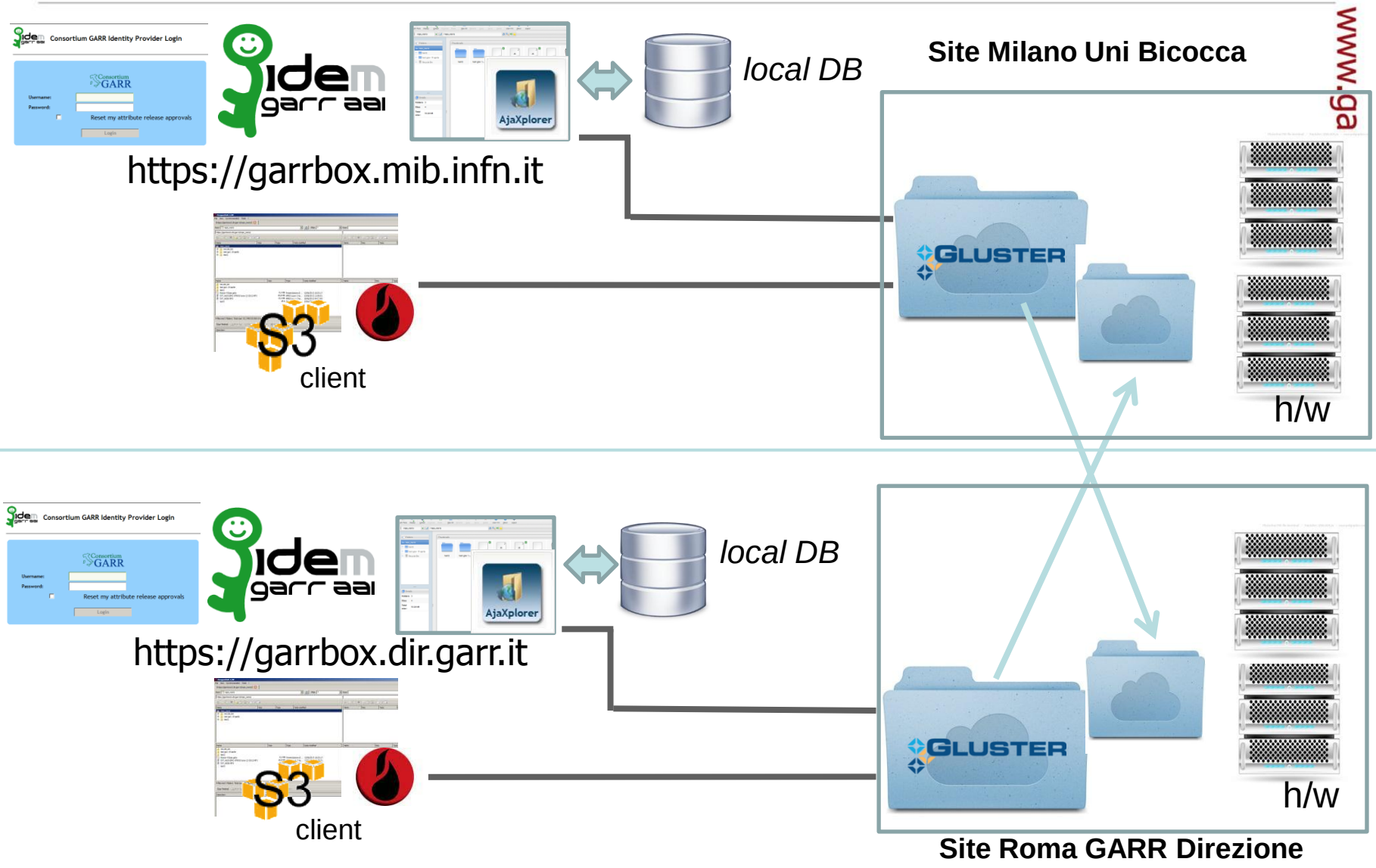
- Identity Federation: National Italian Federation GARR IDEM

Current status of the prototype

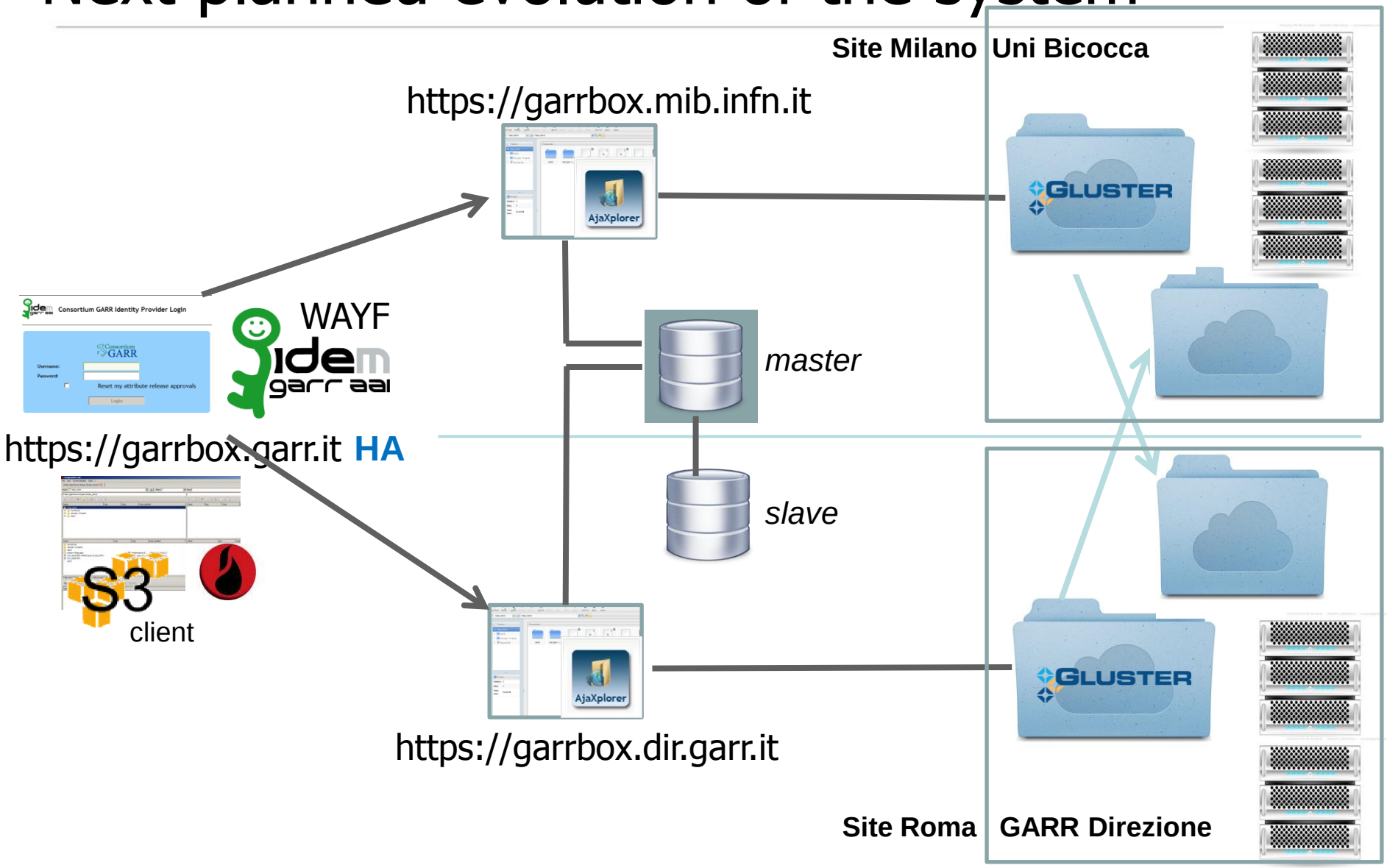
- Available for staff GARR: 15-20 users, 10 GB each.
- Available functionality:
  - Upload / download, search and manipulate data from client browsers and S3
  - Sharing files and folders
  - Use of IDEM
  - Sharing files with anyone, shared folders by IDEM PrincipalName
- Web authentication and access rights, creation of the S3 keys
- Different quotas based on user, group and organization
- Local and geographical replication of files

- First tests of efficiency and resilience have yielded promising results

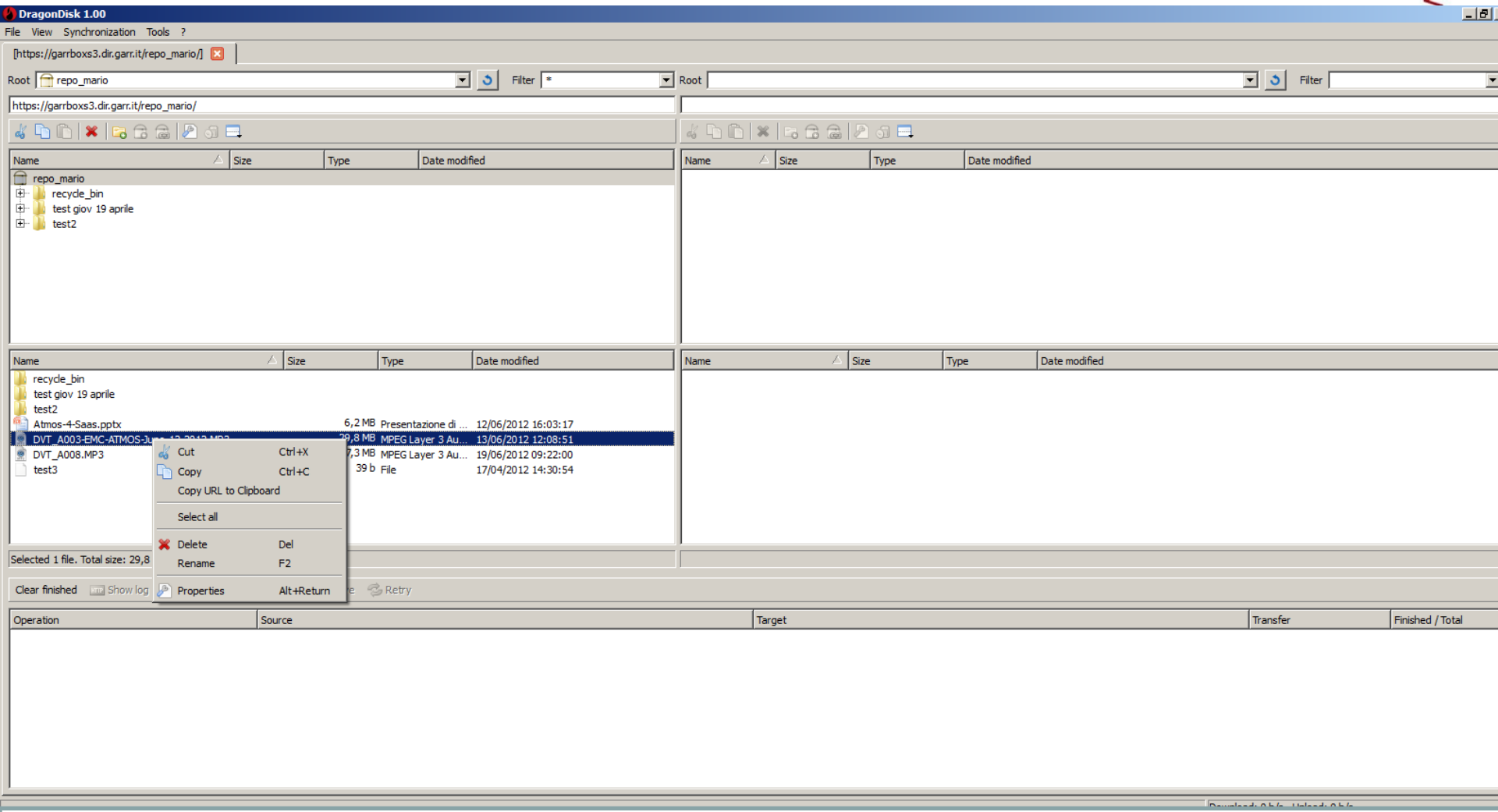
Current GARRbox Implementation



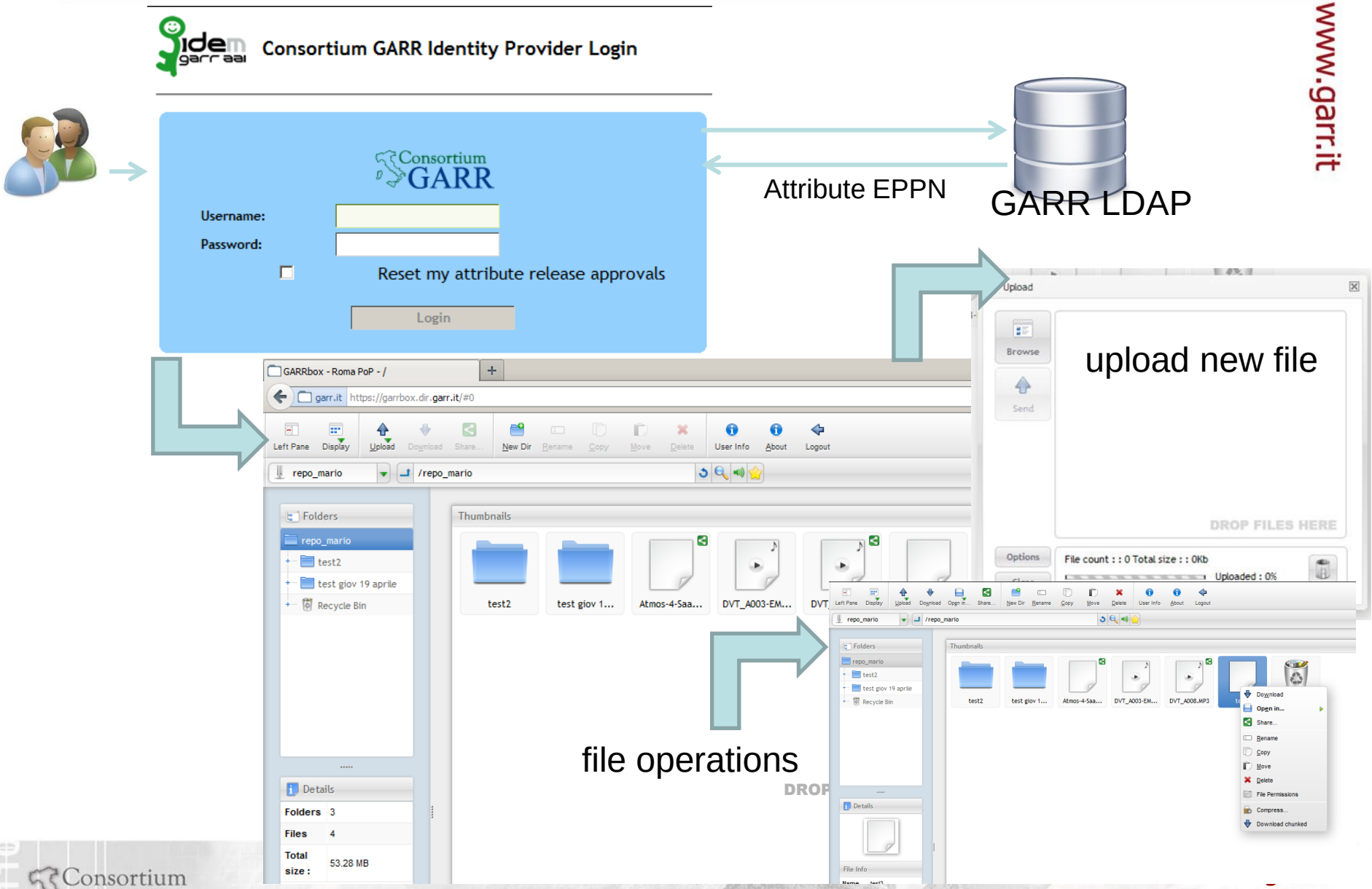
Next planned evolution of the system



Current GARRbox System Usage/1: S3client



Current GARRbox System Usage/2: Web



Steps ahead

- **Identity Management**
  - Harmonisation of credentials management: IDEM vs S3
  - Implement Authorization entirely based on IDEM / Shibboleth
    - Administrative Roles
    - New Attributes or New Attribute Values required ?
  - Integration of WAYF
- **Resilience in case of fault**
  - Transparent migration of user across backend sites
    - Switchover
    - Failover/failback
  - Information System DB implemented in master-slave / multi-master
  - High Availability for the front ends
- **Load Balancing**
  - Geographical DNS between the 2 front ends
  - Front end selection for users based on their IDEM attributes

Further steps ahead

- Global **security assessment**
- **Presentation Layer**
  - Improving Web Interface
  - WebDAV interface
- **Aggregation Layer**
  - Client & Server side encryption
  - Management of complex metadata
  - Decoupling data from metadata
  - Multi-tenancy / Virtual resources management delegation
  - Assessment of alternative technical solutions:
    - NoSQL
    - Newer Gluster version
- **Extend validation phase by users**
  - Extend beta users community

DOGS: Data On Grid Services

A Data Engine for Grid Science Gateways

Riccardo Rotondo (GARR),
Marco Fargetta (Consorzio COMETA)
Roberto Barbera (INFN Catania – Uni Catania)

Challenges related to Grid Data Management

- Make interfaces simple for non expert users
  - CLI-based Grid storage interface is not straightforward
- Grid transactions require user certificates
- Complexity of current protocols to manage grid storage elements
  - Very little or no support for access through modern browsers or others web-based applications

- DOGS Idea: create a Data Management pillar of the Science Gateway (GISELA, CHAIN, DECIDE, GRIDIP..)
  - Adopting the same basic user-friendly working principles

Requirements for DOGS

- Grid Storage complexity hidden to end users
  - Users move files from/to a portal and see it as simple external storage accessible from a web interface and do not care about grid (or any other) technologies behind
- File management smoothly integrated with all the services provided in the Science Gateway
- Underlining architecture exposes a file-system-like view (i.e., a Virtual File System or VFS) through which users can perform the following actions:
  - Create, move, delete files/directories with the desired structure
  - Share files with other users
  - Set the number of backup copies desired

Data On Grid Services: features

- A file browser shows Grid files in a tree
- File system exposed by the SG is virtual
- Easy transfer from/to Grid (by SG) is done in a few clicks
- Users do not need to care about how and where their files are really located

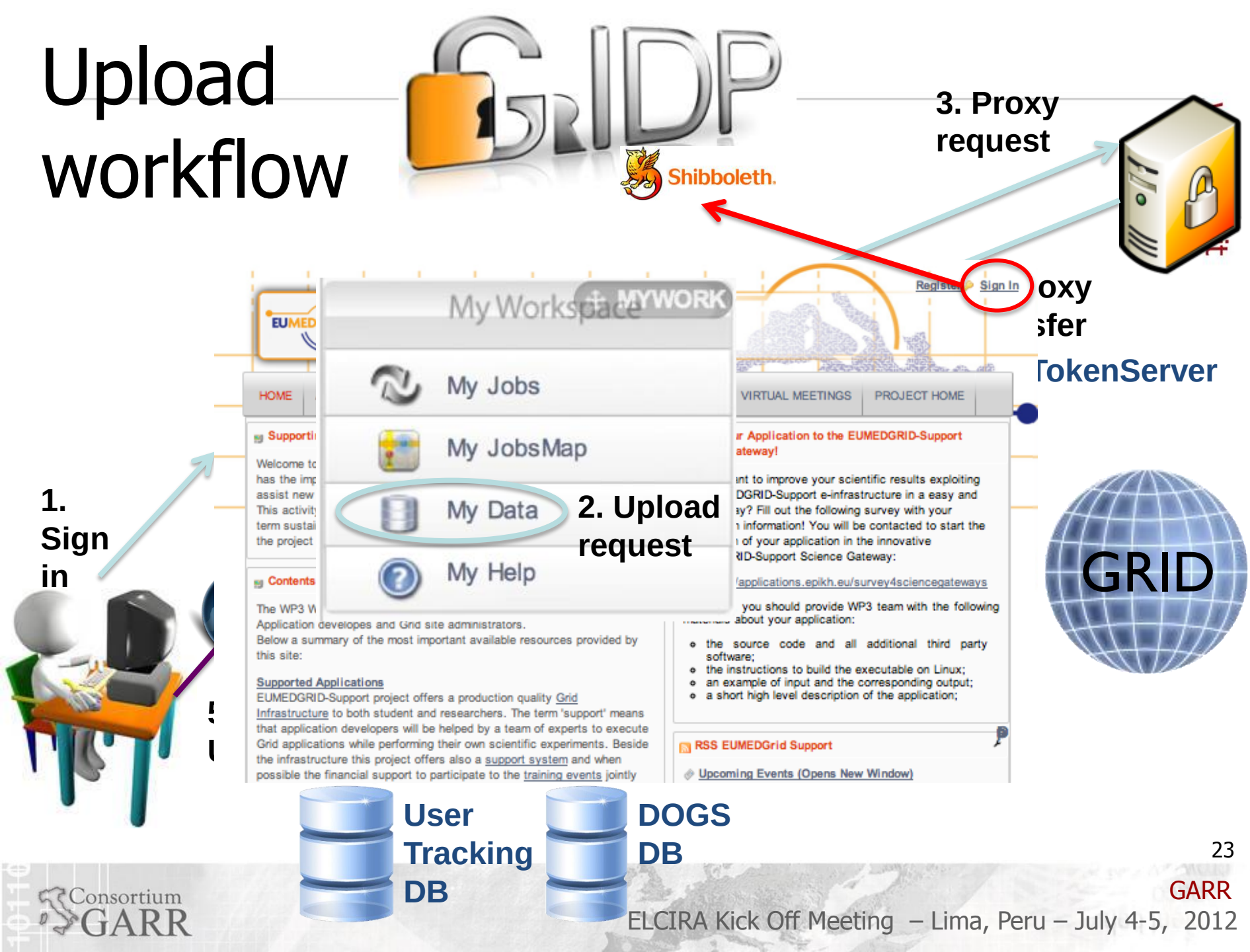
Back-end technical details

- JSAGA API used to transfer data from/to storage elements
- Hibernate to manage the VFS collecting information on files stored on Grid; any changes/actions in the user view affect the VFS
- MySQL as underlying RDBMS
- An additional component has been developed in order to keep track of each transaction in the users tracking DB (to be compliant with the EGI Portal and User Traceability Policies)

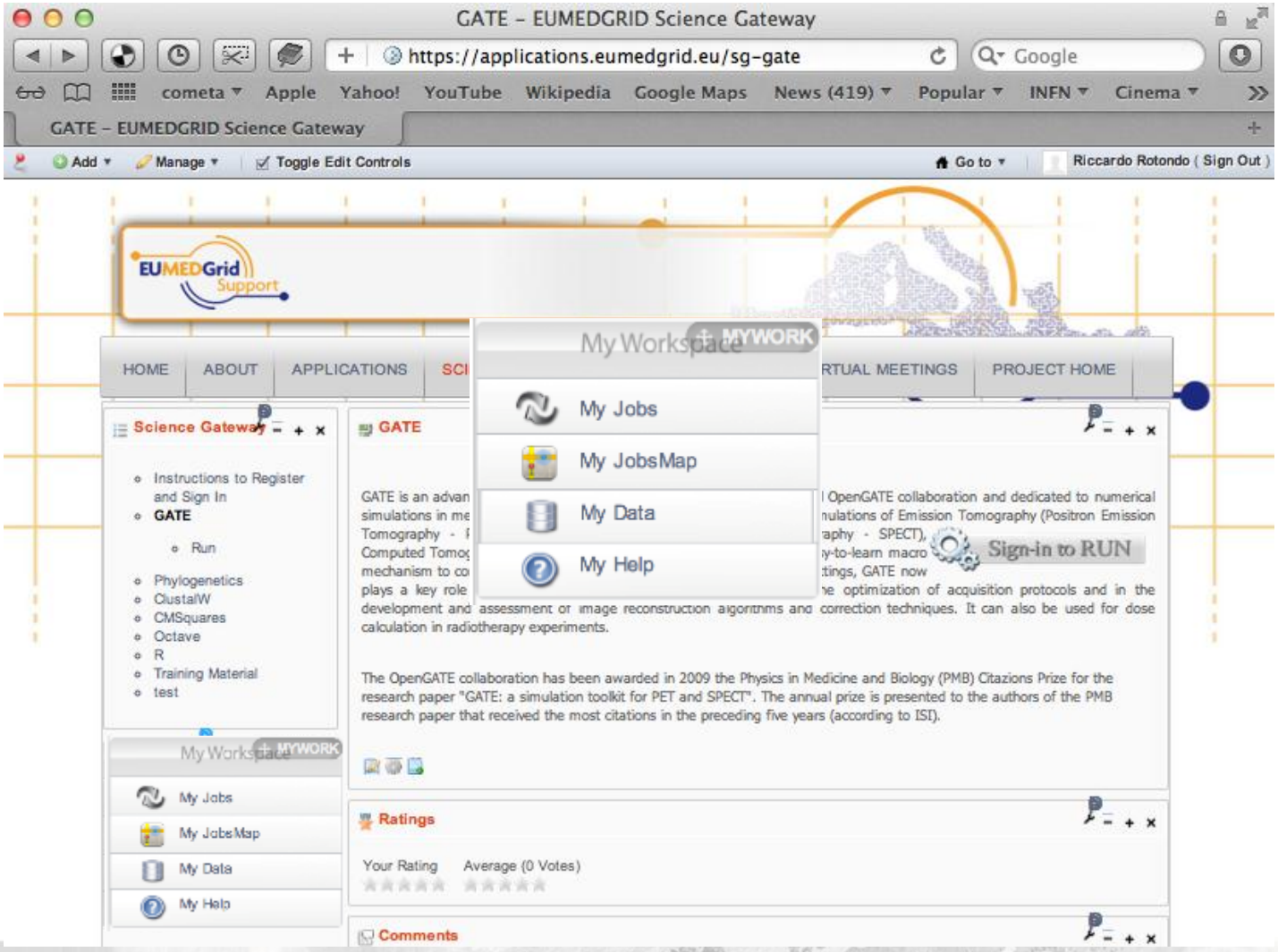
Front-end technical details

- A portlet has been created and deployed in a Liferay-based portal to which access is provided only to federated users with given roles and privileges
  - http://www.liferay.com
- The portlet view component includes elFinder, a web-based file manager developed in Javascript using jQuery UI for a dynamic and user friendly interface
  - http://elrte.org/elfinder

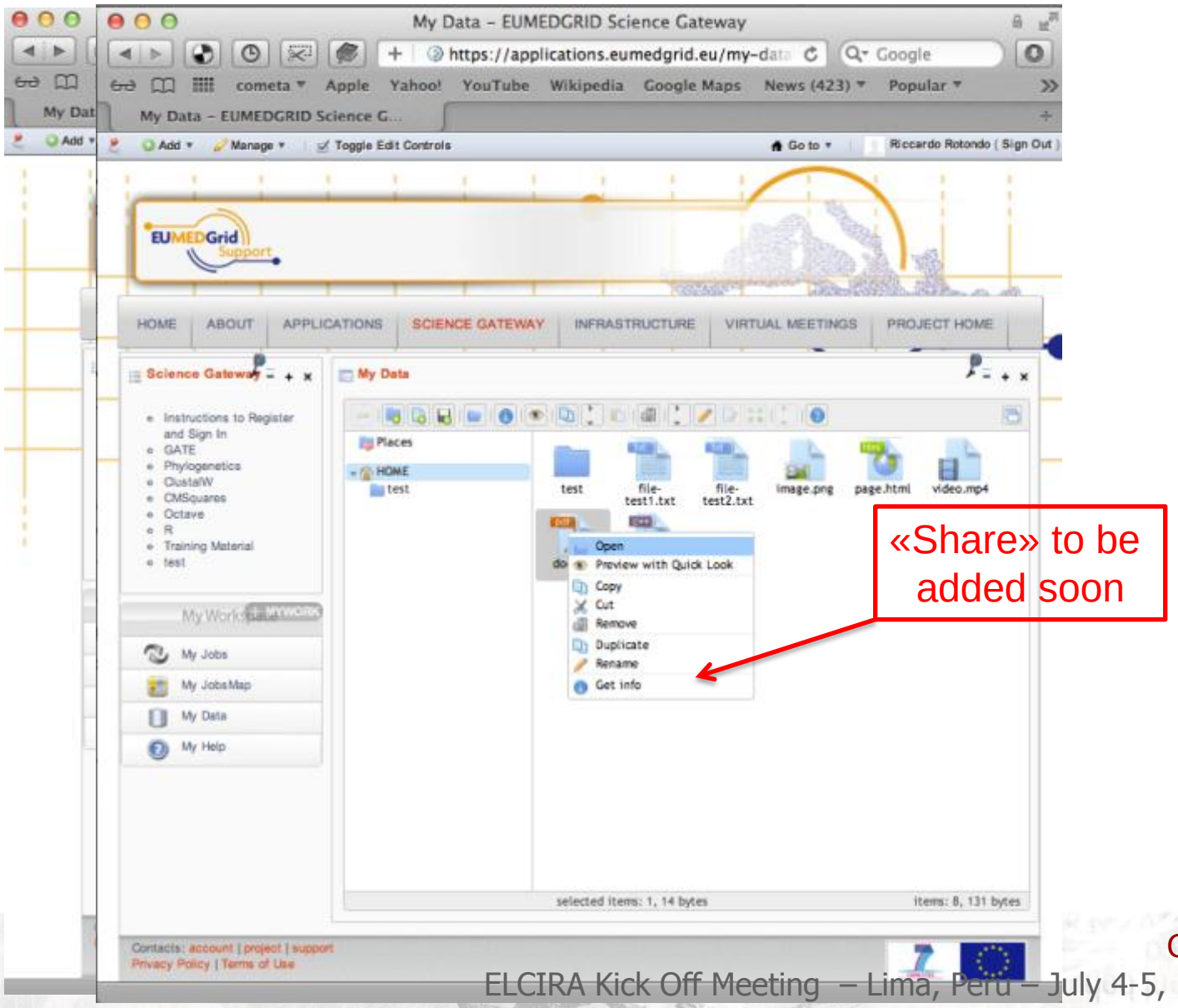
Upload workflow



The Data Engine in action



The Data Engine in action



Wrap Up on DOGS :

- File transfer enabled with **100% browser friendly protocols**
- Using JSAGA makes the Data Engine **middleware independent**
- User friendly interface
- All common file actions provided (**file sharing** among users available soon)

Conclusions

- GARRbox represents main current effort of GARR on Cloud
  - Namely on Cloud Storage through Identity Federation
- Different technologies exploited – hands on experience gained
  - Shibboleth/SAML2 (IdP, SP, WAYF, IDEM attributes)
  - Gluster 3.3
  - Ajaxplorer
  - S3 interface
- DOGS is a Grid Data Management tool accessed through Federation – in the context of the Grid Science Gateway
  - Currently in advanced status of development
- Further exploiting technologies like
  - LifeRay portal
  - Robot Certificates / Automated credential generation
- We're fully available to share know how and commit to ELCIRA DoW as requested

Thank you.