2nd EUDAT User Forum

Data staging to HPC

Giuseppe Fiameni

SuperComputing, Application and Innovation – CINECA, Italy

2nd EUDAT User Forum, London – 11, 12 March 2013

SEVENTH FRAMEWORK PROGRAMME

EUDAT

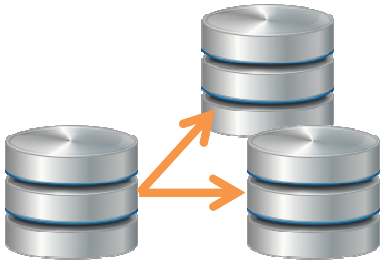
Agenda

	Topic
	Data Staging to HPC - *Conveners: P.Coveney, G. Fiameni*
09:30	**Introducing Data Staging in EUDAT, plus DEMO** - *G.Fiameni, S. Zesada*
10:00	**Euro-VO HPC data needs** - *Sebastien Derriere*
10:15	**EISCAT HPC data needs** - *Ingemar Häggström*
10:30	**Mapper HPC data needs** - *Derek Groen*
11:40	**Open discussion**

Preliminary services

- **Data Staging** to facilitate communities to stage stored data onto external computational facilities, such as HPC resources

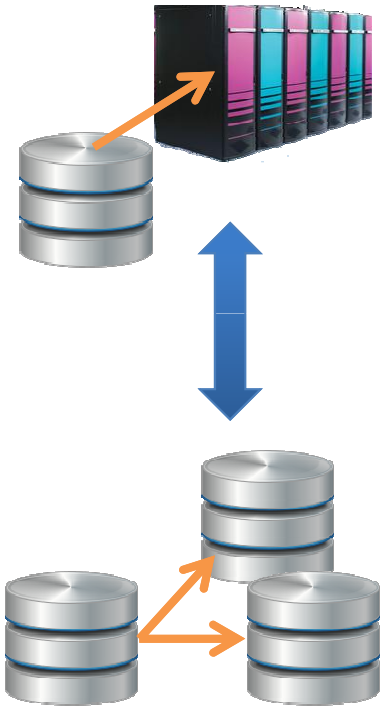
- **Safe Replication** to enable communities easily create replicas of their scientific datasets in multiple data centres for improving data curation and accessibility



Preliminary services

- **Data Staging** to facilitate communities to stage stored data onto external computational facilities, such as HPC resources

- **Safe Replication** to enable communities easily create replicas of their scientific datasets in multiple data centres for improving data curation and accessibility



Building Blocks of the CDI

EUDAT Access Interface

Integrated APIs and harmonized access to EUDAT facilities

Metadata Catalog

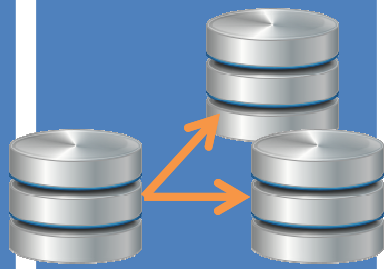
Aggregated EUDAT metadata domain.
Data inventory

Data Staging

Dynamic replication to HPC workspace for processing

Safe Replication

Data curation and access optimization



Simple Store

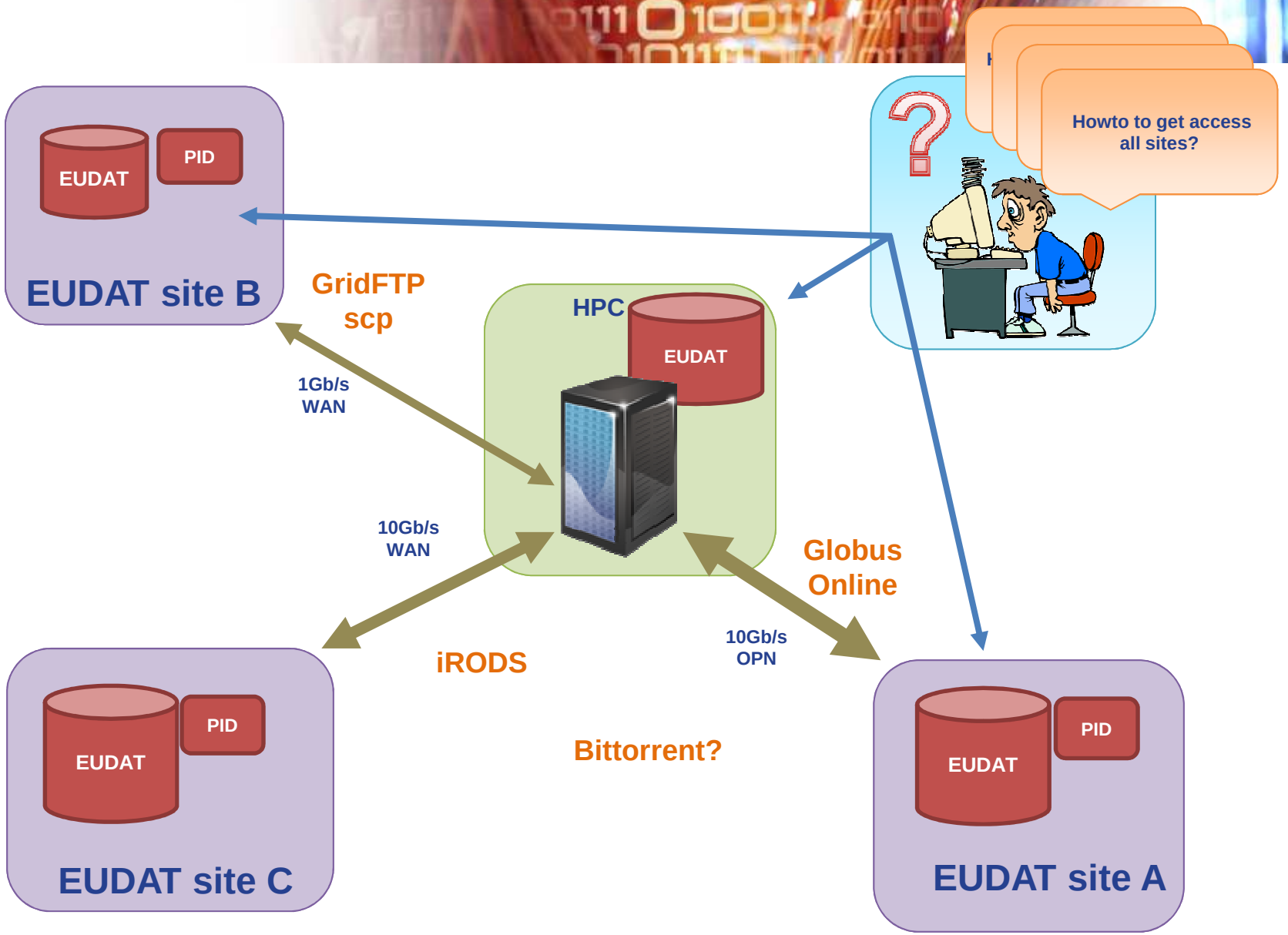
Researcher data store (simple upload, share and access)

AAI

Network of trust among authentication and authorization actors

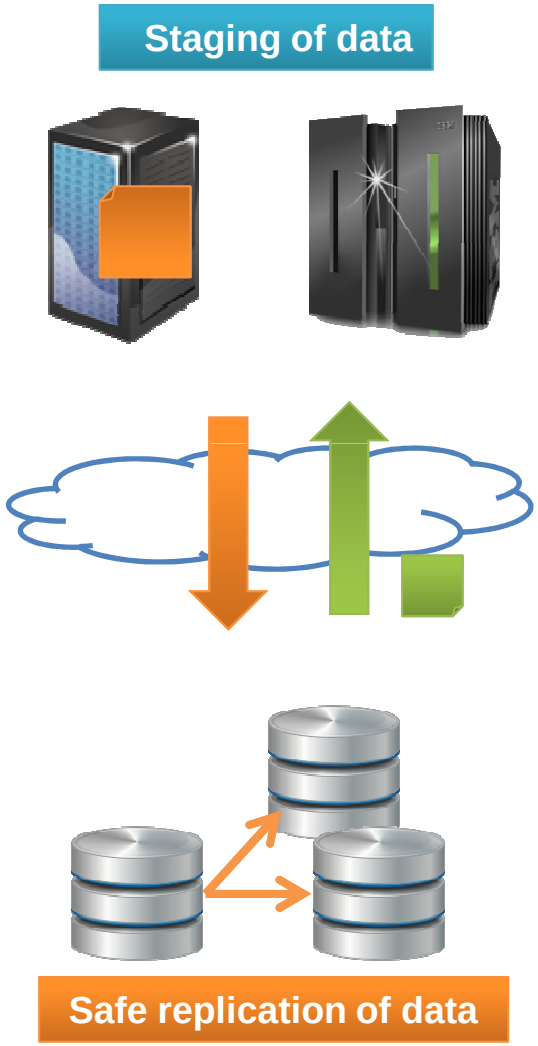
Driving principles

- Work close to scientific communities
- Leverage on existing technologies, experiences, knowledge
- Short term, very frequent, delivers
- Services are meant for the production
- Collaboration with other activities, projects, e-infrastructures is fundamental (i.e. PRACE, EGI)
- Focus on “Low-hanging fruits” - Easy things come first


EUDAT
PID
EUDAT site B
GridFTP
scp
1Gb/s
WAN
HPC
EUDAT
10Gb/s
WAN
iRODS
Globus
Online
10Gb/s
OPN
Bittorrent?
Howto to get access
all sites?
EUDAT
PID
EUDAT site C
EUDAT
PID
EUDAT site A

Goals

- Allow communities move easily large amounts of data between EUDAT storage resources and workspace areas on HPC systems to be further processed
- Offer reliable, efficient, easy-to-use tools to manage data transfers
- Provide the means to re-ingest computational results back into the EUDAT infrastructure
- Permit integration with existing infrastructure (i.e. PRACE), data services



Challenges

- Many technologies are available outside
- Communities prefer to extend their existing solutions rather than acquiring new ones
- Computational infrastructures already have their own data services which EUDAT should comply with
- Transferring large amount of data across the public network is not a trivial task
- Limited effort to be allocated on developing new software

Who might benefit from it?

- The **Data Staging** service is aimed at researchers, who**:**
  - need access to both large-scale data storage and high-performance computing systems;
  - wish to move data easily between the EUDAT data stores and remote HPC facilities such as those provided by the PRACE distributed infrastructure.

How it works

- **Server side**
  - the data staging functionality is realized by extending the **iRODS system with a GridFTP interface using the Griffin technology so as to permit** the transfer of data through a reliable, high-performance protocol

- **Client side**
  - any existing client, supporting the GridFTP protocol can be employed – globus-url-copy, Globus On Line, UberFTP, gTransfer, etc.

- Among available clients, EUDAT recommends the **XSEDE-EUDAT File Manager** which supports a range of transfer protocols (i.e. GridFTP, FTP, native iRODS, etc.) and provides an intuitive and friendly interface

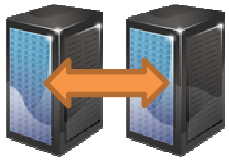
XSEDE
Extreme Science and Engineering Discovery Environment

- Users need a personal certificate (X.509) to fully exploit the service

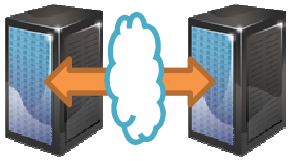
EUDAT

Different flavors of the same service

- **Local data staging**: staging of data among the resources of the same site using low level tools

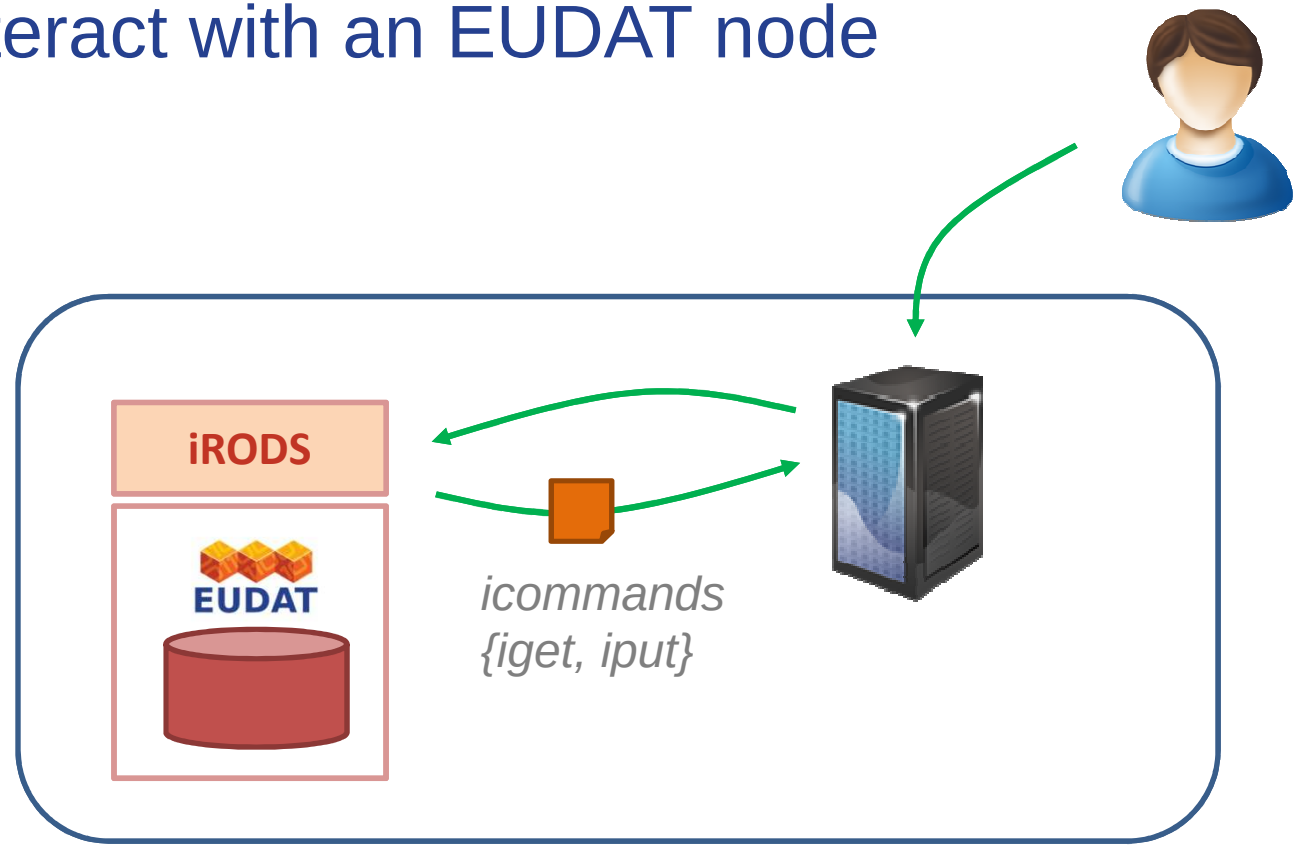


- **Remote data staging**: staging of data using remote data services (e.g. external GridFTP service). Many more implications and constraints!



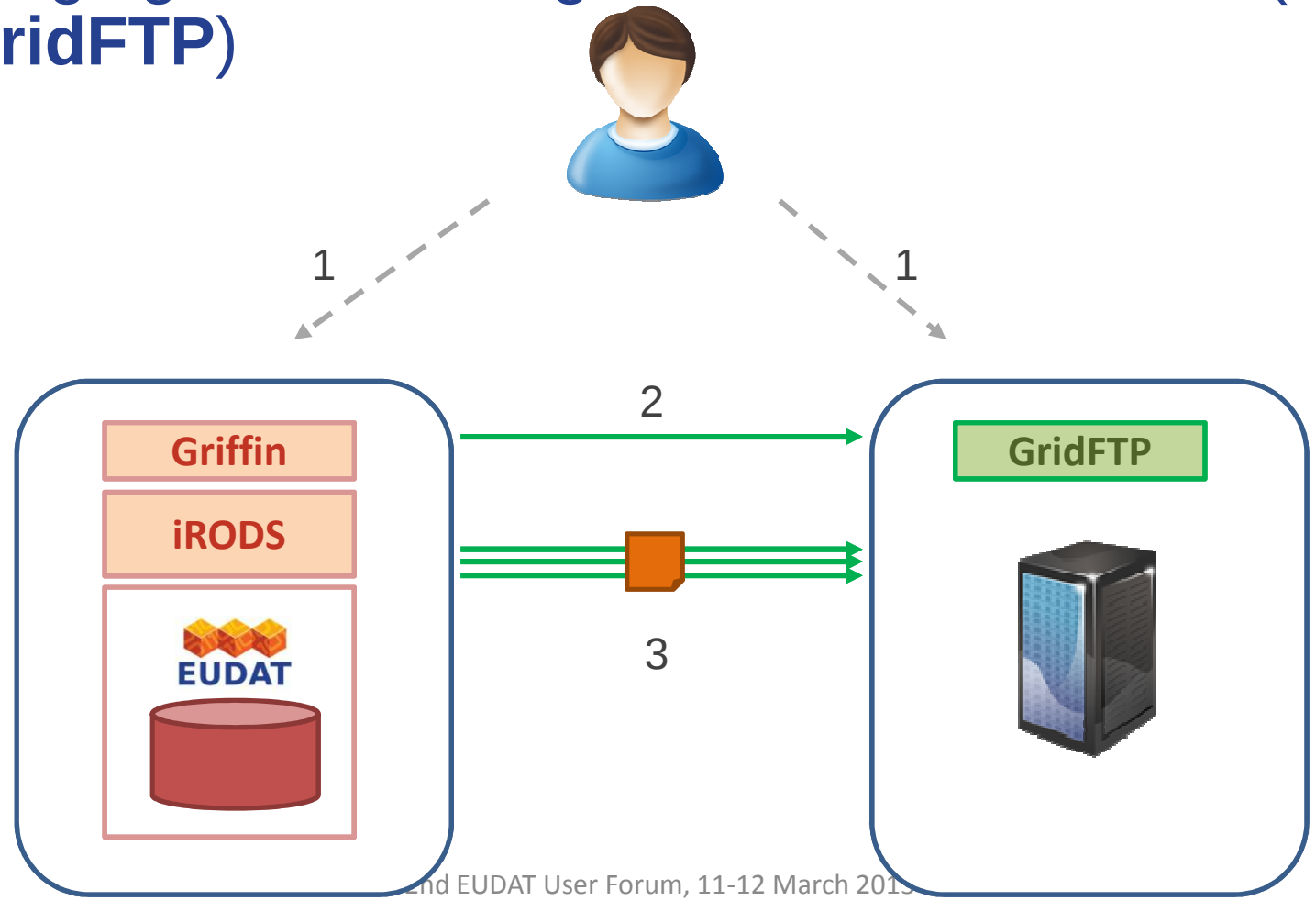
Local data staging

- ***iRODS icommands*** provide a low level interface to interact with an EUDAT node



Remote data staging

- Staging of data using remote data services (**e.g. GridFTP**)



Requirements mapping							
	Griffin	UNICORE FTP	FTS	Globus Online	Parrot	iRODS + iCommands	gTransfer
	Functional						
Capability to stage entire directory	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stage large data sets without big performance penalty	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Search mechanism	N.A.	N.A.	N.A.	Yes	N.A.	Yes	No
Multi-point transfers (i.e. from many sources to one destination)	N.A.	N.A.	No	No	N.A.	No	Yes
API	Yes	No	No	Yes (beta)	No	Yes (Jargon, PyRords)	No
Automatic deletion of staged data sets	No	Unknown	No	No	Unknown	Yes	No
Compatibility with GridFTP (to permit interaction with PRACE)	Yes	Yes	Yes	Yes	Yes	No	Yes
Support for X.509 credentials	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Non-functional						
Ease of use	N.A.	Good	N.A.	Very good	Medium	Medium	Medium
Support for third-party transfers	Yes	Yes	Yes	Yes	No	Yes	Yes
Possibility to tune network parameters	Manual	No	Manual	Automatic	Manual	No	Semi-automatic
Compatibility with iRODS	Yes	No	Through Griffin	Through Griffin	Through Griffin	Yes	Through Griffin
Transfer restart/resume	Yes (only for third-party)	No	Unknown	Yes	No	No	Yes (only for third-party)
Ability in managing many transfers simultaneously	No	No	Yes	Yes	No	No	No

Some examples

iRODS icommands

- Low level command-line tools to manage data which are stored onto iRODS resources
- Reliable and programmatic
- Provide high performance
- Easy to setup and portable on many systems

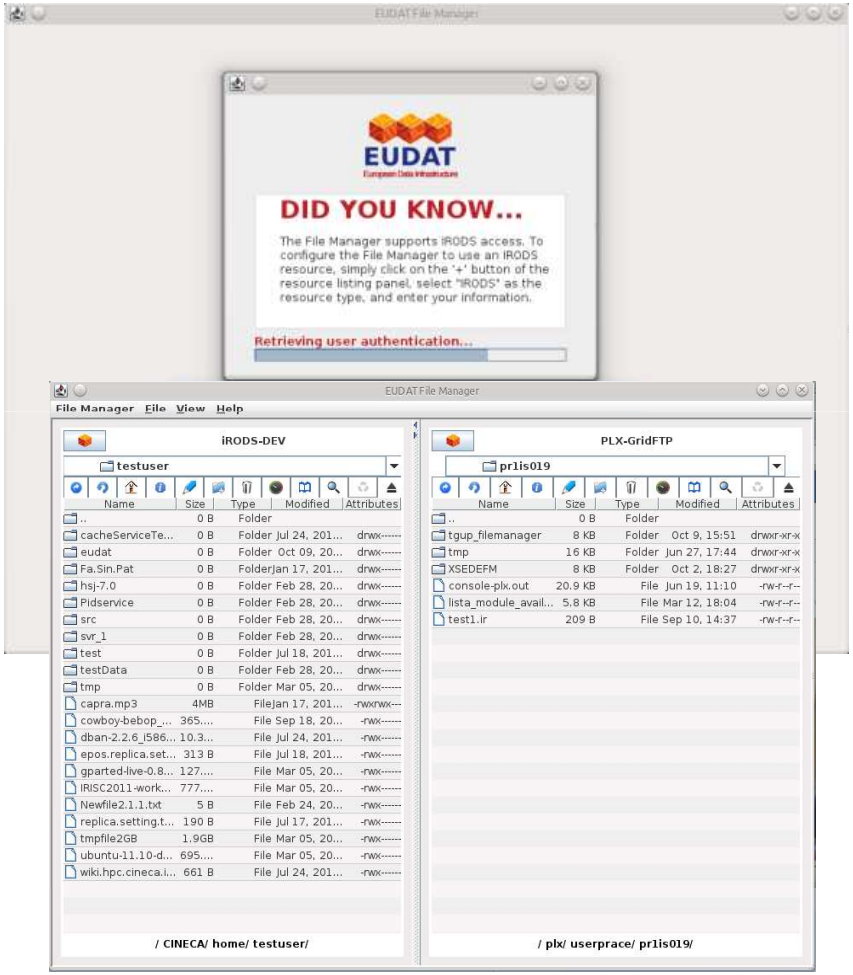
```
prompt$ iget –N 4 /home/irods/data/archive /shared/data/userprace/tmp
prompt$ iput –N 4 /shared/data/userprace/tmp /home/irods/data/archive
```

Griffin (GridFTP interface)

- GridFTP server, entirely written in JAVA, being able to access iRODS resources as well as plain file systems
- Supports most of GridFTP features, including multiple streams, tcp-buffer size tuning and files pipelining
- Unfortunately multiple-stripe is not supported
- Several tests were performed at CINECA and SARA where large data sets, of the order of hundreds of GBs, have been transferred back and forth the two sites with success

XSEDE/EUDAT File Manager

- Client interface
- Developed within the **XSEDE** (Extreme Science and Engineering Discovery Environment) project
- Provides users with an easy-to-use, drag-and-drop interface for managing data transfers over GridFTP/iRODS servers.
- Heavily tested within **EUDAT** in collaboration with **Texas Advanced Computing Center**



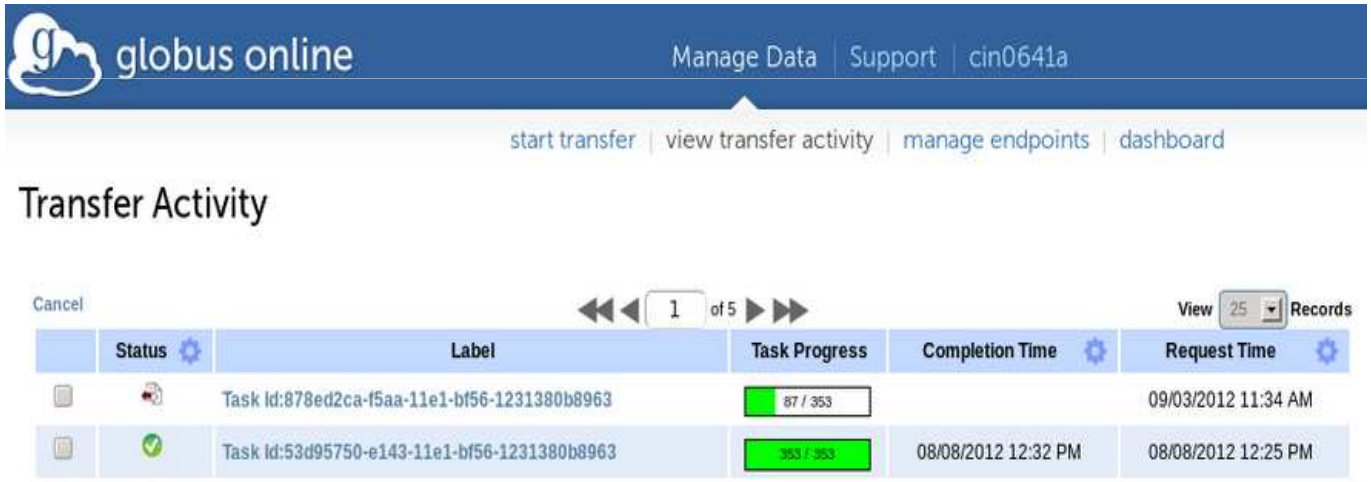
Data Staging Script

- A simple python modular staging script to help communities integrate the data staging service within their exiting solutions
- Based on Globus Online API and iRODS rule mechanism for data selectio

```
prompt$ ./datastager.py -p /home/irods/data/archive -y 2004 -n MN
-s AQU –c BHE -u cin0641a --ss ingv --ds GSI-PLX –dd
/shared/data/userprace/tmp
```

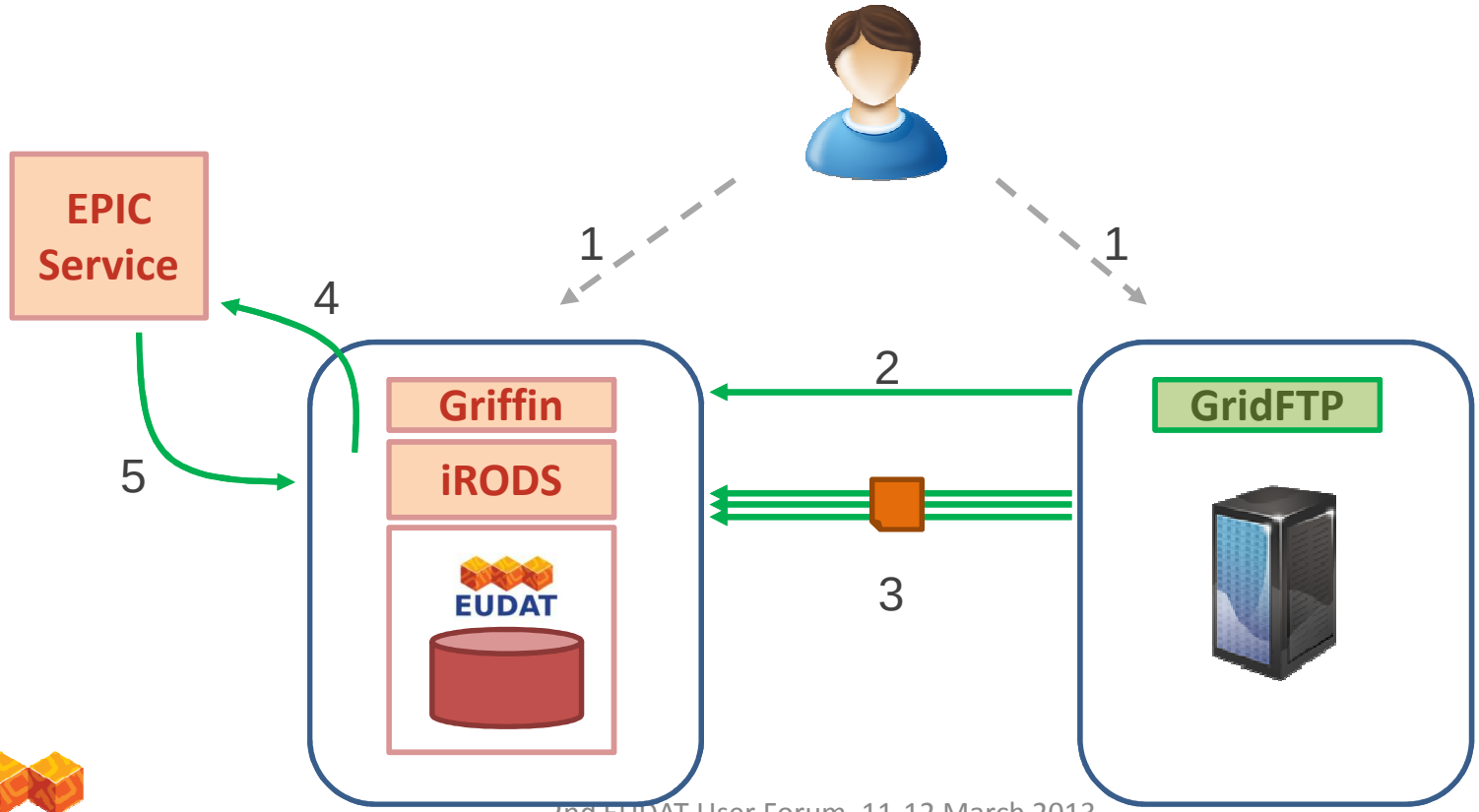
Globus Online

- **Globus Online** functionalities were evaluated with success and the Griffin component extended in order to support it



PIDs to staged data sets

- Data ingested back onto EUDAT resources are regularly registered through PID



Conclusions

- Data staging building blocks are right now available!
- Some extensions could be needed to better fit with the upcoming Authentication/Authorization Infrastructure
- XSEDE/EUDAT File Manager valuable interface to move data across different infrastructures (EUDAT, PRACE, XSEDE)
- Other transfer protocols, such HTTP or WebDav, are under investigation

Keep involved…

- **Visit:** http://www.eudat.eu/data-staging
- **Email:** eudat-datastaging@postit.csc.fi
- **Periodic news:** http://www.eudat.eu/newspublications

Useful links

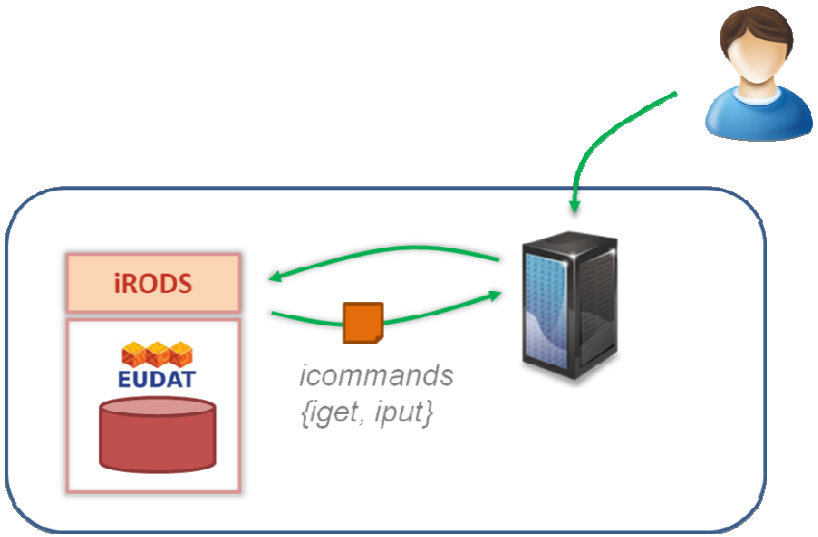
- **XSEDE-EUDAT File Manager**
  - github.com/TACC/filemanager/tree/eudat
- **gTransfer**
  - github.com/fr4nk5ch31n3r/gtransfer
- **iRODS**
  - www.irods.org
- **GridFTP**
  - www.globus.org/toolkit/data/gridftp
- **Globus OnLine**
  - www.globusonline.org
- **Griffin**
  - https://projects.arcs.org.au/trac/griffin

Any question?

Final Wrap-up

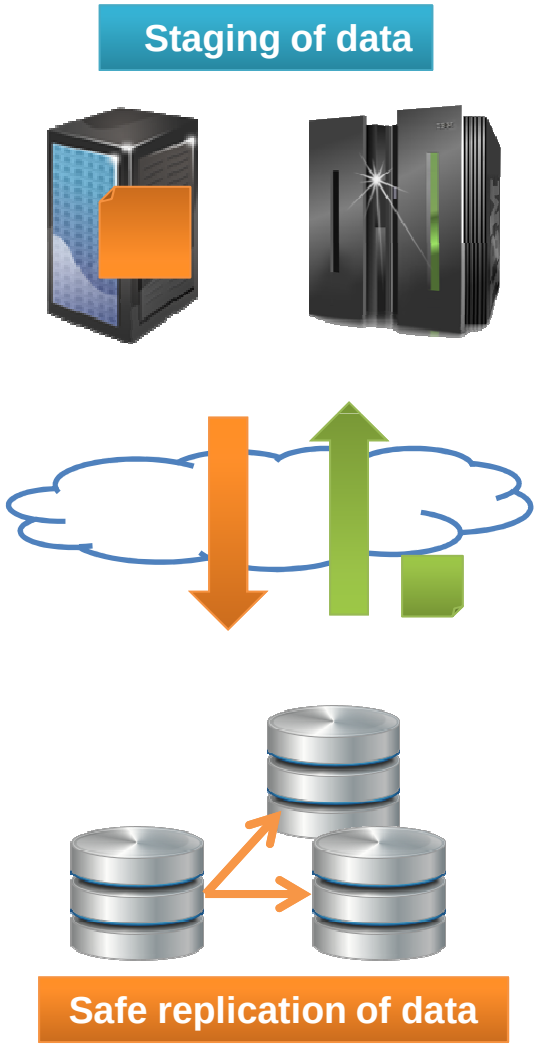
Demonstration

- Staging of data within the same site (CSC) through the VPH web portal
- ***Stefan Zasada (VPH@UCL)***



Data staging Session Wrap-up

- **Data Staging** to facilitate communities to stage stored data onto external computational facilities, such as HPC resources



Overall comments

- Authentication/Authorization/Accounting
  - Access to data replicas
  - Use of certificate
  - Harmonization of existing solutions/systems!
- Checksum of transferred data
  - Protocol specific
- SRM (Storage Resource Manager) interface
- How to get new communities onboard?

EURO-VO

- Astronomical data collected worldwide
- Heterogeneous, large data sets
- Metadata in XML (OAI-PMH)
- Using iRODS for managing data
- Many similarities with EUDAT
  - Simple store
  - Meta-data
  - Data Staging
- Move computation to data

EISCAT_3D

- Data from radar
  - atmospheric studies of the Fenno-Scandinavian Arctic
- Moving from 2D to 3D images
- Need of computational power
  - Targeting:100000 Pflops!
- Evolution roadmap
  - EISCAT 2D – 60TBs
  - EISCAT 3D 1st Phase (2018) – 1PB
  - EISCAT 3D 2nd Phase (2023) – 10 PB (EUDAT?)

Mapper

- Multi-scale models
  - Different communities being involved
- Distributed simulations
  - HPC, HTC resources
- GridFTP for data staging
- Efficiently organize data between resources
- Performance is a key
  - Small data sets but transferred frequently
- Data service for long term preservation and analysis