



## GENERAL INTRODUCTION TO DATA MANAGEMENT TOOLS: Globus GridFTP and iRODS

Michele Carpené – CINECA SCAI  
m.carpen@cineca.it  
Liverpool 03/09/2012



- **Introduction to data management**
  - Main problems
  - Critical features
  - Technologies (GridFTP/iRODS)
- **GridFTP**
  - What is GridFTP?
  - How can I use GridFTP?
  - third Party Transfer with Stripe
  - GridFTP clients
  - usage example
  - Ready available at CINECA
- **iRODS**
  - Overview of iRODS Data System
  - CINECA iRODS installation
- **References**

**Data Management:** a very crucial problem worldwide.

**Two fundamental issues:**

(from e-IRG “Blue Paper” on Data Management)

- Enormous growth of data
- missing common infrastructure for long term archiving

How to answer these requisites?

Experts working on this topic (e-IRG “Blue Paper) have isolated **important and critical features** about data access, data archiving, searching and movement in a “**data e-Infrastructure**”

## **Critical features:**

- **Reliability and Replications**
- **Metadata**
- **Unified Access and Interoperability**
- **Security**

Necessity to incorporate all aspects of data management providing services that meet the critical features.

European projects (example **EUDAT**, <http://www.eudat.eu/>) are going to provide a **Collaborative Data Infrastructure** whose design is driven by the needs of various user communities.

In this context two main technologies emerged for data management activities:

**GridFTP (Grid file transfer protocol)**

**iRODS (integrated Rule Oriented Data System)**

GridFTP is a protocol which extends established technologies like FTP (File Transfer Protocol) and SCP (Secure CoPy) with the following improvements:

- Authentication via GSI (ssh also supported)
- Multiple parallel channels with streams and stripes
- Third-party transfers
- Ability to tune network and I/O parameters

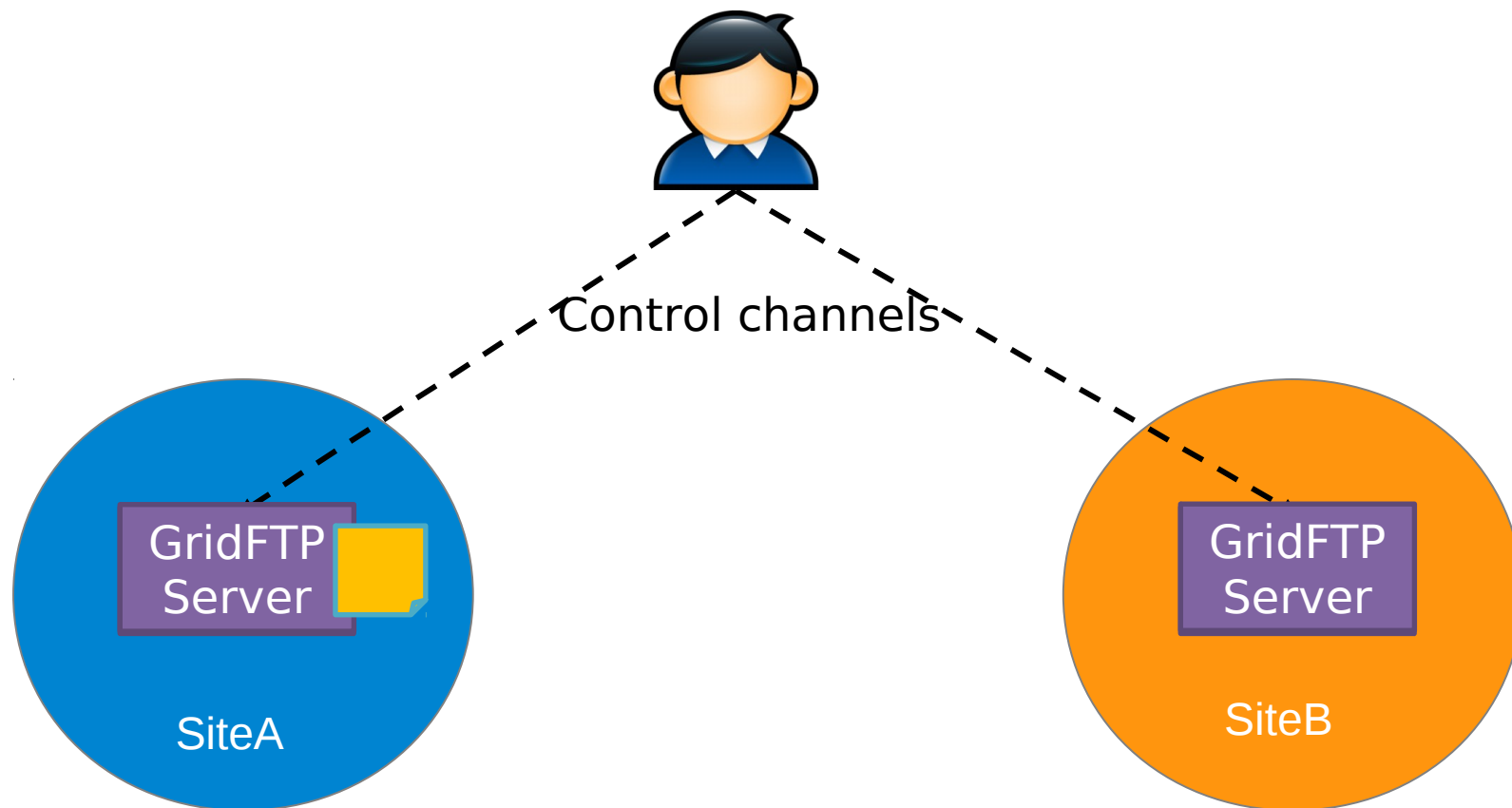
- GridFTP main client (globus-url-copy) is used like a normal FTP client
- It may only be necessary to define some variables

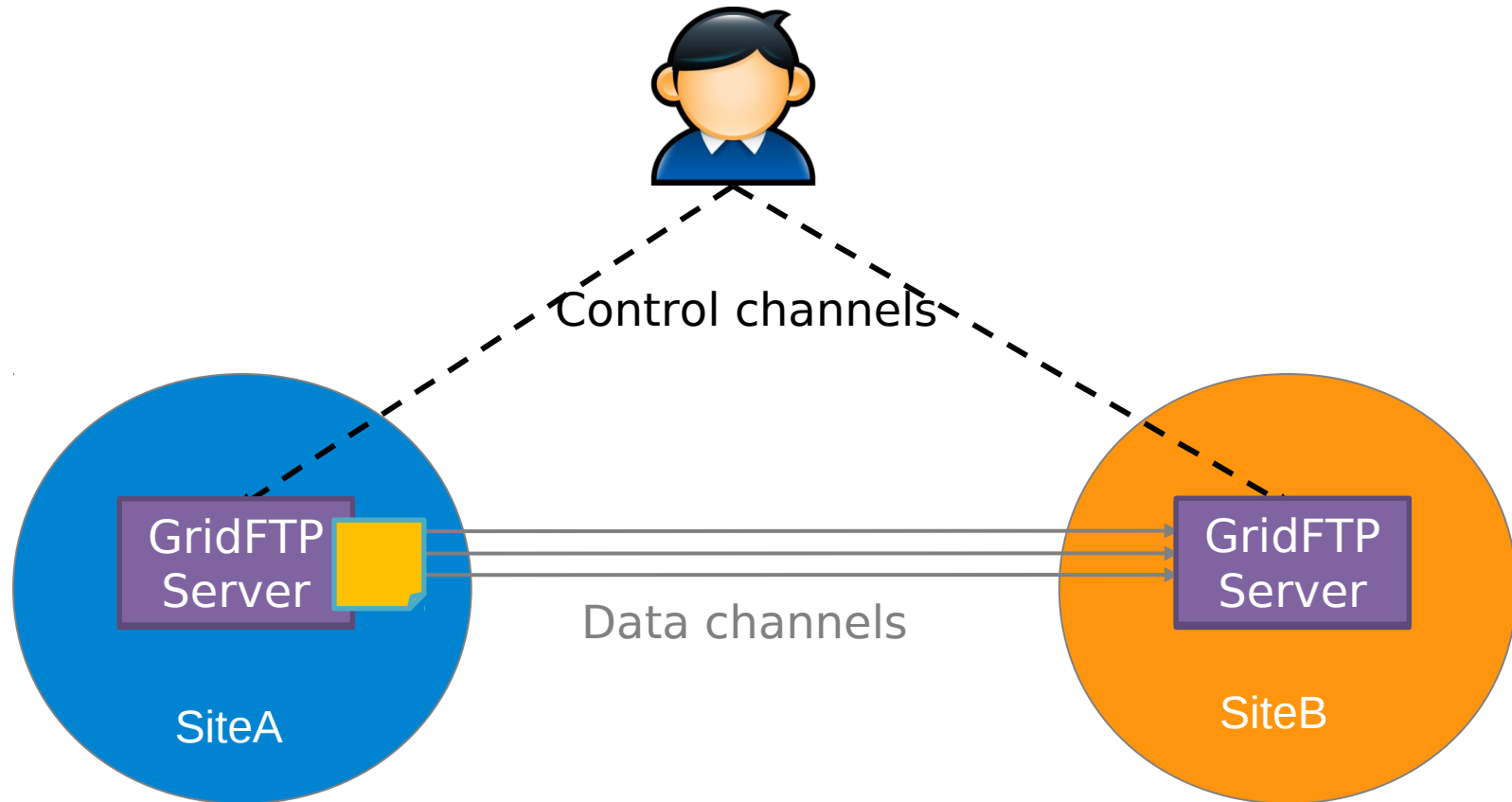
```
user$ globus-url-copy  
sshftp://<username>@login.plx.cineca.it/<remote_path/to/yourfile>  
file:///home/user/<local_path/to/yourfile>
```

```
user$ globus-url-copy gsiftp://gftp-plx.cineca.it/<remote_path/to/yourfile>  
file:///home/user/<local_path/to/yourfile>
```

- **p** <number>: number of streams, i.e. parallel tcp channels
- **stripe**: enable striped transfer
- **tcp-bs** <size>: dimension (in bytes) of the tcp buffer
- **pp**: allow pipelining
- **list**: used with one argument, lists the given directory
- **v**: verbose output
- **dbg**: very verbose output







- **UberFTP**: is the most “official” client and improves globus-url-copy making it more interactive and easier to use

## Interactive Commands

This listing is generated by typing 'help' at the command prompt.

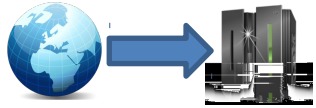
!	?	active	ascii	binary
blksize	bugs	bye	cat	cd
cdup	chgrp	chmod	cksum	close
dcau	debug	dir	family	get
glob	hash	help	keepalive	lcat
lcd	lcdup	lchgrp	lchmod	lclose
ldir	lls	lmkdir	lopen	lpwd
lquote	lrename	lrm	lrmdir	ls
lsize	lstage	mget	mkdir	mode
mput	open	order	parallel	passive
pbsz	pget	pput	prot	put
pwd	quit	quote	rename	resume
retry	rm	rmdir	runique	size
stage	sunique	tcpbuf	versions	wait

<http://dims.ncsa.illinois.edu/set/uberftp>

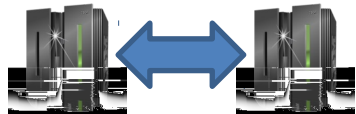
## GlobusOnline: GridFTP Client in the Cloud

The screenshot displays the GlobusOnline web interface for a file transfer. The browser address bar shows <https://www.globusonline.org/xfer/StartTransfer>. The interface includes a navigation bar with the GlobusOnline logo, a "Go To:" dropdown set to "Start Transfer", and a "Sign Out" button. Below the navigation bar, there are two panels for managing endpoints. The left panel shows a list of endpoints with columns for "select all", "none", "up one folder", "refresh list", and "Folder". The right panel shows a similar list for the destination endpoint. At the bottom, there is a "Get Globus Connect" section with a description: "Turn your computer into an endpoint. The easiest and most convenient way to send and receive files on your machine." and a "Label This Transfer" section with a text input field and a note: "This will be displayed in your tra".

<http://www.globusonline.org>



- **user@ws**\$globus-url-copy /path/file sshftp://user@remote-hpc/path/



- **user@plx**\$globus-url-copy -pp -restart gsiftp://grid.cineca.it/path/dir sshftp://user@remote-hpc/path/
- **user@plx**\$globus-url-copy -p 8 -restart gsiftp://grid.cineca.it/path/file sshftp://user@remote-hpc/path/

- A public installation (without -stripe option) is available for both CINECA and PRACE users. It is reachable at:
  - **GSI authentication at <gsiftp://gftp-plx.cineca.it:2812/>**
- An installation only for PRACE users (without -stripe option) is reachable at:
  - **GSI only at <gsiftp://gftp-prace.plx.cineca.it:2811/>**
- All of them are on PLX and let users read and write file systems on the base of their permissions

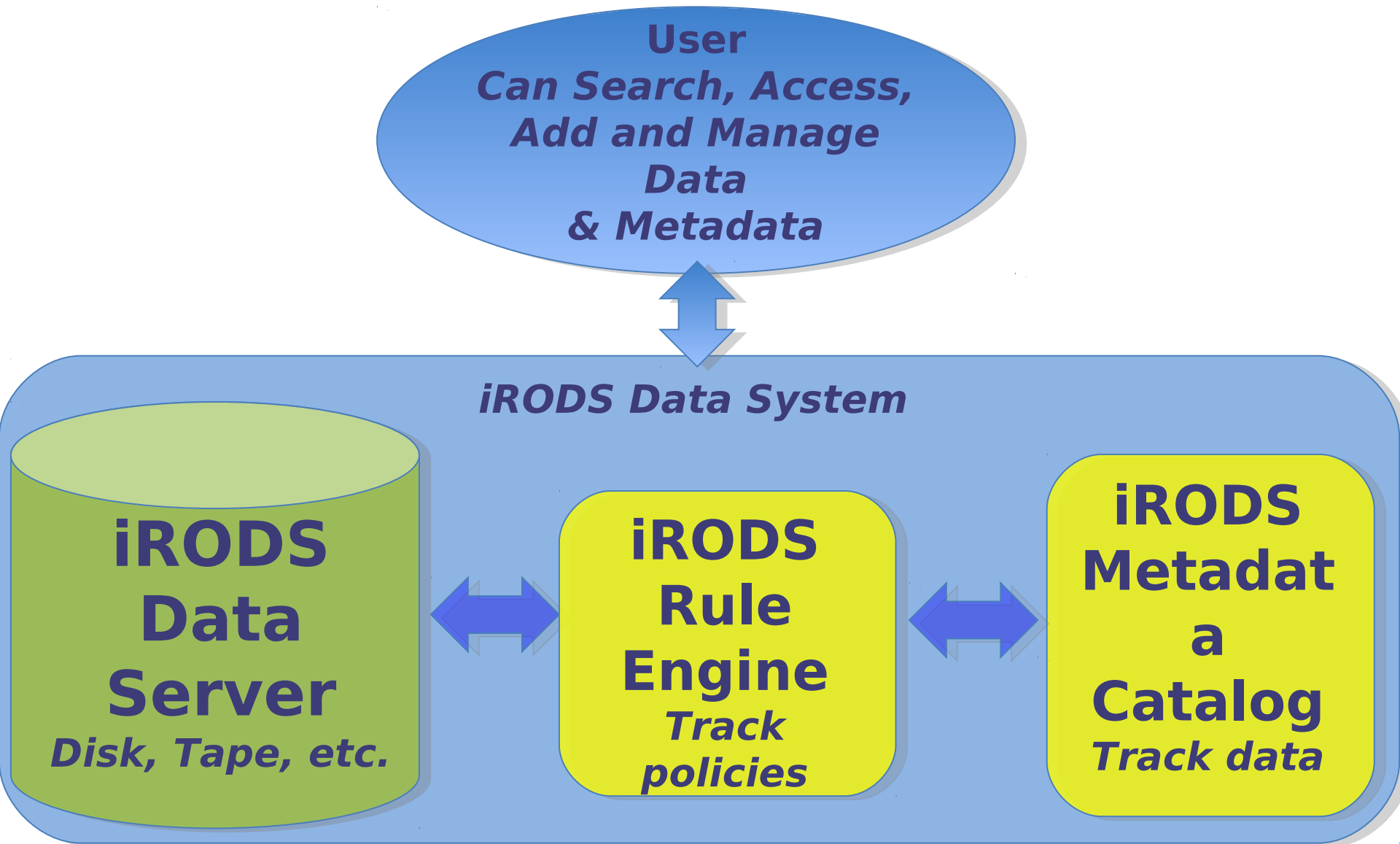
# iRODS: integrated Rule Oriented Data System

## Key Features:

- Build sharable collections from data distributed across file systems and tape archives;
- **iRODS iCAT Metadata Catalog** stores state information and descriptive metadata in a database;
- iRODS allows search, management, controlling and tracking of data access and manipulation;

**The Rule Engine** applies user-defined Policies and Rules to data to automate administrative tasks.

# Overview of iRODS Data System







The installation of iRODS in CINECA supports three authentication mechanisms:

- **Username-password;**
- **GSI authentication;**
- **OS Authentication.**

Users can use iRODS **icommands** (iput/iget) by typing:

**\$ module load profile/advanced irods**

You can access CINECA iRODS instance via web with your browser pointing at: `irods-dev.cineca.it` or, for WebDav, `irods-dev.cineca.it:8080`

The webdav web interface offers the possibility to share your files through the internet with only one click, as in the following picture:

The screenshot shows a web browser window with the URL `irods-dev.cineca.it:8080/#1331563138347`. The page header includes the CINECA logo and the text "You are logged in as <testuser>". Below the header, the breadcrumb path is `/CINECA > home > testuser > test`. There are two buttons: "Upload File" and "Create Directory". A table lists the files in the directory:

Name	Last Modified	Size	QuickShare	Select all
... Parent Directory				
<input checked="" type="checkbox"/> Newfile.txt	Fri, 17 Jan 2012 16:41:34 GMT	0 bytes		

On the right side of the interface, there is a vertical menu with the following options: "Access Control", "Metadata", "Replicas", "Rename", "Copy", "Move", "Delete", and "Share/Unshare". A red arrow points to the "Share/Unshare" button.

- [Blue Paper on Data Management](#)
- [GridFTP page at Globus official site](#)
- [GridFTP page at CINECA](#)
- [UberFTP official web-page](#)
- [Globus GUI web-page](#)
- [OFFIS GUI web-page](#)
- [iRODS Official Web Site](#)
- [CINECA iRODS Web Page](#)