

# Managing Data

**Bordeaux 28-29/01/2015**



## Data management components

Storage Elements

File Catalogs

DIRAC conventions for user data

Data operation commands

Data bookkeeping with the File Catalog CLI

Replica Catalog

Metadata Catalog

## Storage Elements

### gLite/EGI Storage Elements

Standard SRM interface

Gridftp protocol

Need Globus libraries, limited number of platforms

Allow third party transfers between them

Managed by the site managers within EGI SLAs

### DIRAC Storage Elements

- ◆ DASET based components

DIPS (Dirac Secure Protocol)

Does not allow third party transfers

Replication through local cache

Third party transfers will be available in the future

### Irods Storage Elements

▶ File Catalogs

LCG File Catalog (LFC)

Part of the EGI middleware

Service provided by the NGI

ORACLE backend

Client tools: command line, Python API

- Need Globus libraries

No User Metadata support

DIRAC File Catalog

DISET based components

Part of the DIRAC set of services

Community service

MySQL backend

Client tools: command line, CLI, Python API

Support of the User Metadata

For DIRAC users the use of any Storage Element or File Catalog is transparent

Community choice which components to use

Different SE types can be mixed together

Several File Catalogs can be used in parallel

- ▶ Complementary functionality
- ▶ Redundancy

Users see depending on the DIRAC Configuration

Logical Storage Elements

e.g. DIRAC-USER, M3PEC-disk, MCIA-irods

Logical File Catalog

Each file is identified by its Logical File Name (LFN)

Primary unique identifier

GUIDs are supported but their uniqueness is under the responsibility of user applications

This is different from LFC

Mostly for support of some applications, e.g. ROOT I/O

## LFN construction

Starts always with the VO name

`/vo.formation.idgrilles.fr/...`

User data

`/vo.formation.idgrilles.fr/user/v/vhamar/...`

## PFN (Physical File Name) construction

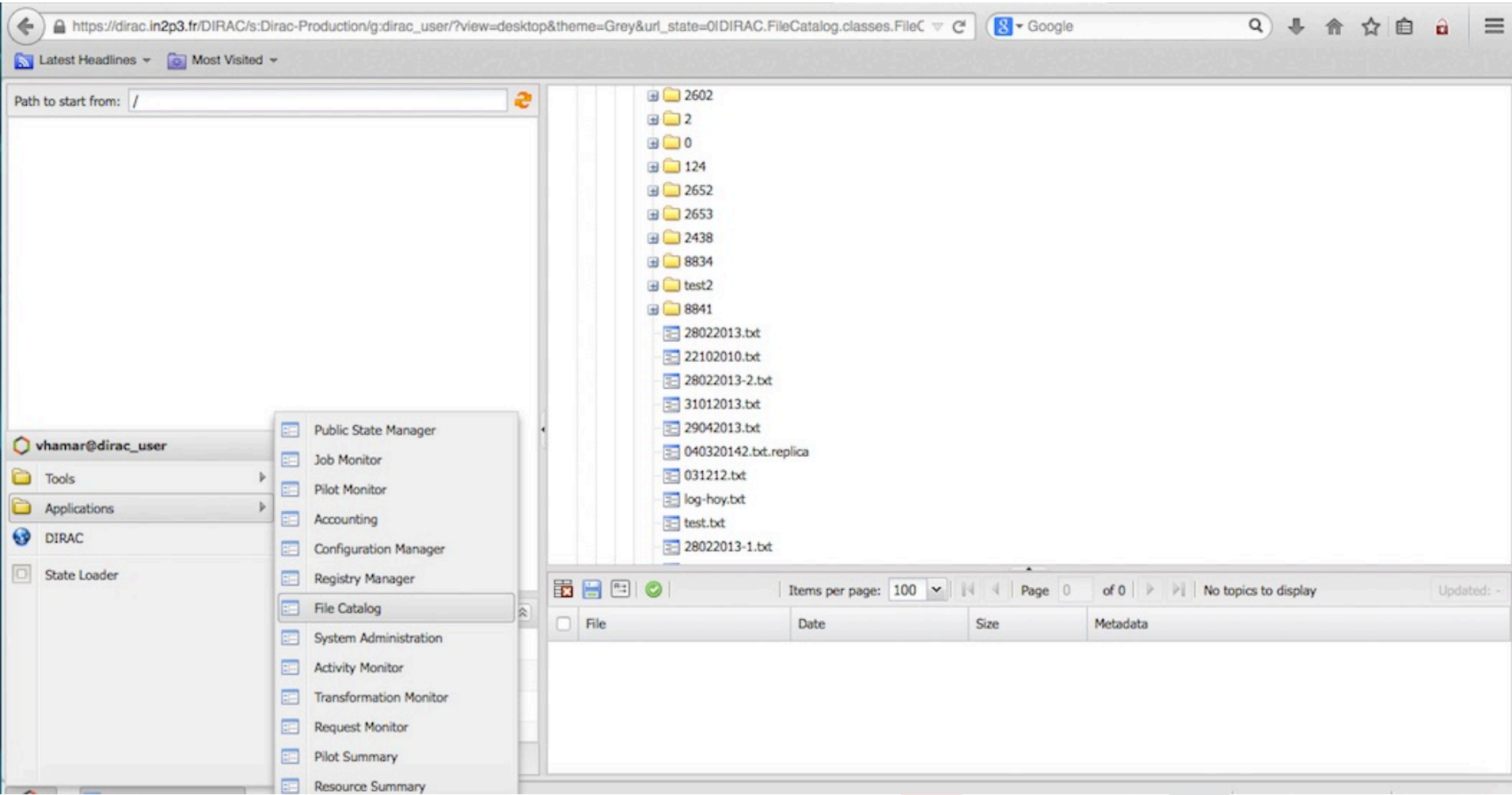
Always contains LFN as its trailing part

- ▶ **dirac-dms-add-file**
  - ▶ Upload file to the grid SE (lcg-cr)
- ▶ **dirac-dms-get-file**
  - ▶ Download file to the grid SE (lcg-cp)
- ▶ **dirac-dms-replicate-lfn**
  - ▶ Make another replica of a file (lcg-rep)
- ▶ **dirac-dms-lfn-replicas**
  - ▶ List replicas of a given file (lcg-lr)
- ▶ **dirac-dms-user-lfns**
  - ▶ Get a list of all the user files
- ▶ Plus others ...
  - ▶ See tutorial materials

- ▶ **dput** - works like a kind of scp command (in the local->remote direction):
  - ▶ `dput <local-source-path> [<lfn-destination-file-or-directory>]`
  - ▶ `dput <local-source-path>... <lfn-destination-directory>`
- ▶ **dget** - works in the opposite direction (remote->local):
  - ▶ `dget <lfn-source-path>`
  - ▶ `dget <lfn-source-path>... <local-destination-directory>`



- ▶ `dpwd` – shows current directory
- ▶ `dcd` – change directory
- ▶ `dls` - list FC directory contents
- ▶ `dmkdir` - create directory
- ▶ `drmdir` - remove directory
- ▶ `dreplicas` – file replication
- ▶ `drm` – file and directory removal



The screenshot shows a web browser window displaying the DIRAC File Catalog. The address bar shows the URL: [https://dirac.in2p3.fr/DIRAC/s:Dirac-Production/g:dirac\\_user/?view=desktop&theme=Grey&url\\_state=0](https://dirac.in2p3.fr/DIRAC/s:Dirac-Production/g:dirac_user/?view=desktop&theme=Grey&url_state=0). The interface includes a navigation menu on the left with the user 'vhamar@dirac\_user' and categories like Tools, Applications, DIRAC, and State Loader. A dropdown menu is open, listing various system management tools, with 'File Catalog' selected. The main content area displays a file tree with folders (e.g., 2602, 2, 0, 124, 2652, 2653, 2438, 8834, test2, 8841) and files (e.g., 28022013.txt, 22102010.txt, 28022013-2.txt, 31012013.txt, 29042013.txt, 040320142.txt.replica, 031212.txt, log-hoy.txt, test.txt, 28022013-1.txt). At the bottom, a table header is visible with columns for File, Date, Size, and Metadata. The table currently contains no data, with a message 'No topics to display' and 'Page 0 of 0'.

## Usefull links

Follow this tutorial:

<https://github.com/DIRACGrid/COMDIRAC/wiki/Data-Management>

## Other DIRAC data management tools

<https://github.com/DIRACGrid/DIRAC/wiki/DataManagement>

<https://github.com/DIRACGrid/DIRAC/wiki/FileCatalog>

## With DIRAC command line tools

- ▶ Getting data files to the grid
- ▶ Downloading data files from the grid
- ▶ Replicating files
- ▶ Exploring the File Catalog console



**Backup slides**

- ▶ File Catalog operations are generally synchronous
  - ▶ Quick, can wait for the prompt
- ▶ Physical data operations can take very long time
  - ▶ And even fail in the end
- ▶ For example, consider removing data:
  - ▶ Delete replicas on all the SEs
  - ▶ Delete files (lfns)
  - ▶ Delete directories ( recursively )
- ▶ Long operations are performed asynchronously
  - ▶ Do not wait for completion
  - ▶ Make sure the operation is accomplished despite possible problems

Metadata can be associated with each directory as key:value pairs to describe its contents

Int, Float, String, DateTime value types

Some metadata variables can be declared indices

Those can be used for data selections

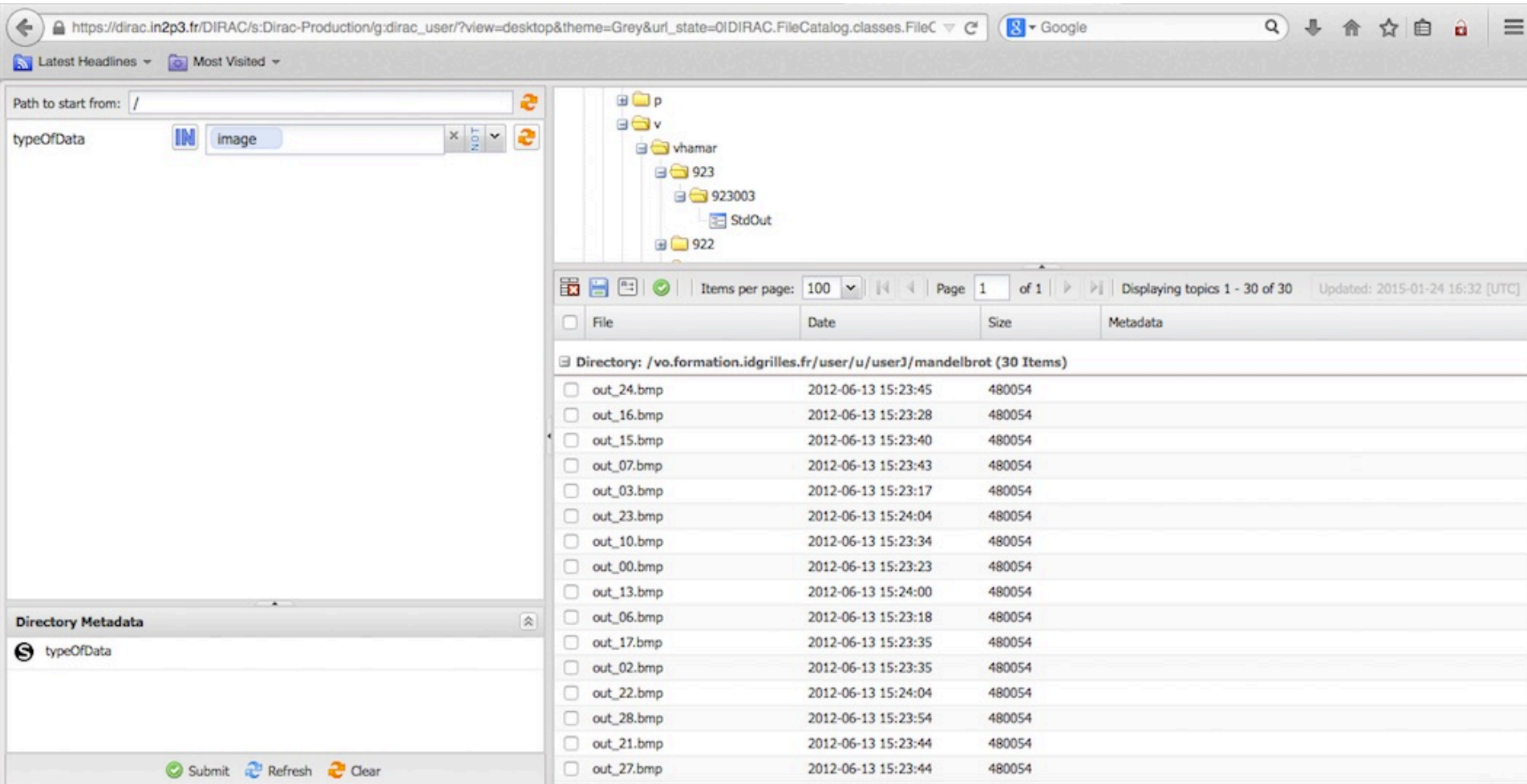
Subdirectories are inheriting the metadata of their parents

Data selection with metadata queries

Example:

```
find /dirac/user Meta1=Value1 Meta2>3 Meta2<5 Meta3=2,3,4
```

File metadata can also be defined



Path to start from: /

typeOfData image

Directory: /vo.formation.idgrilles.fr/user/u/user3/mandelbrot (30 Items)

File	Date	Size	Metadata
<input type="checkbox"/> out_24.bmp	2012-06-13 15:23:45	480054	
<input type="checkbox"/> out_16.bmp	2012-06-13 15:23:28	480054	
<input type="checkbox"/> out_15.bmp	2012-06-13 15:23:40	480054	
<input type="checkbox"/> out_07.bmp	2012-06-13 15:23:43	480054	
<input type="checkbox"/> out_03.bmp	2012-06-13 15:23:17	480054	
<input type="checkbox"/> out_23.bmp	2012-06-13 15:24:04	480054	
<input type="checkbox"/> out_10.bmp	2012-06-13 15:23:34	480054	
<input type="checkbox"/> out_00.bmp	2012-06-13 15:23:23	480054	
<input type="checkbox"/> out_13.bmp	2012-06-13 15:24:00	480054	
<input type="checkbox"/> out_06.bmp	2012-06-13 15:23:18	480054	
<input type="checkbox"/> out_17.bmp	2012-06-13 15:23:35	480054	
<input type="checkbox"/> out_02.bmp	2012-06-13 15:23:35	480054	
<input type="checkbox"/> out_22.bmp	2012-06-13 15:24:04	480054	
<input type="checkbox"/> out_28.bmp	2012-06-13 15:23:54	480054	
<input type="checkbox"/> out_21.bmp	2012-06-13 15:23:44	480054	
<input type="checkbox"/> out_27.bmp	2012-06-13 15:23:44	480054	

Submit Refresh Clear