

# Managing metadata

DIRAC Tutorial



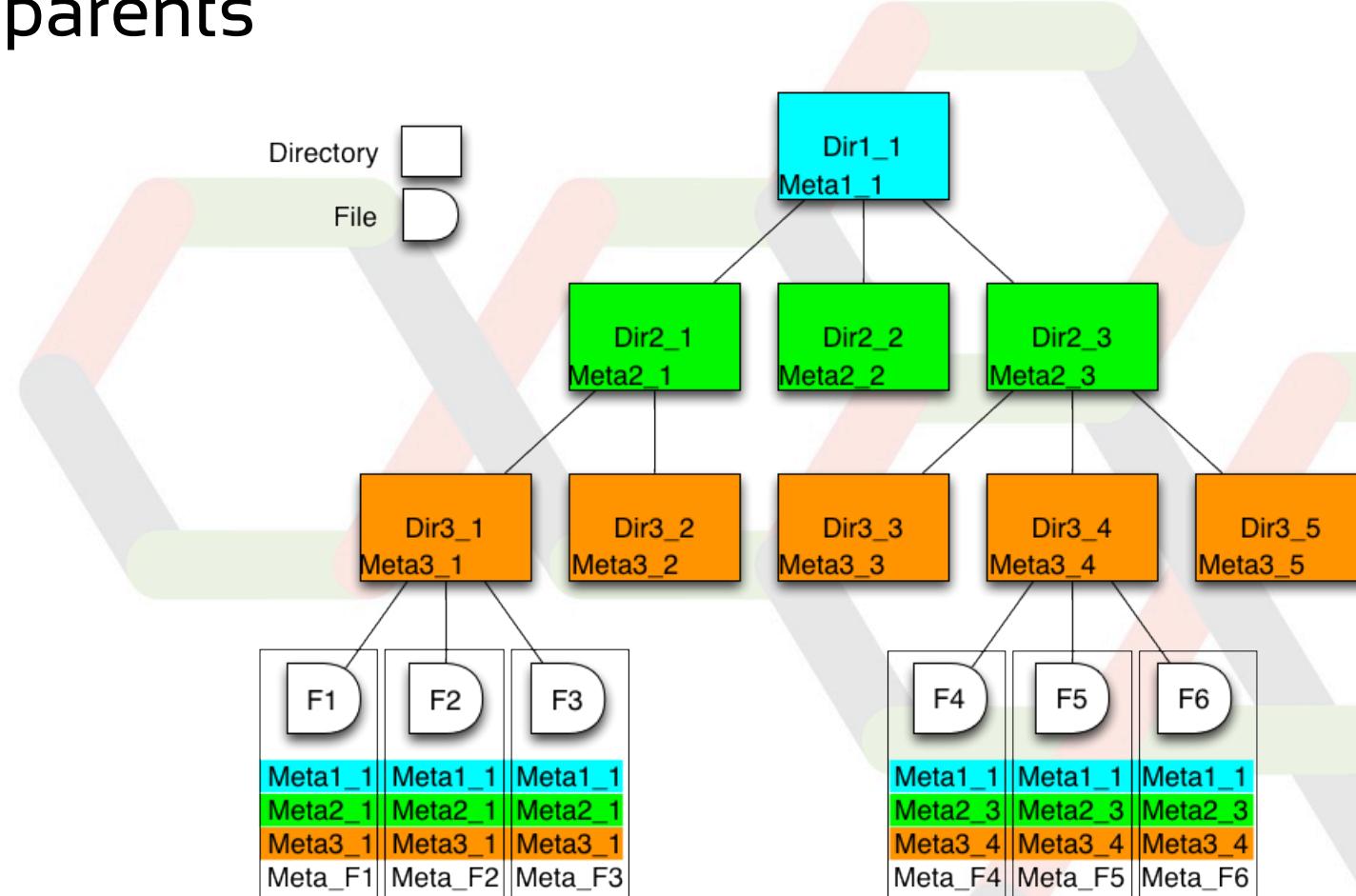
# FileCatalog Metadata

---

- User metadata can be associated to files and directories as key:value pairs
  - Value types: integer, float, string, date
- Some metadata can be declared globally as indices for further use in file selection / discovery
  - If not indexed, metadata can be considered as annotations

# FileCatalog Metadata

- Files and subdirectories inherit metadata of their parents



## ▶ Data selection with metadata queries

- ▶ Example with COMDIRAC

```
dfind . Year=2010 Version=v1r0,v1r1 SE=CERN-disk  
"CreationDate>10-10-2013"
```

## ▶ Web interfaces to make queries easier

- ▶ Dynamically updated selectors as a function of already selected values
  - ▶ Available for directory metadata only

## ▶ COMDIRAC ***dmeta*** command

### ▶ Examples:

- ▶ Add metadata to file or directory

```
dmeta add ./some_lfn_file some_meta="some_value"
```

- ▶ Look up metadata for a given path

```
dmeta ls ./some_lfn_file
```

- ▶ Remove metadata

```
dmeta rm ./some_lfn_file some_meta
```

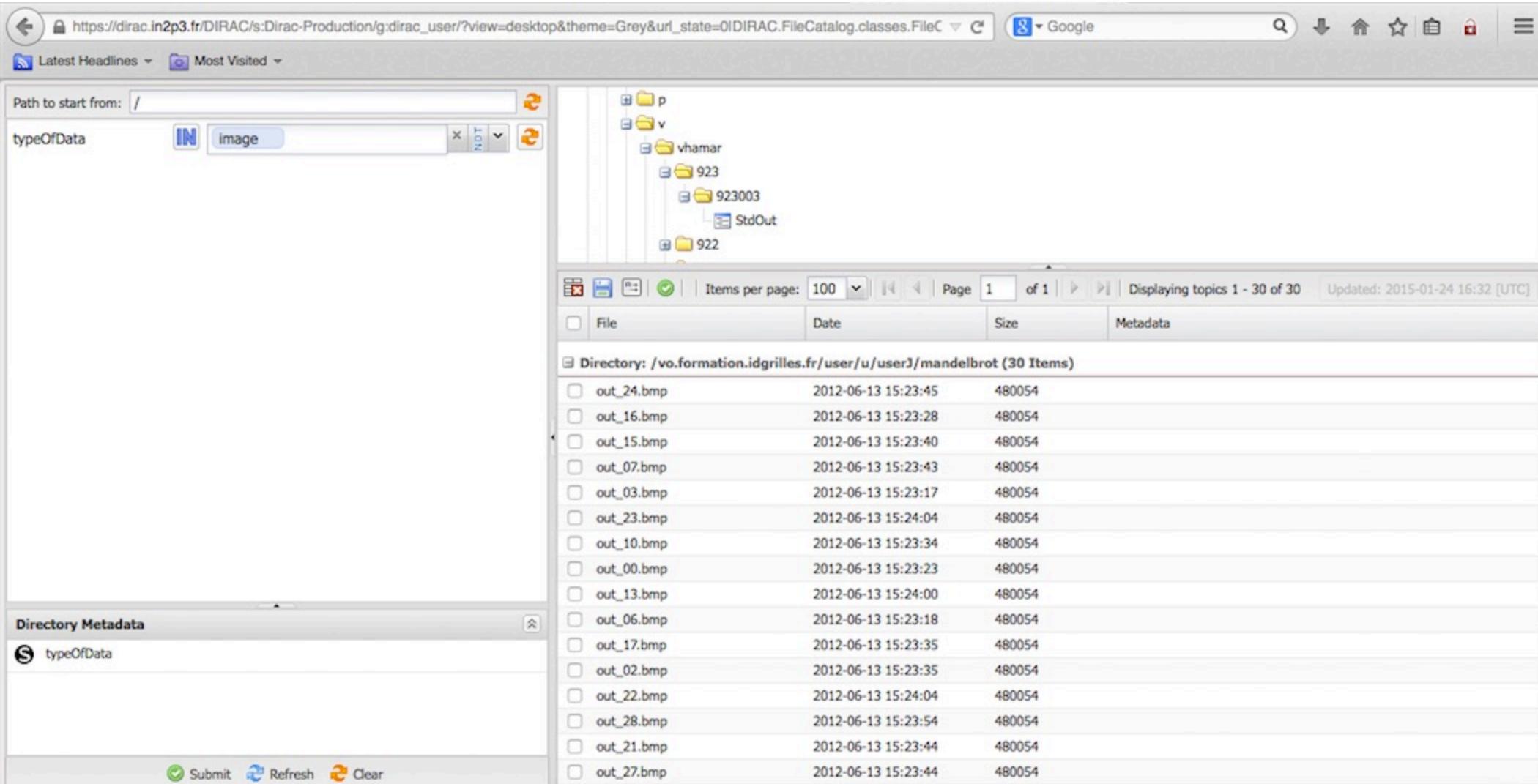
- ▶ Create metadata index

```
dmeta -l f|d meta=(int|float|string|date)
```

- ▶ Show all metadata indices

```
dmeta -l
```

# File Catalog Wep App



The screenshot shows a web browser window displaying the DIRAC File Catalog application. The URL in the address bar is [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 page title is "DIRAC File Catalog".

The interface includes:

- A top navigation bar with links for "Latest Headlines" and "Most Visited".
- A search bar and a "Google" button.
- A sidebar on the left with a "Path to start from:" input field containing "/", a "typeOfData" dropdown set to "image", and a "Search" button.
- A main content area showing a hierarchical file tree:
  - /
  - p
  - v
    - vhamar
      - 923
        - 923003
          - StdOut
      - 922
- A table below the file tree showing 30 items in the directory:

File	Date	Size	Metadata
out_24.bmp	2012-06-13 15:23:45	480054	
out_16.bmp	2012-06-13 15:23:28	480054	
out_15.bmp	2012-06-13 15:23:40	480054	
out_07.bmp	2012-06-13 15:23:43	480054	
out_03.bmp	2012-06-13 15:23:17	480054	
out_23.bmp	2012-06-13 15:24:04	480054	
out_10.bmp	2012-06-13 15:23:34	480054	
out_00.bmp	2012-06-13 15:23:23	480054	
out_13.bmp	2012-06-13 15:24:00	480054	
out_06.bmp	2012-06-13 15:23:18	480054	
out_17.bmp	2012-06-13 15:23:35	480054	
out_02.bmp	2012-06-13 15:23:35	480054	
out_22.bmp	2012-06-13 15:24:04	480054	
out_28.bmp	2012-06-13 15:23:54	480054	
out_21.bmp	2012-06-13 15:23:44	480054	
out_27.bmp	2012-06-13 15:23:44	480054	
- A "Directory Metadata" panel on the left showing "typeOfData" and buttons for "Submit", "Refresh", and "Clear".

# FileCatalog special metadata

- Functionality is similar to the AMGA gLite service or iRODS metadata
- Some extra features of DFC
  - Support the data provenance information
    - Ancestor<->descendent relationships
  - Support for efficient storage usage reports
    - Realtime
    - Necessary for the storage quota policies

# Storage Usage

```
FC:/> size -l /lhcb/user/a/atsareg/1
directory: /lhcb/user/a/atsareg/1
Logical Size: 134,756,846 Files: 498 Directories: 500

  StorageElement    Size      Replicas
=====
  1 IN2P3-USER     20,254,050    75
  2 CNAF-USER      18,363,672    68
  3 RAL-USER       16,473,294    61
  4 CERN-USER      19,443,888    72
  5 GRIDKA-USER    21,064,212    78
  6 SARA-USER      20,254,050    75
  7 PIC-USER       18,903,780    70
-----
  Total           134,756,946   499
Query time 0.98 sec
```

Report of storage usage for any directory

- ▶ Whole community data
- ▶ Per user data
- ▶ “Logical” storage
  - ▶ LFNs, sum of the LFN sizes
- ▶ “Physical” storage
  - ▶ Physical replicas, total volume per Storage Element

# Support for datasets

- ▶ A relatively new feature developed for the BES III experiment in IHEP, Beijing
  - ▶ Datasets defined as a resulting list of files from a given metaquery
    - ▶ Particular case: all the files under a given directory
  - ▶ Dataset objects are stored in the same directory hierarchy as files
    - ▶ ACLs, ownership, show up in the output of *ls* command as for files
  - ▶ Datasets can be frozen in order not to change from one query to another
    - ▶ Can be refreshed by an explicit command, a quick check if changed since the last update
  - ▶ Datasets can be annotated
  - ▶ Operations on datasets
    - ▶ Replica lookup for all the files in a dataset
    - ▶ Total size, number of files report
    - ▶ More to come, e.g. replication, removal, etc

- Tutorial :
- <https://github.com/DIRACGrid/COMDIRAC/wiki/Metadata-Management>
- Add file and directory metadata
- Remove metadata
- Define metadata indices
- Search files depending on metadata values