

Managing metadata

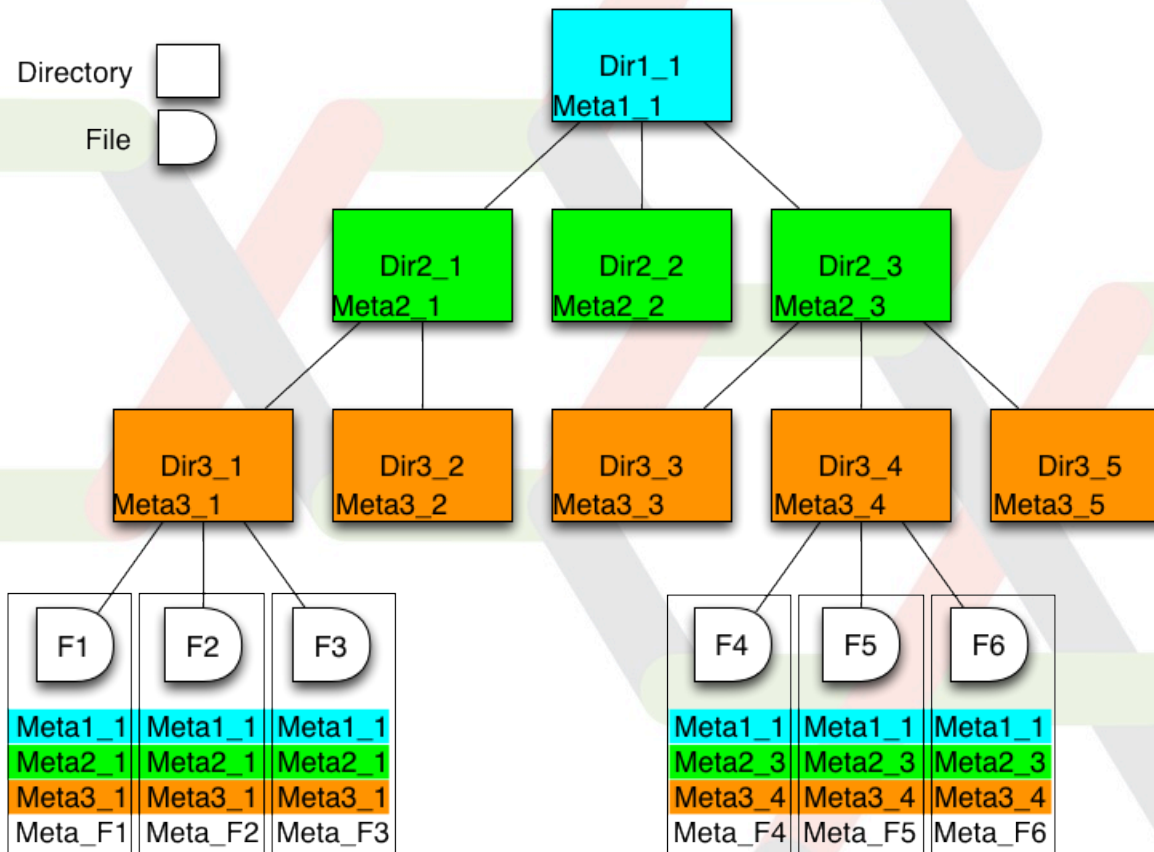
DIRAC Tutorial



- 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



- Files and subdirectories inherit metadata of their parents



- ▶ Data selection with metadata queries

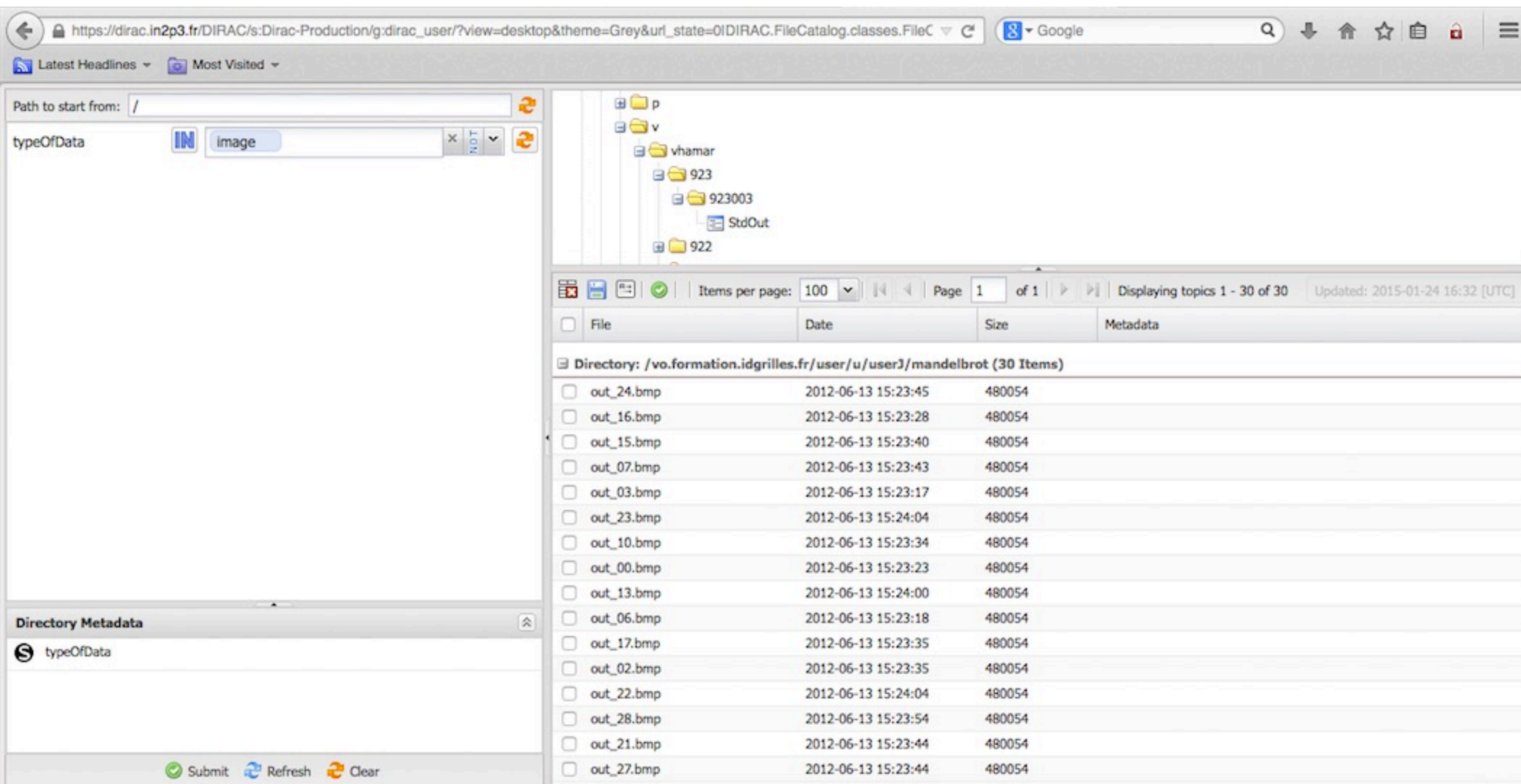
- ▶ Example for DFC shell

```
find . Year=2010 Version=v1r0,v1r1 SE=CERN-disk  
      ProductionTime>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 *dfind* command mimicking the unix *find*



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=0DIRAC.FileCatalog.classes.FileC`. The interface includes a search bar, a breadcrumb trail, and a table of files.

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	

Directory Metadata: typeOfData

Submit Refresh Clear

- 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



```
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

- ▶ 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

