

**EGI Hands On Training for AEGIS Users  
Institute of Physics Belgrade**

## GRID Data Management

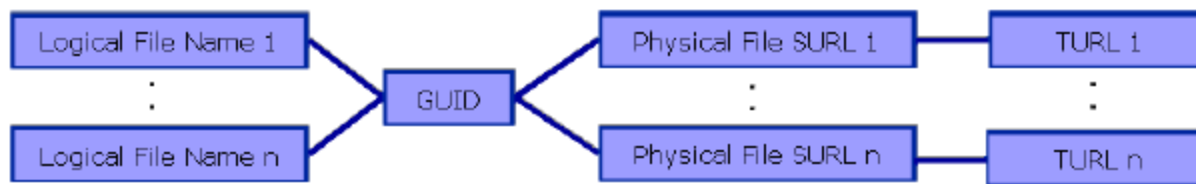
Danica Stojiljkovic  
Scientific Computing Laboratory  
Institute of Physics Belgrade  
[danica@ipb.ac.rs](mailto:danica@ipb.ac.rs)



- Storage elements types
- Grid file names
- LFC Concept
- Data management client tools
  - File catalogue interaction
  - Replica tool

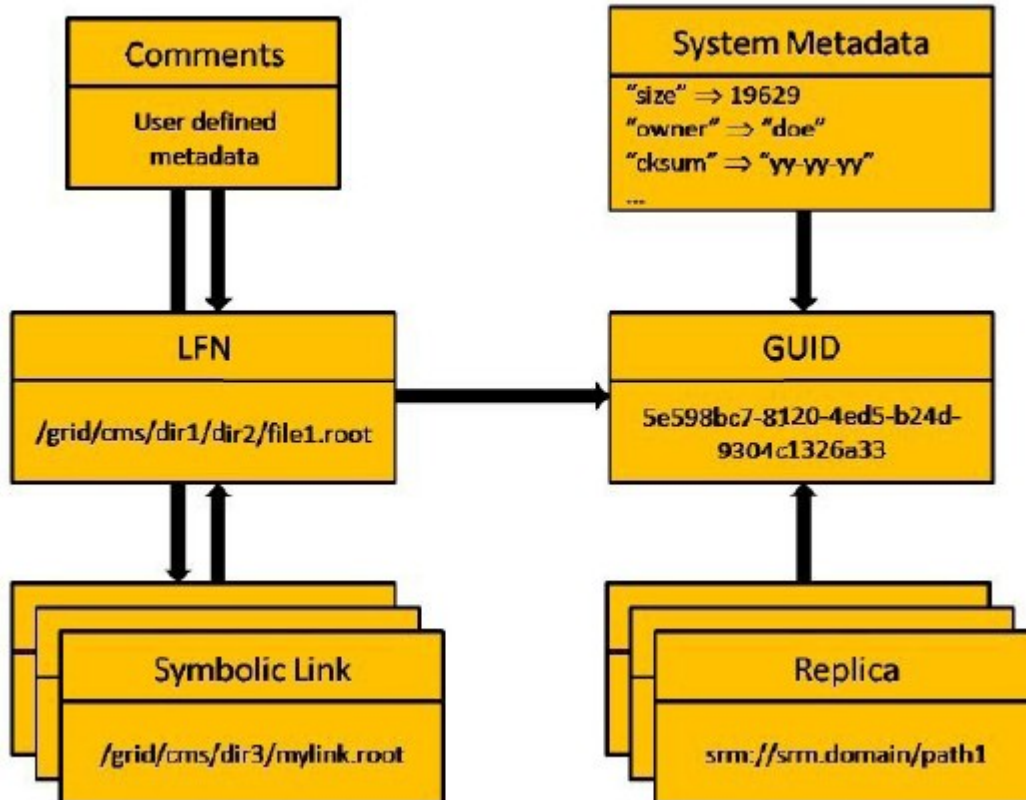
- Classic SE (obsolete)
- LCG Disk pool manager (DPM)
- CASTOR
- dCache
- Resource management is done through Storage Resource Manager (SRM) interface
  - Hides the complexity of the resources

- LCG File Catalogue (LFC)
  - Maintains mappings between LFN(s), GUID and SURL(s)
  - Provides System metadata
  - Provides single string user metadata
- GRID file
  - Both physically present in a SE and registered in the file catalogue
  - Files are write-once, read-many
  - Files cannot be changed unless remove or replaced



- Grid Unique Identifier (GUID)
  - Identifies a file uniquely, is of the form:
    - `guid:<36_bytes_unique_string>`
    - `guid:38ed3f60-c402-11d7-a6b0-f53ee5a37e1d`
- Logical File Name (LFN) or User Alias
  - Used to refer to a file in human readable form
    - `lfn:/grid/<MyVO>/<MyDirs>/<MyFile>`
    - `lfn:/grid/aegis/vlada/outputs/output1.txt`

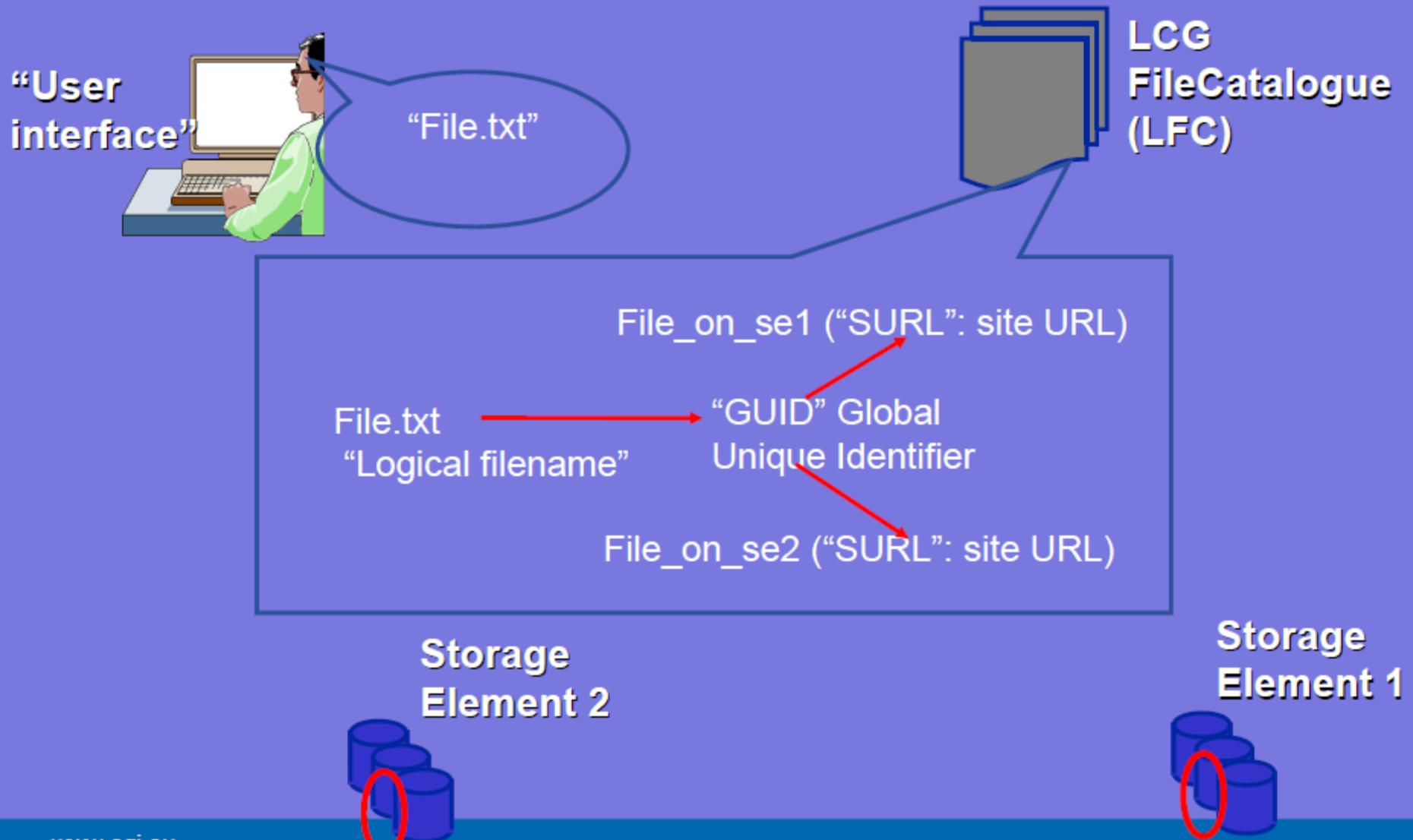
- Storage URL (SURL)
  - Identifies a replica in a SE
    - `<sfn|srm>://<SE_hostname>/<some_string>`
    - `srm://se.ipb.ac.rs/dpm/ipb.ac.rs/home/aegis/generated/2010-10-10/fileb3469b96-d956-4062-9662-4da7e203557d`
- Transport URL (TURL)
  - Valid URI with the necessary information to access a file in a SE
    - `<protocol>://<some_string>`
    - `gsiftp://se2.ipb.ac.rs/se2.ipb.ac.rs:/storage8/aegis/2010-10-10/fileb3469b96-d956-4062-9662-4da7e203557d.2015029.0`



- Directory structure
  - /grid/<vo>/<you\_create\_it>
  - /grid/aegis/vlada
- All members of given VO have rw permissions in their directory



# Data management scheme



- Unix-like LFC commands:
  - echo \$LFC\_HOST
  - lfc-ls List file/directory entries in a directory
  - lfc-mkdir Create directory
  - lfc-ln Make a symbolic link to a file/directory
  - lfc-ls List file/directory entries in a directory
  - lfc-chmod Change access mode of a LFC file/directory
  - lfc-chown Change owner and group of a LFC file/directory
  - lfc-getacl Get file/directory access control lists
  - lfc-setacl Set file/directory access control lists
- User needs a valid proxy for using these commands

- Listing the entries of a LFC directory
  - `$ lfc-ls /grid/aegis/vlada/misanu`  
result1  
result2
- Creating directories in the LFC
  - `$ lfc-mkdir /grid/aegis/vlada/misanu/new_dir`
  - `$ lfc-ls -l /grid/aegis/vlada`  
drwxrwxr-x 0 165 101 0 Oct 10 15:53 new\_dir  
-rw-rw-r-- 1 165 101 12 Oct 10 15:51 result1  
-rw-rw-r-- 1 165 101 12 Oct 10 15:51 result1

- Creation of symbolic links

```
$ lfc-ln -s /grid/aegis/vlada/misanu/result1 /grid/aegis/vlada/misanu/result1_link
```

```
$ lfc-ls -l /grid/aegis/vlada/misanu/result1_link
```

```
  rwxrwxrwx 1 165 101 0 Oct 10 16:05
```

- `grid/aegis/vlada/misanu/result1_link -> /grid/aegis/vlada/misanu/result1`

- Adding metadata information to LFC entries

```
$ lfc-setcomment /grid/aegis/vlada/misanu/result2 "best result"
```

```
$ lfc-ls --comment /grid/aegis/vlada/misanu/result2
```

```
  /grid/aegis/vlada/misanu/result2 best result
```

- Removing LFNs from the LFC

```
$ lfc-rm -r /grid/aegis/vlada/misanu/novi_dir
```

- Print the ACL of a directory

```
$ lfc-getacl /grid/aegis/vlada/misanu
```

```
file: /grid/aegis/vlada/misanu
```

```
# owner: /C=RS/O=AEGIS/OU=Institute of Physics Belgrade/CN=Vladimir Slavnic
```

```
# group: aegis
```

```
user::rwx
```

```
group::rwx
```

```
#effective:rwx
```

```
other::r-x
```

```
default:user::rwx
```

```
default:group::rwx
```

```
default:other::r-x
```

- LCG Data Management tools ( lcg\_utils)
- High level tools
- Hide the complexities of catalogue and SEs interaction
- CLI, API

- Check environment
  - echo \$LCG\_GFAL\_INFOSYS
  - echo VO\_<VO>\_DEFAULT\_SE
- lcg-cp Copies a Grid file to a local destination (download)
- lcg-cr Copies a file to a SE and registers the file in the catalogue (upload)
- lcg-del Deletes one file (either one replica or all replicas)
- lcg-rep Copies a file from one SE to another SE and registers it in the catalogue (replicate)
- lcg-gt Gets the TURL for a given SURL and transfer protocol

- `lcg-aa` Adds an alias in the catalogue for a given GUID
- `lcg-ra` Removes an alias in the catalogue for a given GUID
- `lcg-rf` Registers in the catalogue a file residing on an SE
- `lcg-uf` Unregisters in the catalogue a file residing on an SE
- `lcg-la` Lists the aliases for a given LFN, GUID or SURL
- `lcg-lr` Lists the replicas for a given LFN, GUID or SURL
- `lcg-ls` Lists file information for given SURLs or LFNs



- Upload a file
  - `$ lcg-cr file:/home/vlada/misanu/mpi.jdl  
guid:830e7a82-f9fb-4c04-aafd-e24a5f25acdb`
- Specify lfn:
  - `$ lcg-cr -l lfn:/grid/aegis/vlada/misanu/mpi  
file:/home/vlada/misanu/mpi.jdl guid:25fdcd16-  
bc76-4458-9c1a-8c597fb33194`
- Replicate a file (one replica per SE)
- `$ lcg-rep -d grid15.rcub.bg.ac.rs  
lfn:/grid/aegis/vlada/misanu/mpi`

- List replicas
  - `$ lcg-lr lfn:/grid/aegis/vlada/misanu/mpi`  
`srm://grid15.rcub.bg.ac.rs/dpm/rcub.bg.ac.rs/home/aegis/generated/2010-10-10/file531a6e7f-ab4a-4a9b-bb9f-a2d9a693e615`  
`srm://se.ipb.ac.rs/dpm/ipb.ac.rs/home/aegis/generated/2010-10-10/fileb3469b96-d956-4062-9662-4da7e203557d`
- Copying files out of the Grid (download)
  - `$ lcg-cp lfn:/grid/aegis/vlada/misanu/mpi`  
`file:/home/vlada/misanu/mpi`
- Deleting replicas
  - `lcg-del`  
`srm://grid15.rcub.bg.ac.rs/dpm/rcub.bg.ac.rs/home/aegis/generated/2010-10-10/file531a6e7f-ab4a-4a9b-bb9f-a2d9a693e615`
  - `$ lcg-del -a lfn:/grid/aegis/vlada/misanu/mpi`

- List guid

```
$ lcg-lg lfn:/grid/aegis/vlada/misanu/mpi  
guid:25fdcd16-bc76-4458-9c1a-8c597fb33194
```

- Listing files and directories

```
$ lcg-ls -l lfn:/grid/aegis/vlada/misanu/  
-rw-rw-r-- 1 165 101 606 mpi  
-rw-rw-r-- 1 165 101 12 result1  
-rw-rw-r-- 1 165 101 12 result1_link  
-rw-rw-r-- 1 165 101 18 result2
```

- the input files do not have to be on the host from which the job is submitted;
- the output files are immediately available when the job ends, without having to issue a command to retrieve them;
- the sandbox files do not have to go through the WMS host, which otherwise can easily become a bottleneck.

- If the job input files are stored on a GridFTP server, it is possible to specify those files as GridFTP URI in the InputSandbox attribute:
  - InputSandbox =  
`{"gsiftp://lxb0707.cern.ch/cms/doe/data/fileA, "fileB"};`

- In order to store the output sandbox files to a GridFTP server, the `OutputSandboxDestURI` attribute must be used
  - `OutputSandbox = {"fileA", "data/fileB", "fileC"};`
  - `OutputSandboxDestURI = {"gsiftp://lxb0707.cern.ch/cms/doe/fileA", "gsiftp://lxb0707.cern.ch/cms/doe/fileB", "fileC"};`

- Another possibility is to use the `OutputSandboxBaseDestURI` attribute to specify a base URI on a GridFTP server where the files listed in `OutputSandbox` will be copied.

```
OutputSandbox = {"fileA", "fileB"};
```

```
OutputSandboxBaseDestURI = "gsiftp://lxb0707.cern.ch/cms/does/";
```

- Note: the directory on the GridFTP where the files have to be copied must already exist

- <https://edms.cern.ch/file/722398/1.3/gLite-3-UserGuide.pdf>
- [http://wiki.egee-see.org/index.php/SEE-GRID\\_Wiki](http://wiki.egee-see.org/index.php/SEE-GRID_Wiki)
- [http://wiki.egee-see.org/index.php/SG\\_Data\\_Managemet\\_High\\_Level\\_Tools](http://wiki.egee-see.org/index.php/SG_Data_Managemet_High_Level_Tools)