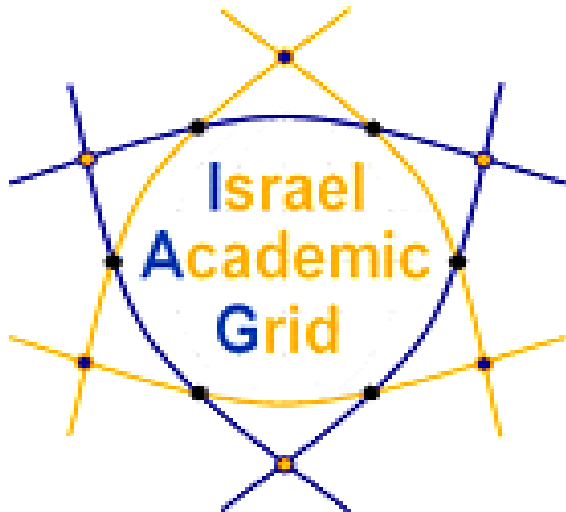


Grid Data Management

www.eu-egee.org

Assaf Gottlieb
Tel-Aviv University
assafgot @ tau.ac.il

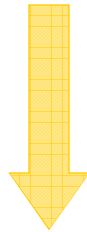


- **Introduction**
- **Grid Data Management Services**
- **File catalogues**
- **Data Management commands**
- **Hands on**

- The Input / Output Sandbox is limited to small files (< 10 MB)
- Large files are stored in permanent resources called **SE = Storage Elements.**
- SE are present at almost every site together with the computing resources
- No periodical deletion! User responsible for his files.

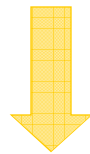
Grid Data Management Services enable users to:

- move files in and out of the Grid
- Replicate files on different SE's
- Locate files on various SE's



Data Management means movement and replication of files on grid elements

- Data transfer is done by a number of protocols (gsiftp, rfiio, file, etc`)
- Usage of a central **File catalogue**



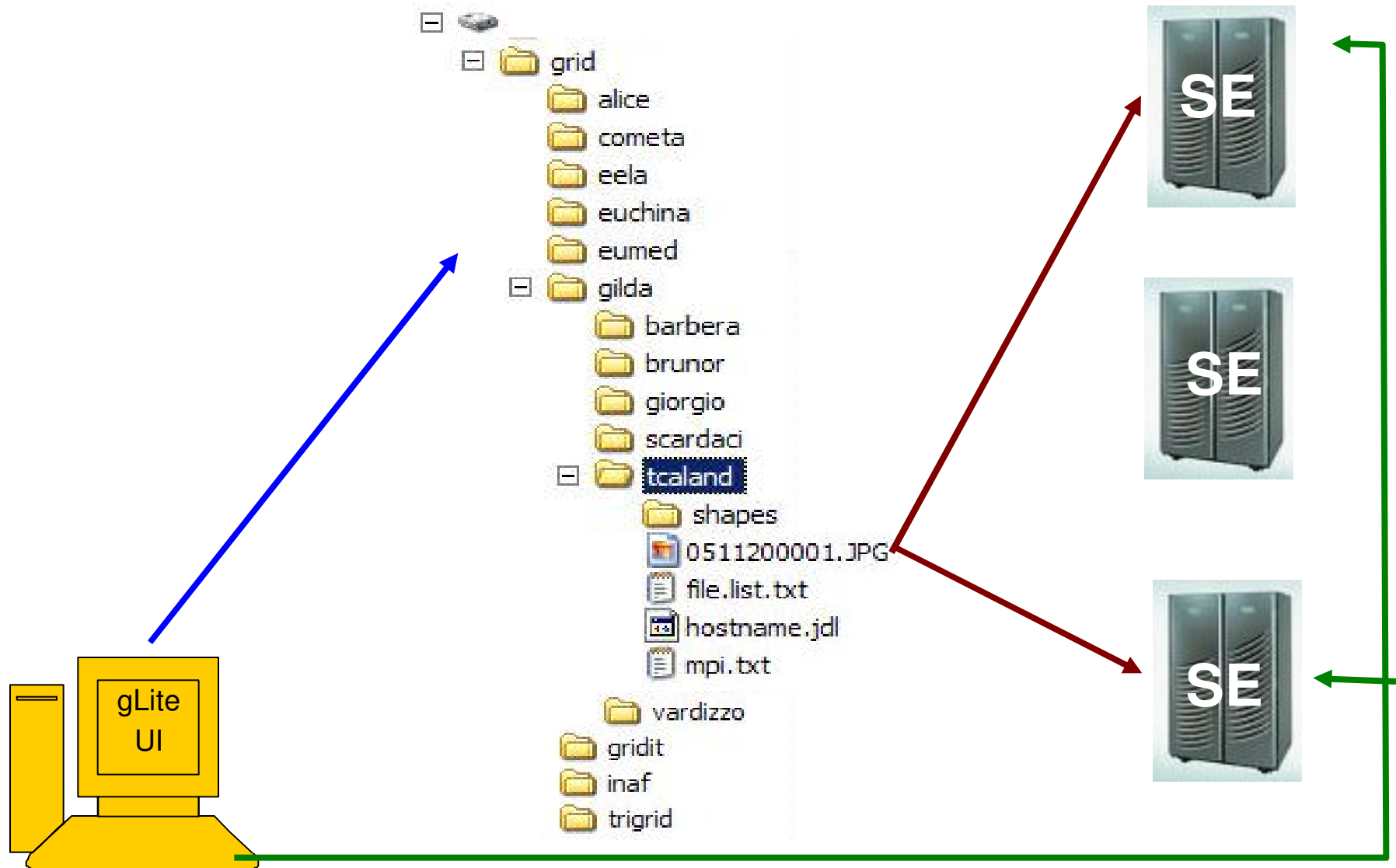
By using high level data management tools which enable transparency of the transport layer details (protocols) , storage location and the internal structure of the SE's



The SE is a “black box”

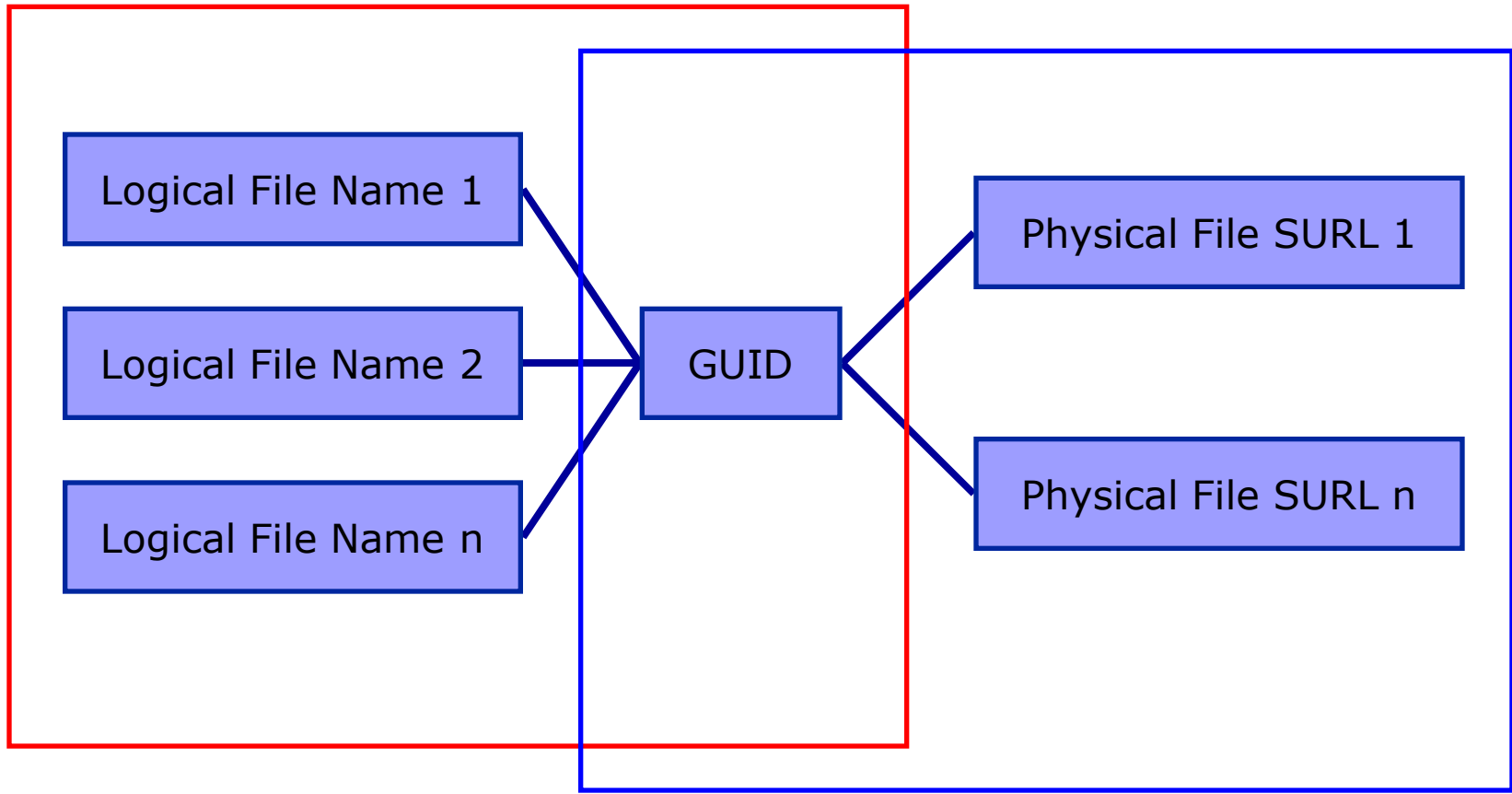
- How do I keep track of all of the files I have on the Grid ?
- Even if I remember all the lfn's of my files, what about someone else's files ?
- How does the Grid keep track of lfn-guid-surl associations ?
- **Well... for that we have a FILE CATALOG**

File Catalogs



File Catalogs – cont'd

RMC = Replica Metadata Catalog



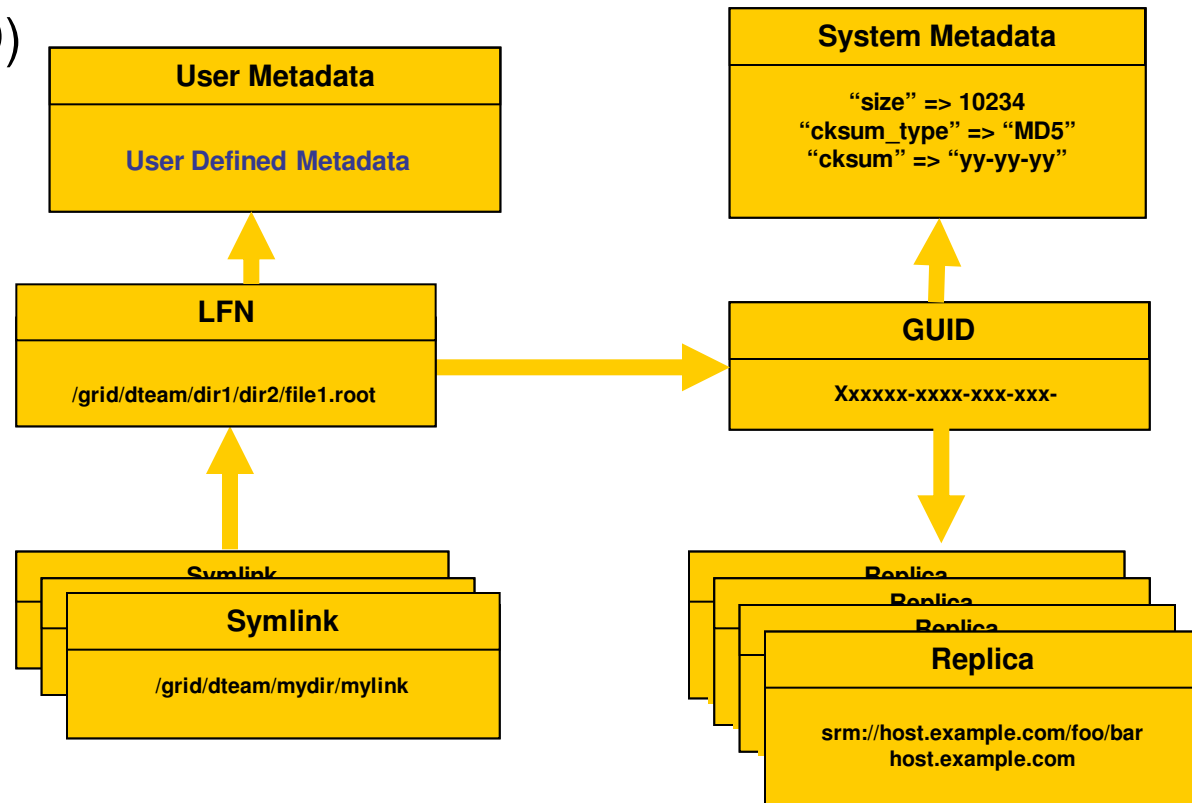
LRC = Local Replica Catalog

File Catalogs – cont'd

- The LFN acts as a main key in the database.

It has:

- Symbolic links to it (additional LFNs)
- Unique Identifier (GUID)
- System metadata
- Information on replicas



Files : name conventions

■ Logical File Name (LFN)

- An alias created by the user to refer to some file
- A LFN is of the form: **lfn:/grid/<MyVO>/<MyDir>/<MyFile>**
- **Example:** lfn:/grid/gilda/importantResults/Test1240.dat

■ Globally Unique Identifier (GUID)

- A file can always be identified by its GUID (based on UUID)
- A GUID is of the form: **guid:<unique_string>**
- All replicas of a file will share the same GUID
- **Example:** guid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6

both lfn's and guid's refer to files (not replicas)



Replicas : name conventions

- **Storage URL (SURL)**
- **(AKA: Physical/Storage File Name (PFN/SFN))**
 - Used by the LRC to find where the replica is physically stored
 - A SURL is of the form:
sfn://<SE_hostname>/<VO_path>/<file_name>
 - **Example:** sfn://tbed1.cern.ch/flatfiles/SE00/gilda/project1/testSUTL.dat
- **Transport URL (TURL)**
 - Temporary locator of a physical replica including the access protocol understood by a SE
 - A TURL is of the form:
<protocol>://<SE_hostname>/<VO_path>/<filename>
 - **Example:** gsiftp://tbed1.cern.ch/gilda/project1/testTURL.dat

provide info about the physical location of the **replica**



- **lcg-cp** Copies a Grid file to a local destination
- **lcg-cr** Copies a file to a SE and registers the file in the LRC
- **lcg-del** Deletes one file (either one replica or all replicas)
- **lcg-rep** Copies a file from SE to SE and registers it in the LRC

Data Management commands – cont'd

- **lcg-lg** Gets the guid for a given lfn or surl
- **lcg-aa** Adds an alias in RMC for a given guid
- **lcg-la** Lists the aliases for a given LFN, GUID or SURL
- **lcg-gt** Gets the turl for a given surl and transfer protocol

Data Management commands – cont'd

- **lcg-lr** Lists the replicas for a given lfn, guid or surl
- **lcg-ra** Removes an alias in RMC for a given guid
- **lcg-rf** Registers a SE file in the LRC (optionally in the RMC)
- **lcg-uf** Un-registers a file residing on an SE from the LRC

File catalog commands

- **lfc-ls** List file/directory entries in a directory.
- **lfc-mkdir** Create directory.
- **lfc-rename** Rename a file/directory.
- **lfc-rm** Remove an empty directory.
- **lfc-chmod** Change access mode of a file/directory
- **lfc-chown** Change owner and group of a file/directory

