



Enabling Grids for E-scienceE

Data Management in Grid Environment

Marek Ciglan

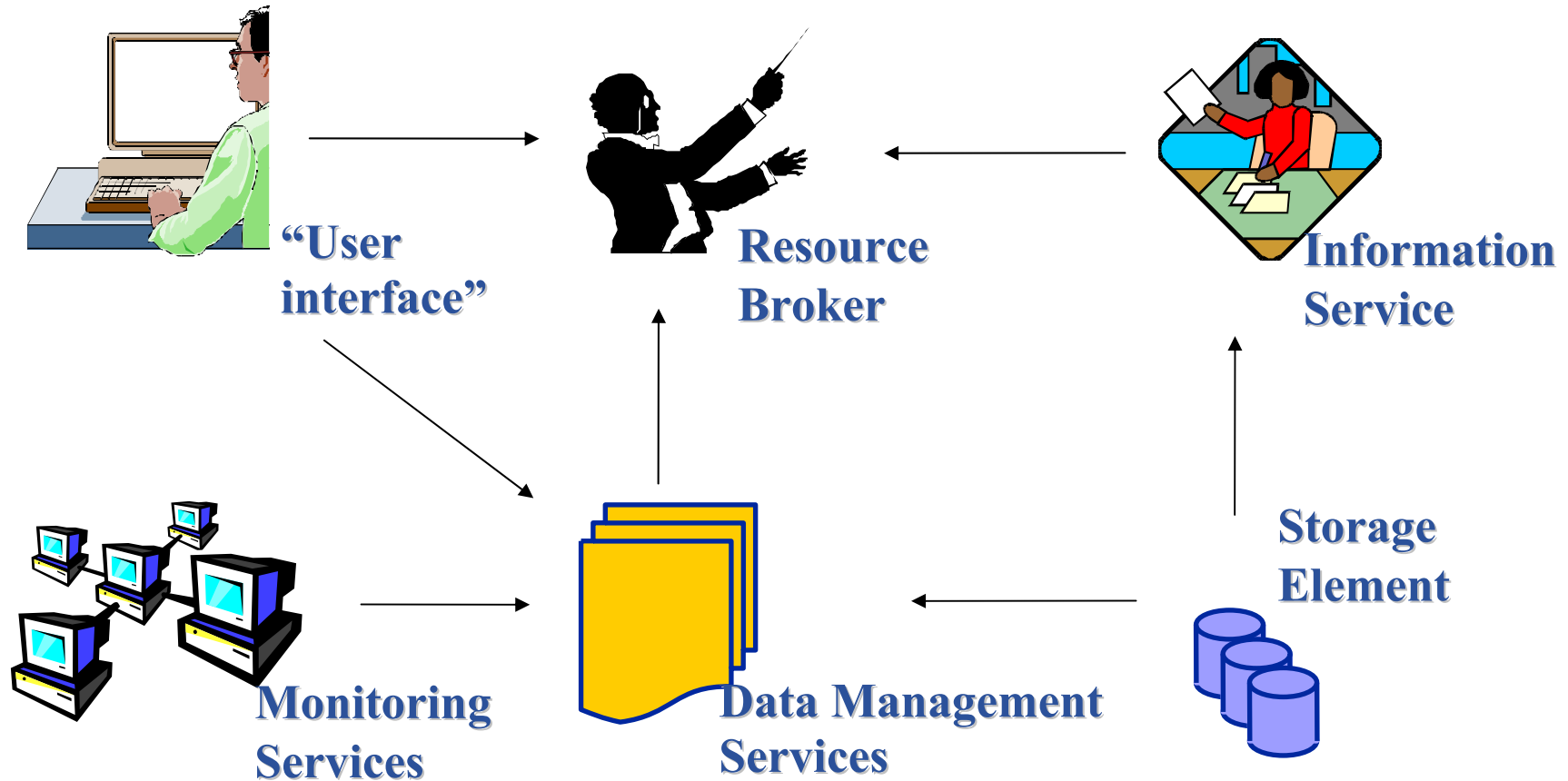
*Ústav informatiky, SAV
Dúbravská cesta 9, Bratislava
Slovakia*

www.eu-egee.org



INFSO-RI-508833

- **Safe and reliable file transfer between grid nodes**
- **Gathering and maintenance of information about data sources in the grid**
- **Data access optimization**
- **Providing of metadata services**
- **Data sets consistency maintenance**



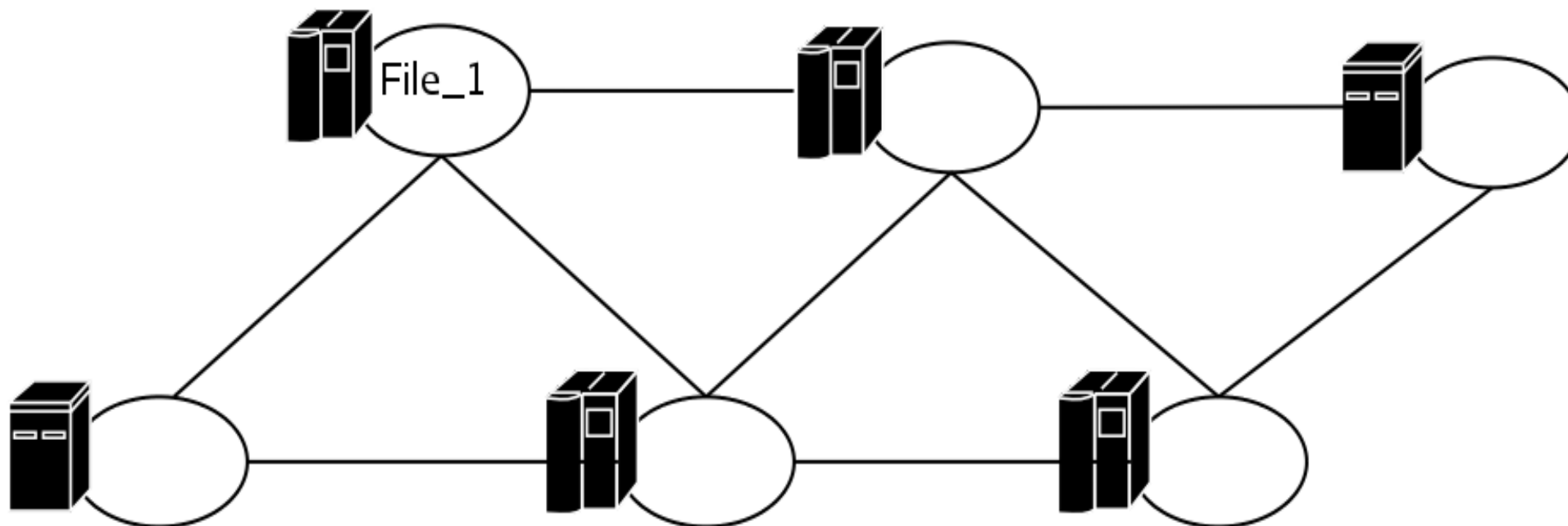
GridFTP - based on FTP:

- Widely implemented and well-understood standard protocol
- Well-defined architecture for protocol extensions

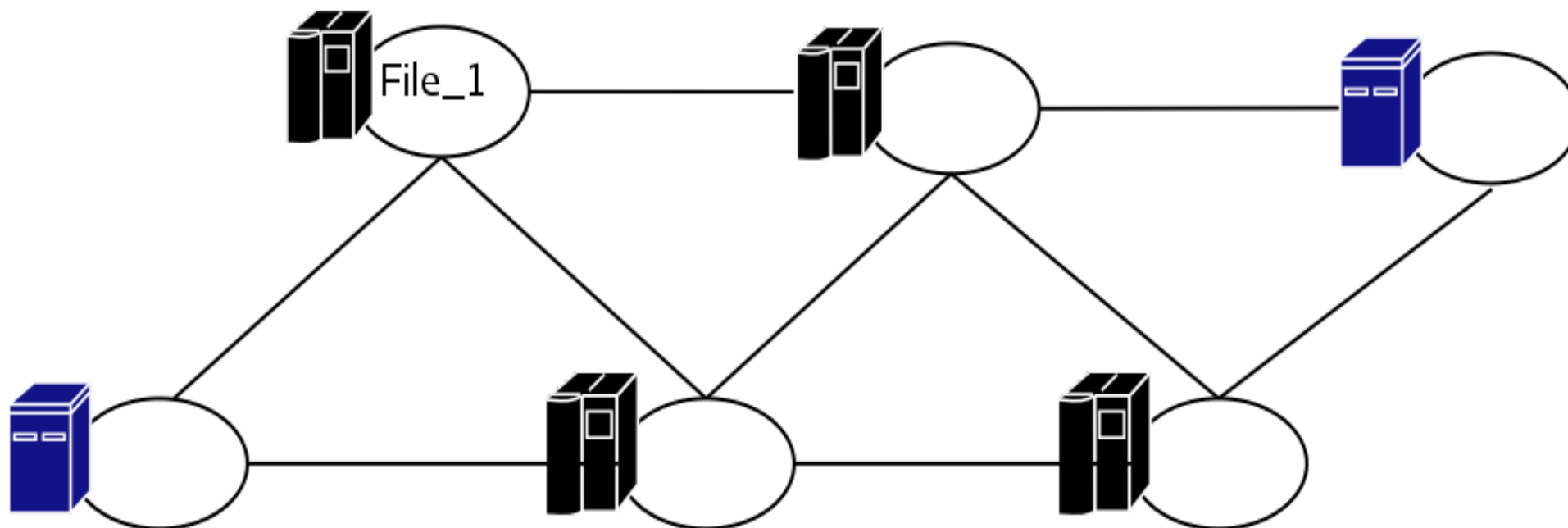
GridFTP - extension of FTP:

- Grid Security Infrastructure (GSI)
- Third-party control of data transfer
- Parallel data transfer
- Partial file transfer

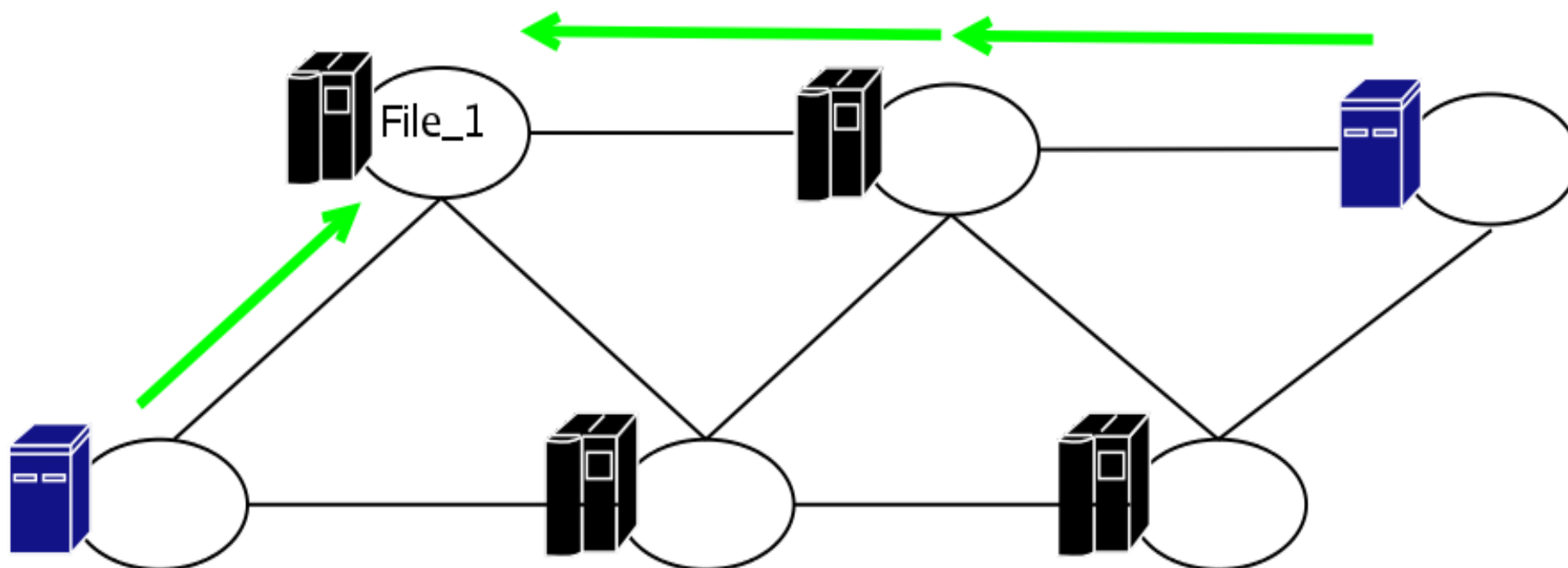
Replication of Data Sets - key concept to increase data availability



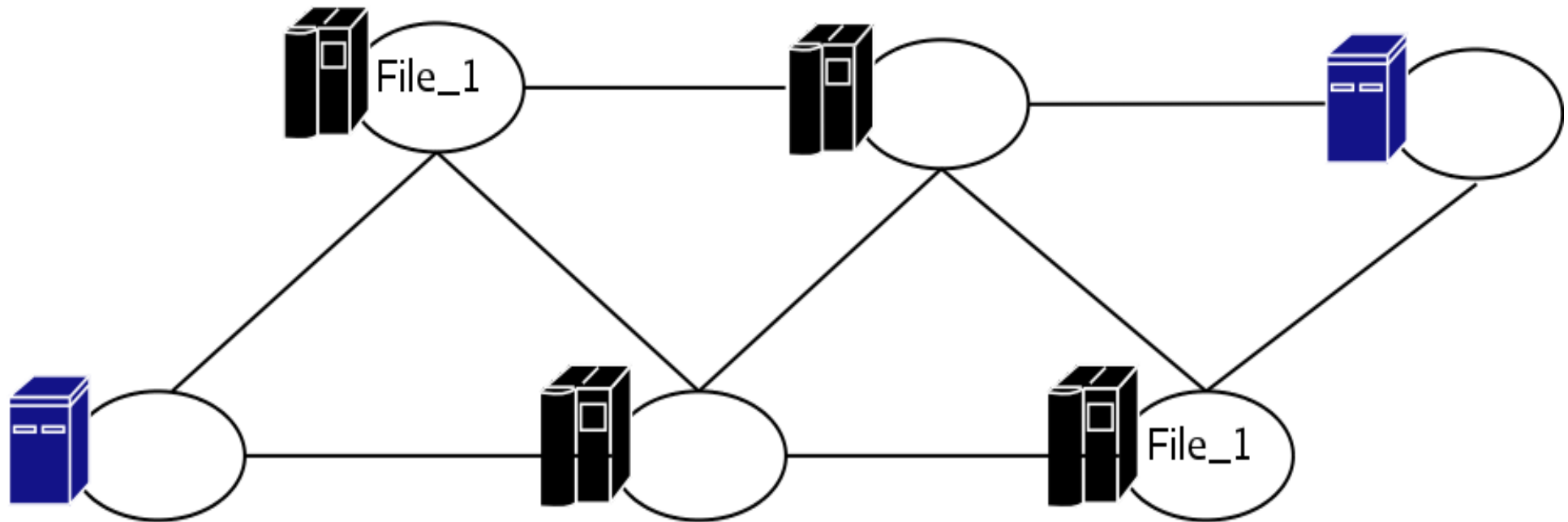
Replication of Data Sets - key concept to increase data availability



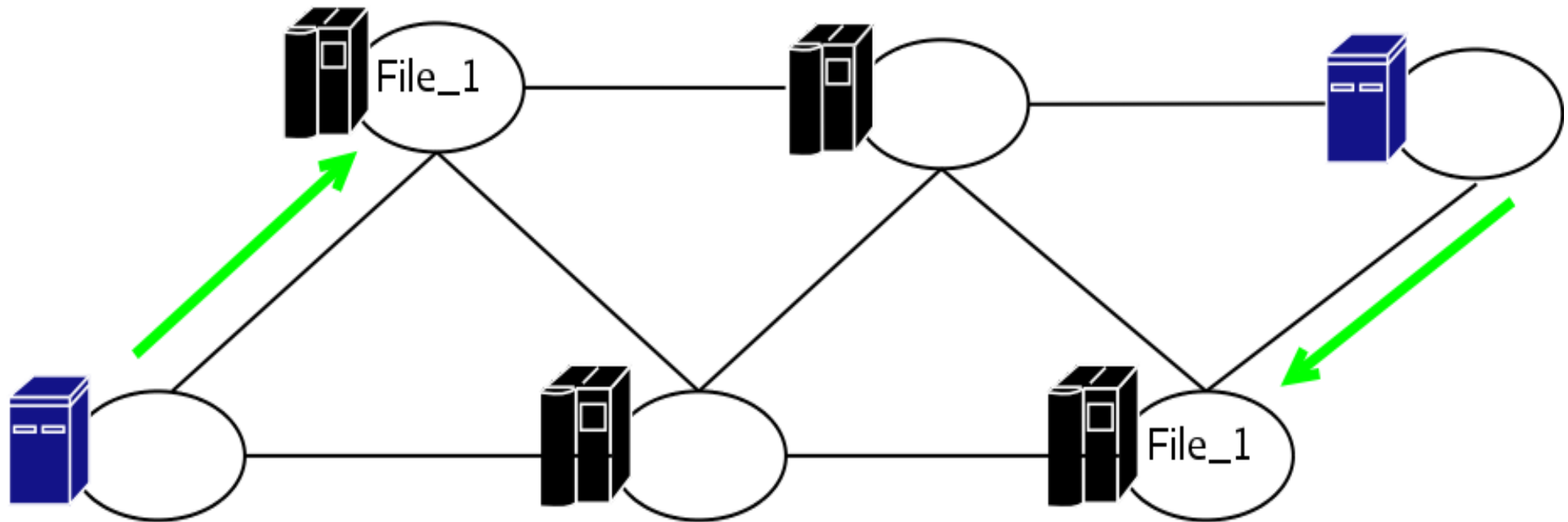
Replication of Data Sets - key concept to increase data availability



Replication of Data Sets - key concept to increase data availability



Replication of Data Sets - key concept to increase data availability



- **Provide information about replicas location**
- **Replicas are identified by LFN and PFN**

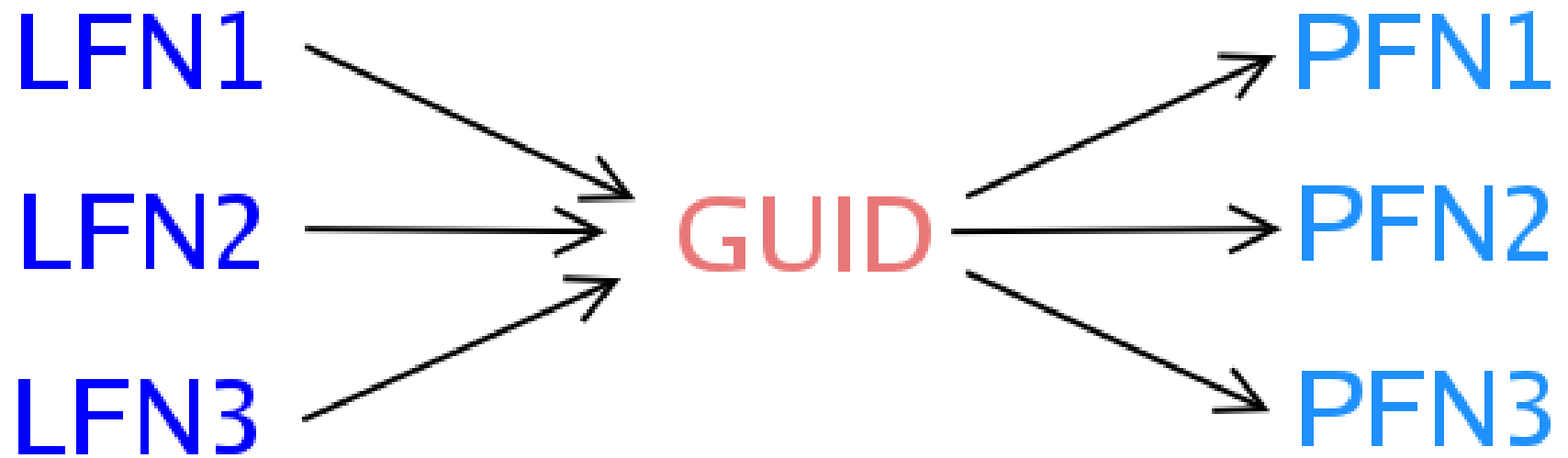
LFN - Logical File Name

- **global unique file identifier**

GUID – Grid Unique ID

PFN – Physical File Name

- **replica name on concrete Storage Element**

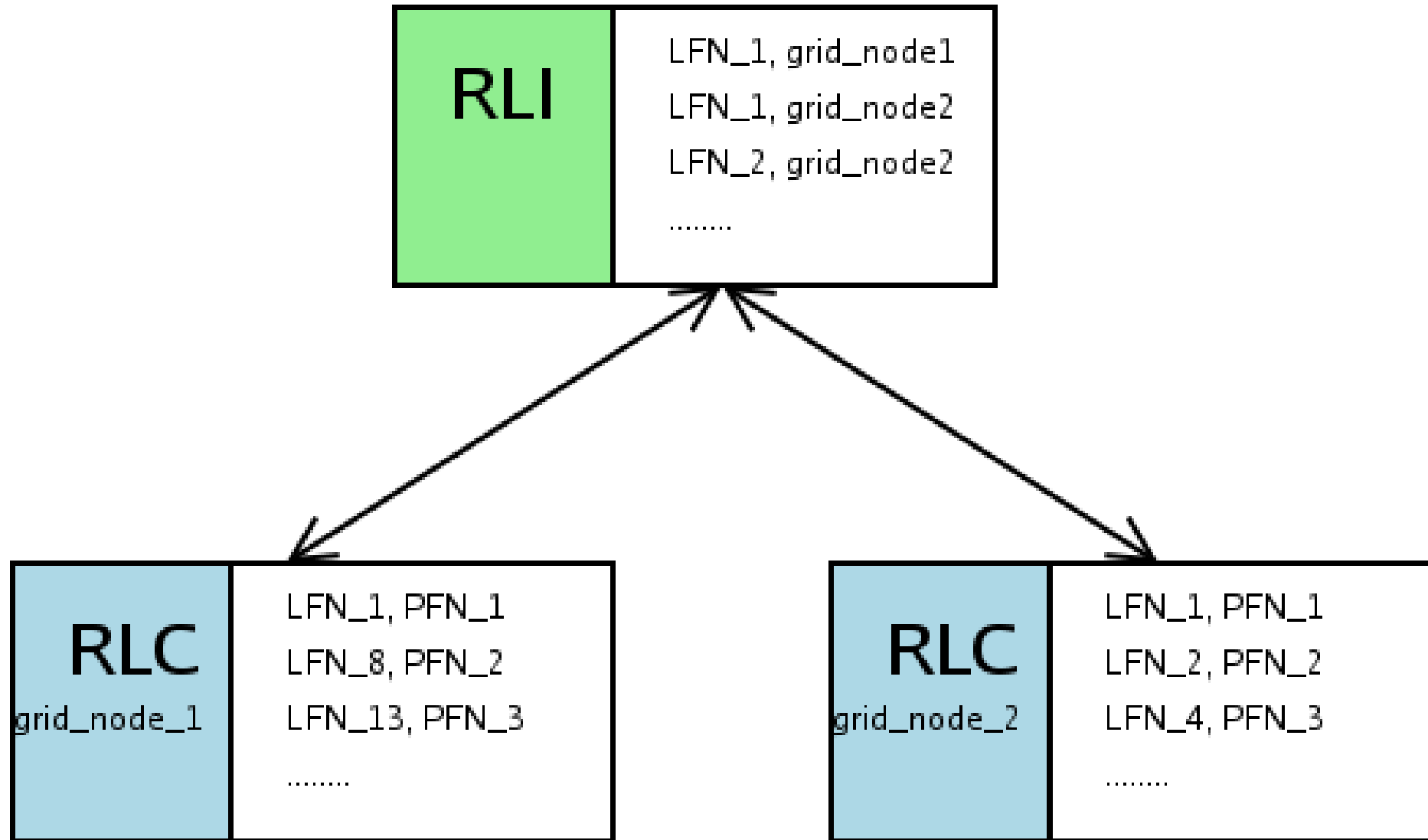


Replica Location Catalog

- **Service run at each grid node**
- **Stores information about files stored at the grid node**
- **Provides mapping between LFN and PFN**
- **Propagate information to Replica Location Index**
- **Soft state update mechanism**

Replica Location Index

- **Central grid service**
- **Gathers information from grid RLCs**
- **Provides mapping between LFN and a set of grid nodes containing given file**



Replica Manger Service

- Integrates GritFTP and catalog services

```
[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \  
> copyAndRegisterFile file:///home/ciglan/file3 -l lfn:testfile4 ; echo;echo  
guid:5c87e983-28db-11d9-8660-d310fe46c7ba  
  
[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \  
> replicateFile lfn:testfile4 -d grid009.ct.infn.it ; echo;echo  
sfn://grid009.ct.infn.it/flatfiles/SE00/gilda/generated/2004-10-28/file881cae84-28db-11d9-afaf-fb89c48a8151  
  
[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \  
> listReplicas lfn:testfile4 ; echo;echo  
sfn://dgt02.ui.savba.sk/storage/gilda/generated/2004-10-28/file5b77b202-28db-11d9-8660-d310fe46c7ba  
sfn://grid009.ct.infn.it/flatfiles/SE00/gilda/generated/2004-10-28/file881cae84-28db-11d9-afaf-fb89c48a8151
```

Short-term optimization

- select best file replica for given grid node and LFN
- selection based on current network status & replica location

Replica manager

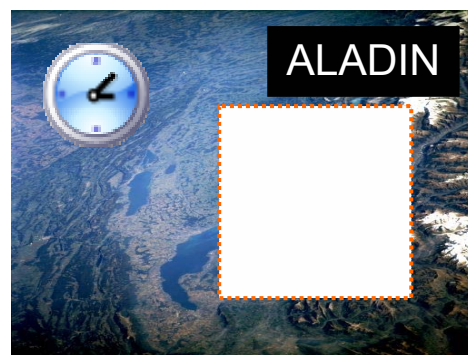
- Integrates catalog services & optimization services

```
[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \
> listReplicas lfn:testfile4 ;                                echo;echo
sfn://grid009.ct.infn.it/flatfiles/SE00/gilda/generated/2004-10-28/file881cae84-28db-11d9-afaf-fb89c48a8151
sfn://setilab.tilab.com/flatfiles/gilda/generated/2004-10-29/file37582d23-299d-11d9-88d8-adb3f4a8eae

[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \
> listBestFile lfn:testfile4 ;                                echo;echo
sfn://grid009.ct.infn.it/flatfiles/SE00/gilda/generated/2004-10-28/file881cae84-28db-11d9-afaf-fb89c48a8151

[ciglan@dgt03 ciglan]$ edg-rm --vo gilda \
> getBestFile lfn:testfile4 ;                                echo;echo
sfn://dgt02.ui.savba.sk/storage/gilda/generated/2004-10-29/fileae67c2d6-299d-11d9-8f09-b967102c5b4b
```

- **descriptive information about data**
- **lot of metadata types**
- **general metadata (file type, file size, creation date)**
- **security metadata**
- **application specific metadata**
- **metadata description by ontologies**



```
AF01-BE1227-3318A1  
CC58-48FA03-0054A0  
0322-BB1829-101504
```



www.globus.org

www.eu-datagrid.org



Enabling Grids for E-scienceE

Thank you for your attention.

www.eu-egee.org



INFSO-RI-508833