



The EU DataGrid Data Management (EDG release 1.4.x)

**The European
DataGrid Project Team**

<http://www.eu-datagrid.org>

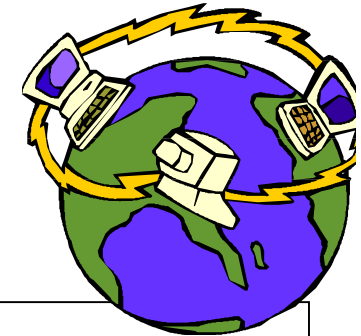


EDG Tutorial Overview

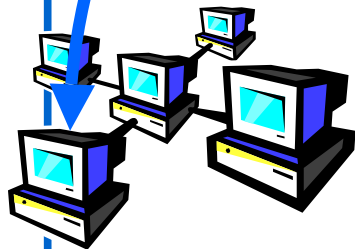


Workload Management Services

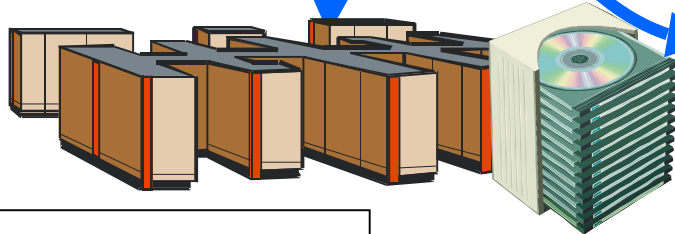
Data Management Services



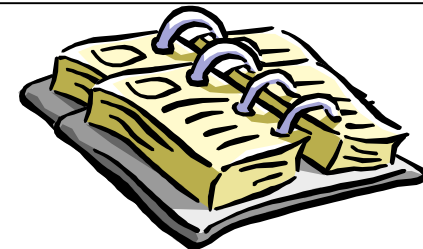
Networking



Fabric Management



Information Service



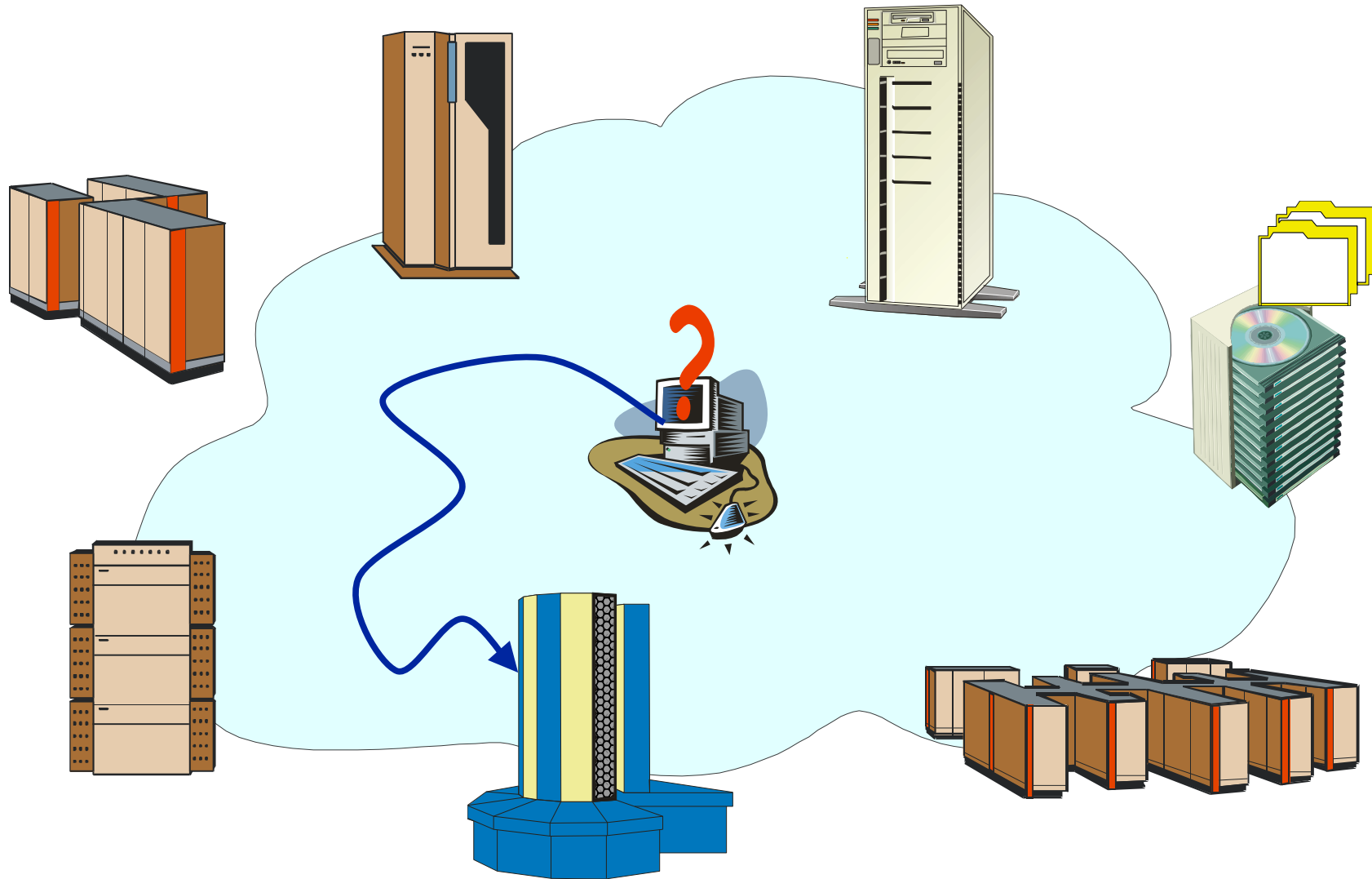


Overview

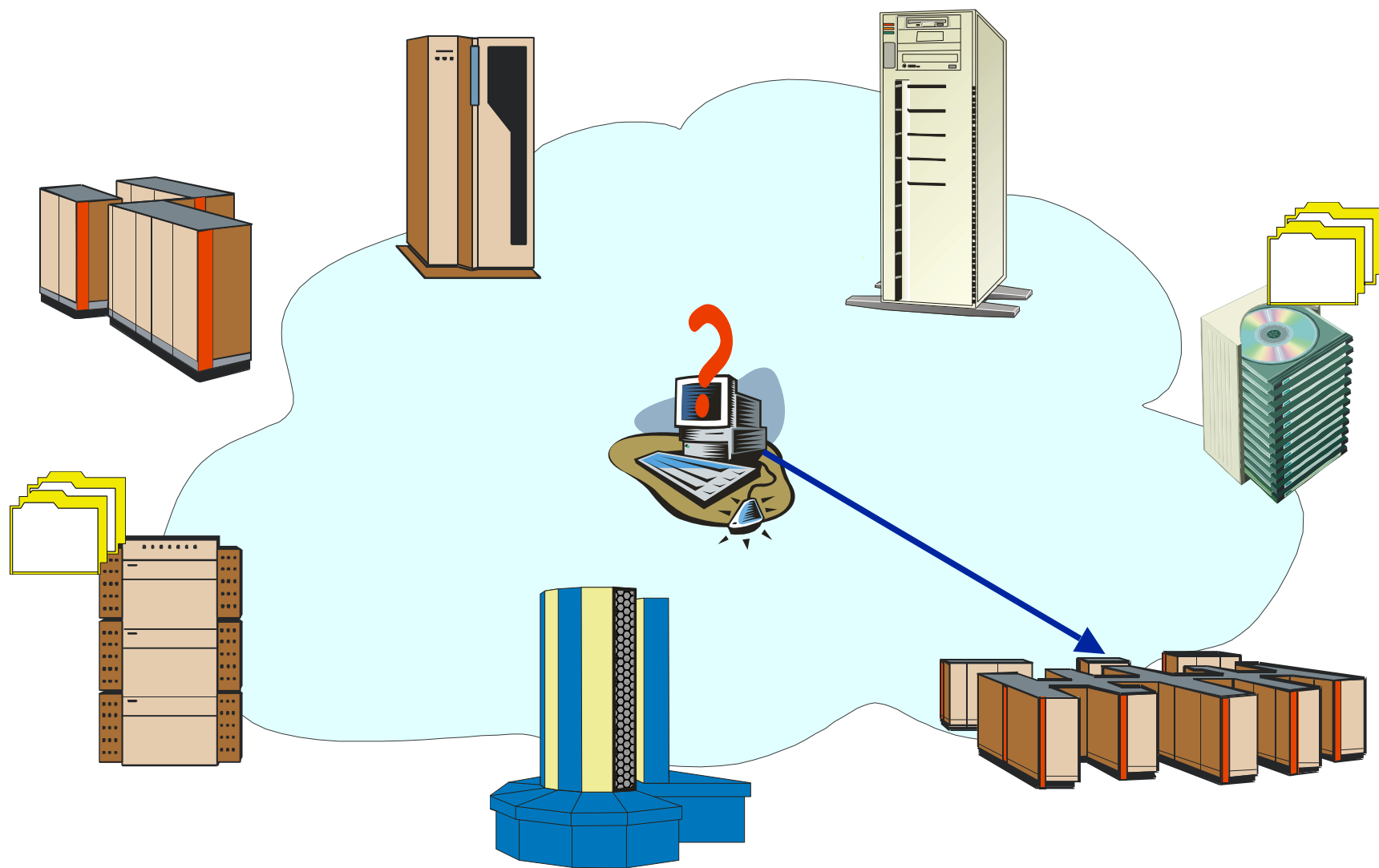
- ◆ Data Management Issues

- ◆ Main Components
 - EDG Replica Catalog
 - EDG Replica Manager

Data Management Issues



Data Management Issues





Data Management Tools

- ◆ Tools for
 - Locating data
 - Copying data
 - Managing and replicating data
 - Meta Data management

- ◆ On EDG Testbed you have
 - EDG Replica catalog
 - EDG Replica Manager
 - Spitfire



EDG Replica Catalog

- ◆ Based upon the Globus LDAP Replica Catalog
- ◆ Stores LFN/PFN mappings and additional information (e.g. filesize):
 - Physical File Name (PFN): host + full path & and file name
 - Logical File Name (LFN): logical name that may be resolved to PFNs
 - LFN : PFN = 1 : n
- ◆ Only files on storage elements may be registered
- ◆ Each VO has a specific *storage dir* on an SE
- ◆ Example PFN:
`lxshare0222.cern.ch/flatfiles/SE1/iteam/file1.dat`

└──────────────────────────┘	└──────────────────────────┘
host	storage dir
- ◆ LFN must be full path of file starting from storage dir
LFN of above PFN: `file1.dat`



The EDG Replica Manager

- ◆ Extends the Globus replica manager
- ◆ Only client side tool
- ◆ Allows replication (copy) and registering of files in RC
 - works with LDAP based RC and RLS (see release 2.0 next day)
- ◆ Keeps RC consistent with stored data.



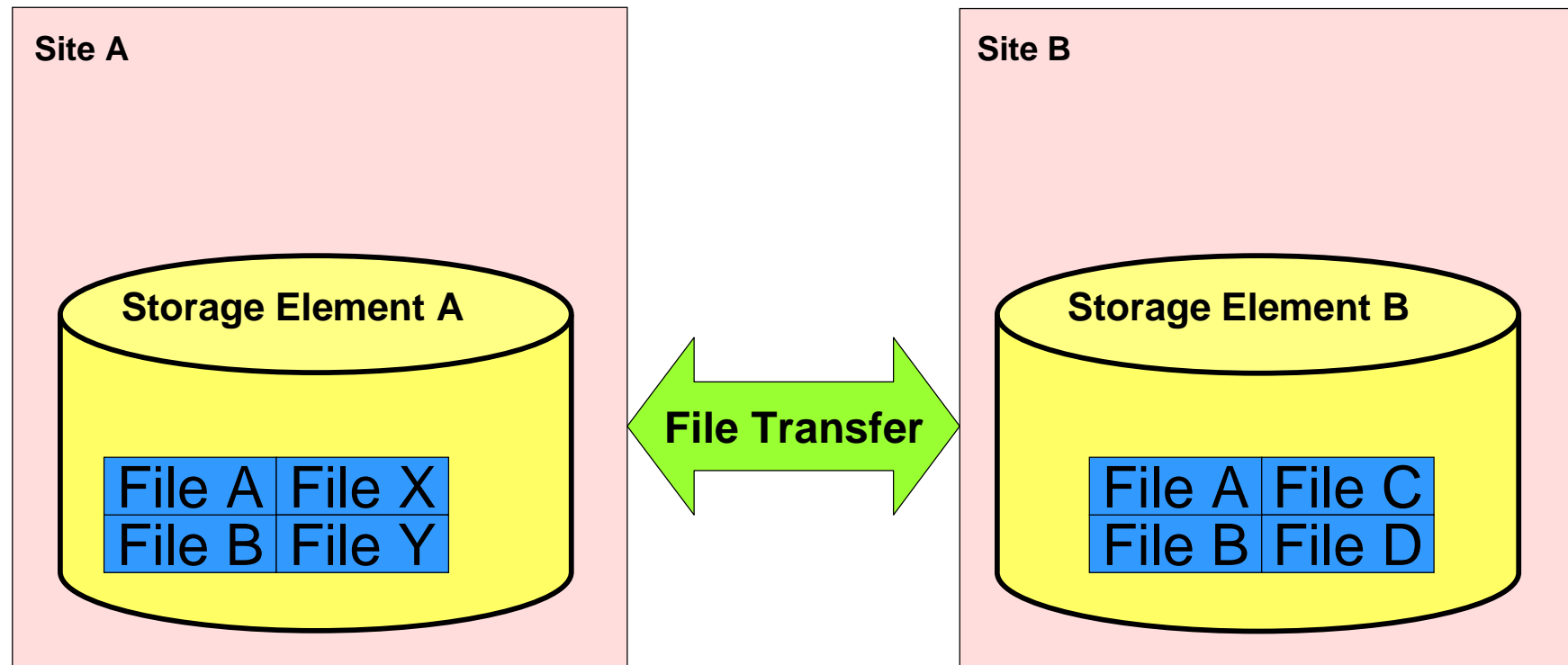
The Replica Manager APIs

- ◆ **(un)registerEntry**(LogicalFileName *lfn*,
FileName *source*)
 - Replica Catalogue operations only - no file transfer
- ◆ **copyFile**(FileName *source*,
FileName *destination*,
String *protocol*)
 - allows for third-party transfer
 - transfer between:
 - two StorageElements or
 - ComputingElement and Storage Element
 - Space management policies under development
 - all tools support **parallel streams** for file transfers

The Replica Manager APIs

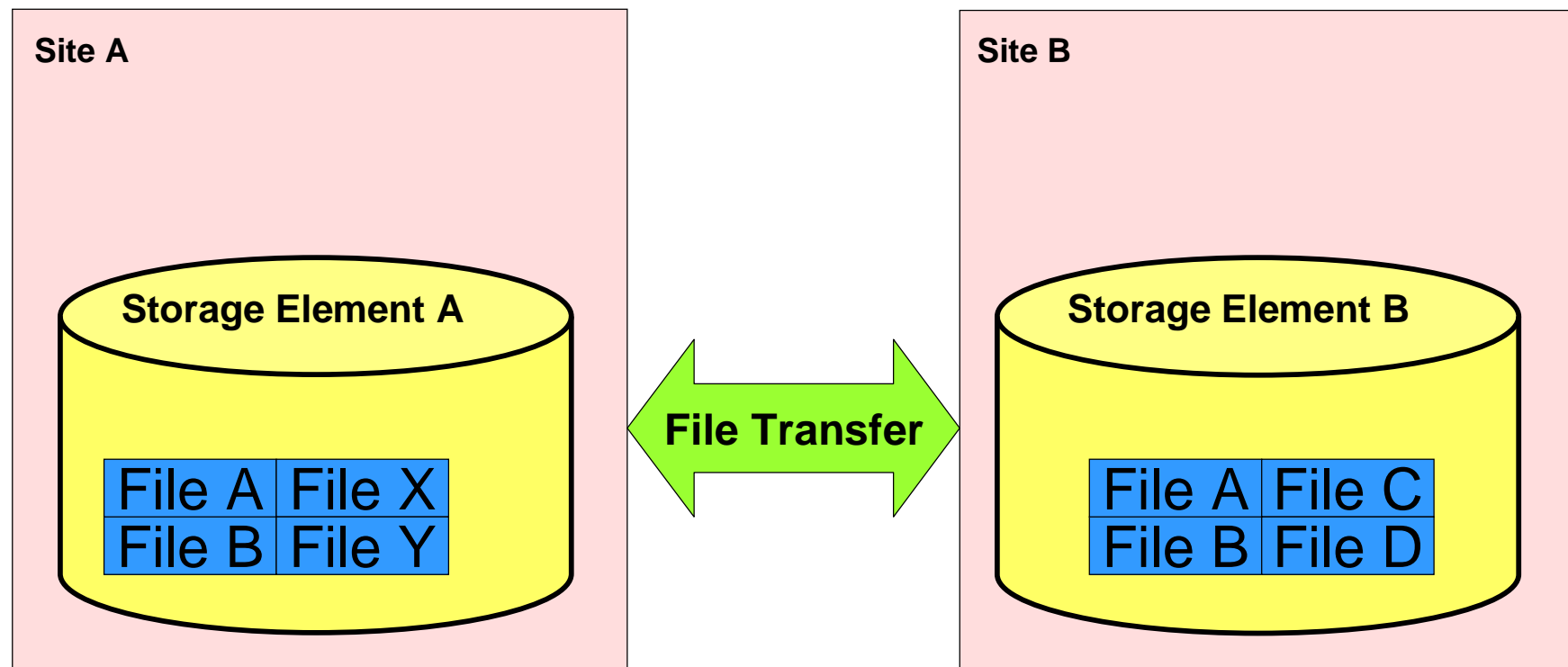
- ◆ ***copyAndRegisterFile***(LogicalFileName *lfn*,
 FileName *source*,
 FileName *destination*,
 String *protocol*)
 - third-party transfer but :
 - files can only be registered in Replica Catalogue if **destination** PFN contains a valid **SE** (i.e. needs to be registered in the RC)!
- ◆ ***replicateFile***(LogicalFileName *lfn*,
 FileName *source*,
 FileName *destination*,
 String *protocol*)
- ◆ ***deleteFile***(LogicalFileName *lfn*,
 FileName *source*)

File Management Summary



File Management Summary

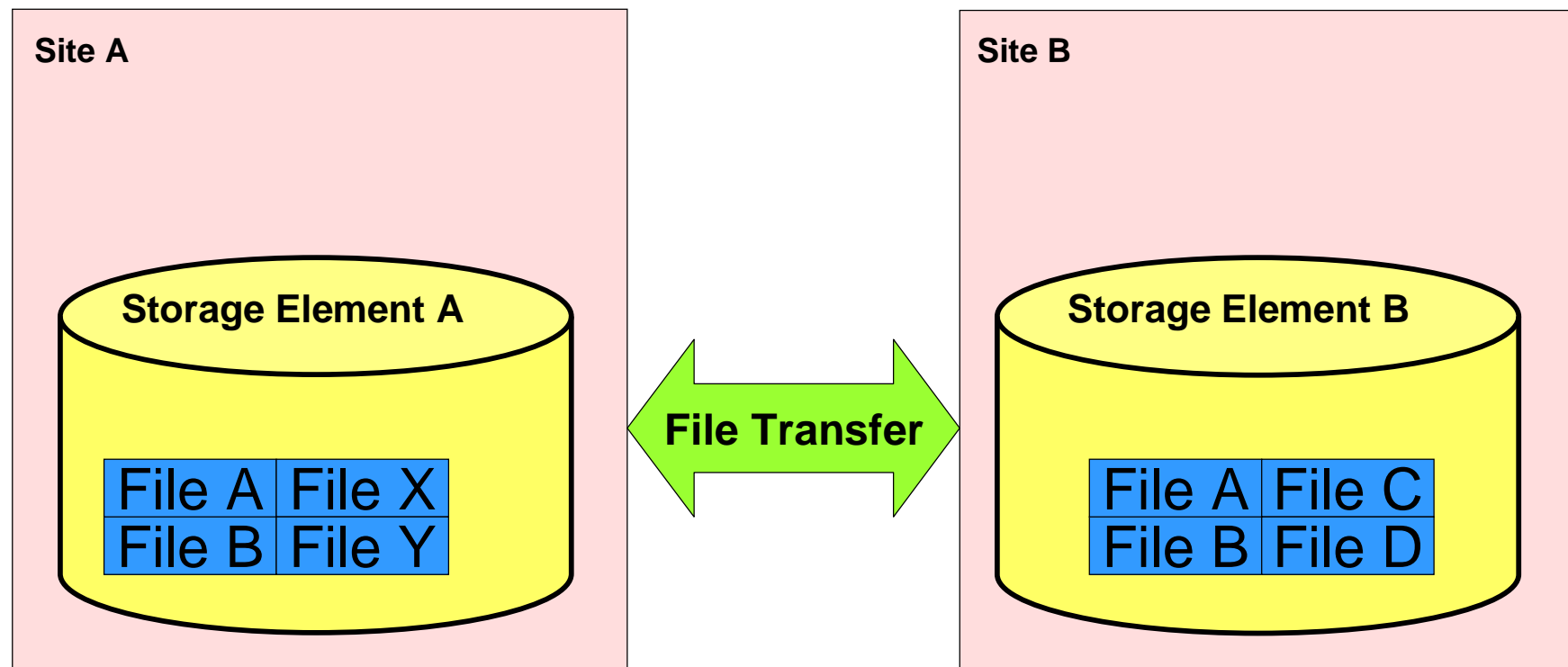
Replica Catalog:
Map Logical to Site files



File Management Summary

Replica Catalog:
Map Logical to Site files

Replica Selection:
Get 'best' file

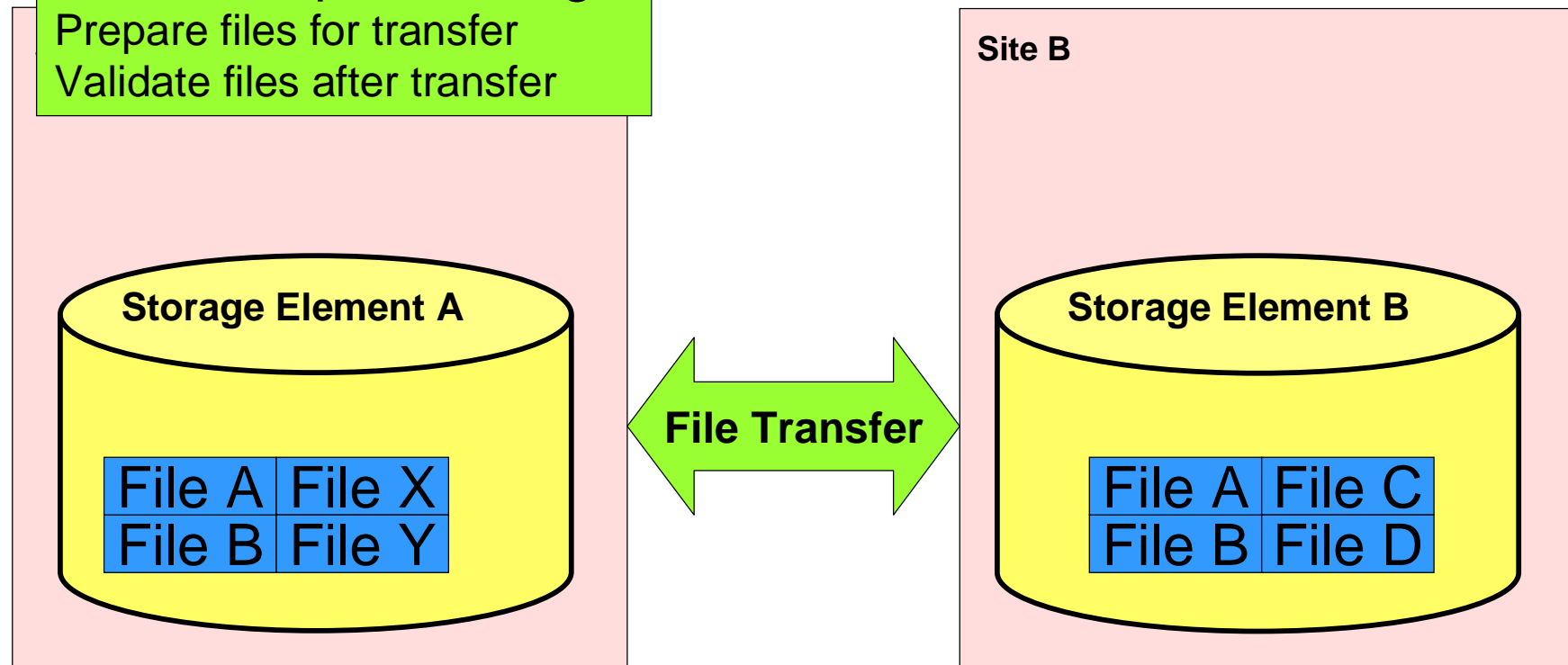


File Management Summary

Replica Catalog:
Map Logical to Site files

Replica Selection:
Get 'best' file

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer



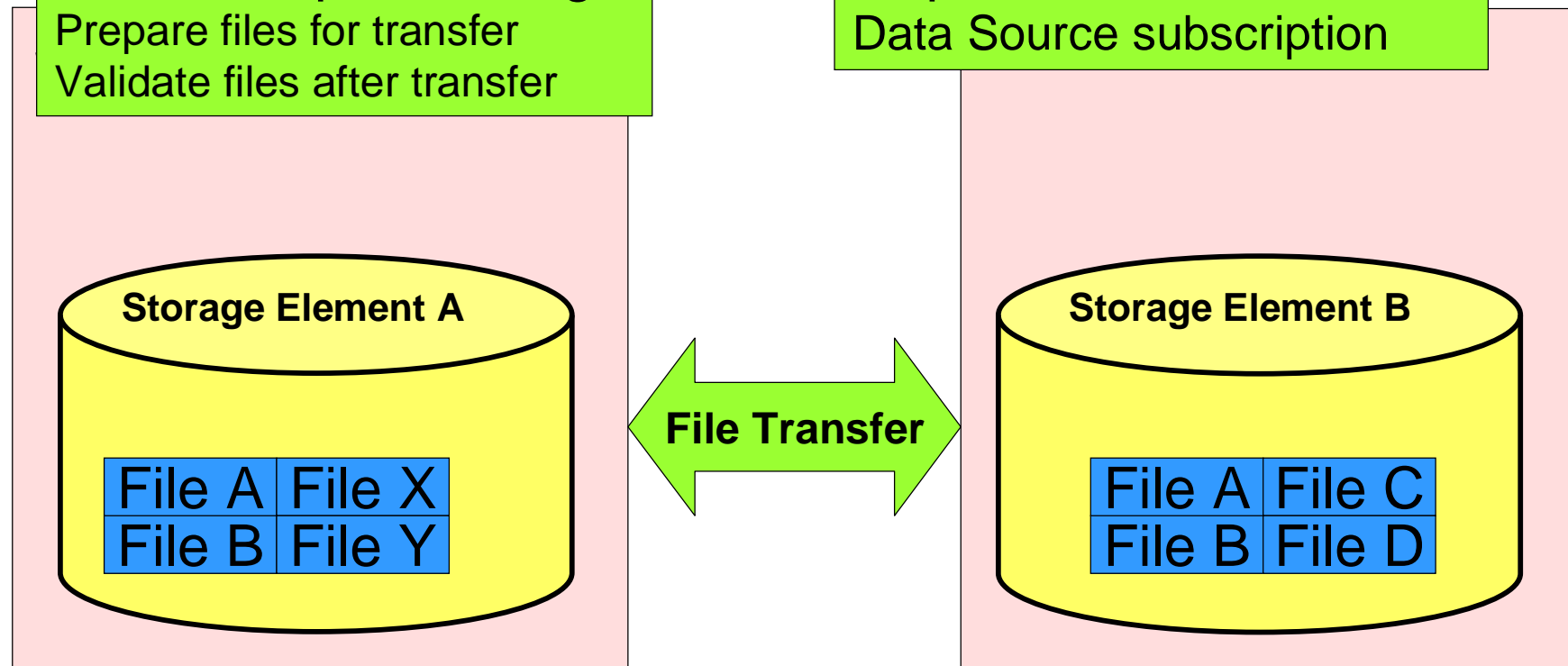
File Management Summary

Replica Catalog:
Map Logical to Site files

Replica Selection:
Get 'best' file

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer

Replication Automation:
Data Source subscription



File Management Summary

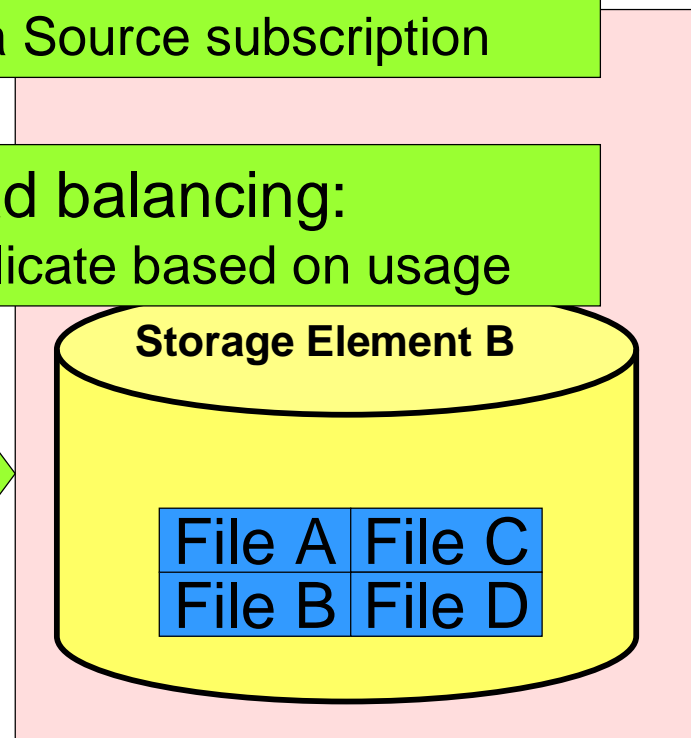
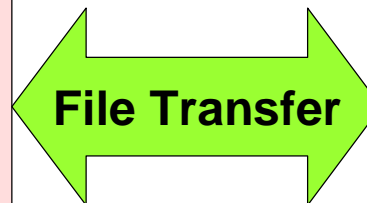
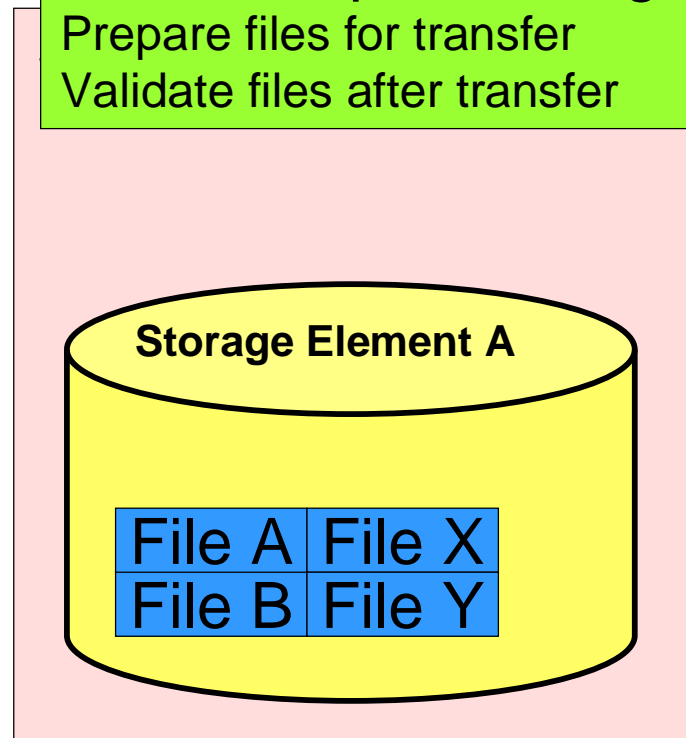
Replica Catalog:
Map Logical to Site files

Replica Selection:
Get 'best' file

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer

Replication Automation:
Data Source subscription

Load balancing:
Replicate based on usage



File Management



Replica Manager:
'atomic' replication operation
single client interface
orchestrator

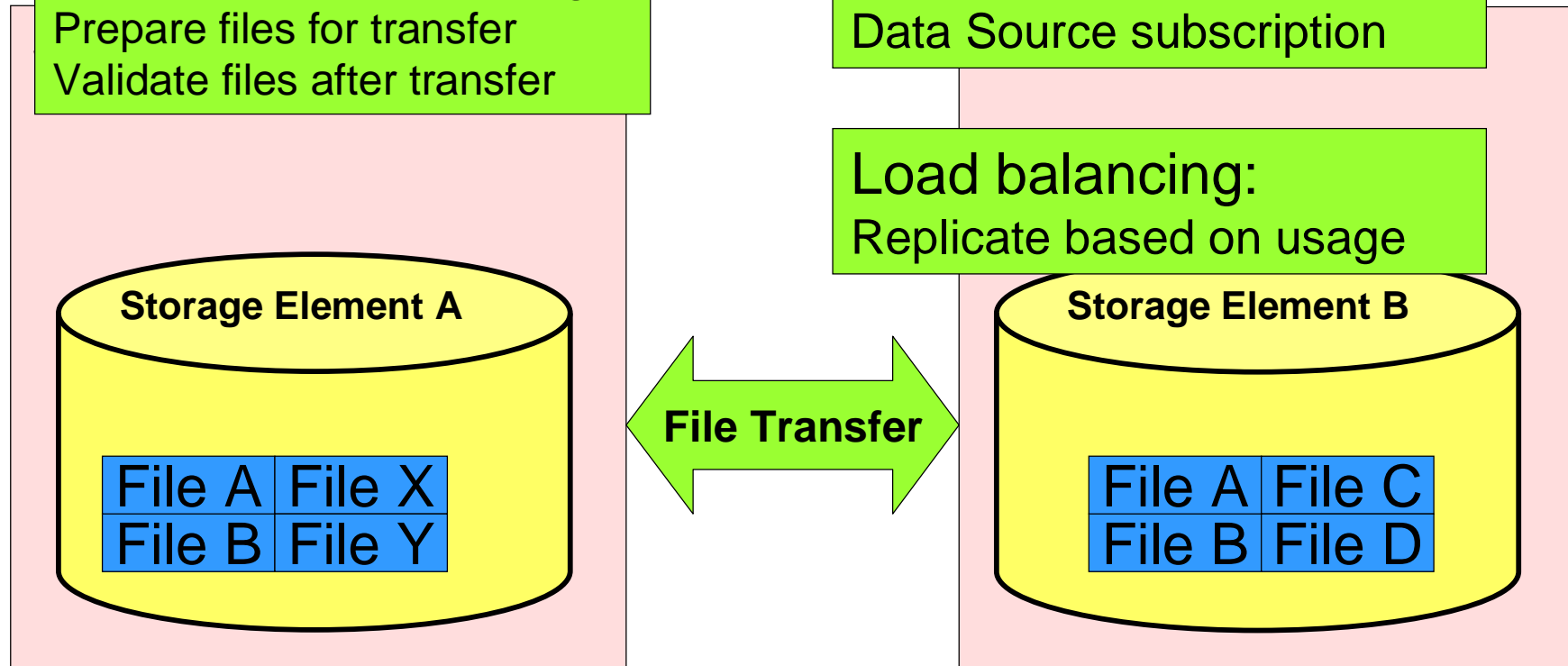
Replica Catalog:
Map Logical to Site files

Replica Selection:
Get 'best' file

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer

Replication Automation:
Data Source subscription

Load balancing:
Replicate based on usage



File Management



Replica Manager:
'atomic' replication operation
single client interface
orchestrator

Replica Catalog:
Map Logical to Site files

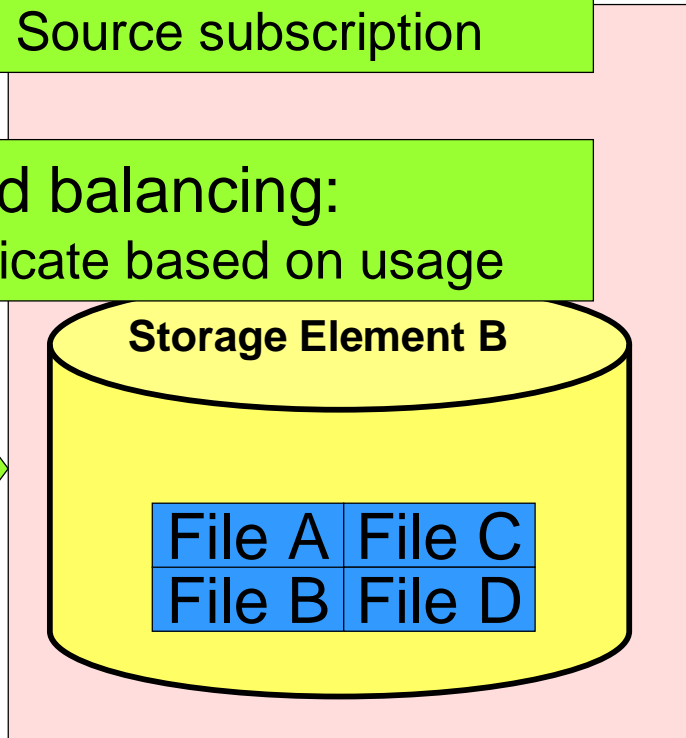
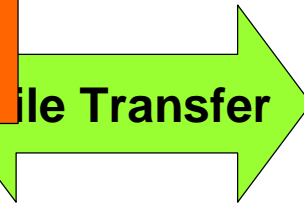
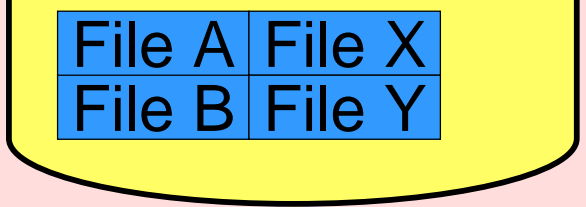
Replica Selection:
Get 'best' file

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer

Replication Automation:
Data Source subscription

Load balancing:
Replicate based on usage

Metadata:
LFN metadata
Transaction information
Access patterns



File Management



Replica Manager:
'atomic' replication operation
single client interface
orchestrator

Replica Catalog:
Map Logical to Site files

Replica Selection:
Get best replica

Pre- Post-processing:
Prepare files for transfer
Validate files after transfer

Replication Automation:
Data Source subscription

Metadata:
LFN metadata
Transaction information
Access patterns

Load balancing:
Replicate based on usage

File A	File X
File B	File Y

File Transfer

Storage Element B

File A	File C
File B	File D

Security