EMI Data, the unified European Data Management Middleware

Patrick Fuhrmann (DESY)

EMI Data Area lead

(on behalf of many people and slides stolen from all over the place)

Credits

- Alejandro Alvarez
- Alex Sim
- Claudio Cacciari
- Christian Bernardt
- Christian Loeschen
- Elisabetta Ronchieri
- Fabrizio Furano
- Giuseppe Fiameni
- Giacinto Donvito
- Giuseppe Lo Presti
- Jon Kerr Nilsen
- Jan Schaefer
- Jean-Philippe Baud
- Dmitry Ozerov
- Yves Kemp
- Karsten Schwank
- Michele Carpene
- Michele Dibenedetto
- Michail Salichos
- Mischa Salle
- Oscar Koeroo
- Oliver Keeble
- Paul Millar
- Ralph Mueller-Pfefferkorn
- Ricardo Rocha
- Riccardo Zappi
- Tigran Mkrtchyan
- Zsolt Molnar
- Zsombor Nagy

Our wiki : https://twiki.cern.ch/twiki/bin/view/EMI/EmiJra1T3Data

European Middleware Initiative
EMI INFSO-RI-261611

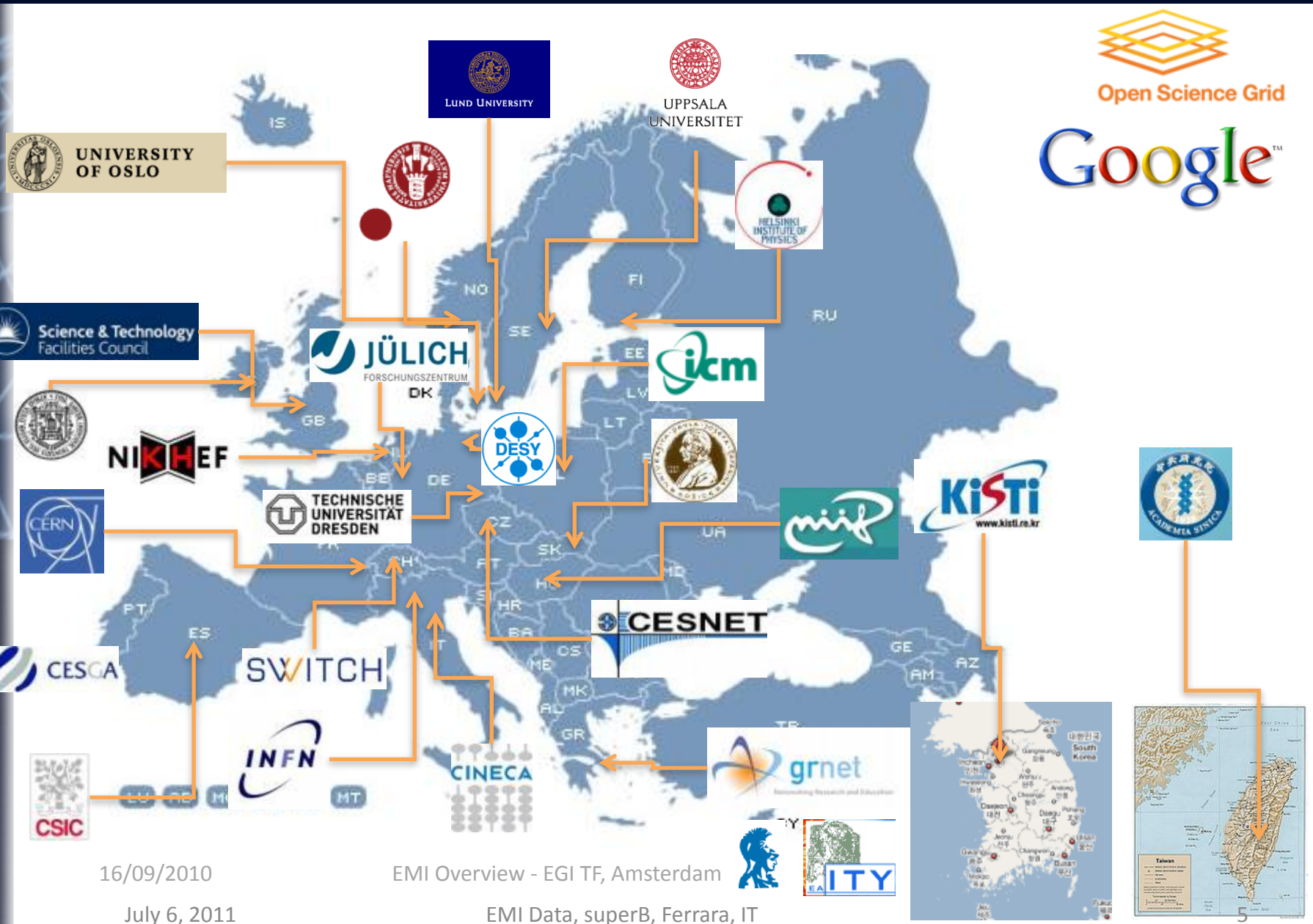
Outline

- EMI, the facts
- EMI-Data, the components
- EMI-Data, the mission
- EMI-Data, selected Topics
  - Topic 1 : Interoperability and reduction of components
    - Client library consolidation
  - Topic 2 : Fixing flaws and evolution
    - Catalogue and SE synchronization
  - Topic 3 : Standardization
    - Example 1 : WebDAV
    - Example 2 : NFSv4.1/pNFS

EMI, the project

EMI, the facts

EMI Factsheet



EMI INFSO-RI-261611

EMI Factsheet

EMI Factsheet

Budget : about 24 Million Euros

Funding : about 50% by EU-FP7, rest by partners

Covers : JRA, SA and NA

Partners : 22

Middlewares: Arc, gLite, UNICORE and dCache

For more facts

For more facts on EMI in general, see Francesco's presentation from this morning on

EMI/EGI Issues

Director's Mantra

According to our Project Director, Alberto Di Meglio :

The European Middleware Initiative (EMI) project represents a close collaboration of the major European middleware providers - ARC, gLite, UNICORE and dCache - to establish a sustainable model to support, harmonise and evolve distributed computing middleware for deployment in EGI, PRACE and other distributed e-Infrastructures

Translates to EMI data as

Maintain and improve the currently existing production WLCG software stack !

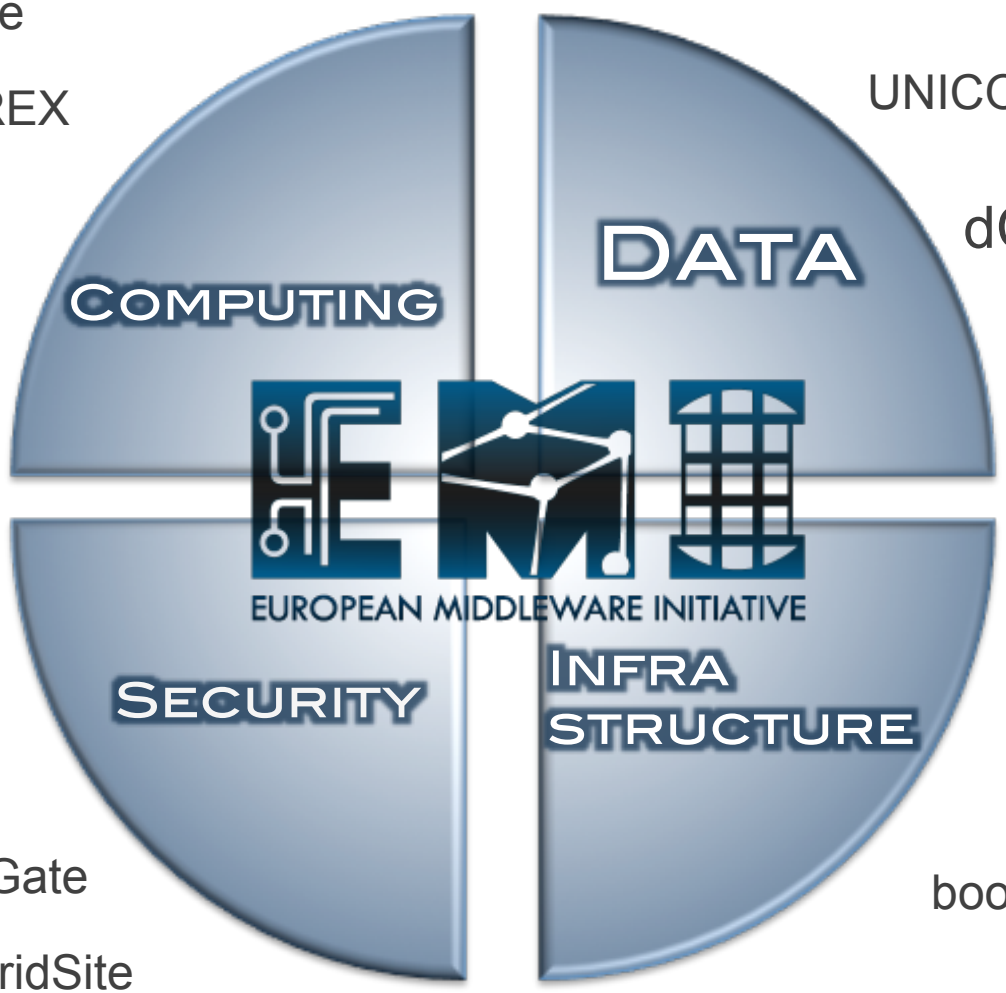
Allow new communities to utilize the EMI middle-wares by applying standards

Components

EMI, the components

The EMI Pie

63 components and about 350 packages



UAS-Compute

A-REX

MPI

WMS

UNICORE-SMS

StoRM

dCache

Arc-libs

FTS, LFC,
DPM, GFAL

Information system

accounting

bookkeeping

ARGUS

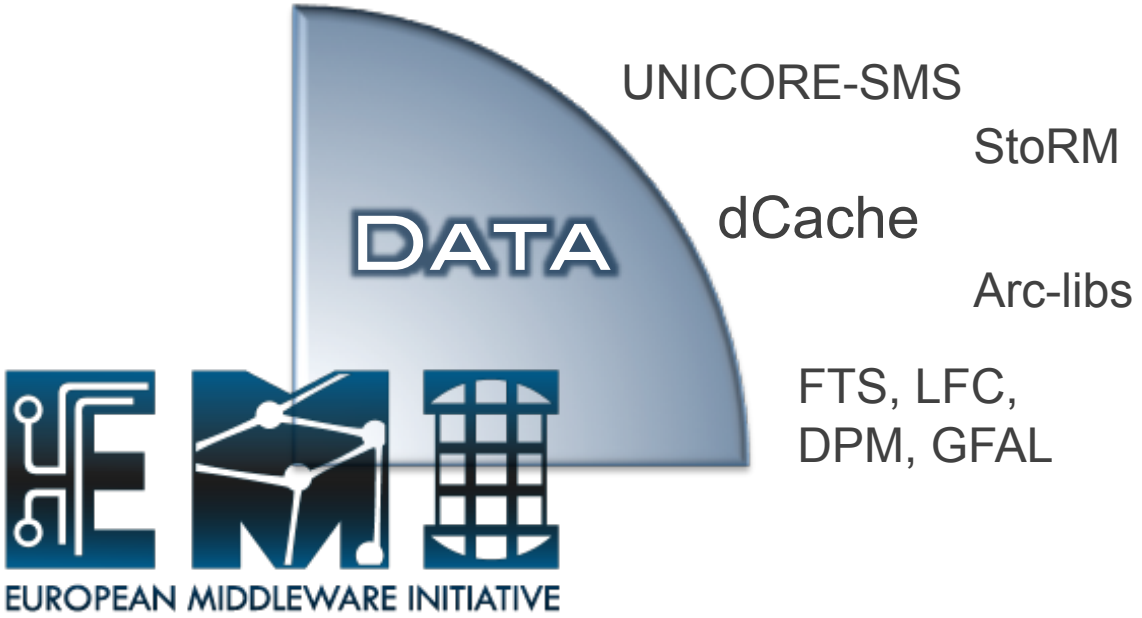
VOMS

UNICORE-Gate

gridSite

The EMI Pie

63 components and about 350 packages



What does EMI-Data provide ?

The EMI-Data shopping ~~basket~~ CART

The EMI shopping cart

Reliable **F**ile
Transport **S**ervice

Client Libraries

File Location and meta data Service (LFC)

Professional Storage Solutions

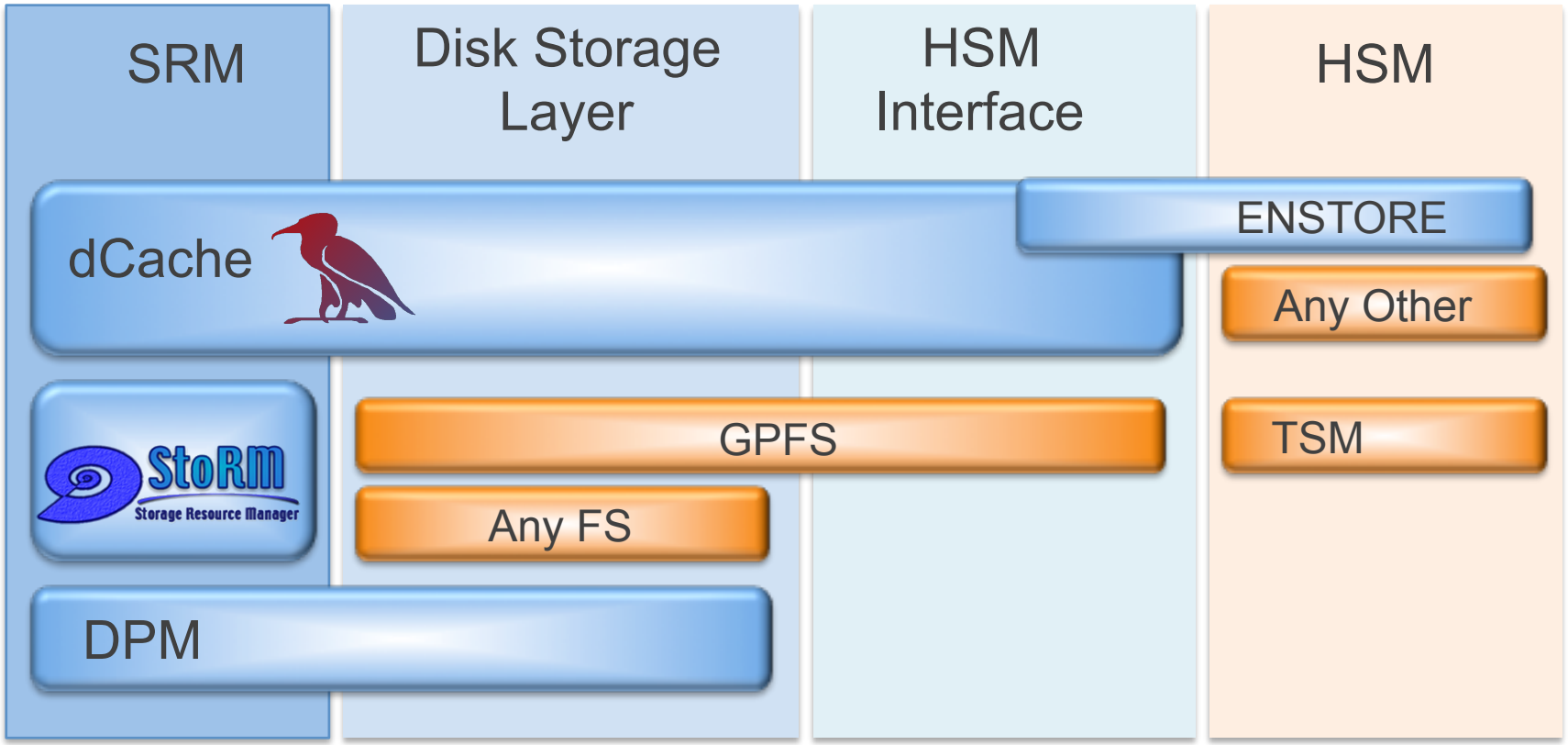
Fits all size

dCache

DPM

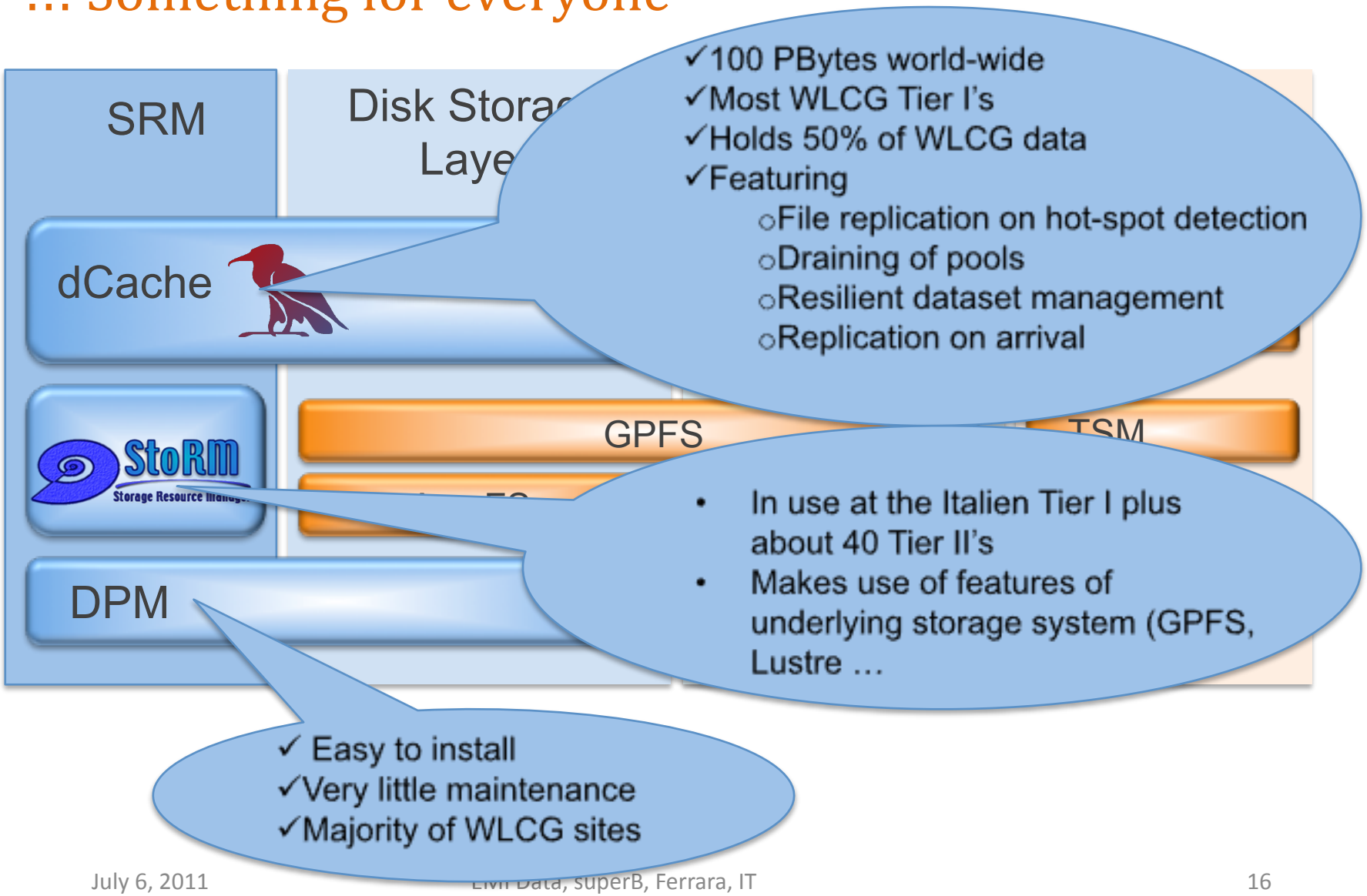
StoRM
Storage Resource Manager

EMI, the storage elements



EMI, the storage elements

!!! Something for everyone



The Mission

EMI Data mission

The Mission

- Fixing of issues based on the experience of operating the infrastructures for some years.
- Improving or creating interoperability between components and middle-wares.
- Reducing components by merging functionality or removing duplication.
- Applying standards where available
- Standardizing EMI-Data mechanisms with “standardization bodies” e.g. OGF
- EGI : Attracting resp. enabling new communities.
- Becoming competitive and attractive by :
  - Standards
  - Professional Support
  - Strict quality monitoring

Some more examples

- Defining (with OGF) and implementing an Storage Accounting Record
- Migrating the security of the Storage Resource Manager protocol from GSI (httpg) to standard SSL/X509.
- Migrating to next version of the information provider schema GLUE2.0
- Improving the File Transfer Service by integrating the load of the network and the storage element backend.
- For the entire list, have a look at :
  - https://twiki.cern.ch/twiki/bin/view/EMI/EmiJra1T3Data

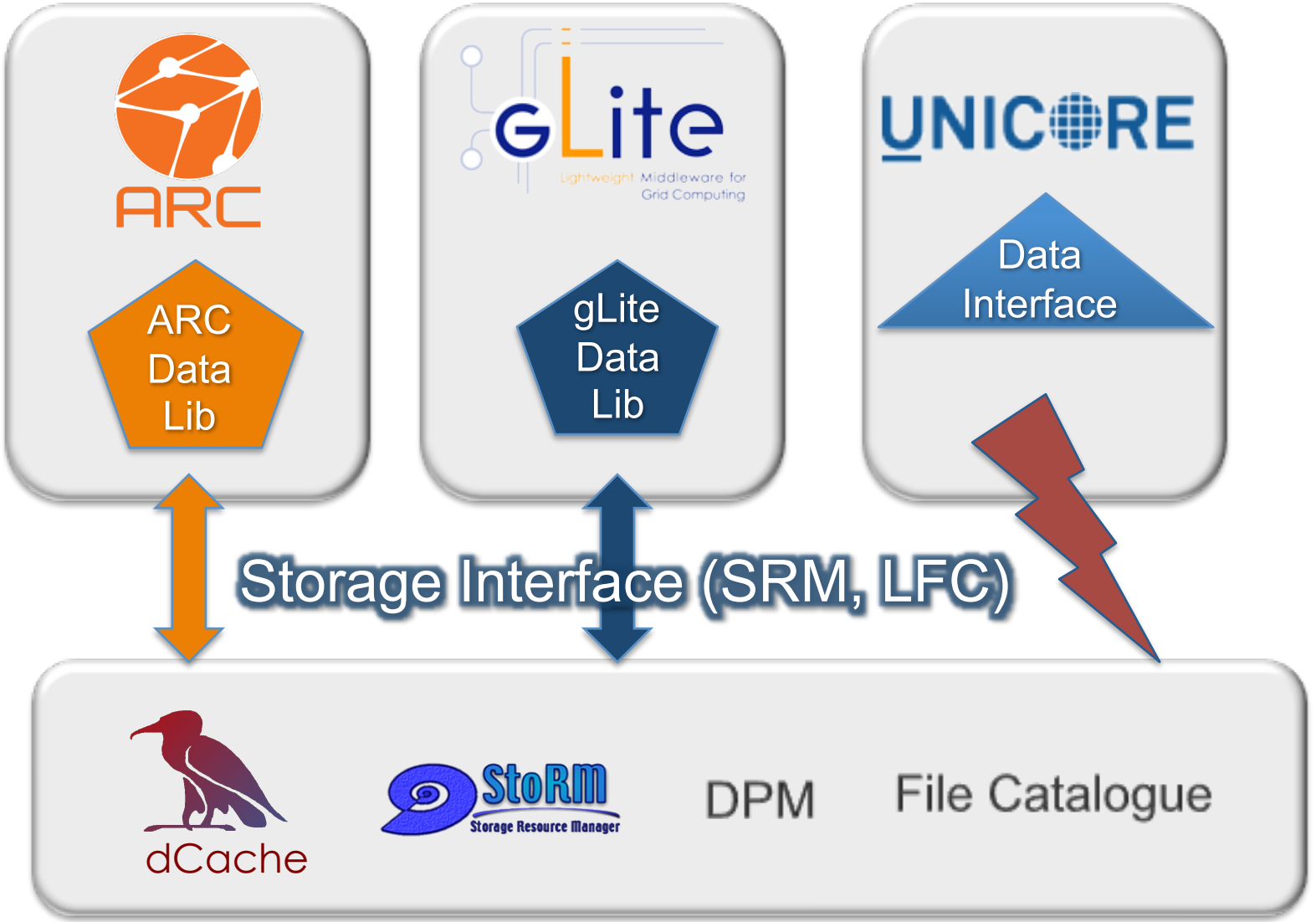
Selected Topics

EMI, some selected topics

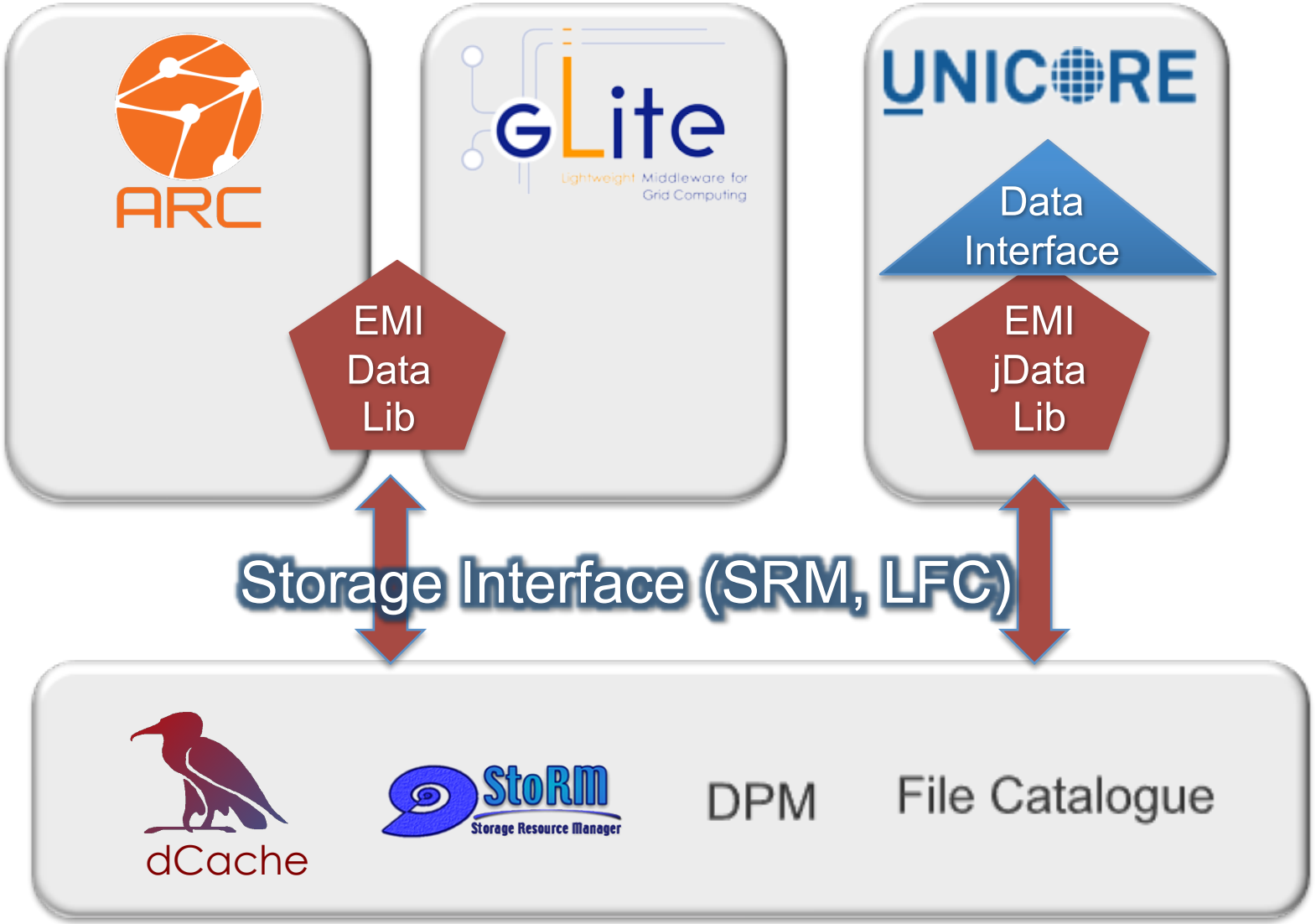
Topic 1

Client library consolidation

Component consolidation



Component consolidation



Topic 2

Catalogue / Storage element synchronization

Catalogue and Storage Elements

Problem

- Storage Elements are storing the actual data locally.
- Catalogues map logical filenames to actually locations URLs.
- Typical store operation :
  - Store file in SE
  - Add entry (URL) to the catalogue under the logical file name
- May possibilities for SE's and catalogues to get out of sync as the store and 'add entry' operations are not transactional in WLCG.
- This results in
  - Dark data (files in SE but not in catalogue)
  - Dangling references (Pointer in catalogue, pointing into the void)
- Over time, catalogues and SE's diverge more and more.

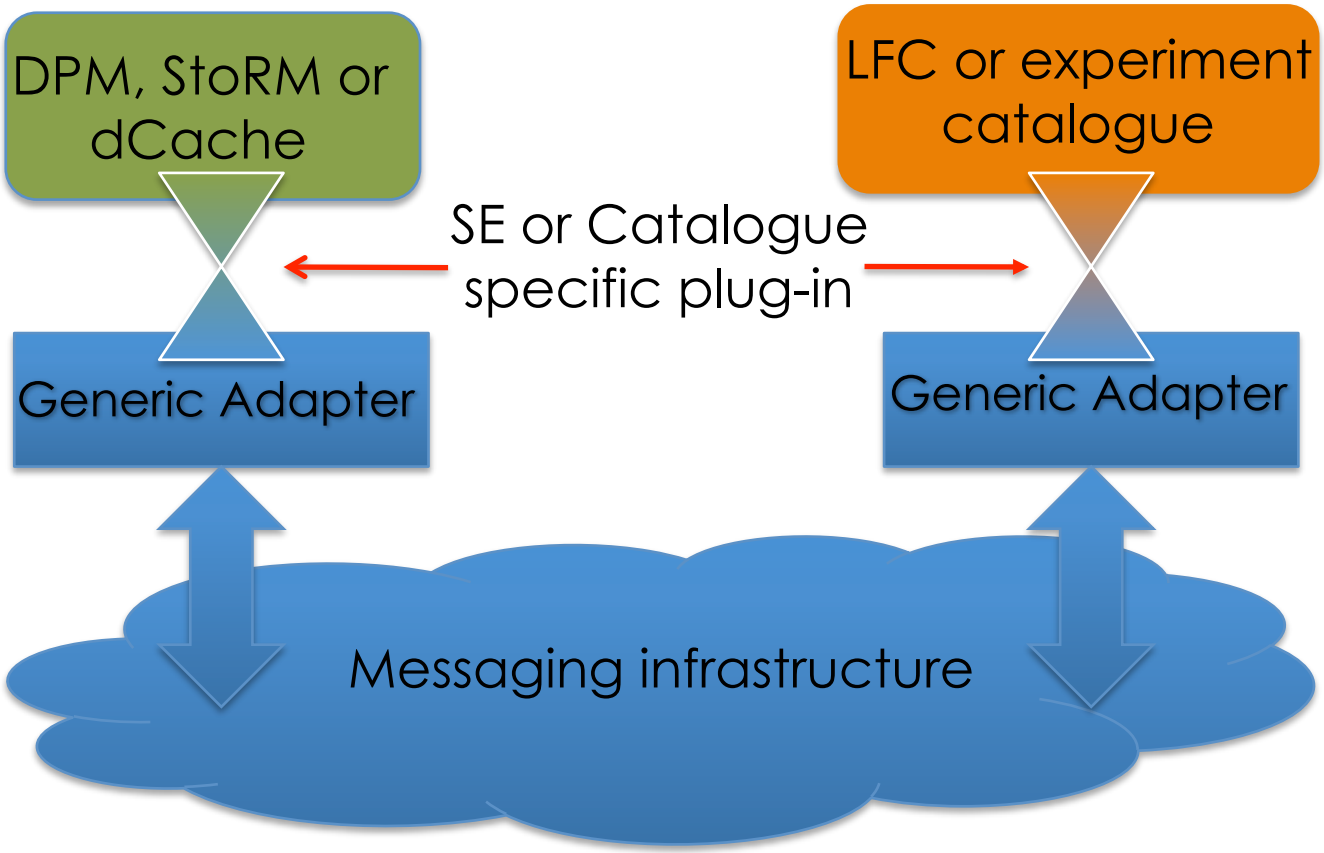
How to solve this issue

Two different approaches

Two possible solutions

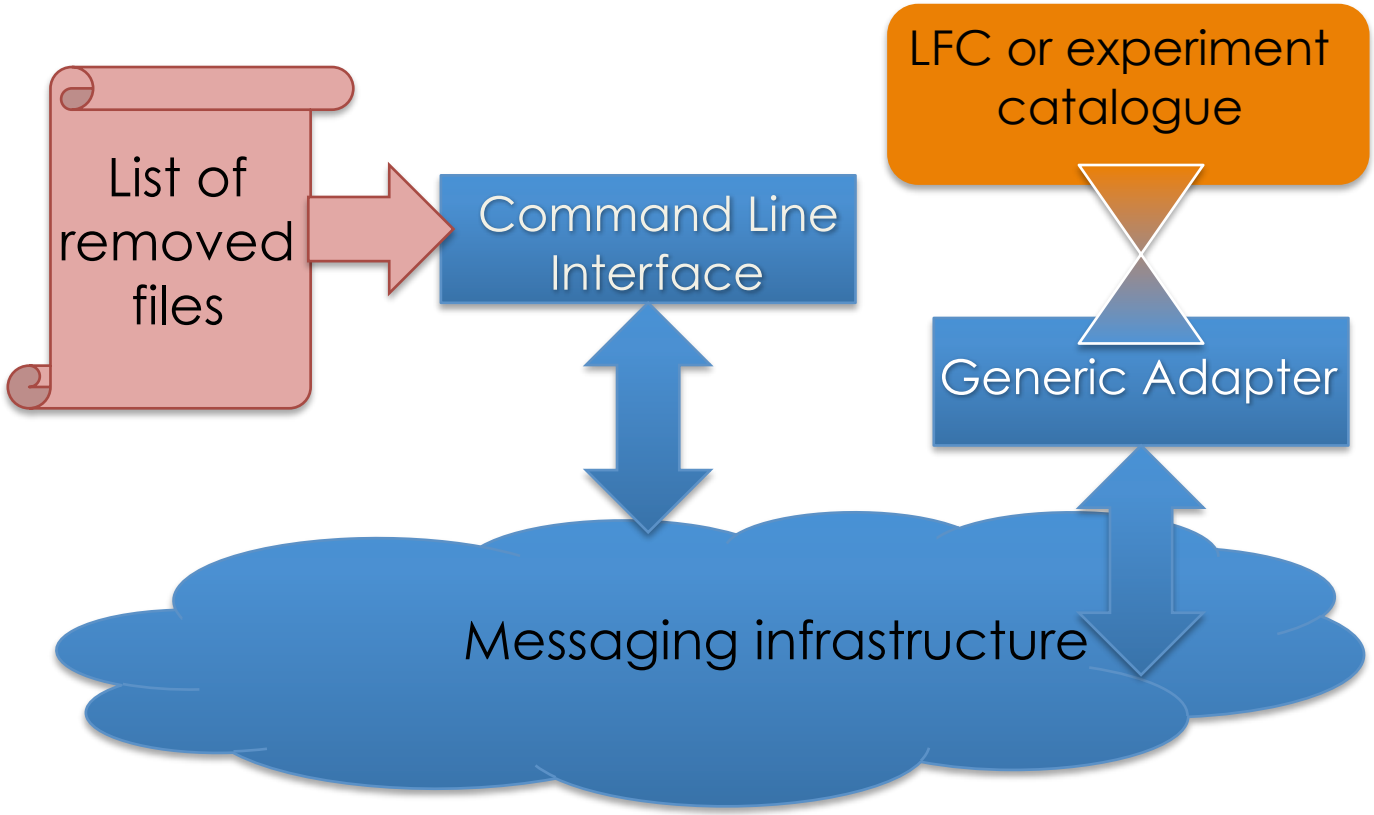
Synchronizing catalogues and SE's using message passing

Talk to Fabrizio on details.



Already in EMI-1

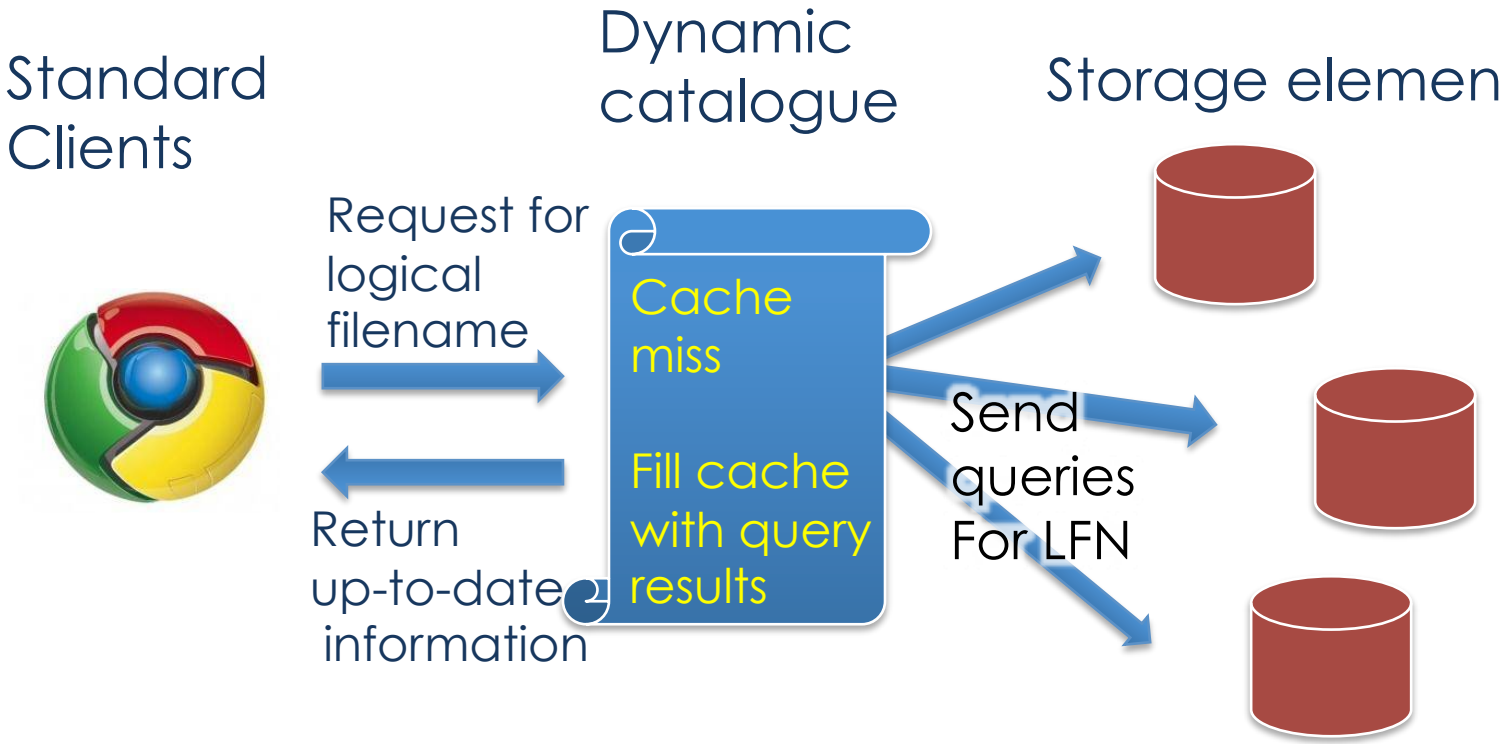
Manual interaction



Second approach (in discussion)

Dynamic location catalogues

The catalogue, on request, builds a list of URLs for logical filenames. The URLs automatically invalidate after some time (hours).



Example on how EMI works

Synchronizing catalogues and SE's guaranties continuity of production environment.

Dynamic catalogues is part of the data management evolution.

Selected topics

Topic 3

Standards

WebDAV
NFS 4.1 / pNFS

Standardization

Applying industry standards

Standards are the key for sustainability of Open Source Projects

Selected Topics

Example 1

WebDAV

Standardization : WebDAV

WebDAV (for data access)

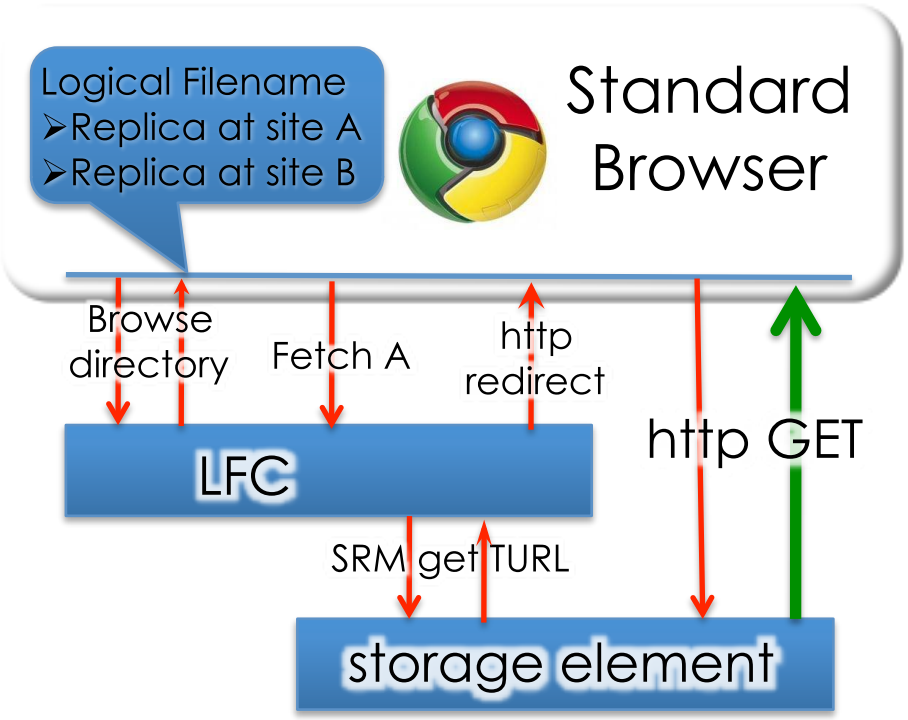
- Very useful for new (non-LHC) communities.
- ITEF Standard
- Allows “File system like” access with
  - Mac OS
  - Linux
  - Windows

With EMI-2 (mostly already with EMI-1) we provide WebDAV support from our SE’s

Standardization : WebDAV

WebDAV (for data browsing)

On request of the Irish Grid community EMI evaluates a WebDAV front-end for the file location catalogue.



Selected Topics

Example 2

NFS v4.1 / pNFS

What's NFSv4.1/pNFS ?

CITI, at the University of Michigan, is funded by major storage providers to coordinate the pNFS effort and provide reference implementations.

Industry Support - Implementations

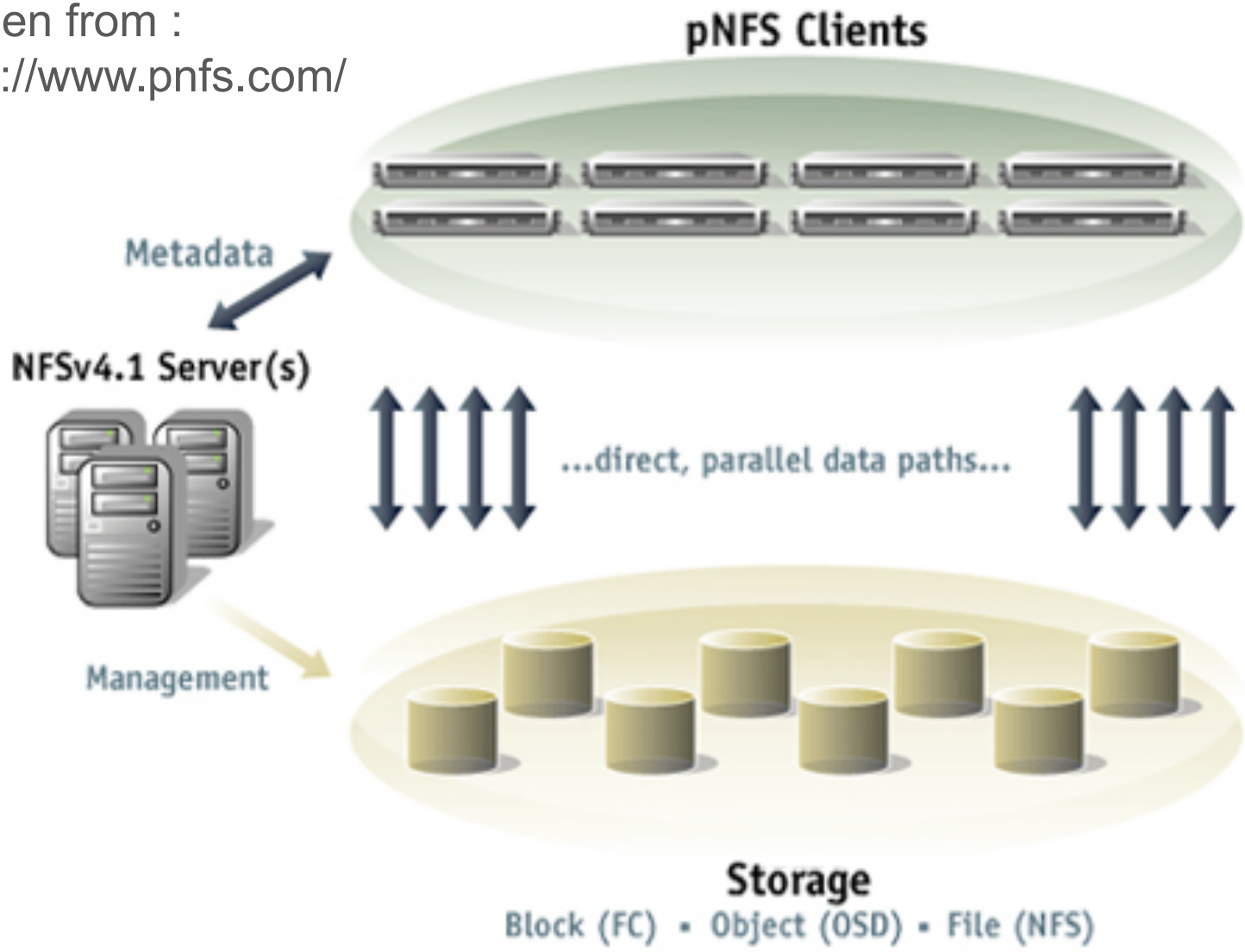
- Clients
  - Linux
  - Sun (Solaris)
- Servers
  - Desy
  - EMC
  - IBM
  - Linux
  - NetApp
  - Panasas
  - Sun (Solaris)

Several other implementations have been tested at Bake-a-thons and Connectathons

Group meets three times a year to check interoperability.

What is NFS4.1/pNFS and how does it work ?

Stolen from :
http://www.pnfs.com/



Why would one need it ?

Stolen from :
http://www.pnfs.com/

Benefits of Parallel I/O

- ✓ Delivers Very High Application Performance
- ✓ Allows for Massive Scalability without diminished performance

Benefits of NFS (or most any standard)

- Ensures Interoperability among vendor solutions
- Allows Choice of best-of-breed products
- Eliminates Risks of deploying proprietary technology

Why would we need it ?

Simplicity

- ✓ Regular mount-point and real POSIX I/O
- ✓ Can be used by unmodified applications (e.g. Mathematica..)
- ✓ Data client provided by the OS vendor
- ✓ Smart caching (block caching) development done by OS vendors

Performance

- ✓ pNFS : parallel NFS (first version of NFS which support multiple data servers)
- ✓ Clever protocols , e.g. Compound Requests

NFS v4.1 / pNFS availability

- ✓ EMI server
  - dCache : production version with EMI 1
  - DPM : prototype, ready for EMI-2
- ✓ Linux Kernel
  - Completed in 2.6.39
  - Back-port of pNFS into RH 6.2
- ✓ Industry
  - NetApp OnTab 8.1
  - Blue Arc in August
  - Other vendors : code ready but not officially available

Testing and deployment

- NFS 4.1 / pNFS tested for more than 9 month at the DESY GridLab (50% Tier II).
- Stable and performing well.
- First production dCache NFS 4.1 / pNFS system at DESY for photon science.
- Other WLCG sites including Tier I's start with testing NFS 4.1 / pNFS.

Competition

NFS 4.1 / pNFS is a great opportunity for Open Source Projects (EMI) to compete with industry and of course the other way around.

Conclusions

- *EMI Data* is a good opportunity to get our storage management middleware into a maintainable shape and at the same time provides a platform for evolution.
- Standardization is the way to get broader acceptance by other communities, which is especially important for EGI.
- EMI-Data will become THE competitor in Storage Management in Europe ☺.
- Everybody can join or may provide suggestions through EGI.eu.

Further reading

https://twiki.cern.ch/twiki/bin/view/EMI/EmiJra1T3Data

EMI is partially funded by the European Commission under Grant Agreement INFSO-RI-261611