WP5

Mass Storage Management

J Jensen
j.jensen@rl.ac.uk

DataGrid is a project funded by the European Commission
under contract IST-2000-25182

Outline

- Objectives
- Achievements
- Lessons learned
- Future & Exploitation
- Summary

Objectives

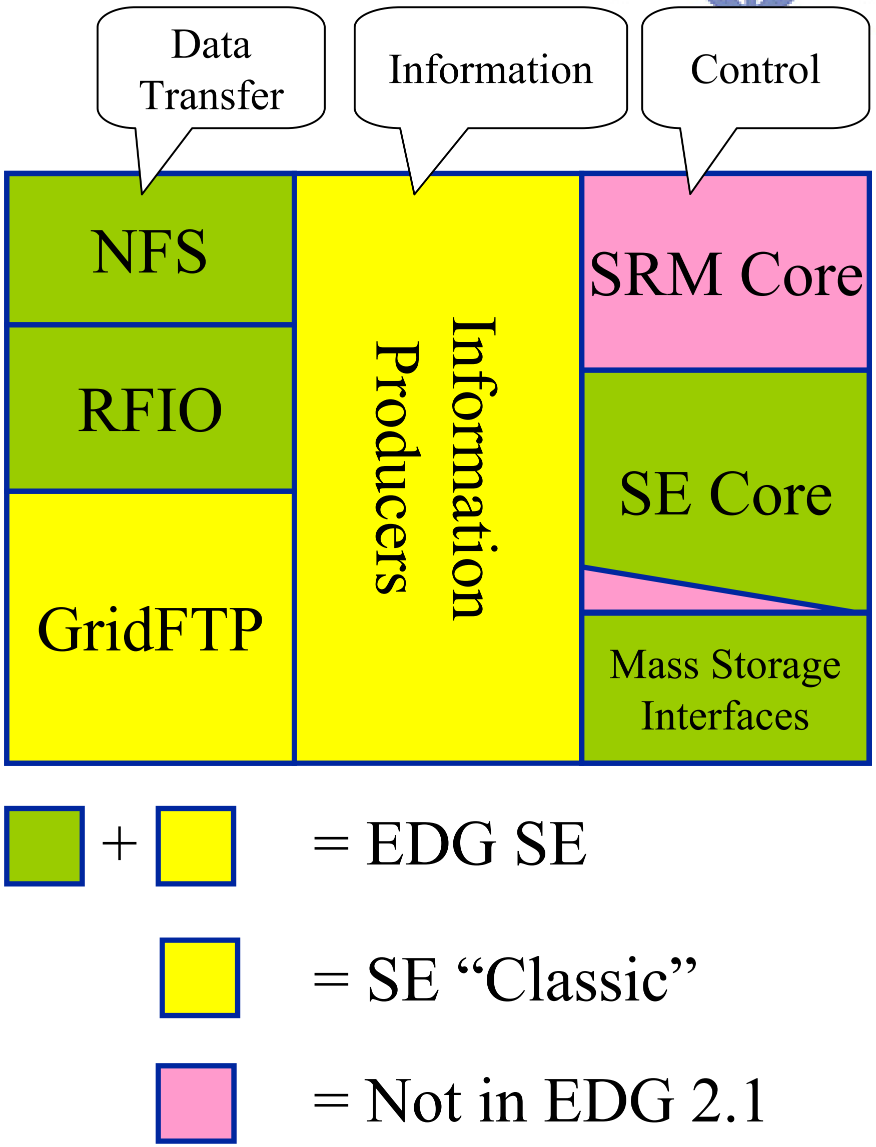
- Develop uniform interfaces to mass storage
  - Independent of underlying storage system
- Integrate with EDG Replica Management services
  - “Normally” users access SE via RM
- Develop back-end support for mass storage systems
  - Provide “missing” features, e.g. directory support
  - Provide Grid access control
- Publish information

Objectives – uniform interface

- Control interface
  - Original objective was "develop uniform interface to mass storage"
  - Must work with proxies ("Single sign-on")
  - Interface changed to be a web service for compatibility with other WPs halfway through the project
  - SRM version 1 was adopted as an alternative API for compatibility with other projects and LCG
- Data Transfer interface
  - Globus GridFTP required
  - Must support both encrypted and unencrypted transfers
- Information interface
  - Publish to MDS – later, to R-GMA

Achievements – Storage Element

- EDG Storage Element meets these objectives
- Flexible architecture
  - Cope with changing requirements
  - Pluggable features such as access control
  - Easy to extend
- Security
  - Secure interfaces
  - File level access control (not in EDG 2.1 though)
- Currently supports CASTOR, HPSS, ADS, as well as disk



Achievements – Storage Element

- SE's performance is acceptable
  - Performance dominated by data transfer times
    - E.g. 0.7 second per file for small files via GridFTP
  - Performance dominated by mass storage access
    - 10 minutes to stage in file from ADS
    - 30 minutes to stage in file from CASTOR
  - Basic core performance – 0.3 seconds per command
- Scalability
  - Scalability an issue, particularly for EO with many small files
  - Release 2.1 : 10000 files ok, 10000000 files not
  - Limits reached in underlying file system
  - Being addressed in new metadata implementation

Achievements – SE deployment

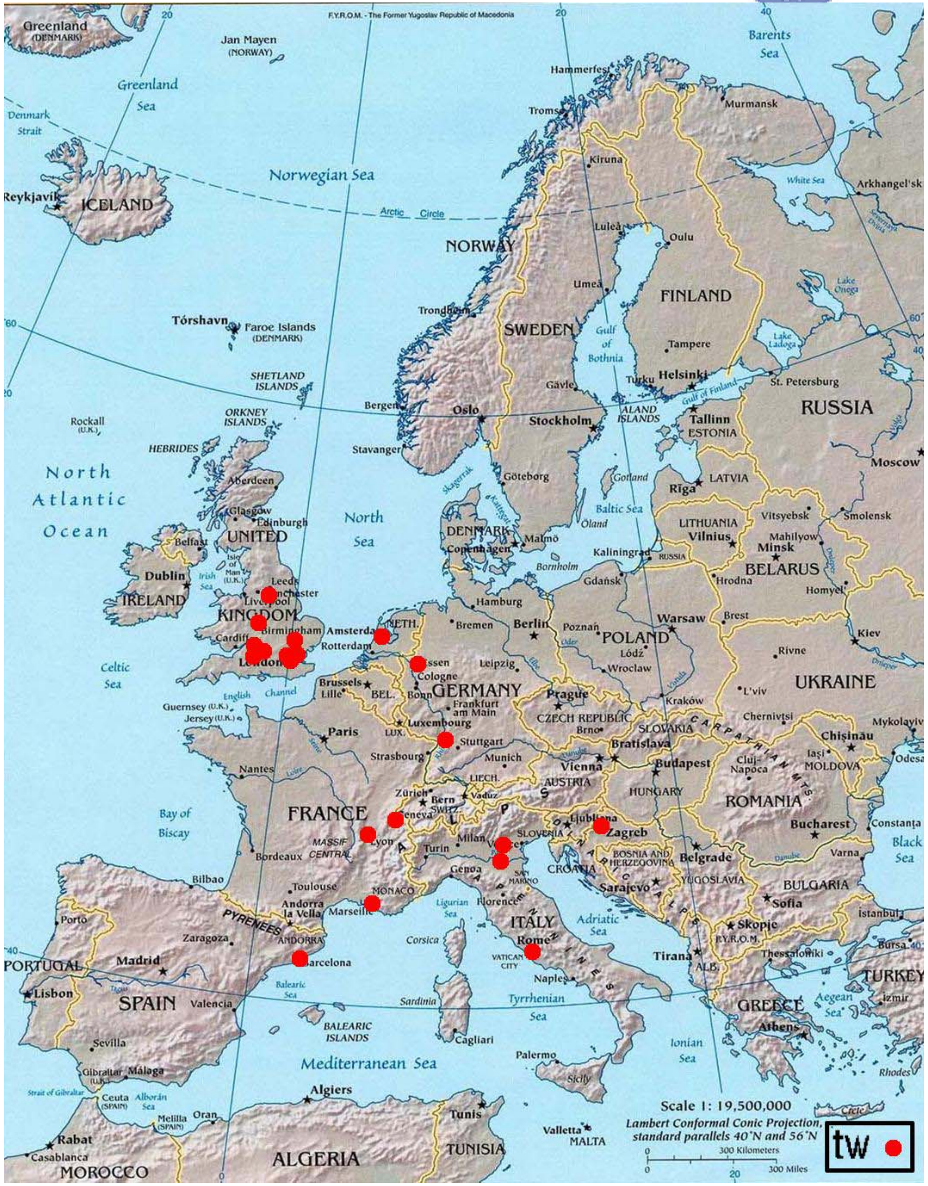
EDG SEs as of 17 Feb 2004

Note Taiwan !

Data from R-GMA (WP3) and mapcenter (WP7)

Many sites have more than one SE – a few sites have only Classic SE

London alone has three sites: IC, UCL, QMUL



Achievements – site specific

- CASTOR SRM
  - Provided an SRM interface to CASTOR at CERN
  - Interoperability demonstrated with FermiLab
  - SRMCopy implemented
- CASTOR GridFTP
  - Provided a GridFTP interface to CASTOR’s cache
  - Based on the Globus wu-ftpd GridFTP server
  - Files must be staged in before access
  - Transfer rates up to 30 MB/s (with specially tuned TCP settings)
- SARA
  - Porting SE to Irix, developing cache management tools

Achievements – collaborations

- Contributions to international standards and fora
  - SRM
    - Collaboration between Fermilab, Jefferson Lab, Lawrence Berkeley, RAL, CERN
    - Contributed to the design of the SRM version 2 protocol
  - GLUE
    - Contributed to the design of GLUE storage schema
  - GGF
    - Tracked developments in appropriate working groups
    - SRM not currently part of GGF
  - Dissemination
    - Talks at conferences and in working groups, publications,…
- EDG
  - Participated in ITeam, ATF, SCG, QAG,…

Achievements beyond release 2.1

- Access Control Lists (ACL)
  - Based on GACL
  - Fine-grained: Access based on user, file, and operation
  - Files can share ACLs
  - Work required to make more usable and user-friendly
- Improvements to metadata system
  - Toward a more scalable system
  - Two phases: first replace current metadata plugins (“handlers”)
  - Second: hook up to metadata database
  - First phase nearly complete, second phase expected concluded by April

Lessons learned

- Choice of architecture was definitely right
  - Architecture has successfully coped with changing requirements
- Look for opportunities for component reuse
  - Used web services deployment and security components provided by WP2
  - Deployed and developed further information producers supplied by WP3
  - Almost all parts of the Data Transfer components developed externally
- Prototype implementations live longer than expected
  - SE’s metadata system was implemented as prototype
  - Scalability issues discovered on application testbed

Lessons learned

- Inter-WP integration requires a lot of effort !
  - At times, nearly 100% of WP5 devoted to ITeam work and site installation support
  - Storage interface machines are heterogeneous
    - More installation support was required
  - For example, effort required to support DICOM servers was significantly underestimated
    - Requires significant effort from WPs 2, 3, 5, 10 – plus of course SCG, ATF, and, eventually, ITeam
- Need to agree standard protocols
  - Standards must be open and well-defined

Exploitation

- Used yesterday in middleware demo to access mass storage
- Used successfully on EDG testbeds by all EDG applications WPs
- “Atlas Data Challenge 1.5”
  - SE is currently used by Atlas to transfer data between ADS at RAL and CASTOR at CERN
  - About 1500 files; 2 TB in total
  - Files are copied by EDG RM and registered in an RC at RAL
  - This work is being done by Atlas outside the EDG testbeds
- The SE provides the Grid interface to ADS at RAL
  - This is important because ADS is being used by a large variety of scientific applications groups

Future and exploitation

- Storage Element SRM
  - SE will provide *generic* SRM 1 interface
  - This work is almost finished
  - Learning from the experience with CASTOR SRM
  - Work will be carried on by RAL; later in GridPP 2
  - Will investigate whether to build SRM version 2
    - Depends on uptake of protocol in international community
    - Current SRM implementation is built with also SRM 2 in mind
    - Some additional features required
- Storage Element – further mass storage systems
  - Scope for implementing support for AMS, DICOM?
  - Support for UK Tier-2 sites to be developed by GridPP2

Future and exploitation

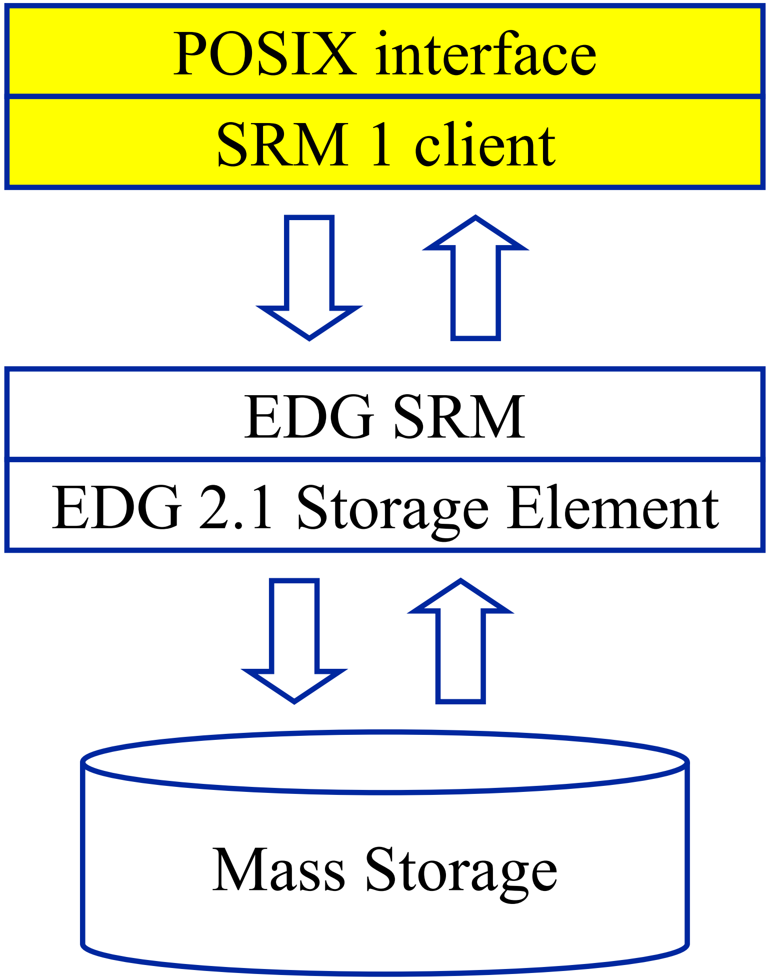
- Storage Element and VOMS
  - Integrate VOMS support into SE – SE already works with VOMS proxies
  - Will enable more scalable access control
  - Fairly easy task – accomplished again by reusing components
  - May need to VOMS-enable GridFTP server – integrate LCAS and LCMAPS
- Integration with GFAL
  - LCG's "Grid File Access Library" – POSIX style interface
  - Planned integration using SRM 1 interface
- Automatic Grid mirroring
  - Edinburgh and Glasgow looking into using SE for automatic mirroring of data

Summary

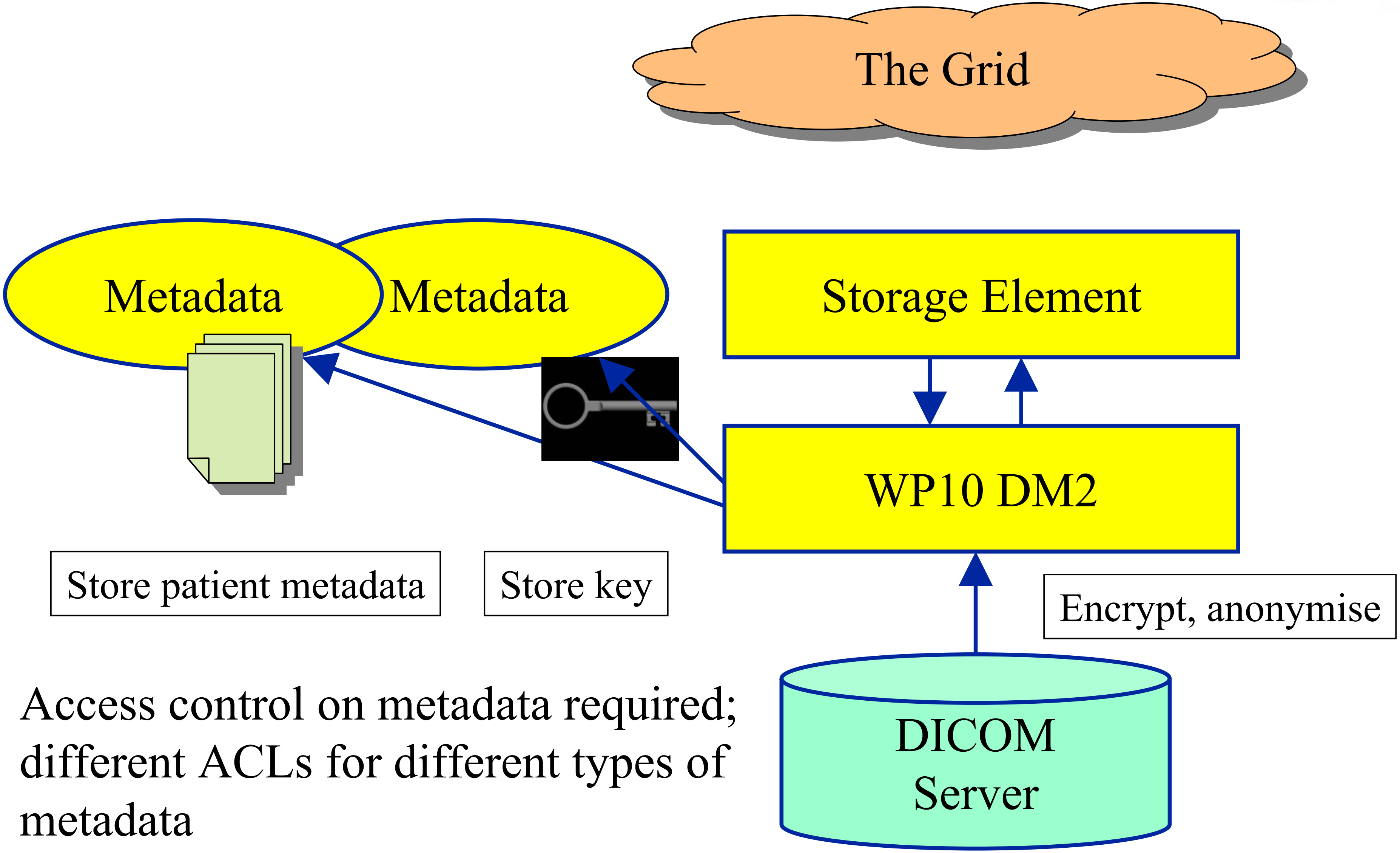
- EDG Storage Element
  - Meets the requirements; in some cases exceeds them
  - Provides a uniform Grid interface to mass storage
  - Interfaces with EDG Replica Management system
  - Dual solution – lightweight “SE classic” and full-featured SE
  - SRM 1 to CASTOR, other systems being prepared
  - Commitment to resolve open issues
- Applications
  - SE being used by middleware WPs
  - Applications in follow-on and external projects
    - E.g. UK e-Science programme projects
    - For example, SE is Grid interface to ADS

GFAL, SRM, and Storage Element

- LCG decided to use GFAL – the "Grid File Access Library"
- It was decided to interface to EDG SE using SRM 1 interface
- SRM 1 can also be used for interoperability with DoE Labs
- We are integrating the EDG SRM layer with the EDG SE
- Some complications → not in 2.1
- We are committed to completing the task



DICOM server support



Access control on metadata required; different ACLs for different types of metadata