



On Convergence

Replica Catalogue, Replica Manager, Metadata Catalogue

Heinz.Stockinger@cern.ch
for EDG-WP2 (Data Management)

European DataGrid - DataTAG meeting
29 May 2002



Outline

- Replica Catalogue
 - centralised, LDAP based vs distributed (Replica Location Service (Giggle))
- Replica Manager
 - edg-replica-manager vs Reptor
- Metadata Catalogue (RepMeC)



Replica Catalogue (RC) - Status

- Current state:
 - in EDG testbed we use Globus Replica Catalogue (**LDAP** based)
 - does not fulfill all requirements but is a short term solution
 - WP2 provides some **patches for security** to the Replica catalogue
 - new RPMs should be available this week
 - centralised LDAP based RC will remain **until testbed 2**

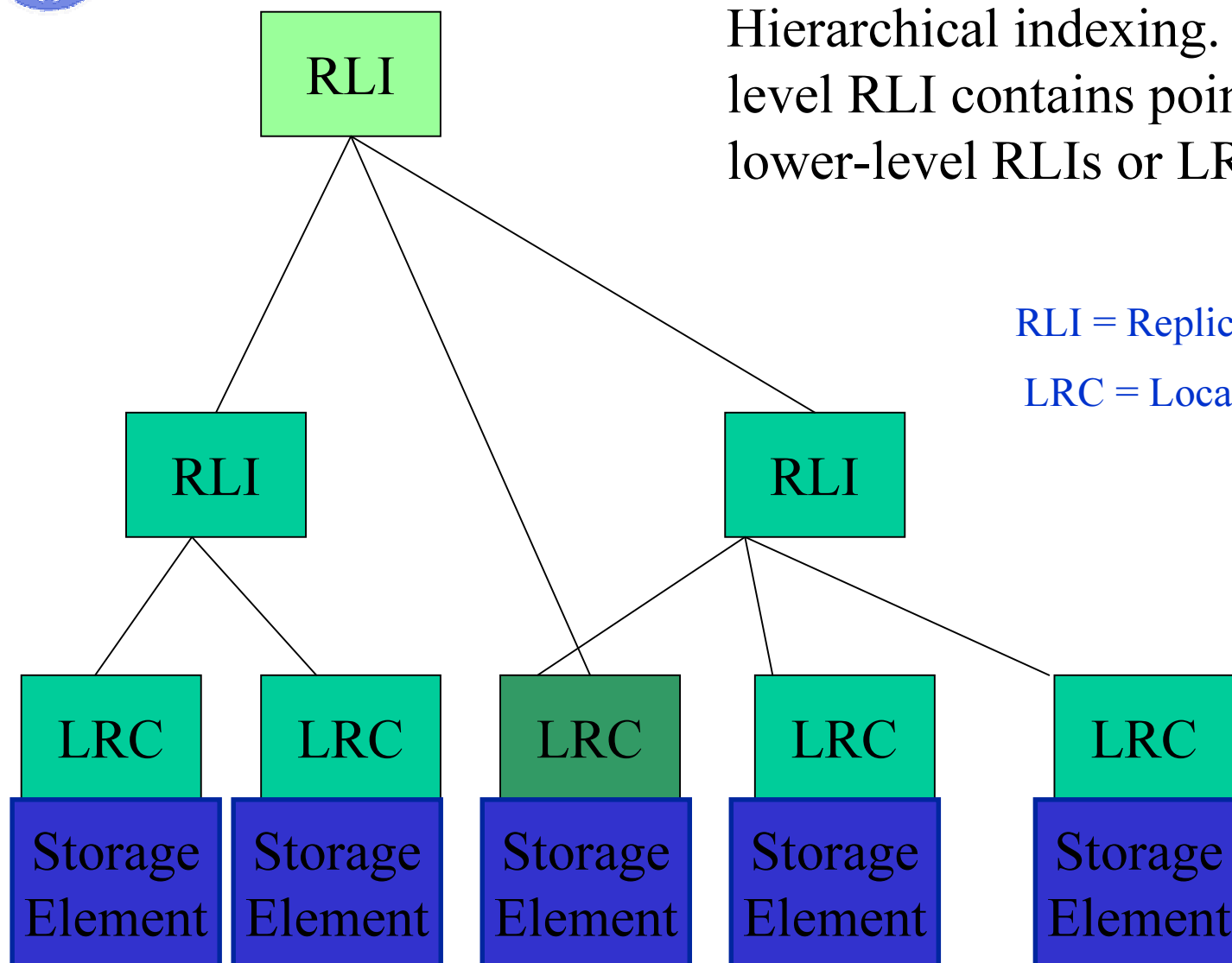


Replica Location Service (Giggle)

- Distributed Replica Catalogue
- Giggle is a **joint effort** between
 - Globus (ISI, Argonne)
 - WP2 (Data Management) – European DataGrid
 - some PPDG members (e.g. SLAC)

[http://www.cern.ch/grid-data-management/replica-location service/RLS.html](http://www.cern.ch/grid-data-management/replica-location%20service/RLS.html)

Hierarchical indexing. The higher-level RLI contains pointers to lower-level RLIs or LRCs.



RLI = Replica Location Index
LRC = Local Replica Catalog



Replica Location Service - Status

- Mutual design
 - bi-weekly phone conferences
- Development mainly Globus
- testing in EDG
- **prototype available** on some WP2 testbed machines
- more stability tests and documentation will be provided
- **will replace** current **LDAP** based Replica Catalogue



Replica Manager - Status

- **edg-replica-manager-1.0** available in testbed
 - <http://www.cern.ch/grid-data-management/edg-replica-manager>
 - based on current Globus replica management libraries
- prototype replica manager
 - client package only: on server side GridFTP
 - EDG should shall get familiar with replica manager concept
 - no Mass Storage System interface yet
- an additional solution to GDMP
 - GDMP 3.0-1 available on testbed



Reptor - new Replica Manager

- Currently in design and development phase
 - an internal beta prototype already available
 - <http://www.cern.ch/grid-data-management/reptor>
- Java based and will have web service interface
- already **partly integrated with Replica Location Service**
- **will replace edg-replica-manager** in testbed 2
- WP2 works on **white paper with Globus** on general replica management
 - a broad scope for a general design is planned



Replica Metadata Catalog - RepMeC

- Store **metadata for Replica Manager** and its modules.
- Usage expected to be **simple inserts and lookups**, NOT extensive queries returning thousands of rows.
- Optimized for **many concurrent clients**.
- **Extensible** for user-defined metadata on Logical File Names.
- May handle object-to-file mapping and LFN aliases.
- Based on Spitfire



RepMeC functionality

- Store and retrieve **information on**
 - Security data: Original owner and associated ACL
 - Master LRC site
 - Uniform file attributes (filesize, CRC, creation date..)
- **User-defined attributes**
 - add user-defined columns like Collections
 - insert and lookup on user-defined columns
- Fine granularity data
 - capability to **store and retrieve OIDs** for objects or data elements that are contained in the file. Assumes n:1 relation.
 - capability to store **aliases for LFNs** along the same lines as symbolic links for files in the unix file system. Aliases are clearly distinguishable by their name.



Conclusion

- Several software packages are already in place
- convergence with US projects is well under way
 - close collaboration on several fields
 - including Mass Storage System interface
- upgrade within EDG from current prototype and intermediate solutions to EDG