

# Next-Generation EU DataGrid Data Management Services

James Casey  
CERN

[James.Casey@cern.ch](mailto:James.Casey@cern.ch)

On behalf of EU DataGrid WP2





# Talk Outline

- ◆ Introduction to EU DataGrid workpackage 2
- ◆ WP2 Service Design and Interactions
  - Spitfire
  - Replication Services
  - Security
- ◆ Conclusions and outlook

## Authors

Diana Bosio, James Casey, Akos Frohner, Leanne Guy, Wolfgang Hoschek, Peter Kunszt, Erwin Laure, Levi Lucio, Heinz Stockinger, Kurt Stockinger - **CERN**

Giuseppe Andronico, Federico DiCarlo, Andrea Domenici, Flavia Donno, Livio Salconi - **INFN**

William Bell, David Cameron, Gavin McCance, Paul Millar, Caitriona Nicholson - **PPARC, University of Glasgow**

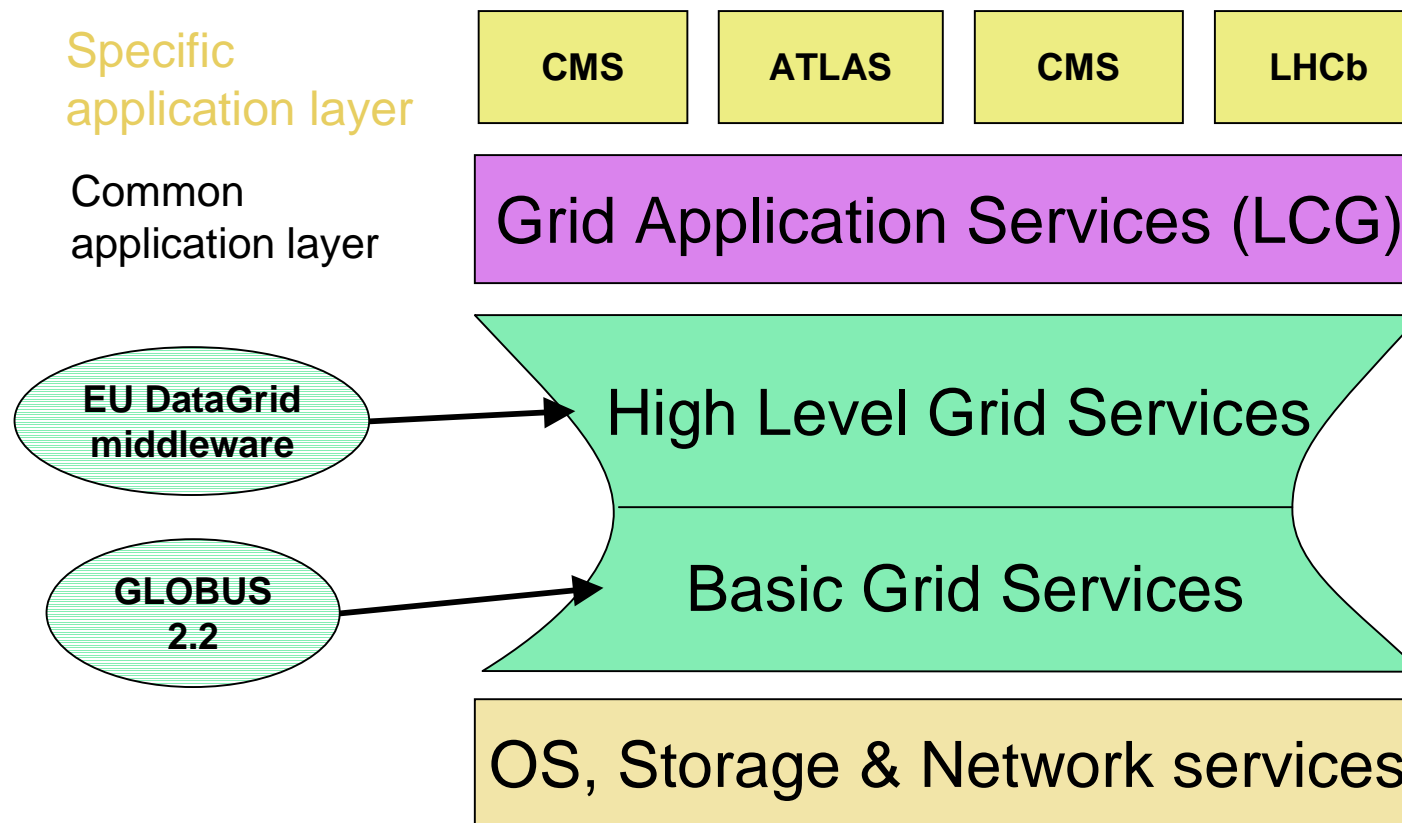
Joni Hahkala, Niklas Karlsson, Ville Nenonen, Mika Silander, Marko Niinimaki - **Helsinki Institute of Physics**

Olle Mulmo, Gian Luca Volpato - **Swedish Research Council**

# Grid middleware architecture hourglass



Current Grid architectural functional blocks:



# EU DataGrid WP2

## Data Management Work Package



### **Responsible for**

- ◆ Transparent data location and secure access
- ◆ Wide-area replication
- ◆ Data access optimization
- ◆ Metadata access

### **NOT responsible for** (but partially relying on other WPs for)

- ◆ Data storage
- ◆ Proper Relational Database bindings
- ◆ Remote I/O
- ◆ Security infrastructure



## WP2 Service Paradigms

### ◆ Choice of technology:

- Java-based servers using Web Services
  - Tomcat, Oracle 9iAS
- Interface definitions in WSDL
- Client stubs for many languages (Java, C, C++)
  - Axis, gSOAP
- Persistent service data in Relational Databases
  - MySQL, Oracle

### ◆ Modularity

- Modular service design for pluggability and extensibility
- No vendor specific lock-ins

### ◆ Evolvable

- Easy adaptation to OGSA foreseen, based on the same technology
- Largely independent of underlying OS, RDBMS

# Spitfire: Grid-enabling RDBMS



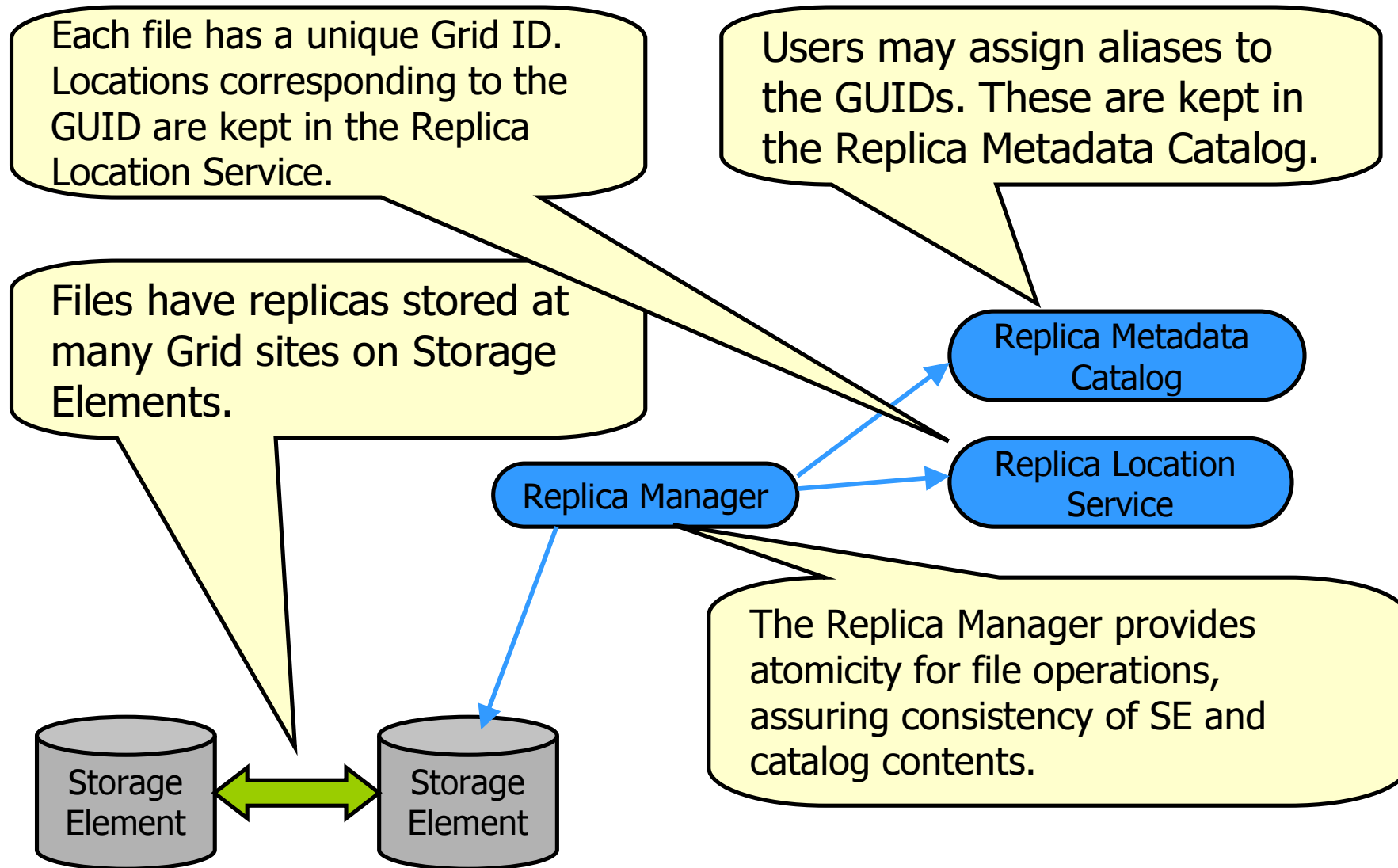
## ◆ **Capabilities:**

- Simple Grid enabled front end to any type of local or remote RDBMS through secure SOAP-RPC
- Sample generic RDBMS methods may easily be customized with little additional development, providing WSDL interfaces
- Browser integration
- GSI authentication
- Hooks in place for local authorization

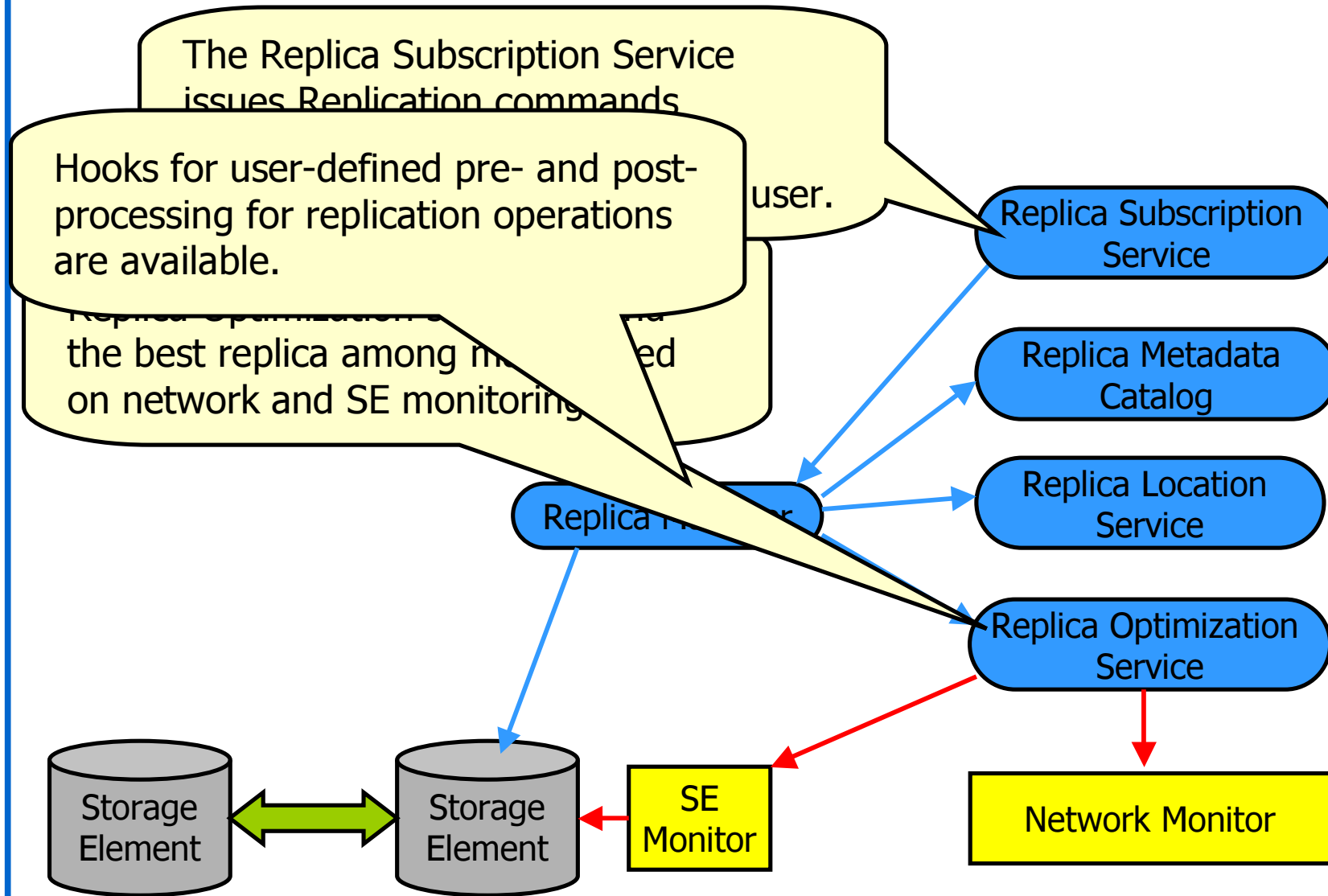
## ◆ **Status:** current version 2.1

- Used by EU DataGrid Earth Observation and Biomedical applications.
- Not suitable for the retrieval of LARGE result sets

# Replication Services: Basic Functionality

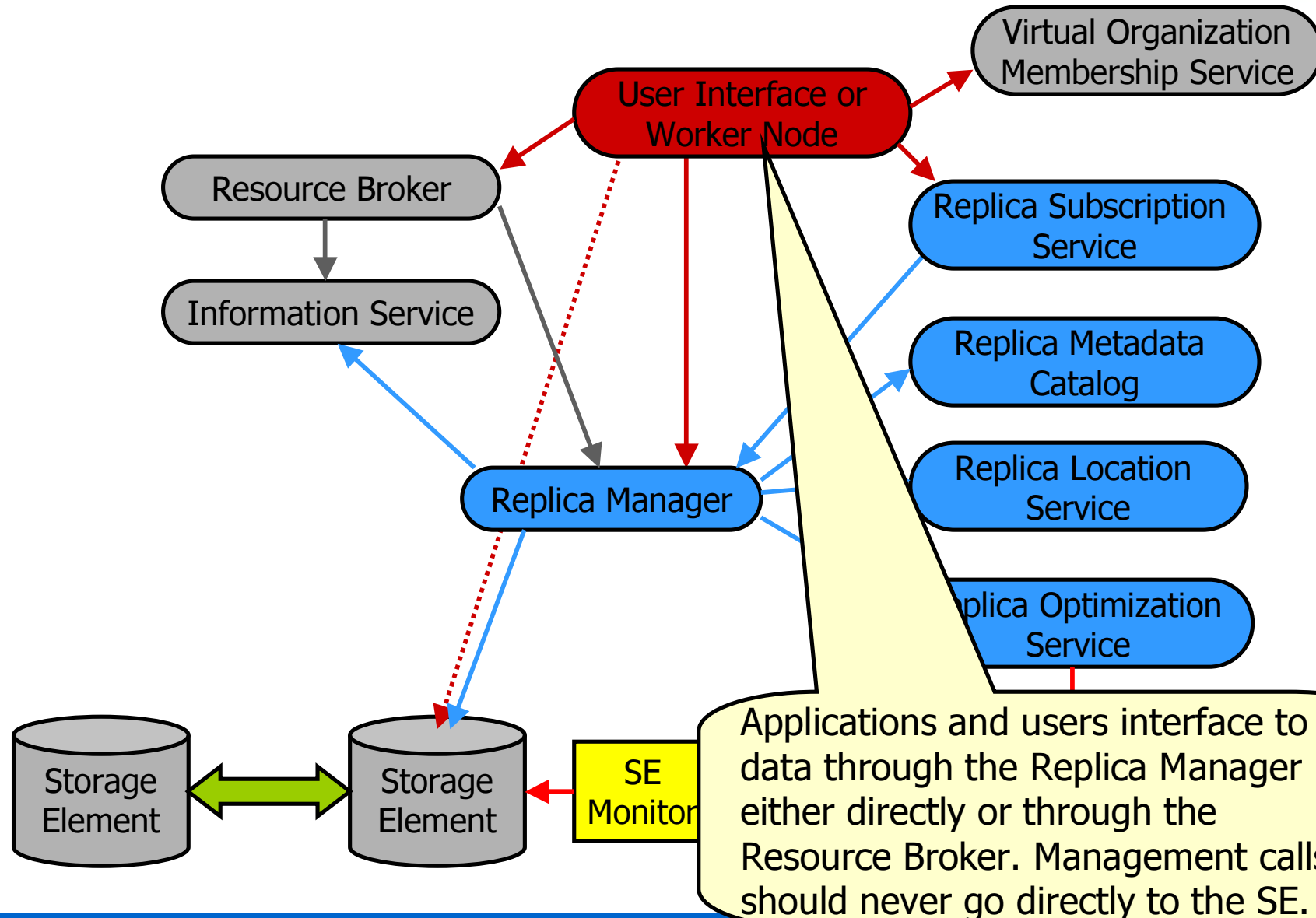


# Higher Level Replication Services





# Interactions with other Grid components





# Replication Services Status

## ◆ Current Status

- All components are deployed *right now*
- Initial tests show that expected performance can be met
- Need proper testing in a 'real user environment' – EDG2; LCG1

## ◆ Features for next release

- Currently Worker Nodes need outbound connectivity – Replica Manager Proxy Service needed. Needs proper security delegation mechanism.
- Logical collections support
- Service-level authorization
- Subscription Service does not handle individual users – due to missing delegation.

# Security: Infrastructure for Java-based Web Services



## ◆ Trust Manager

- Mutual client-server authentication using GSI (ie PKI X509 certificates) for all WP2 services
- Supports everything transported over SSL

## ◆ Authorization Manager

- Supports coarse grained authorization:  
Mapping user->role->attribute
- Fine grained authorization through policies, role and attribute maps
- Web-based Admin interface for managing the authorization policies and tables

## ◆ Status:

- Fully implemented, authentication is enabled on the service level
- Delegation implementation needs to be finished
- Authorization needs more integration, waiting for deployment of VOMS



## Conclusions and outlook

- ◆ The second generation Data Management services have been designed and implemented based on the **Web Service paradigm**
- ◆ **Flexible, extensible** service framework
- ◆ Deployment **choices** : robust, highly available commercial products supported (eg. Oracle) as well as open-source (MySQL, Tomcat)
- ◆ First experiences with these services show that their performance **meets the expectations**
- ◆ Real-life usage will show its strengths and weaknesses on the LCG-1 and EDG2.0 testbeds during **the rest of this year.**