The gLite FiReMan Catalog

Peter Kunszt on behalf of JRA1-DM

gLite
Lightweight Middleware for Grid Computing

Information Society

www.eu-egee.org

Service Oriented Architecture

gLite
Lightweight Middleware for Grid Computing

Interoperability

Portability

Modularity

Scalability

Web Services

Building on existing components in a lightweight manner

| AliEn | LCG | Condor |
| Globus | SRM | ... |
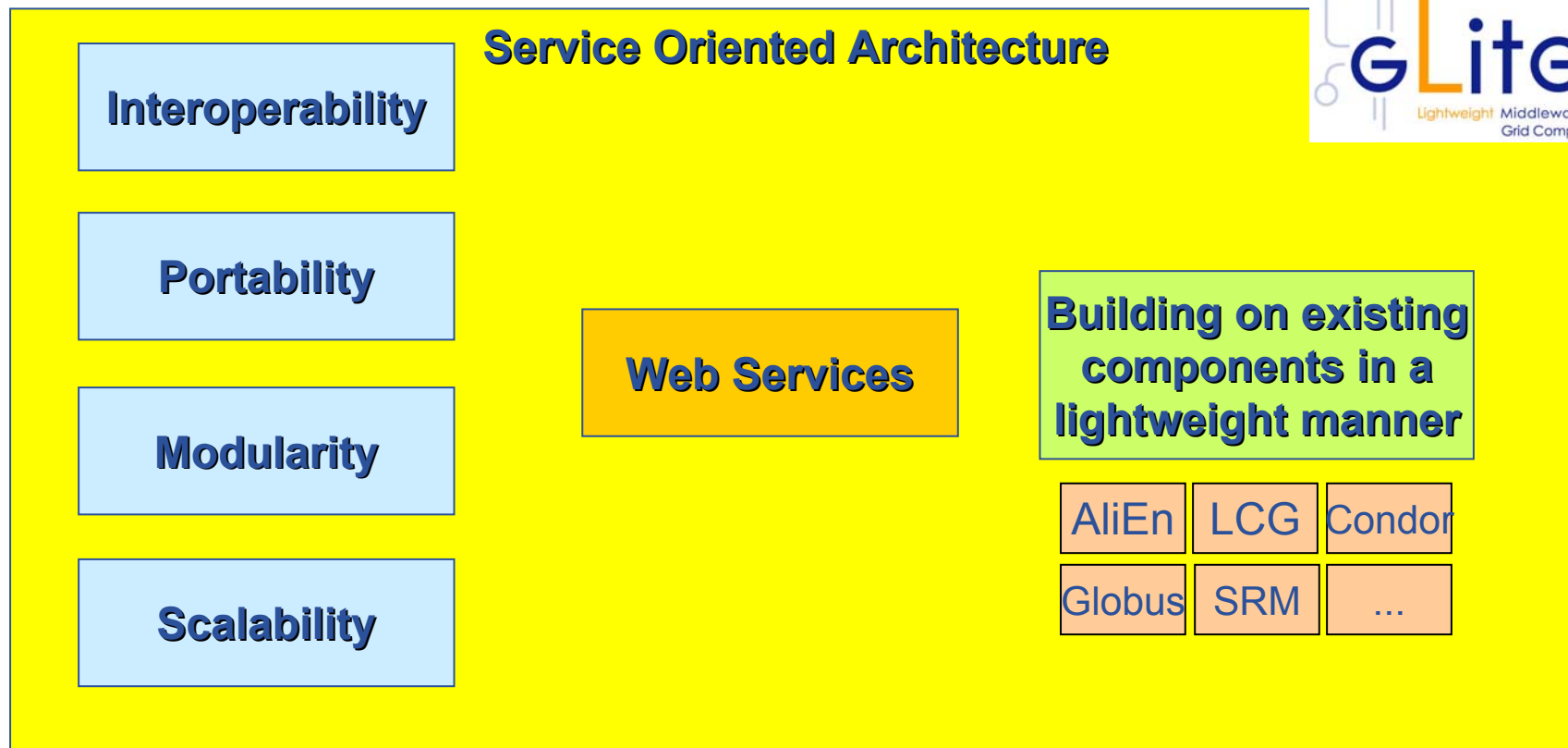
- **Catalogs built based on requirements from HEP experiments and the Biomedical EGEE community**
- **Started design from AliEn File Catalog**
  - Logical namespace management
  - Virtual Filesystem view (DataSets via directory hierarchy)
  - Support Metadata attached to files
  - Bulk Operations
  - Strong security: basic unix permissions and fine-grained ACLs (i.e. not just directory but file-granularity)
  - Support flexible deployment models
    - Single central catalog model
    - Site local catalogs connected to a single central catalog model
    - Site local catalogs without single central catalog model
  - Scalable to many clients and to a large number of entries; address performance issues seen with EDG RLS

Enabling Grids for E-sciencE

- **Oracle implementation**
- **Hierarchy: Filesystem view**
- **Full Security support**
  - VOMS integration
  - Fine grained ACL control
  - Minimal performance penalties
- **Bulk operations**
  - Necessary to meet performance requirements
- **WS-tuning**
  - Web services are not as bad as they first seem
- **Useful and intuitive interfaces**
  - Logical interface decomposition for well-defined feature-sets
- **Metadata support**

egee
Enabling Grids for E-sciencE
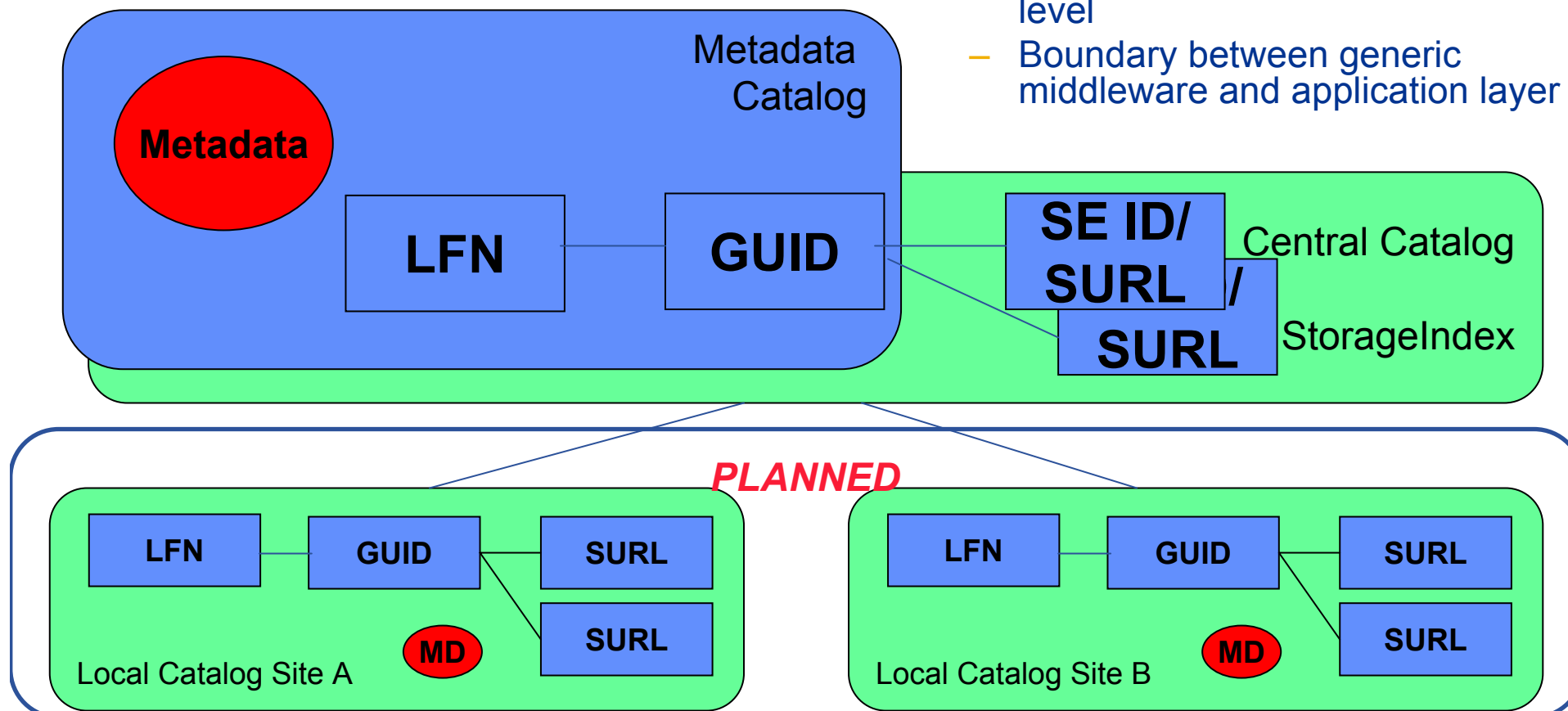
- **File Catalog and StorageIndex**
  - Filesystem-like view on logical file names
  - Keeps track of sites where data is stored
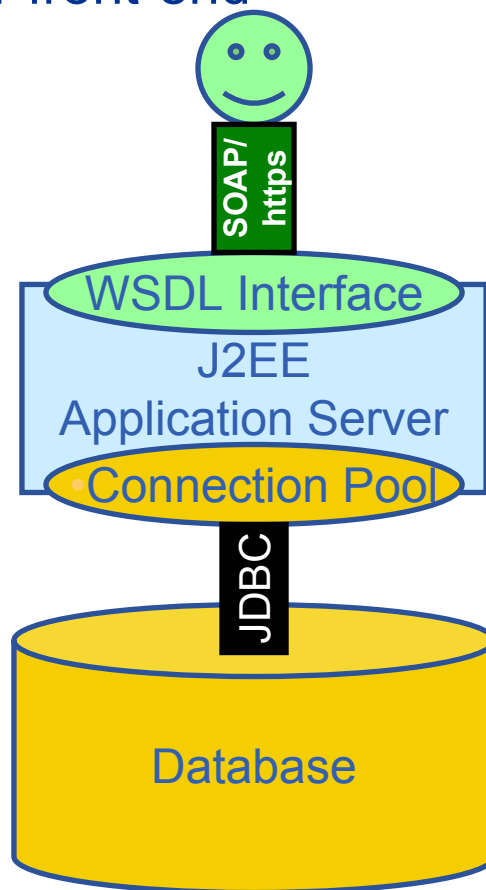 - Conflict resolution

- **Replica Catalog**
  - Keeps information at a site
- **Meta Data Catalog**
  - Attributes of files on the logical level
  - Boundary between generic middleware and application layer

- **2-tier architecture**
  - Database backend
 - Application Server front-end

SOAP/ https

WSDL Interface

J2EE
Application Server

Connection Pool

JDBC

Database

- **Both tiers scalable**
  - Database Clustering (Oracle RAC)
  - Any number of stateless Application Servers

eGee

Enabling Grids for E-sciencE
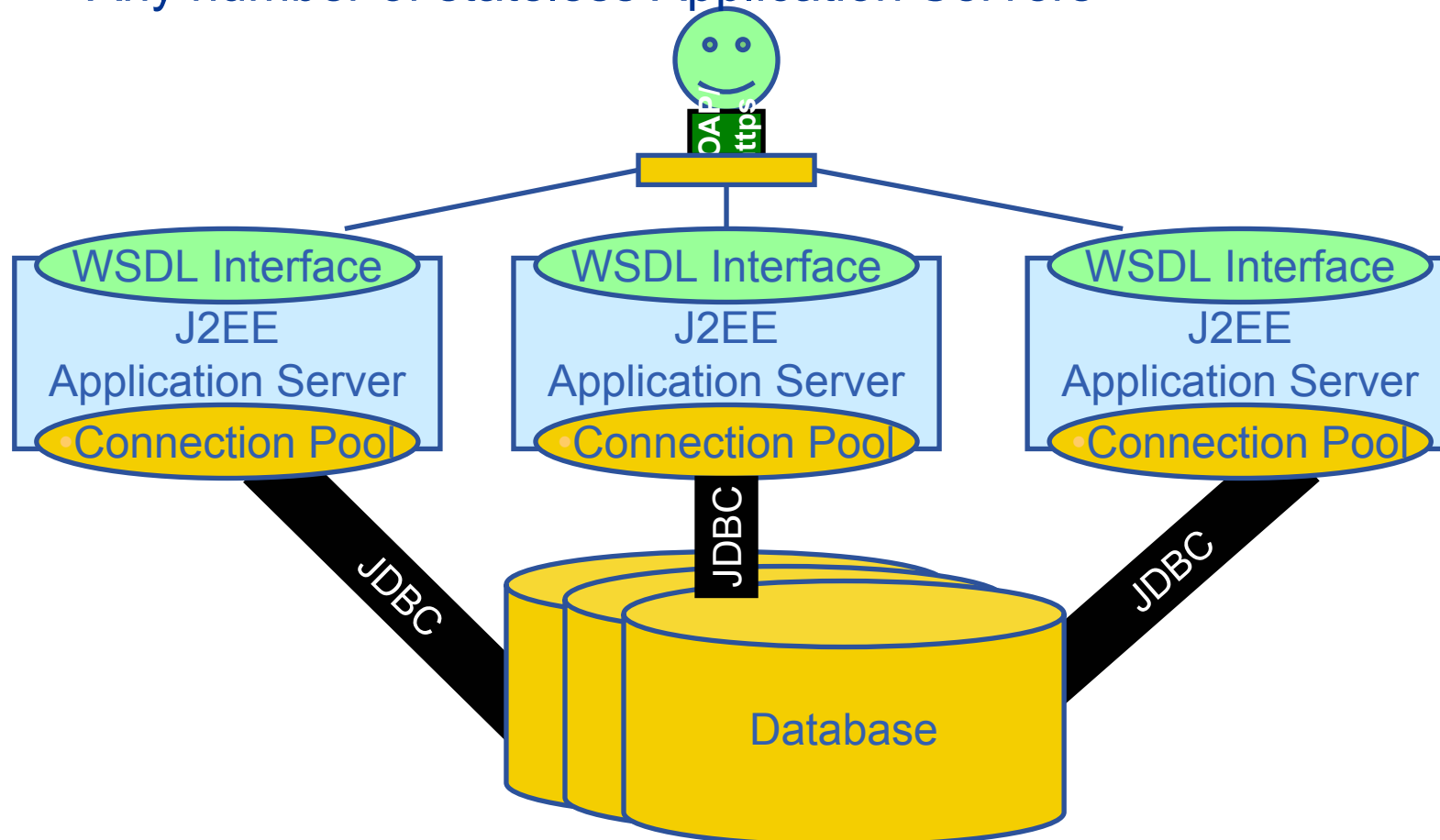
- **2 independent implementations exist**

Oracle Implementation

- **Catalog Logic lives inside Oracle as Stored Procedures**
- **Tomcat parses credential only, passes operations through to DB**

TOMCAT5
J2EE
Application Server

Database
ORACLE
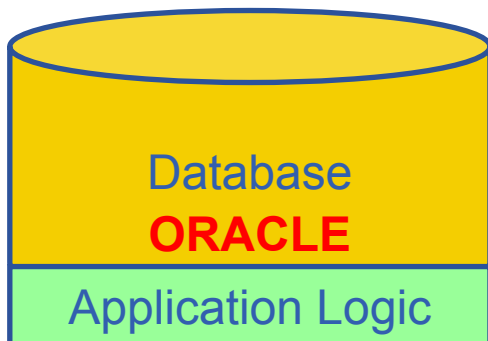
Application Logic

MySQL Implementation

- **Simple Table Structure using InnoDB tables**
- **Credential parsing and all of the logic is in Tomcat**

TOMCAT5
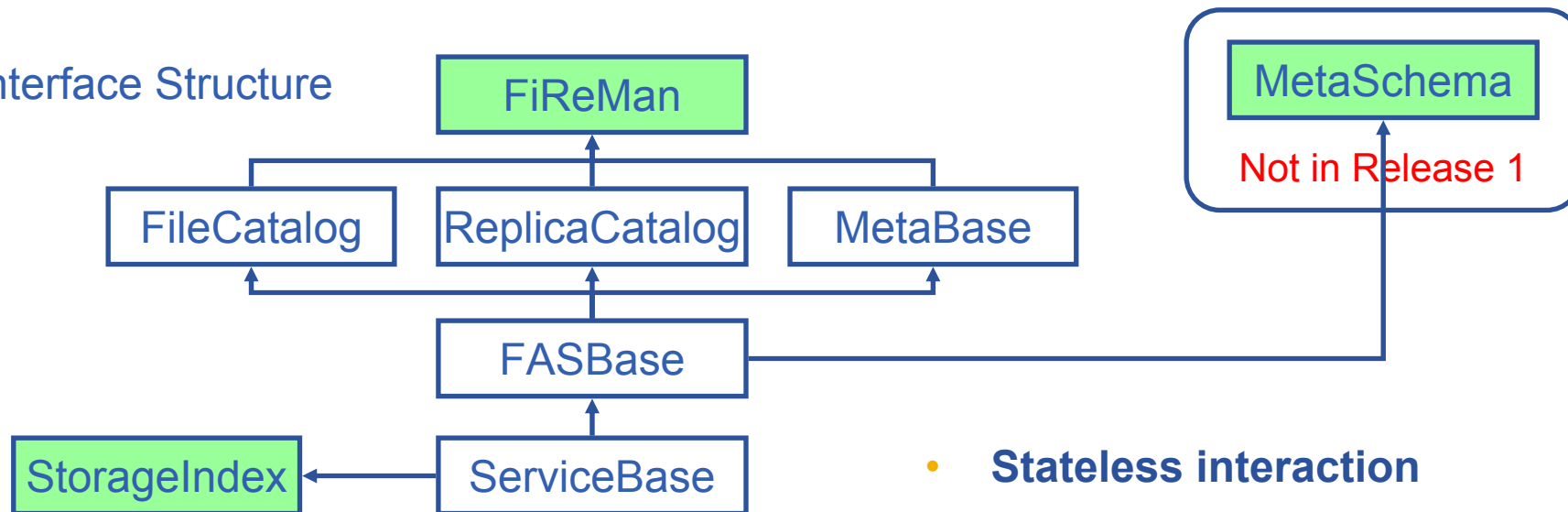J2EE
Application Server

Application Logic

Database
MYSQL
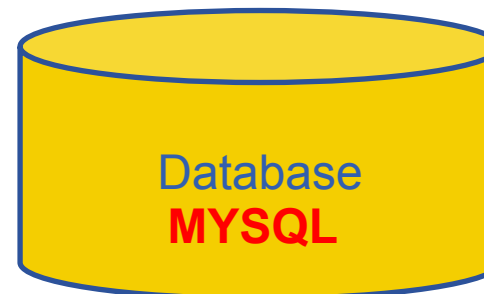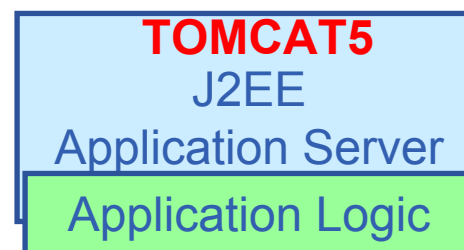
- **Logical File Namespace management**   **FileCatalog**

- **Replica locations**   **ReplicaCatalog**

- **File-based metadata**   **MetaBase**

- **Metadata Management**   **MetaSchema**

- **Authentication and Authorization information (ACLs)**   **FASBase**

- **Service Metadata**   **ServiceBase**

- **WMS interaction and global file location**   **ServiceIndex**

Interface Structure

```
                    FiReMan                          MetaSchema

                                                    Not in Release 1
   FileCatalog   ReplicaCatalog   MetaBase

                    FASBase

StorageIndex  ←  ServiceBase
```

- **Stateless interaction**

- **No transactions outside Bulk**

Enabling Grids for E-sciencE

- **Web-services interface**: Guarantees client support on many platforms and in many languages. **Standardization effort ongoing**. It is being managed through the EGEE PTF, where HEP has their representatives. Provided by us are:
  - **Linux Command Line tools**
  - **C/C++ API**
  - **Java API**
  - **Perl modules**
  - **JavaScript (for web clients)**
  - **gLite integrated bash (glitesh) – prototype**
- **Security:** Fine-grained ACL support with minimal performance penalty.
  - DNs own the files
  - VOMS group support
  - Basic Unix security (ugo rwx)
  - Additional ACLs for setPermission, list, remove, setMetadata, getMetadata

Enabling Grids for E-sciencE

- **ServiceIndex:** Interface also implemented in AliEn FC.
  - Will support the DLI method as well by end of April
- **Metadata support**: The final interface (not in current Fireman) is a joint effort between EGEE PTF, ARDA and the Experiment Metadata Group.
- **POOL integration:** A first version is available and has been delivered to the POOL team. POOL transactions are translated into bulk operations.
- **Performance**: If the single-shot operation performance is important, a non-WS interface a la LFC can easily be added. *See next talk.*
- **Distribution**: Planned in two modes:
  - **Based on reliable messaging. A proof-of-concept prototype exists.**
  - **Based on Namespace partitioning**

eGee

Enabling Grids for E-sciencE

- **Usage is of course optional.**
  - Interfaces are very basic. In order to use gLite tooling they may also be implemented on top of any other catalog if deemed useful.
- **As File Catalog**
  - Global or local deployment
  - Locally just as policy and authorization enforcement point (ACL control)
- **As Replica Catalog**
- **As StorageIndex/DLI (interface for WMS interaction)**
- **To explore**
  - Fireman is basically a hierarchical namespace catalog with metadata and ACL capabilities. Might be used to some extent as
    - Dataset or collection catalog
    - File Metadata catalog
  - Distributed catalog technology (Messaging) usage for experiment-specific (meta)data

- **Fireman is the result of a long iterative effort to get a complete file/replica catalog with the necessary features, meeting most requirements. We had a very fruitful close interaction with ARDA. It was delivered according to the gLite release schedule.**
- **Integrated with all other gLite components and with POOL**
- **As mature as can be expected**
  - Code freeze was christmas 2004
  - INTENSE testing has started in january
  - A lot of bugs have been discovered and fixed in his period, more to come – depends on how heavily it is being used
- **Single central catalog is ready to be moved to the next level of testing/deployment by LCG**
- **We are available to work out the details of possible usage by the experiments individually**
  - Directly
  - Or through ARDA
- **Next release: distributed catalogs – and more according to priority**

Enabling Grids for E-sciencE

- **JRA1 Data Management homepage**
  http://cern.ch/egee-jra1-dm

- **gLite FiReMan user guide**
  - Overview
    https://edms.cern.ch/file/570643/1/EGEE-TECH-570643-v1.0.pdf
  - Command Line tools
    https://edms.cern.ch/file/570780/1/EGEE-TECH-570780-v1.0.pdf
  - C/C++ API
    https://edms.cern.ch/file/570780/1/EGEE-TECH-570780-C-CPP-API-v1.0.pdf
  - Java API
    https://edms.cern.ch/file/570780/1/EGEE-TECH-570780-JAVA-API-v1.0.pdf

- **gLite Release 1**
  - http://glite.web.cern.ch/glite/packages/R1.0/R20050331
  - http://glite.web.cern.ch/glite/documentation

eGee

Backup Slides

www.eu-egee.org

Information Society

Scaling: Namespace Partitioning

- **Local Catalogs having parts of the namespace**

Site 1

Site 2

Site 3
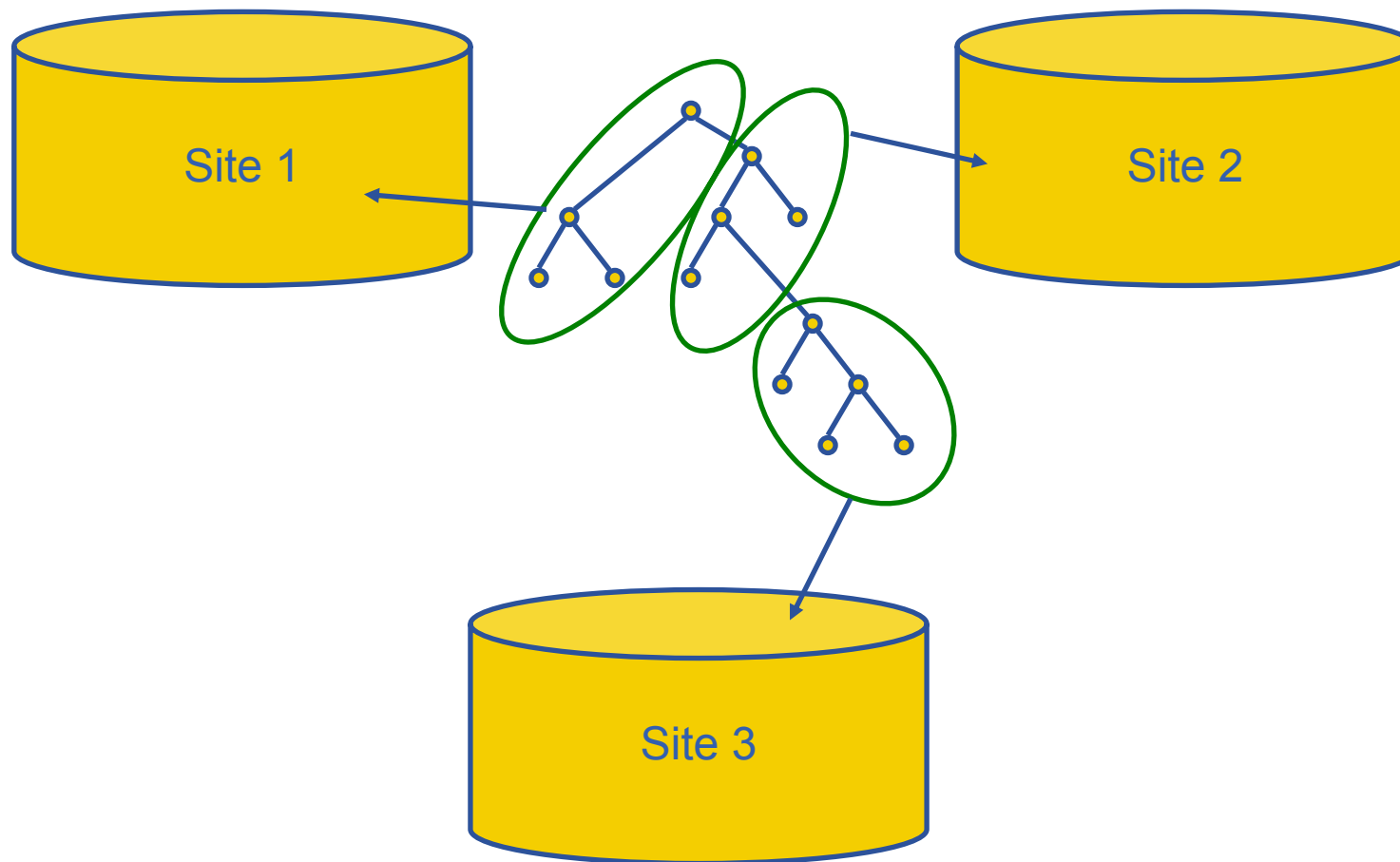
- **Local Catalogs**
- **Catalog Indices**
- **Message Oriented Middleware (JMS Messaging) to link catalogs and indices**
  - Both Open Source and Commercial implementations exist (just like for RDBMSs) – we use JORAM for open source
  - Reliable, managed queuing and delivery of messages
  - Flexible setup

Catalog

Index

Catalog

Index

Catalog

MOM Infrastructure
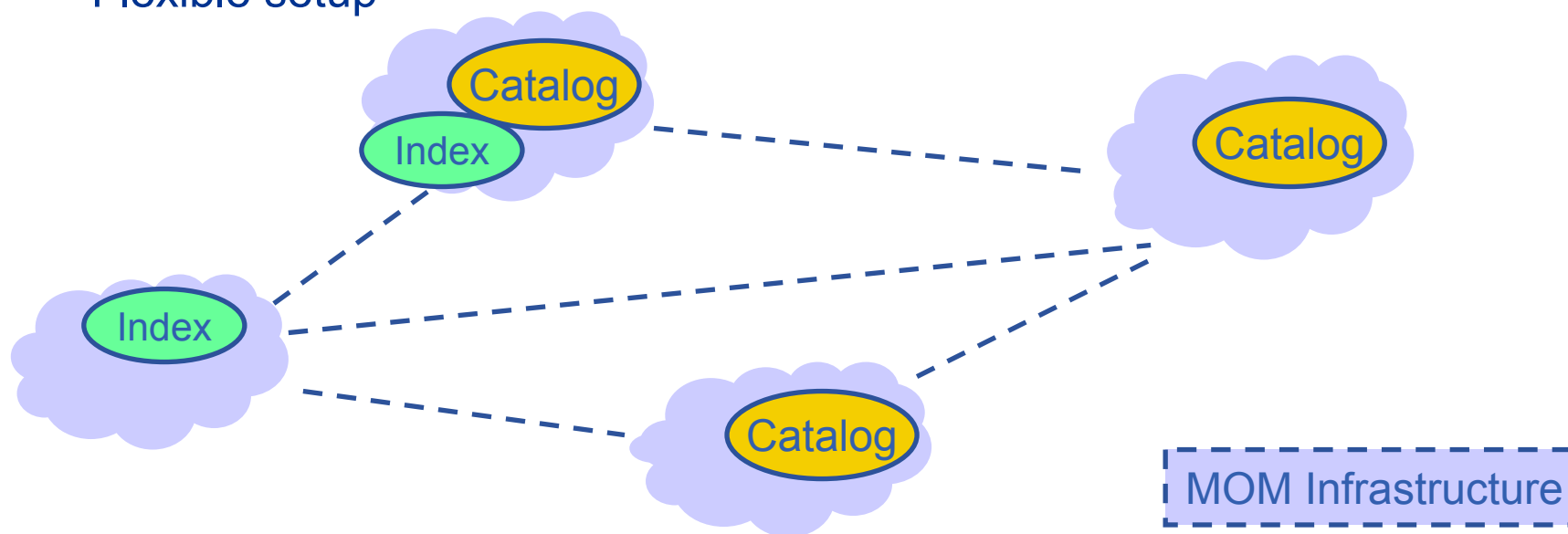
- **File Access**
  - glite-get, glite-put, glite-rm   on LFN and GUID
  - glite-IO API - C

- **Logical Namespace Management**
  - glite-catalog-* commands (like ls, create, rename, ..)
  - Fireman API - C, C++, Java, Perl
  - POOL File Catalog API (GliteCatalog implementation)

- **Transfer and Replication**
  - glite-transfer-* commands (submit, status, cancel, ..)
  - FPS API - C, C++, Java, Perl

Enabling Grids for E-sciencE

- **FiReMan Catalog**

  – Release 1: Single Central deployment model only

  – Release 2: Distributed catalog according to design using Java Messaging Services to propagate updates between catalog instances

- **Storage Index**

  – Already in Release 1

  – Main interaction point with Workload Management

- **Metadata Catalog**

  – Release 1: Base Implemented by FiReMan

  – Also a standalone service, single central instance

  – Release 2: distribution using a messaging infrastructure