

# The AMGA Metadata Service



**Antonio Calanducci**  
National Institute of Nuclear Physics  
INFN Catania  
EGEE NA3 Training & Dissemination  
CYCLOPS Second Training Workshop  
Chania (Crete), 05th-07th May 2008



# Contents

- Metadata services background and motivation
- Architecture and features of AMGA
- Grid DB Access with AMGA
- Use cases



# Why Grid needs Metadata?



# Why Grid needs Metadata?

- Grids allow to save **millions of files** spread over several storage sites.



# Why Grid needs Metadata?

- Grids allow to save **millions of files** spread over several storage sites.
- Users and applications need an efficient mechanism
  - to **describe** files
  - to **locate** files based on their contents





# Why Grid needs Metadata?

- Grids allow to save **millions of files** spread over several storage sites.
- Users and applications need an efficient mechanism
  - to **describe** files
  - to **locate** files based on their contents
- This is achieved by
  - associating descriptive attributes to files
    - Metadata is **data about data**
  - answering user queries against the associated information



# Basic Metadata Concept





# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them





# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair



# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)



# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)
  - Name/Key – The name of the attribute



# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)
  - Name/Key – The name of the attribute
  - Value - Value of an entry's attribute





# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)
  - Name/Key – The name of the attribute
  - Value - Value of an entry's attribute
- **Schema** – A set of attributes







# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)
  - Name/Key – The name of the attribute
  - Value - Value of an entry's attribute
- **Schema** – A set of attributes
- **Collection** – A set of entries associated with a schema





# Basic Metadata Concept

- **Entries** – Representation of real world entities which we are attaching metadata to for describing them
- **Attribute** – key/value pair
  - Type – The type (int, float, string,...)
  - Name/Key – The name of the attribute
  - Value - Value of an entry's attribute
- **Schema** – A set of attributes
- **Collection** – A set of entries associated with a schema
- **Metadata** - List of attributes (including their values) associated with entries



# Analogies





# Analogies

- Analogy to the RDBMS world:
  - schema  $\leftrightarrow$  table schema
  - collection  $\leftrightarrow$  db table
  - attribute  $\leftrightarrow$  schema column
  - entry  $\leftrightarrow$  table row/record





# Analogies

- Analogy to the RDBMS world:
  - schema  $\leftrightarrow$  table schema
  - collection  $\leftrightarrow$  db table
  - attribute  $\leftrightarrow$  schema column
  - entry  $\leftrightarrow$  table row/record





# Analogies

- Analogy to the RDBMS world:
  - schema  $\leftrightarrow$  table schema
  - collection  $\leftrightarrow$  db table
  - attribute  $\leftrightarrow$  schema column
  - entry  $\leftrightarrow$  table row/record
- Analogy to a file system:
  - **Collection**  $\leftrightarrow$  **Directory**
  - **Entry**  $\leftrightarrow$  **File**





# Example: Movie Trailers



# Example: Movie Trailers

- Movie trailers files (**entries**) saved on Grid Storage Elements and registered into File Catalogue



# Example: Movie Trailers

- Movie trailers files (**entries**) saved on Grid Storage Elements and registered into File Catalogue
- We want to add **metadata** to describe movie content.





# Example: Movie Trailers

- Movie trailers files (**entries**) saved on Grid Storage Elements and registered into File Catalogue
- We want to add **metadata** to describe movie content.
- A possible **schema**:
  - Title -- varchar
  - Runtime -- int
  - Cast -- varchar
  - LFN -- varchar





# Example: Movie Trailers

- Movie trailers files (**entries**) saved on Grid Storage Elements and registered into File Catalogue
- We want to add **metadata** to describe movie content.
- A possible **schema**:
  - Title -- varchar
  - Runtime -- int
  - Cast -- varchar
  - LFN -- varchar
- A metadata catalogue will be the repository of the movies' metadata and will allow to find movies satisfying users' queries



# Trailer's example

## *Entry names*

<i>Entry names</i>	Title	Ru	Cast	LFN
8c3315c1-811f-4823-a778-60a203439689	My Best Friend's wedding	80	Julia Roberts	lfn:/grid/gilda/movies/mybfwed.avi
51a18b7a-fd21-4b2c-aa74-4c53ee64846a	Spider-man 2	120	Kirsten Dunst	lfn:/grid/gilda/movies/spiderman2.avi
401e6df4-c1be-4822-958c-ce3eb5c54fcb	The God Father	113	Al pacino	lfn:/grid/gilda/movies/godfather.avi





# Trailer's example

Attribute

*Entry names*

	Title	Ru	Cast	LFN
8c3315c1-811f-4823-a778-60a203439689	My Best Friend's wedding	80	Julia Roberts	lfn:/grid/gilda/movies/mybfwed.avi
51a18b7a-fd21-4b2c-aa74-4c53ee64846a	Spider-man 2	120	Kirsten Dunst	lfn:/grid/gilda/movies/spiderman2.avi
401e6df4-c1be-4822-958c-ce3eb5c54fcb	The God Father	113	Al pacino	lfn:/grid/gilda/movies/godfather.avi





# Trailer's example

Schema

Attribute

*Entry names*

	Title	Ru	Cast	LFN
8c3315c1-811f-4823-a778-60a203439689	My Best Friend's wedding	80	Julia Roberts	lfn:/grid/gilda/movies/mybfwed.avi
51a18b7a-fd21-4b2c-aa74-4c53ee64846a	Spider-man 2	120	Kirsten Dunst	lfn:/grid/gilda/movies/spiderman2.avi
401e6df4-c1be-4822-958c-ce3eb5c54fcb	The God Father	113	Al pacino	lfn:/grid/gilda/movies/godfather.avi



# Trailer's example

Schema

Attribute

*Entry names*

Title

Ru

Cast

LFN

8c3315c1-811f-4823-a778-60a203439689

My Best Friend's  
wedding

80

Julia  
Roberts

lfn:/grid/gilda/movies/  
mybfwed.avi

51a18b7a-fd21-4b2c-aa74-4c53ee64846a

Spider-man 2

120

Kirsten  
Dunst

lfn:/grid/gilda/movies/  
spiderman2.avi

401e6df4-c1be-4822-958c-ce3eb5c54fcb

The God Father

113

Al pacino

lfn:/grid/gilda/movies/  
godfather.avi

Entries





# Trailer's example

Schema

Attribute

*Entry names*

Title

Ru

Cast

LFN

8c3315c1-811f-4823-a778-60a203439689	My Best Friend's wedding	80	Julia Roberts	lfn:/grid/gilda/movies/mybfwed.avi
51a18b7a-fd21-4b2c-aa74-4c53ee64846a	Spider-man 2	120	Kirsten Dunst	lfn:/grid/gilda/movies/spiderman2.avi
401e6df4-c1be-4822-958c-ce3eb5c54fcb	The God Father	113	Al pacino	lfn:/grid/gilda/movies/godfather.avi



Collection

Entries



# Example

```
Query> selectattr /trailers:Title Runtime FILE 'Runtime > 80'
```

```
>> Amelie of Montmartre
>> 122
>> 004405ac-da9a-1417-92db-c1ced08dbeef
>> American Pie 2
>> 108
>> 006d56b4-d7d1-1417-8417-c1ced08dbeef
>> Batman Begins
>> 141
>> 0072f510-db33-1417-b12e-c1ced08dbeef
>> The Fast and The Furious
>> 106
>> 00737e72-d8cb-1417-871f-c1ced08dbeef
>> Madagascar
>> 86
>> 0069b608-d95c-1417-9fd1-c1ced08dbeef
>> The Matrix
```

```
Query> ls
```

```
>> 004405ac-da9a-1417-92db-c1ced08dbeef
>> 006d56b4-d7d1-1417-8417-c1ced08dbeef
>> 0072f510-db33-1417-b12e-c1ced08dbeef
>> 00737e72-d8cb-1417-871f-c1ced08dbeef
>> 0069b608-d95c-1417-9fd1-c1ced08dbeef
>> 0010bf6c-d9cc-1417-a38c-c1ced08dbeef
>> 002e3966-d877-1417-8b9c-c1ced08dbeef
```

```
Query> listattr /trailers
```

```
>> Title
>> varchar(200)
>> Runtime
>> int
>> Country
>> varchar(25)
>> ReleaseDate
>> int
>> Director
>> varchar(80)
>> PlotOutline
>> text
>> Cast
>> varchar(2048)
>> Genre
>> varchar(100)
>> Image
>> text
```



# Metadata on the Grid





# Metadata on the Grid

- Information about files -- but not only!







# Metadata on the Grid

- Information about files -- but not only!
- metadata can **describe** any grid entity/object
  - ex: JobIDs - add logging information to your jobs





# Metadata on the Grid

- Information about files -- but not only!
- metadata can **describe** any grid entity/object
  - ex: JobIDs - add logging information to your jobs
- **monitoring** of running applications:
  - ex: ongoing results from running jobs can be published on the metadata server



# Metadata on the Grid

- Information about files -- but not only!
- metadata can **describe** any grid entity/object
  - ex: JobIDs - add logging information to your jobs
- **monitoring** of running applications:
  - ex: ongoing results from running jobs can be published on the metadata server
- **information exchanging** among grid peers
  - ex: producers/consumers job collections: master jobs produce data to be analyzed; slave jobs query the metadata server to retrieve input to “consume”





# Metadata on the Grid

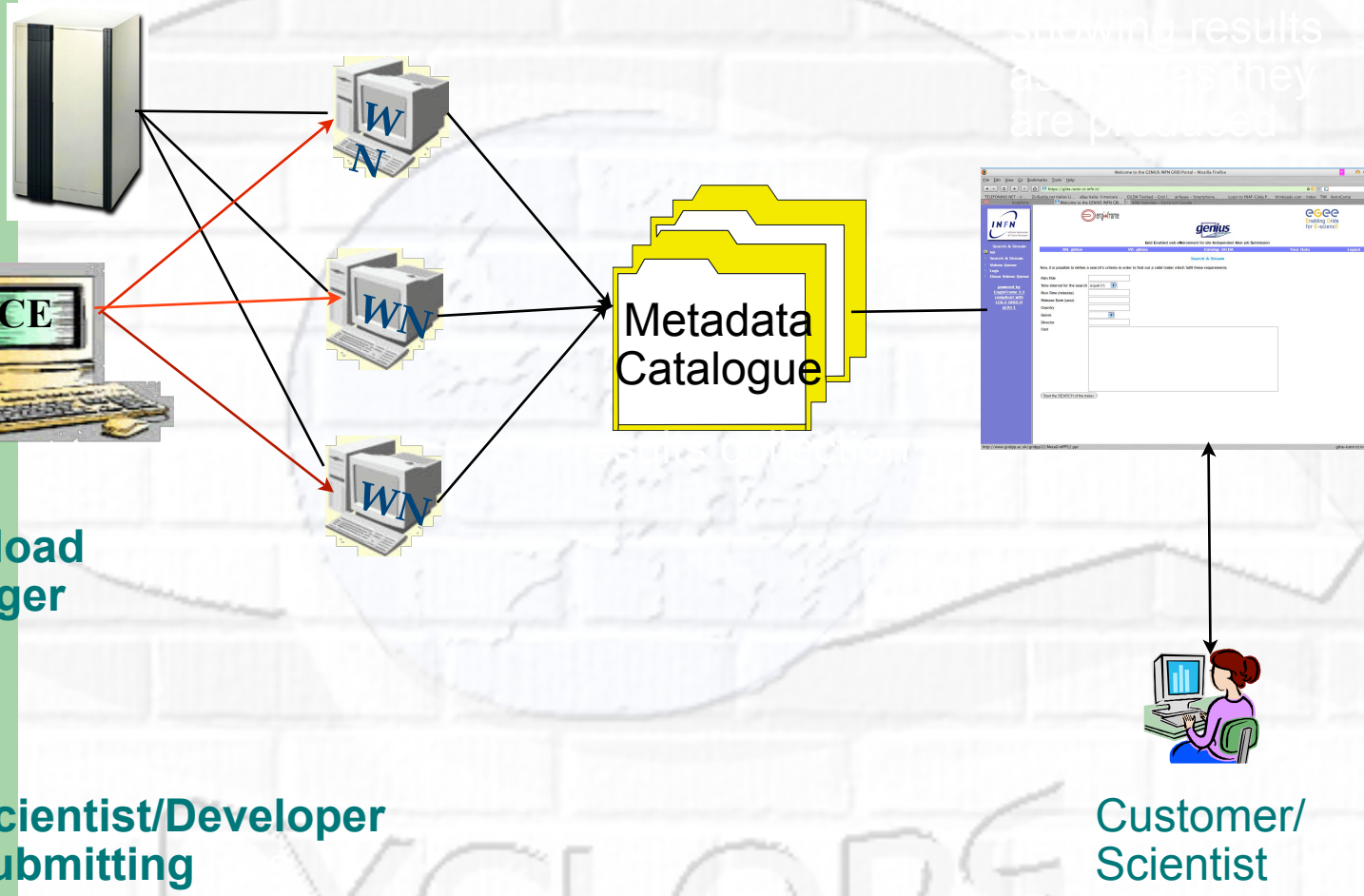
- Information about files -- but not only!
- metadata can **describe** any grid entity/object
  - ex: JobIDs - add logging information to your jobs
- **monitoring** of running applications:
  - ex: ongoing results from running jobs can be published on the metadata server
- **information exchanging** among grid peers
  - ex: producers/consumers job collections: master jobs produce data to be analyzed; slave jobs query the metadata server to retrieve input to “consume”
- Simplified **DB access** on the grid
  - Grid applications that needs structured data can model their data schemas as metadata







# Monitoring of running application





# Use a Metadata services to exchange data among running jobs



## Use a Metadata services to exchange data among running jobs

- Suppose we have two sets of jobs:
  - **Producers**: they generate a file, store on a SE, register it onto the LFC File Catalogue assigning a LFN
  - **Consumers**: they will take a LFN, download the file and elaborate it



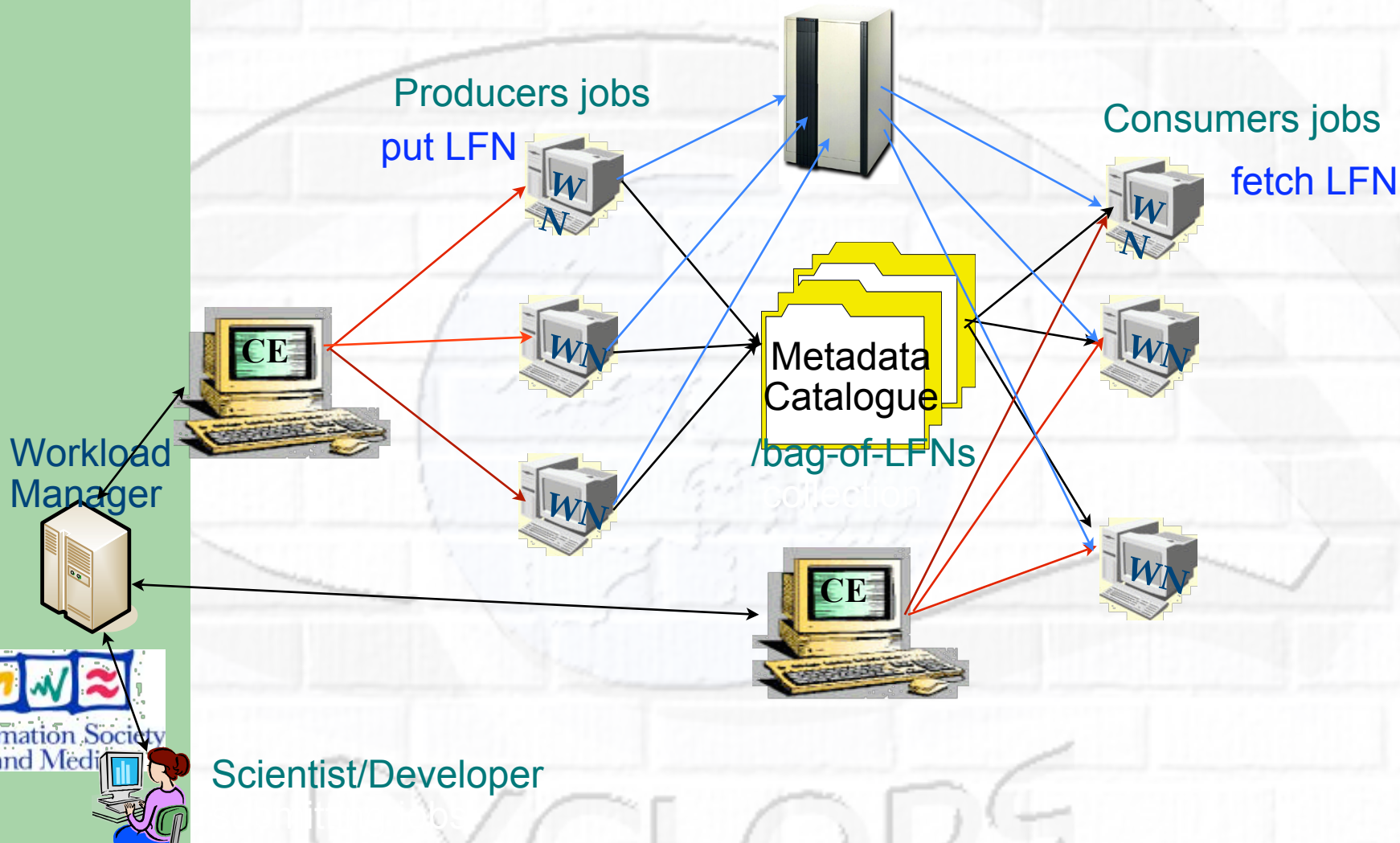
## Use a Metadata services to exchange data among running jobs

- Suppose we have two sets of jobs:
  - **Producers**: they generate a file, store on a SE, register it onto the LFC File Catalogue assigning a LFN
  - **Consumers**: they will take a LFN, download the file and elaborate it
- A Metadata collection can be used to share the information generated by the **Producers**; it could act as a “bag-of-LFNs” (bag-of-task model) from which **Consumers** can fetch file for further elaboration





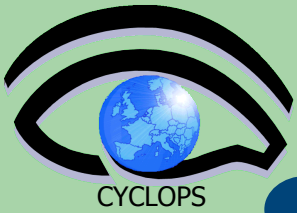
# Information exchanging among grid peers





# The AMGA Metadata Catalogue





# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware



# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: Arda Metadata Grid Application







# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.



# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.
- Designed with **scalability** in mind in order to deal with **large number of entries**



# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.
- Designed with **scalability** in mind in order to deal with **large number of entries**
- **Grid security** is provided to grant **different access levels** to different users.





# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.
- Designed with **scalability** in mind in order to deal with **large number of entries**
- **Grid security** is provided to grant **different access levels** to different users.
- **Flexible** with support to dynamic schemas in order to serve several application domains



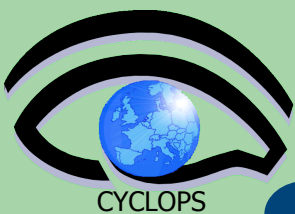




# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.
- Designed with **scalability** in mind in order to deal with **large number of entries**
- **Grid security** is provided to grant **different access levels** to different users.
- **Flexible** with support to dynamic schemas in order to serve several application domains
- Allow **hierarchical metadata** schemas





# The AMGA Metadata Catalogue

- Official metadata service for the gLite middleware
- AMGA: **Arda Metadata Grid Application**
- Provide a **complete** but **simple interface**, in order to make all users able to use it easily.
- Designed with **scalability** in mind in order to deal with **large number of entries**
- **Grid security** is provided to grant **different access levels** to different users.
- **Flexible** with support to dynamic schemas in order to serve several application domains
- Allow **hierarchical metadata** schemas





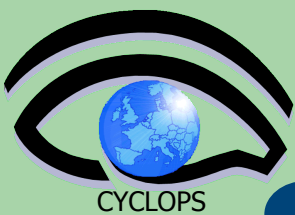
# AMGA Features



# AMGA Features

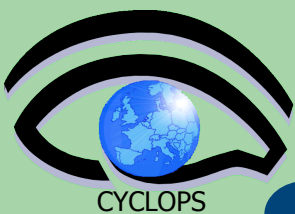
- **Dynamic** Schemas
  - Schemas can be modified at runtime by client
    - Create, delete schemas
    - Add, remove attributes





# AMGA Features

- **Dynamic** Schemas
  - Schemas can be modified at runtime by client
    - Create, delete schemas
    - Add, remove attributes
- AMGA collections are **hierarchical** organized
  - Collections can contain sub-collections
  - Sub-collections can inherit/extend parent collection's schema



# AMGA Features

- **Dynamic Schemas**
  - Schemas can be modified at runtime by client
    - Create, delete schemas
    - Add, remove attributes
- AMGA collections are **hierarchical** organized
  - Collections can contain sub-collections
  - Sub-collections can inherit/extend parent collection's schema
- **Flexible Queries**
  - SQL-like query language
  - Different join type (inner, outer, left, right) between schemas are provided



```
selectattr /gLibrary:FileName /gLAudio:Author /gLAudio:Album  
'/gLibrary:FILE=/gLAudio:FILE and like(/gLibrary:FileName, "%.mp3") '
```



# AMGA Features

- **Dynamic Schemas**
  - Schemas can be modified at runtime by client
    - Create, delete schemas
    - Add, remove attributes
- AMGA collections are **hierarchical** organized
  - Collections can contain sub-collections
  - Sub-collections can inherit/extend parent collection's schema
- **Flexible Queries**
  - SQL-like query language
  - Different join type (inner, outer, left, right) between schemas are provided

```
selectattr /gLibrary:FileName /gLAudio:Author /gLAudio:Album  
'/gLibrary:FILE=/gLAudio:FILE and like(/gLibrary:FileName, "%.mp3") '
```

► Support for **Views, Constraints, Indexes**





# AMGA Security







# AMGA Security

- Unix style permissions - users and groups



# AMGA Security

- Unix style permissions - users and groups
- **ACLs** – Per-collection or per-entry.



# AMGA Security

- Unix style permissions - users and groups
- **ACLs** – Per-collection or per-entry.
- Secure client/server connections – SSL



# AMGA Security

- Unix style permissions - users and groups
- **ACLs** – Per-collection or per-entry.
- Secure client/server connections – SSL
- Client Authentication based on
  - Username/password
  - General X509 certificates (DN based)
  - Grid-proxy certificates (DN based)





# AMGA Security

- Unix style permissions - users and groups
- **ACLs** – Per-collection or per-entry.
- Secure client/server connections – SSL
- Client Authentication based on
  - Username/password
  - General X509 certificates (DN based)
  - Grid-proxy certificates (DN based)
- VOMS support:
  - VO attribute maps to defined AMGA user
  - VOMS Role maps to defined AMGA user
  - VOMS Group maps to defined AMGA group





# AMGA Implementation

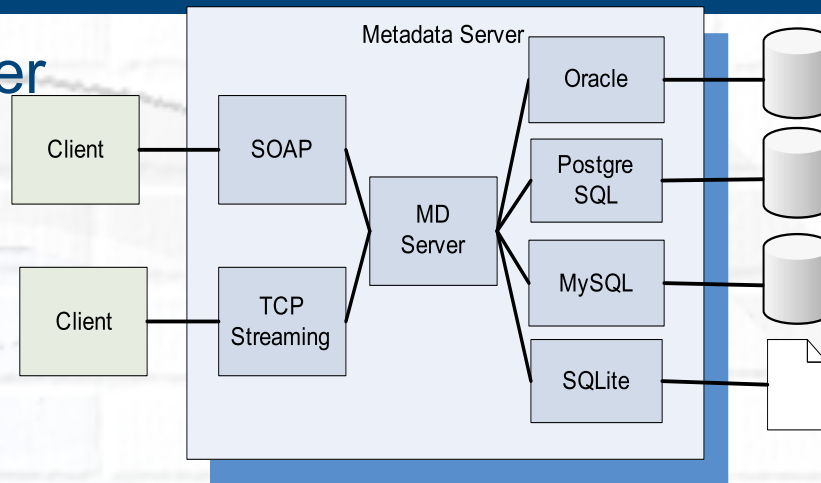
## ✦ C++ multiprocess server

### - Backends

- Oracle, MySQL 4/5, PostgreSQL, SQLite

### - Front Ends

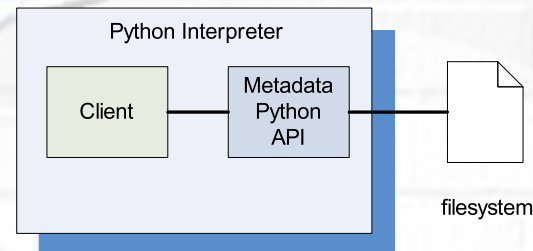
- **TCP text streaming**
  - High performance
  - Client API for C++, Java, Python, Perl, PHP
- **SOAP (web services)**
  - Interoperability
  - Scalability



- AMGA server runs on SLC3/4, Fedora Core, Gentoo, Debian

## ✦ Standalone Python Library implementation

- Data stored on file system





# AMGA Datatypes

	PostgreSQL	MySQL	Oracle	SQLite	Python
<b>int</b>	integer	int	number(38)	int	int
<b>float</b>	double precision	double precision	float	float	float
<b>varchar(n)</b>	character varying(n)	character varying(n)	varchar2(n)	varchar(n)	string
<b>timestamp</b>	timestamp w/o TZ	datetime	timestamp(6)	unsupported	time (unsupp.)
<b>text</b>	text	text	long	text	string
<b>numeric(p,s)</b>	numeric(p,s)	numeric(p,s)	numeric(p,s)	numeric(p,s)	float



# AMGA Datatypes

	PostgreSQL	MySQL	Oracle	SQLite	Python
<b>int</b>	integer	int	number(38)	int	int
<b>float</b>	double precision	double precision	float	float	float
<b>varchar(n)</b>	character varying(n)	character varying(n)	varchar2(n)	varchar(n)	string
<b>timestamp</b>	timestamp w/o TZ	datetime	timestamp(6)	unsupported	time (unsupp.)
<b>text</b>	text	text	long	text	string
<b>numeric(p,s)</b>	numeric(p,s)	numeric(p,s)	numeric(p,s)	numeric(p,s)	float

- ▶ Using the above datatypes you are sure that your metadata can be easily moved to all supported back-ends





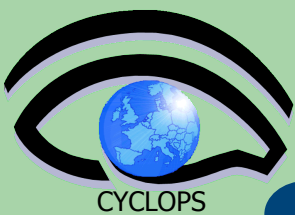
# AMGA Datatypes

	PostgreSQL	MySQL	Oracle	SQLite	Python
<b>int</b>	integer	int	number(38)	int	int
<b>float</b>	double precision	double precision	float	float	float
<b>varchar(n)</b>	character varying(n)	character varying(n)	varchar2(n)	varchar(n)	string
<b>timestamp</b>	timestamp w/o TZ	datetime	timestamp(6)	unsupported	time (unsupp.)
<b>text</b>	text	text	long	text	string
<b>numeric(p,s)</b>	numeric(p,s)	numeric(p,s)	numeric(p,s)	numeric(p,s)	float

- ▶ Using the above datatypes you are sure that your metadata can be easily moved to all supported back-ends
- ▶ If you do not care about DB portability, you can use, in principle, as entry attribute type ALL the datatypes supported by the back-end, even the more esoteric ones (PostgreSQL Network Address type or Geometric ones)



# Accessing AMGA from UI/WNs



# Accessing AMGA from UI/WNs

- TCP Streaming Front-end
  - **mdcli & mdclient** CLI and C++ API (md\_cli.h, MD\_Client.h)
  - **Java** Client API and command line mdjavaclient.sh & mdjavacli.sh (also under Windows !!)
  - **Python** and Perl Client API
  - **PHP Client API – NEW**
    - developed by the GILDA team – INFN CT
  - **AMGA Web Interface (AMGA WI) ---NEW**
    - Developed totally by the GILDA team – INFN CT
    - Based on JAVA AMGA Standard APIs
    - Web Application using standard as JSP Custom Tags, Servlet



# Accessing AMGA from UI/WNs

- TCP Streaming Front-end
  - **mdcli** & **mdclient** CLI and C++ API (md\_cli.h, MD\_Client.h)
  - **Java** Client API and command line mdjavaclient.sh & mdjavacli.sh (also under Windows !!)
  - **Python** and Perl Client API
  - **PHP Client API – NEW**
    - developed by the GILDA team – INFN CT
  - **AMGA Web Interface (AMGA WI) ---NEW**
    - Developed totally by the GILDA team – INFN CT
    - Based on JAVA AMGA Standard APIs
    - Web Application using standard as JSP Custom Tags, Servlet

## SOAP Frontend (WSDL)

- C++ gSOAP
- AXIS (Java)
- ZSI (Python)







# AMGA Web Interface




Provide your user ID to log on AMGA Server

User ID:

MESSAGE: Proxy file has not been found, upload it.



# Collection Management



## Amga web interface

[Logout](#) [User](#) [Proxy](#)

AMGA Web Interface

Menu

- Collections Management
- Groups Management
- Users Management
- User Credentials

Partners



Team







Powered by

- IR&T engineering s.r.l.





















Support

### Collection Management

Current Collection:    Parent

[Show sub collections](#) [Show entries](#)      

#### Entries List

/grid/trigrig/adat/Plutarco-Vitae-001v.jpg		
/grid/trigrig/adat/Plutarco-Vitae-043r.jpg		
/grid/trigrig/adat/Plutarco-Vitae-043v.jpg		
/grid/trigrig/adat/Plutarco-Vitae-001r.jpg		
/grid/trigrig/adat/Plutarco-Vitae-225v.jpg		
/grid/trigrig/adat/Plinio-Nataralis.Historia-001r.jpg		
/grid/trigrig/adat/Plinio-Nataralis.Historia-001v.jpg		
/grid/trigrig/adat/Plinio-Nataralis.Historia-003r.jpg		
/grid/trigrig/adat/Plinio-Nataralis.Historia-003v.jpg		
/grid/trigrig/adat/Agostino-TF1-007v.inn		

Modify Schema Instance

Delete entry



# Existing DB access with AMGA





# Existing DB access with AMGA

- Since AMGA 1.2.10, a new import feature allow to access existing DB table





# Existing DB access with AMGA

- Since AMGA 1.2.10, a new **import** feature allow to access existing DB table
- Once imported into AMGA the tables from or more DBs you want to access through AMGA, you can exploit many of the features brought to you by AMGA for your existing tables



# Existing DB access with AMGA

- Since AMGA 1.2.10, a new **import** feature allow to access existing DB table
- Once imported into AMGA the tables from or more DBs you want to access through AMGA, you can exploit many of the features brought to you by AMGA for your existing tables
- Advantages: your db tables can be accessed by grid users/applications, using grid authentication (VOMS proxies)/authorization with ACLs





# Early adopters of AMGA



# Early adopters of AMGA

- LHCb-bookkeeping
  - Migrated bookkeeping metadata to ARDA prototype
    - 20M entries, 15 GB
    - Large amount of static metadata
  - Feedback valuable in improving interface and fixing bugs
  - AMGA showing good scalability





# Early adopters of AMGA

- LHCb-bookkeeping
  - Migrated bookkeeping metadata to ARDA prototype
    - 20M entries, 15 GB
    - Large amount of static metadata
  - Feedback valuable in improving interface and fixing bugs
  - AMGA showing good scalability
- Ganga
  - Job management system
    - Developed jointly by Atlas and LHCb
  - Uses AMGA for storing information about job status
    - Small amount of highly dynamic metadata



# Biomed - MDM

- **Medical Data Manager – MDM**

- Store and access medical images and associated metadata on the Grid
- Built on top of gLite 1.5 data management system
- Demonstrated at last EGEE conference (October 05, Pisa)

- **Strong security requirements**

- Patient data is sensitive
- Data must be encrypted
- Metadata access must be restricted to authorized users

Images		
GUID	Date	Patient

Patient	
ID	Doctor

Doctor	
Name	Hospital

- **AMGA used as metadata server**

- Demonstrates authentication and encrypted access
- Used as a simplified DB

More details at

- <https://uimon.cern.ch/twiki/bin/view/EGEE/DMEncryptedStorage>



# gMOD: grid Movie On Demand



# gMOD: grid Movie On Demand

- gMOD provides a Video-On-Demand service





# gMOD: grid Movie On Demand

- gMOD provides a **Video-On-Demand** service
- User chooses among a list of video and the chosen one is streamed in real time to the video client of the user's workstation



# gMOD: grid Movie On Demand

- gMOD provides a **Video-On-Demand** service
- User chooses among a list of video and the chosen one is streamed in real time to the video client of the user's workstation
- For each movie a lot of details (Title, Runtime, Country, Release Date, Genre, Director, Case, Plot Outline) are stored and users can search a particular movie querying on one or more attributes



# gMOD: grid Movie On Demand

- gMOD provides a **Video-On-Demand** service
- User chooses among a list of video and the chosen one is streamed in real time to the video client of the user's workstation
- For each movie a lot of details (Title, Runtime, Country, Release Date, Genre, Director, Case, Plot Outline) are stored and users can search a particular movie querying on one or more attributes
- Two kind of users can interact with gMOD:  
**TrailersManagers** that can administer the db of movies (uploading new ones and attaching metadata to them); **GILDA VO users (guest)** can browse, search and choose a movie to be streamed.





# gMOD under the hood





# gMOD under the hood

- Built on top of gLite services:



# gMOD under the hood

- Built on top of gLite services:
- Storage Elements, sited in different place, physically contain the movie files



# gMOD under the hood

- Built on top of gLite services:
- **Storage Elements**, sited in different place, physically contain the movie files
- **LFC**, the File Catalogue, keeps track in which Storage Element a particular movie is located



# gMOD under the hood

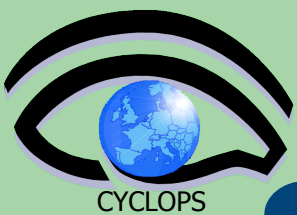
- Built on top of gLite services:
- **Storage Elements**, sited in different place, physically contain the movie files
- **LFC**, the File Catalogue, keeps track in which Storage Element a particular movie is located
- **AMGA** is the repository of the detailed information for each movie, and makes possible queries on them





# gMOD under the hood

- Built on top of gLite services:
- **Storage Elements**, sited in different place, physically contain the movie files
- **LFC**, the File Catalogue, keeps track in which Storage Element a particular movie is located
- **AMGA** is the repository of the detailed information for each movie, and makes possible queries on them
- The **Virtual Organization Membership Service (VOMS)** is used to assign the right role to the different users



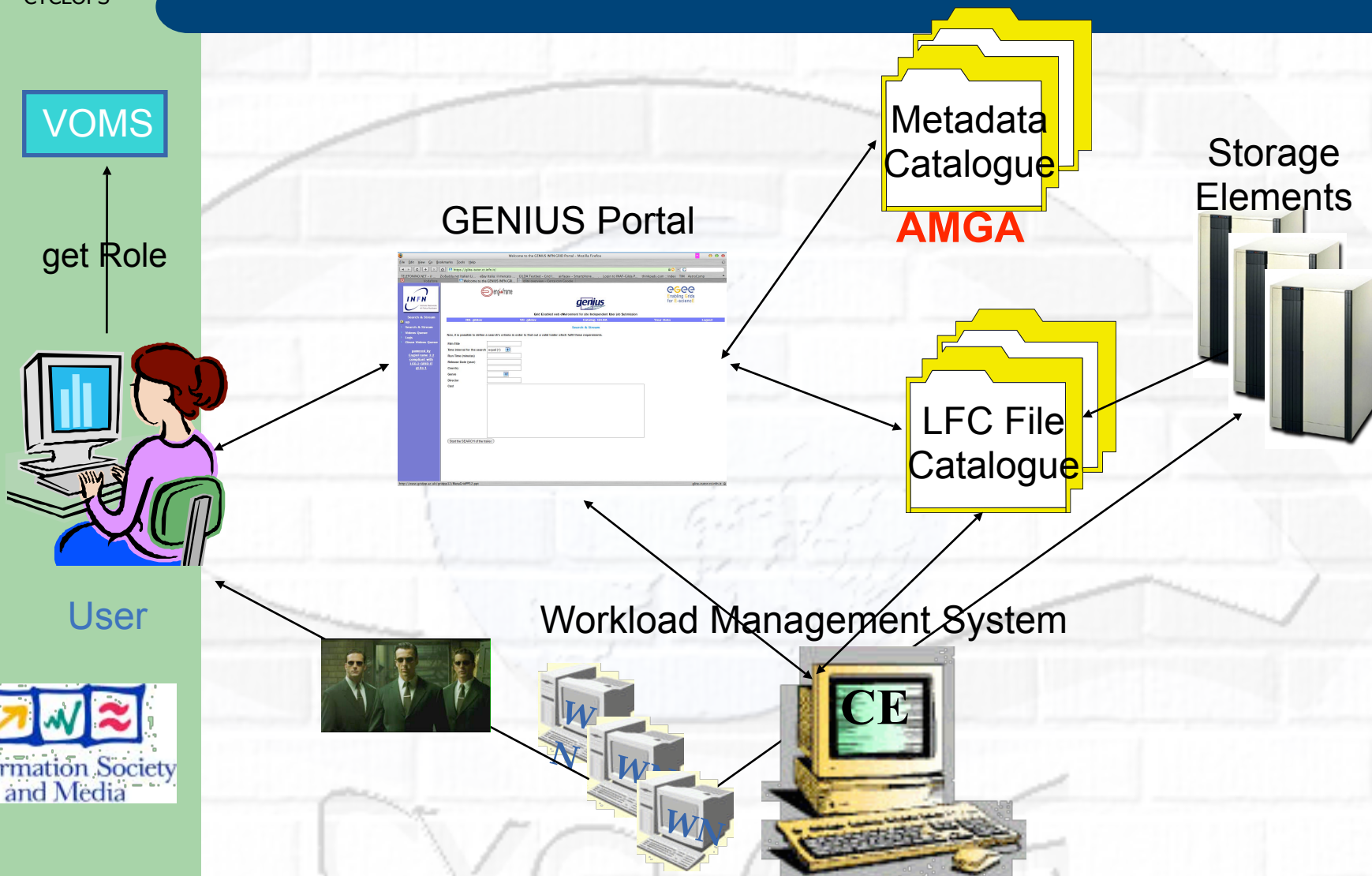
# gMOD under the hood

- Built on top of gLite services:
- **Storage Elements**, sited in different place, physically contain the movie files
- **LFC**, the File Catalogue, keeps track in which Storage Element a particular movie is located
- **AMGA** is the repository of the detailed information for each movie, and makes possible queries on them
- The **Virtual Organization Membership Service (VOMS)** is used to assign the right role to the different users
- The **Workload Management System (WMS)** is responsible to retrieve the chosen movie from the right Storage Element and stream it over the network down to the user's desktop or laptop



CYCLOPS

# gMOD interactions







# gMOD screenshot

gMOD is accesible through the Genius Portal (<https://glite-demo.ct.infn.it>)

Welcome to the GENIUS INFN GRID Portal - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://glite-tutor.ct.infn.it/

TELEFONINO.NET - il ... ZioBudda.net Italian Li... eBay Italia: il mercato ... GILDA Testbed - Grid I... airfagev - Smartphone... Login to INAF-Gilda P... thinkpads.com :: Index TIM AstroComp

Vodafone Welcome to the GENIUS INFN GRI...



gMOD Services

- Search & Stream
- Upload and Add metadata to a Trailer
- Browse, Edit or Remove Metadata
- Back home

powered by  
EnginFrame 3.2  
compliant with  
LCG-2 GRID-IT  
glite-1



Grid Enabled web eNvironment for site Independent User job Submission

RB: gildav VO: gildav Catalog: GILDA Your Data Logout

You have selected the trailer **/trailers/Shrek2.mpg**.  
These are the attributes of the trailer you have choosen.

Title	Shrek 2				
Run Time	92	Country	USA	Release Date	2004
Genre	Action	Director	William Steig		
Cast	Mike Myers .... Shrek (voice) Eddie Murphy .... Donkey (voice) Cameron Diaz .... Fiona (voice) John Cleese .... King (voice) Rupert Everett .... Prince Charming (voice) Jeremy Piven .... Puss in Boots (voice) John Goodman .... Donkey's Father (voice) Jeff Bridges .... Donkey's Mother (voice) Cody Cameron .... Pinocchio/Three (voice) Christopher Knights .... Blind Mouse (voice) David P. Smith .... Herald/Marshmallow				
Plot Outline	The film picks up right where the first movie ended... Shrek and Fiona return from their honeymoon. The only problem is that they have no idea that their daughter is now an ogre.				

LFN: /trailers/Shrek2.mpg

Created: 2005-10-13 17:23:58.000

Modified: 2005-10-13 17:23:58.000

Size: 6100996

Type: File

Expires: Never

GUID: 000f0e2e-7c0d-134e-a731-c1ced08dbeef

Created: 2005-10-13 17:24:03.000

Modified: 2005-10-13 17:24:03.000

Size: 6100996

Checksum: 00000000

Status: 0

User: /C=IT/O=GILDA/OU=Personal Certificate/L=INAF

Group: egee-group

User rights: pdrwl-gs

Group rights: pdrwl-gs

VLC media player

File Visualizza Impostazioni Audio Video Navigazione Aiuto

0:01:21 / 0:02:30 x1.00 C:\Documents and Settings\Tony\I

Done

glite-tutor.ct.infn.it





# What is gLibrary

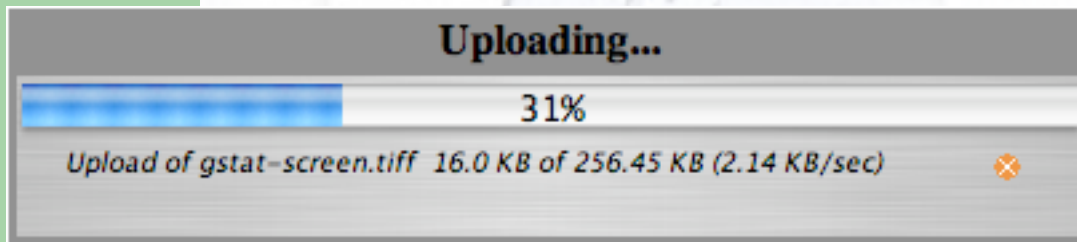
- gLibrary challenge is to offer a **multiplatform**, **flexible**, **secure** and **intuitive** system to handle digital assets on a Grid Infrastructure.
- By Digital Asset, we mean any kind of content and/or media represented as a computer file. Examples:
  - Images
  - Videos
  - Presentations
  - Office documents
  - E-mails, web pages
  - Newsletters, brochures, bulletins, sheets, templates
  - Receipts, e-books
  - ... (only the imagination can make a limit)
- It allows to **store**, **organize**, **search** and **retrieve** those assets on a Grid environment.



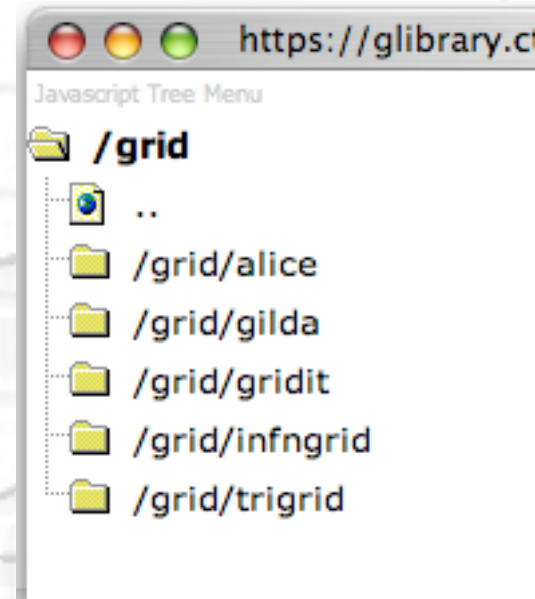


# Store assets on the Grid

- User's local assets are uploaded to one or more (as replicas) Storage Systems the user is authorized on
  - Uploads are managed through Java Applets: a direct GSIFTP copy is done from the local file to the chosen Storage Element



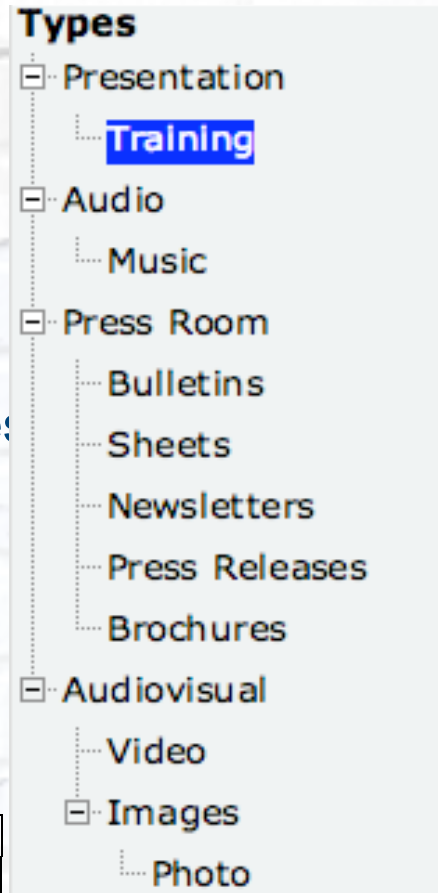
- **File already on the Grid can be managed by gLibrary too**
  - a File Catalogue browser is integrated to select existing grid files.





# Organize assets

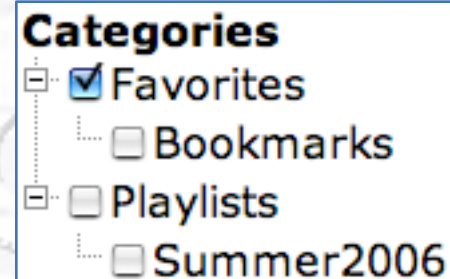
- All entries are organized according to their **type**:
  - a list of specific attributes to describe each kind of assets to be managed by the system;
  - hierarchical (child type shares parent's attributes)
  - defined by the gLibrary administrators
  - queried by users



- **Assets can be organized also by category:**
  - Group together related assets of different types;
  - Useful also to define subset of assets belonging to the same type
  - Multiple category assignment per asset

EXAMPLE OF TYPES AND ATTRIBUTES' LIST

Type	Attributes' list
Audio	Format, Bitrate, Samplerate, Time
Music	(Format, Bitrate, Samplerate, Time), Name, Artist, Album, Genre, Tracknumber, Year, Artwork, Lyric, Rating
Presentation	Format, NumOfPages
Training	(Format, NumOfPages), Title, Runtime, Speaker, Author, Subject, Event, Date, Type
(Root)	FileName, SubmissionDate, Description, Keywords, LastModificationDate, Size





# Search assets

- Assets are browsed selecting a type (or category) and selecting one or more **filters**:
  - type attributes chosen from a defined list, used to narrow the result set
- Filter application is cascading and context-sensitive: the selection of a filter value dynamically influences subsequent filter values (*“à la iTunes”* browser)
- Classic search available too

Speaker

ALL  
Annamaria Muoio  
Emidio Giorgio  
Gianni Ricciardi  
Giuseppe La Rocca  
Giuseppe Platania  
Tony Calanducci  
Valeria Ardizzone

Event

ALL  
Cyclops first workshop  
Tutorial per gli Insegnanti degli Istituti Tecnici Industriali

Subject

ALL  
general presentation

TITLE	RUNTIME	SPEAKER	SUBJECT	EVENT	AUTHOR	DATE	TYPE	SIZE	FORMAT	NUM
the GILDA t-infrastructure	30	Tony Calanducci	general presentation	Cyclops first workshop	Tony Calanducci			7406292	pdf	45





# Retrieve assets from the Grid

- User is presented with a list of asset replicas
- Download from the chosen storage element is matter of a mouse click
- Transfer handled over GridFTP with a Java Applet

FileName	GILDA t-infrastructure_cyclops.pdf
TypeID	Training
CategoryIDs	
SubmissionDate	2007-05-03 11:46:00
Description	GILDA t-infrastructure
Keywords	t-infrastructure
LastModificationDate	2007-04-10 11:14:02
Size	7406292
Encrypted	
Format	pdf
NumOfPages	45
Title	the GILDA t-infrastructure
Runtime	30
Speaker	Tony Calanducci
Author	Tony Calanducci
Subject	general presentation
Event	Cyclops first workshop
Date	
Type	
Thumb	
OWNER	tcaland
FILE	107
PERMISSIONS	rwX
GROUP_RIGHTS	r-X

#### List of replicas:

<srm://aliserv6.ct.infn.it/dpm/ct.infn.it/home/gilda/generated/2007-05-03/file2d8cc915-49e9-4040-ab1c-4a1329a2a8d6>





# Features

- Implemented as Web 2.0 application
  - AJAX and Javascript are strongly used to offer a desktop like user experience
  - Business logic implemented using PHP 5 OOP support

gLibrary

https://glibrary.ct.infn.it/glibrary/browse.php

ry JRA1 All-Hands meeting...

About Browse Upload Search Settings Sign Out

GLibrary

Login name: tcaland - Member of groups: root:glibrarymanagers

Types Categories

Javascript Tree Menu

**Types**

- [-] Presentation
  - [-] Training
- [-] Audio
  - [-] Music
- [-] Press Room
  - [-] Bulletins
  - [-] Sheets
  - [-] Newsletters
  - [-] Press Releases
  - [-] Brochures
- [-] Audiovisual
  - [-] Video
  - [-] Images
    - [-] Photo

Resolution: 1152x774 1280x960 Format: ALL License: ALL Creative Commons - Attribution

FILENAME	DESCRIPTION	KEYWORDS	SIZE	RESOLUTION	FORMAT	TAKENDATE	LICENSE		
EELALogoSet.zip	EELA logo set (vectorial format + font)	EELA logo	226693		ZIP			<a href="#">Edit</a>	<a href="#">Remove</a>
gstat-screen.tiff	GStat Screenshot	Monitoring GStat	262600	1152x774	TIFF			<a href="#">Edit</a>	<a href="#">Remove</a>
P1020804.JPG	Third EELA Workshop	Third EELA Workshop	608741	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020781.JPG	Third EELA Workshop	Third EELA Workshop	622332	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020777.JPG	Third EELA Workshop	Third EELA Workshop	617766	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020800.JPG	Third EELA Workshop	Third EELA Workshop	619262	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020811.JPG	Third EELA Workshop	Third EELA Workshop	601630	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>





# Browsing screenshot

gLibrary

https://glibrary.ct.infn.it/glibrary/browse.php

Google

About Browse Upload Search Settings Sign Out

GLibrary

Login name: **tcaland** - Member of groups: **root:glibrarymanagers**

Types Categories

Javascript Tree Menu

**Types**

- Presentation
  - Training
- Audio
  - Music
- Press Room
  - Bulletins
  - Sheets
  - Newsletters
  - Press Releases
  - Brochures
- Audiovisual
  - Video
  - Images**
    - Photo

Resolution: ALL 1152x774 1280x960

Format: ALL JPG TIFF ZIP

License: ALL Creative Commons - Attribution

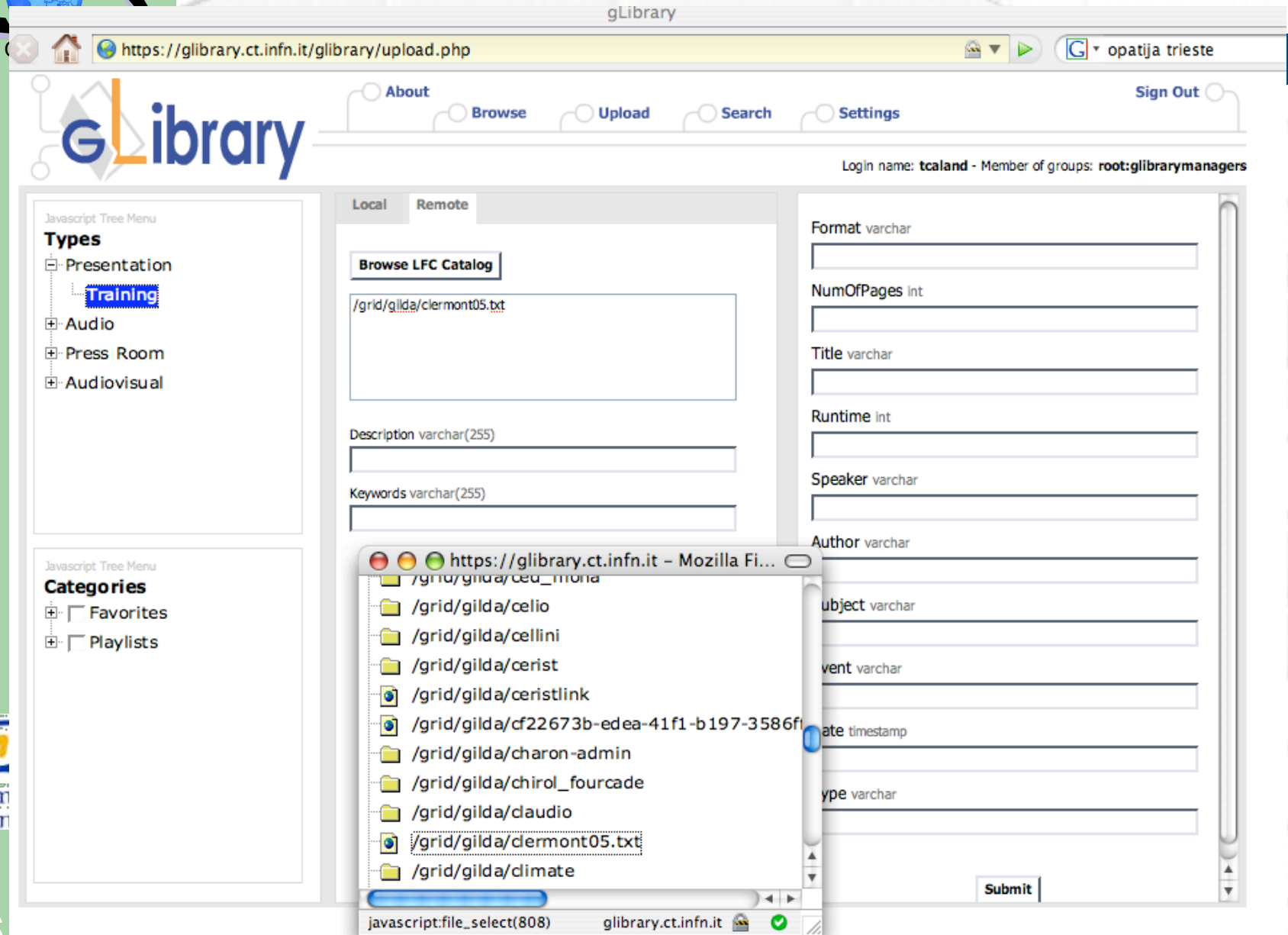
FILENAME	DESCRIPTION	KEYWORDS	SIZE	RESOLUTION	FORMAT	TAKENDATE	LICENSE		
EELALogoSet.zip	EELA logo set (vectorial format + font)	EELA logo	226693		ZIP			<a href="#">Edit</a>	<a href="#">Remove</a>
gstat-screen.tiff	GStat Screenshoot	Monitoring GStat	262600	1152x774	TIFF			<a href="#">Edit</a>	<a href="#">Remove</a>
P1020804.JPG	Third EELA Workshop	Third EELA Workshop	608741	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020781.JPG	Third EELA Workshop	Third EELA Workshop	622332	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020777.JPG	Third EELA Workshop	Third EELA Workshop	617766	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020800.JPG	Third EELA Workshop	Third EELA Workshop	619262	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
P1020811.JPG	Third EELA Workshop	Third EELA Workshop	601630	1280x960	JPG	2007-01-11 00:00:00	Creative Commons - Attribution	<a href="#">Edit</a>	<a href="#">Remove</a>
		Third EELA				2007-01-11	Creative		



Information and Communications Technology



# Upload screenshot



The screenshot shows a web browser window at <https://glibrary.ct.infn.it/glibrary/upload.php>. The page has a header with the "gLibrary" logo and navigation links: About, Browse, Upload, Search, and Settings. A "Sign Out" link is also present. Below the header, the user is logged in as "tcaland" and is a member of the "root:glibrarymanagers" group.

The main content area is divided into two columns. The left column contains a "Javascript Tree Menu" with "Types" (Presentation, Training, Audio, Press Room, Audiovisual) and "Categories" (Favorites, Playlists). The right column contains a form for uploading a file, with fields for Format, NumOfPages, Title, Runtime, Speaker, Author, Subject, Event, Date, and Type. A "Submit" button is at the bottom right.

A file selection dialog is open, showing a list of files in the directory `/grid/gilda/`. The file `/grid/gilda/clermont05.txt` is selected. The dialog also shows a search bar and a "Submit" button.

The browser's address bar shows the URL `https://glibrary.ct.infn.it/glibrary/upload.php`. The browser's status bar shows the file path `javascript:file_select(808)` and the domain `glibrary.ct.infn.it`.



CYCLOPS

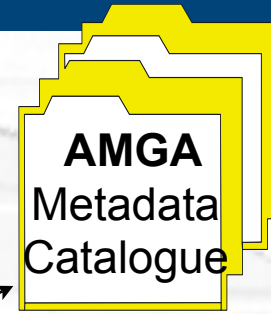
VOMS Server



# Architecture overview



**LFC File Catalogue**



**AMGA Metadata Catalogue**



**SE**



**SE**



**SE**

Login applet

glLibrary Login

Sign up here

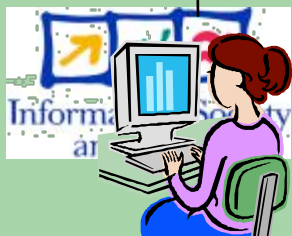
Login:

Virtual Organization:

Certificate (.p12):

Passphrase:

1. local proxy creation



37 User

2. proxy transfer over HTTPS

6. direct transfer from SE

3. get role

4. find the right asset

glLibrary

https://gllibrary.ct.infn.it/gllibrary/browse/

Search: [JRA1 All-Hands meeting...]

Types: Presentation, Training, Audio, Music, Press Room, Bulletins, Sheets, Newsletters, Press Releases, Brochures, Audiovisual, Video, Images, Photo

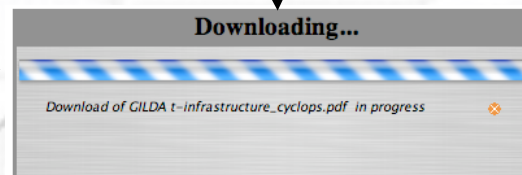
Resolution: ALL, 1152x774, 1280x960

Format: ALL, JPG, TIFF, ZIP

License: ALL, Creative Commons - Attribution

FILENAME	DESCRIPTION	KEYWORDS	SIZE	RESOLUTION	FORMAT	TAKENDATE	LICENSE
EELAlogo001.jpg	EELA logo set (vectorial format + font)		22693		ZIP		
g88t-screen.tif	G88t Screenshots	Monitoring	26200	1152x774	TIFF		
P1020804.JPG	Third EELA Workshop	Third EELA Workshop	688741	1280x960	JPG	2007-05-11 00:00:00	Creative Commons - Attribution
P1020781.JPG	Third EELA Workshop	Third EELA Workshop	622332	1280x960	JPG	2007-05-11 00:00:00	Creative Commons - Attribution
P1020777.JPG	Third EELA Workshop	Third EELA Workshop	617766	1280x960	JPG	2007-05-11 00:00:00	Creative Commons - Attribution
P1020800.JPG	Third EELA Workshop	Third EELA Workshop	610202	1280x960	JPG	2007-05-11 00:00:00	Creative Commons - Attribution
P1020811.JPG	Third EELA Workshop	Third EELA Workshop	601630	1280x960	JPG	2007-05-11 00:00:00	Creative Commons - Attribution

5. proxy retrieved over HTTPS



Upload/Download applet



# Conclusion

- AMGA – Metadata Service of gLite
  - Part of gLite 3.1
  - Useful to realize simple Relational Schemas
  - Integrated on the Grid Environment (Security)
- Replication/Federation features
- Importing existing databases
- Tests show good performance/scalability
- Already deployed by several Grid Applications
  - LHCb, ATLAS, Biomed, gMOD, gLibrary, ADAT





# References

- AMGA Web Site

<http://cern.ch/amga>

- AMGA Manual

[http://amga.web.cern.ch/amga/downloads/amga-manual\\_1\\_3\\_0.pdf](http://amga.web.cern.ch/amga/downloads/amga-manual_1_3_0.pdf)

- AMGA API Javadoc

<http://amga.web.cern.ch/amga/javadoc/index.html>

- AMGA Web Frontend

<http://gilda-forge.ct.infn.it/projects/amgawi/>

- AMGA Basic Tutorial

<https://grid.ct.infn.it/twiki/bin/view/GILDA/AMGAHandsOn>

