AMGA metadata catalogue and high level API

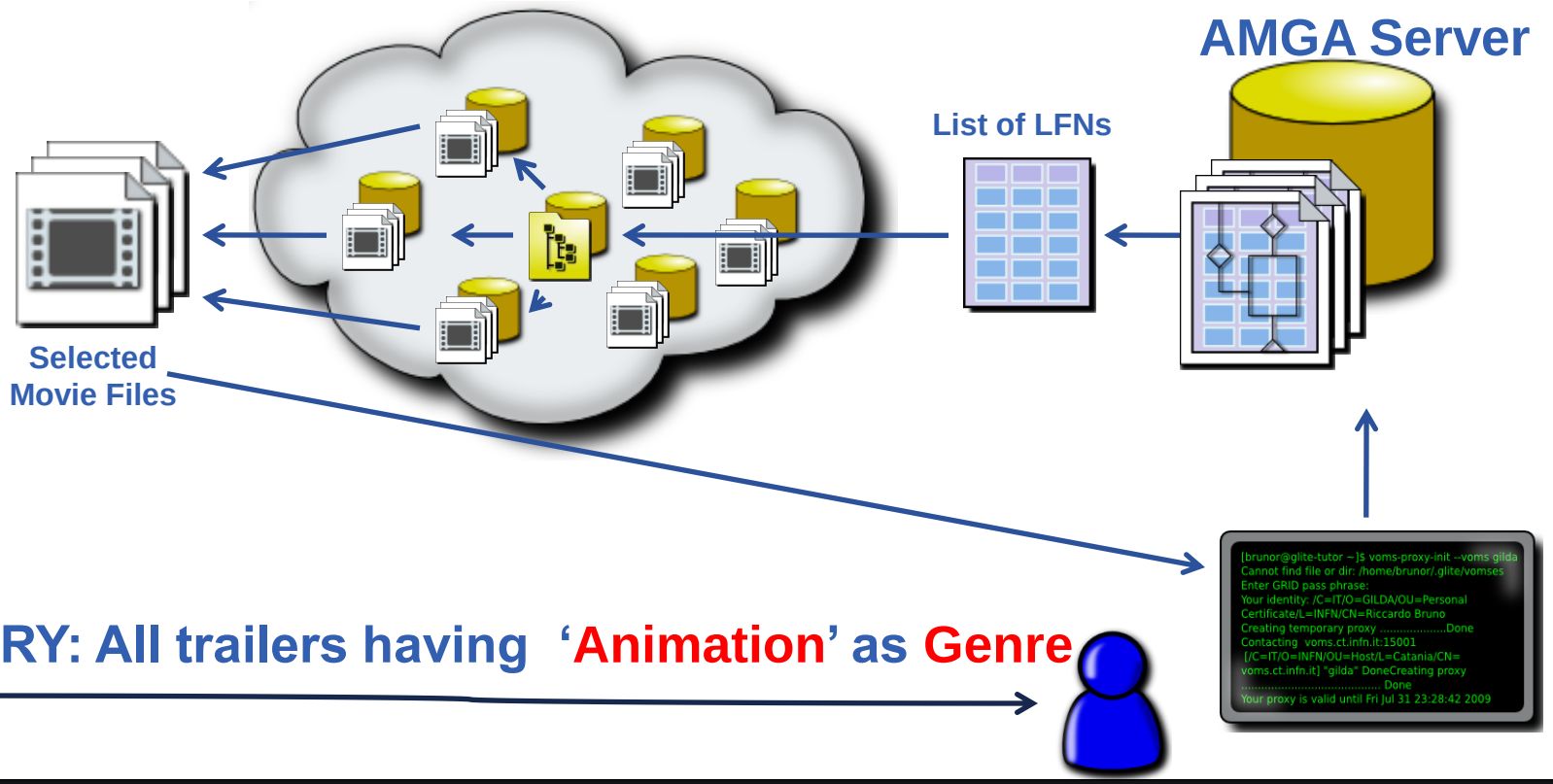
Andrea Cortellese (andrea.cortellese@ct.infn.it)
INFN Catania
Institute of High Energy Physics (IHEP)
6th - 17th September 2010

Outline

- **Introduction to AMGA**
  - What is AMGA
  - How does it work
  - Data types and terminology
  - First access on AMGA
  - The command line
- **Advanced features**
- **High level API overview**
- **Java and Php API usage**
- **User Application sample**

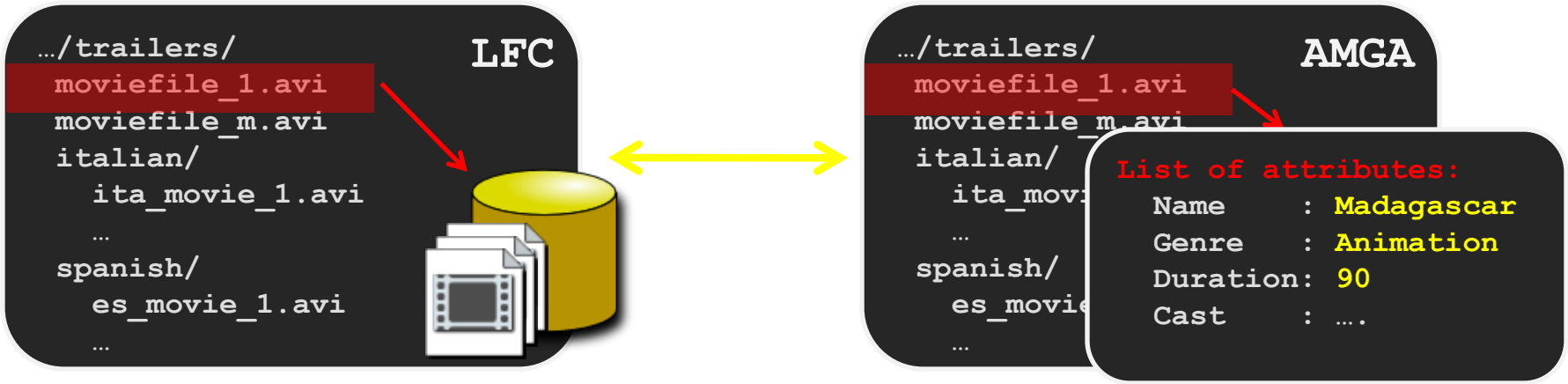
What is AMGA

- **AMGA is a service inside gLite able to store metadata**
  - Metadata is an information which describe a set of data
    - ES: Genre (Action, animation, comic) describe a set of movies
    - User can obtain a set of movies making a query on metadata



How does AMGA work

- **Across the Grid storage elements, AMGA refers to data in files and stores its related metadata, so it's strictly closed to the file organization on GRID.**
  - LFC stores files in storage elements at an higher level that hides users to know the phisical path of the file, using an LFN path.
  - AMGA use the same system, but provides more informations about each file, that are metadata details:



Data types and terminology

- **Entries**
  - List of **entities** having metadata associated
- **Attribute**
  - key **name**, key **type** pair
- **Schema**
  - Set of **attributes**
- **Collection**
  - A set of **entries** associated with a **schema**
- **Metadata**
  - List of **attributes** (including their **values**) associated with **entries**

FS Analogy

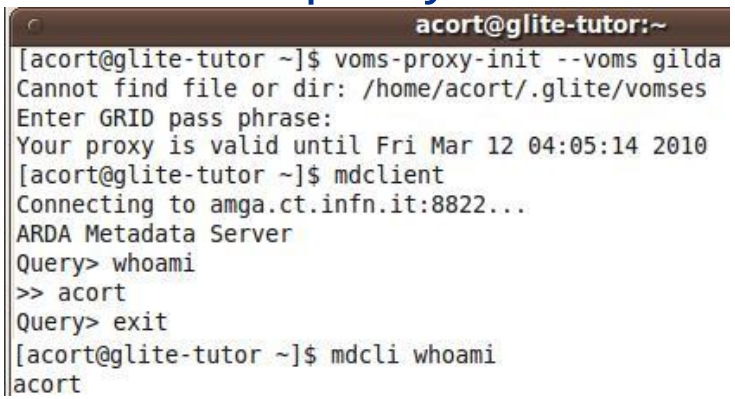
```
collection_1/
  entry_1
  entry_2
      …
collection_2/
  entry_1
  entry_2
      …
```

Integer
Char
Date
…

Entries	Title	Genre	…	LFN
60a203439689	Titanic	Drammatic	…	/grid/gilda/movies/tit.avi
4c53ee64846a	Avatar	Adventure	…	/grid/gilda/movies/av.avi
…	…	…	…	…

First access on AMGA

- **You can access across standard login or certificate**
- **You have to create your personal mdclient.config file:**
  - You need an account on the AMGA server, a valid proxy and personal certificate. Tutorial users will use amga.eela.ufrj.br srvr
  - Use /opt/glite/etc/mdclient.config model and specify:
    - HOST: amga.eela.ufrj.br
    - Port 8822
    - Your login name (NULL)
    - Certificate flag



- **By typing ‘mdclient’ command you can access AMGA console and you can create your data collection on your assigned workspace**
- **You can also use ‘mdcli’ command to execute a single command on the AMGA server.**

Command line overview

- **Create a collection**
  - `createdir <path>/<collection_name> [inherits]`
- **Associate a schema to the collection**
  - `addattr <path>/<collection_name>`
    `<attr_name> <attr_type> [<attr_name> <attr_type>] …`
- **List Attributes**
  - `listattr <path>/<collection_name>`
- **Remove Attributes**
  - `removeattr <path>/<collection_name> <attr_name>`
- **Rename Attributes**
  - `renameattr <path>/<collection_name> <attr_name>`
- **Add entries and attribute values**
  - `addentry <path>/<entry_name> <attr_name> <attr_value>`
    `[<attr_name> <attr_value>] …`
- **Set an attribute value**
  - `setattr <path>/<entry_name> <attr_name> <attr_value>`
    `[<attr_name> <attr_value>] …`
- **List entries**
  - `listentries <path>/<collection_name>`

Advanced Features

- **In order to get back data from AMGA catalogue, is possible to use selectattr command or SQL-like query:**

  Query> selectattr .:FILE .:id .:temp .:hum .:pressure 'id=1'

  Query> SELECT id, temp, hum, pressure FROM . WHERE id=1

  >> 1 ⬅ FILE attribute can be used as DB field instead of ID

  >> 1

  >> 0.5

  >> 34.4

  >> 23.4

  On AMGA available file system notation and commands

- **To fill automatically the DB you can use a shell script:**

  mdcli "addattr $1/meteo/locations code varchar(4)"

  mdcli "addattr $1/meteo/locations name varchar(20)"

  mdcli "addattr $1/meteo/locations latitude varchar(10)"

  mdcli "addattr $1/meteo/locations longitude varchar(10)"

  mdcli "addattr $1/meteo/locations height int"

  php load-parser.php "locations.txt“ "$1/meteo/locations” DEMO

High level API overview

- **To avoid low level catalogue programming, is possible to make everything using high-level languages, across specific API:**
  - 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

- **We will see in detail java and php API usage, followed by practical examples and then a full working sample made using this material**

Java API

- **Java APIs can be used in two ways:**
  - Sending commands directly to the server using the low-level API
    - arda.md.javaclient.MDServerConnection
    - It does not understand the semantics of the commands and does not parse server response into suitable structures
    - But better control on the connection to the server (you can easily abort a query for example)
  - High-level interface (which use also the low level one) across:
    - arda.md.javaclient.MDClient class
    - Provide suitable structures to handle server response
    - Not all the commands are implemented (instead use direct connection)
- **References:**
  - http://amga.web.cern.ch/amga/api_java13.html
  - http://amga.web.cern.ch/amga/downloads/ardamd-apijava-1.2.6RC1.tar.gz

Low level API access

```
import java.io.IOException;
import arda.md.javaclient.*;
public class DirectAMGA {
public static void main(String[] args) throws IOException
{
// Loads default configuration from mdclientjava.config and connects to server
MDServerConnection serverConn = new MDServerConnection(
MDServerConnectionContext.loadDefaultConfiguration());
try {
    serverConn.execute("pwd");
    while (!serverConn.eot()) {
        String row = serverConn.fetchRow();
        System.out.println(">" + row);
    }
} catch (CommandException e) {
    System.out.println("Error executing command: " + e.getMessage());
}
}
```

Middle level API access

```
import arda.md.javaclient.*;
import java.util.Iterator;
public class MDJavaAPI {
public static String TEST_DIR = "/test";
public static void main(String[] args) throws Exception {
/*Alternative configuration, using mdjavaclient.config file:
//setting MDServerConnectionContext
MDServerConnectionContext mdConContext = new
    MDServerConnectionContext();
mdConContext.setHost("amga.gs.ba.infn.it");
mdConContext.setPort(8822);
mdConContext.setCurrentDir("/");
mdConContext.setGroupMask("rwx");
mdConContext.setPermissionMask("rwx");
mdConContext.setLogin("NULL");
mdConContext.setUseSSL(true);
mdConContext.setAuthMode(MDServerConnectionContext.AUTH_GRIDPROXY);
mdConContext.setGridProxyFile("/tmp/x509up_uXXXX");
```

Middle level API access (2)

```
MDServerConnection serverConn = new MDServerConnection(mdConContext);
MDClient mdClient = new MDClient(serverConn);
System.out.println("Listing attribues of " + TEST_DIR);
try {
AttributeDefList attrs = mdClient.listAttr(TEST_DIR);
Iterator<AttributeDef> it = attrs.iterator();
System.out.println("Result: ");
while(it.hasNext())
{
AttributeDef att = it.next();
System.out.println(" >" + att.getName() + ":" + att.getType());
}
} catch(CommandException e) {
System.out.println("Error: " + e.getMessage());
}
```

Java API usage demo

- Set up the CLASSPATH properly:
  - export CLASSPATH=$HOME/amga-api/lib/glite-amga-api-javav1.30.jar:$HOME/amga-api/lib/bcprov-jdk14-128.jar:.

- Set up properly mdjavaclient.config

(if you use loadDefaultConfiguration()):

  - Host = amga.eela.ufrj.br
  - Port = 8822
  - Login = NULL
  - AuthMode = GridProxy
  - GridProxyFile=/tmp/x509up_uXXXX
  - UseSSL = 1

- Compile and run the code:
  - javac MDJavaAPI.java
  - java MDJavaAPI

Php API

- **The right compromise to lock together a strong language like PHP with our innovative storage system**
- **Written by GILDA team on Python's API structure.**
- **Provided as a MDClient class**
  - Can be included in any php code
  - Allow AMGA direct access
  - There are methods to perform any operation
  - Data available quickly and without low level syntax

Php API features

- PHP5 Object Oriented Interface
- Ported from the C++, Python and Perl Client APIs
- Based on the MDClient class
- Can establish plain/SSL connection to AMGA server
- Provides a method for each AMGA command
- Allows a generic *execute($command)*

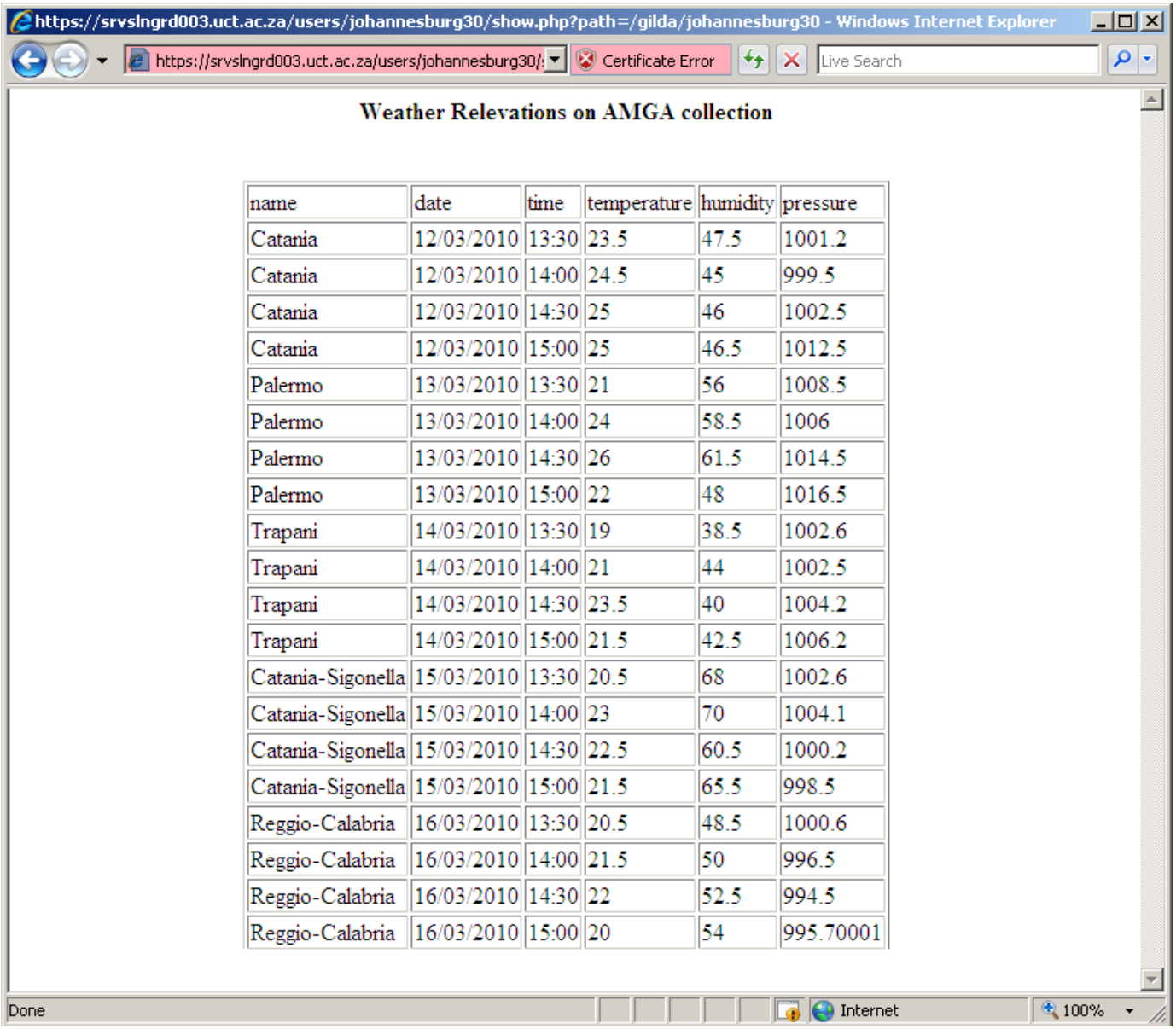
Php API methods

- **Looking inside the Mdclient class, there are:**
  - *void **getattr**(string $file, array $attributes)*
  - *void **setAttr**(string $file, string $keys, string $values)*
  - *void **__construct**(string $host, integer $port, [string $login = "anonymous"], [string $password = ""], [boolean $keepalive = true])*
  - *void **requireSSL**(string $key, string $cert, [$capath = "certificates"])*
  - *void **connect**()*
  - *void **execute**(string $command)*
  - *void **addEntry**(string $file, string $keys, string $values)*
  - *void **addAttr**(string $file, string $name, string $t)*
  - *String **listEntries**(string $pattern)*
  - *array(array(attributes), array(types)) **listAttr**(string $file)*
  - *void **selectAttr**(array $attributes, string $query)*
  - *Array **getSelectAttrEntry**()*
  - ...

Php API sample

```
Require "mdclient.php";
$certkey=trim($argv[2]);
$path=trim($argv[1]);
$capath="/etc/grid-security/certificates";
$conn = new MDClient("devslngrd004.uct.ac.za", 8822,NULL);
$conn->requireSSL($certkey,$certkey,$capath);
$conn->connect();
$conn->execute("createdir " .  $path . "/meteo/locations");
$conn->addattr($path . "/meteo/locations","code","varchar(4)");
$conn->addattr($path . "/meteo/locations","name","varchar(20)");
$conn->addattr($path . "/meteo/locations","latitude","varchar(10)");
$conn->addattr($path . "/meteo/locations","longitude","varchar(10)");
$conn->addattr($path . "/meteo/locations","height","int");
$conn->execute("cd " .  $path . "/meteo/locations");
$list=file("locations.txt");
$att[0]="code"; $att[1]="name";         $att[2]="latitude"; $att[3]="longitude";
$att[4]="height";
for ($i=0;$i<count($list);$i++) {
            $str=explode(" ",trim($list[$i]));
            $conn->addEntry($str[0],$att,$str);
}
```

Php API demo



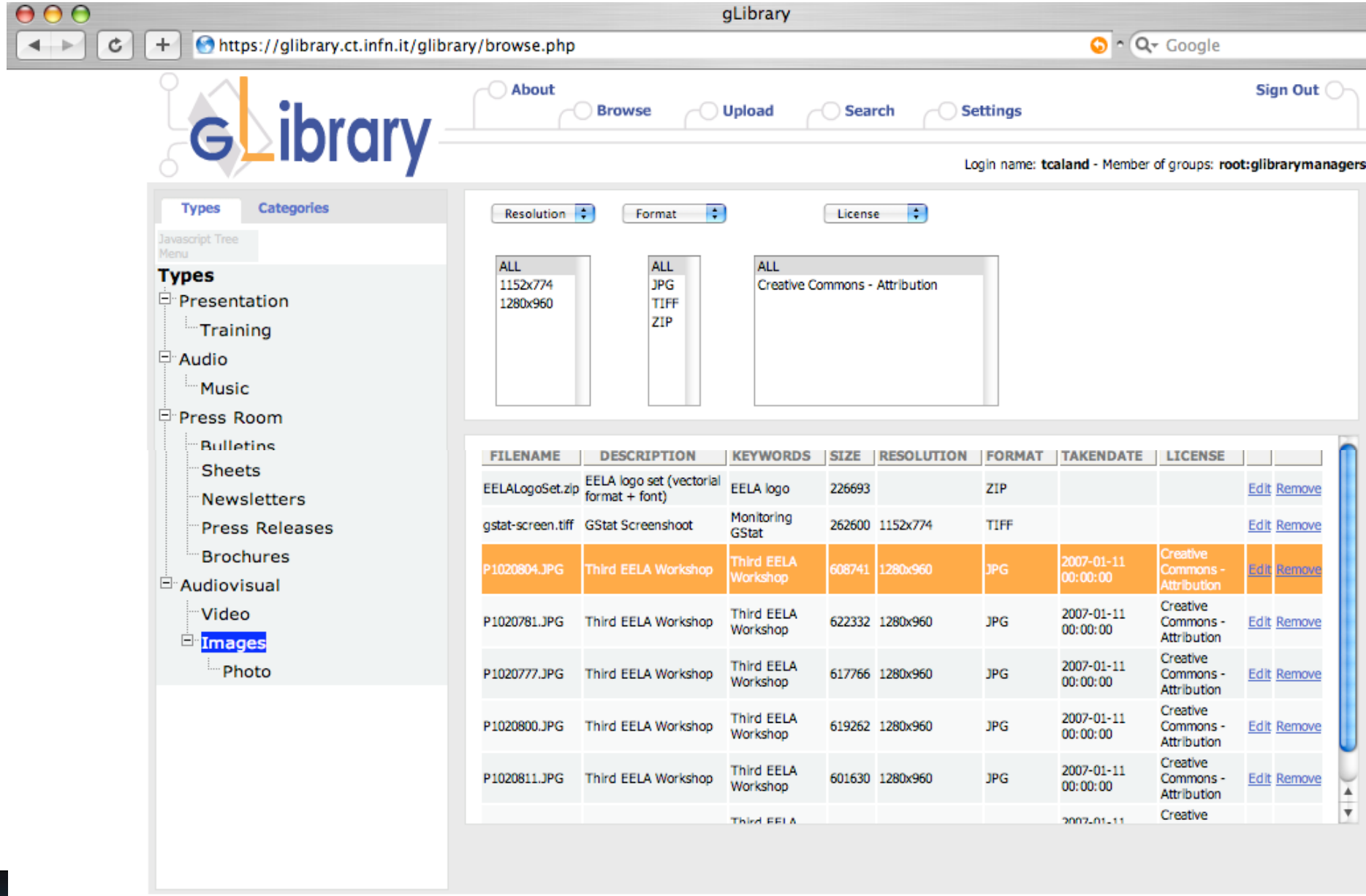
Weather Relevations on AMGA collection

name	date	time	temperature	humidity	pressure
Catania	12/03/2010	13:30	23.5	47.5	1001.2
Catania	12/03/2010	14:00	24.5	45	999.5
Catania	12/03/2010	14:30	25	46	1002.5
Catania	12/03/2010	15:00	25	46.5	1012.5
Palermo	13/03/2010	13:30	21	56	1008.5
Palermo	13/03/2010	14:00	24	58.5	1006
Palermo	13/03/2010	14:30	26	61.5	1014.5
Palermo	13/03/2010	15:00	22	48	1016.5
Trapani	14/03/2010	13:30	19	38.5	1002.6
Trapani	14/03/2010	14:00	21	44	1002.5
Trapani	14/03/2010	14:30	23.5	40	1004.2
Trapani	14/03/2010	15:00	21.5	42.5	1006.2
Catania-Sigonella	15/03/2010	13:30	20.5	68	1002.6
Catania-Sigonella	15/03/2010	14:00	23	70	1004.1
Catania-Sigonella	15/03/2010	14:30	22.5	60.5	1000.2
Catania-Sigonella	15/03/2010	15:00	21.5	65.5	998.5
Reggio-Calabria	16/03/2010	13:30	20.5	48.5	1000.6
Reggio-Calabria	16/03/2010	14:00	21.5	50	996.5
Reggio-Calabria	16/03/2010	14:30	22	52.5	994.5
Reggio-Calabria	16/03/2010	15:00	20	54	995.70001

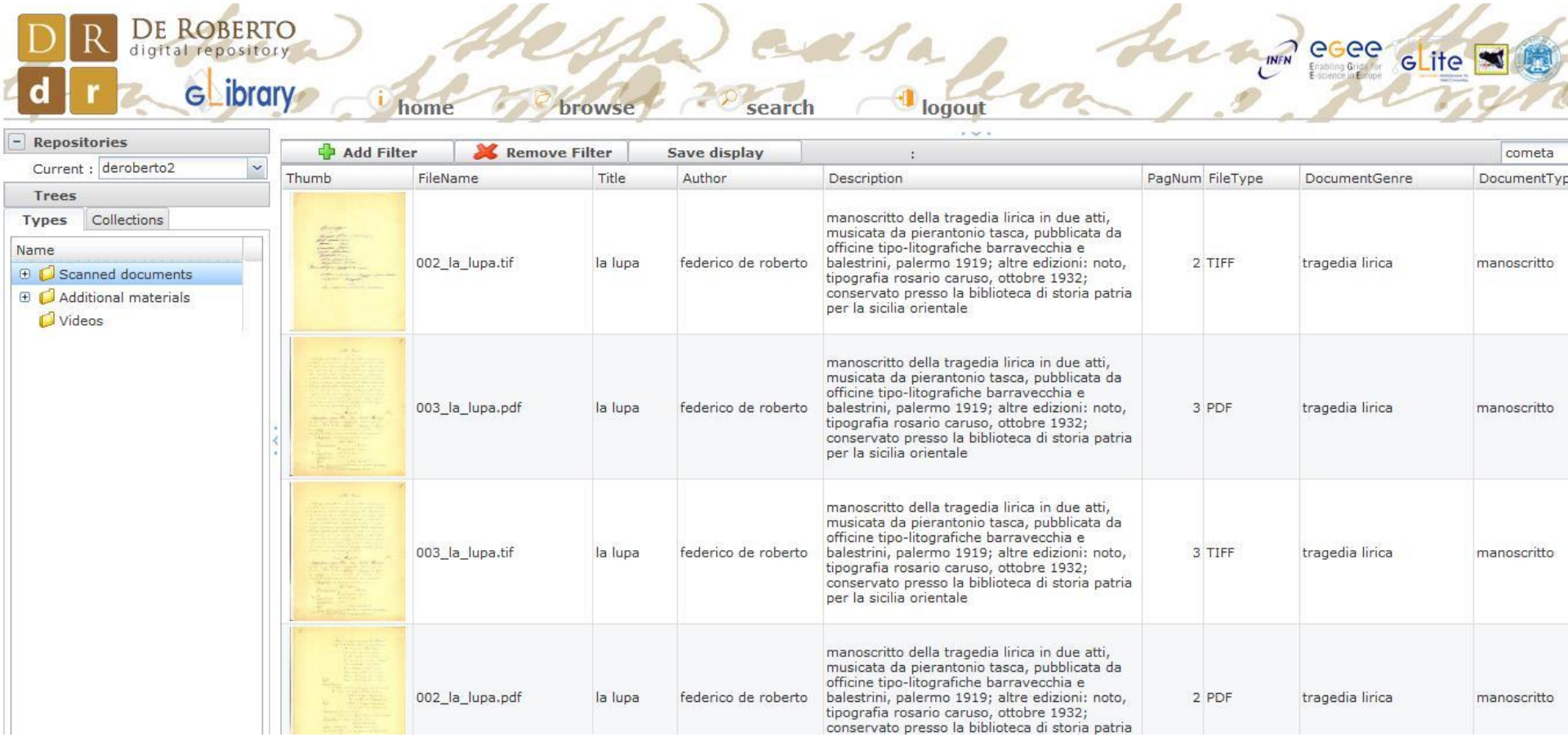
User application sample

- **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, bullettins, sheets
  - ...
- **It allows to store, organize, search and retrieve those assets on a Grid environment.**

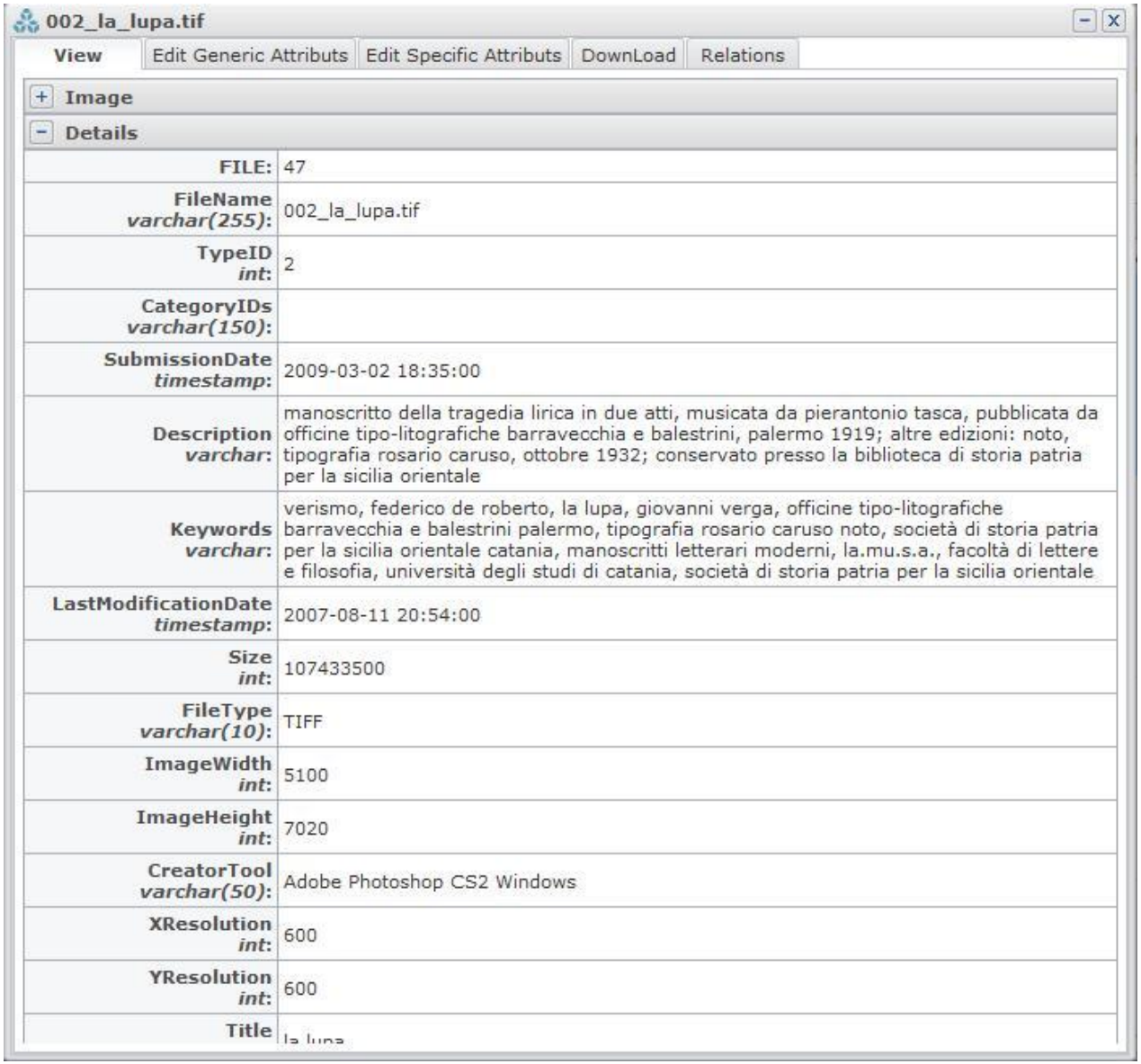
User Application sample



User application features



User application details



FILE:	47
FileName *varchar(255)*:	002_la_lupa.tif
TypeID *int*:	2
CategoryIDs *varchar(150)*:	
SubmissionDate *timestamp*:	2009-03-02 18:35:00
Description *varchar*:	manoscritto della tragedia lirica in due atti, musicata da pierantonio tasca, pubblicata da officine tipo-litografiche barravecchia e balestrini, palermo 1919; altre edizioni: noto, tipografia rosario caruso, ottobre 1932; conservato presso la biblioteca di storia patria per la sicilia orientale
Keywords *varchar*:	verismo, federico de roberto, la lupa, giovanni verga, officine tipo-litografiche barravecchia e balestrini palermo, tipografia rosario caruso noto, società di storia patria per la sicilia orientale catania, manoscritti letterari moderni, la.mu.s.a., facoltà di lettere e filosofia, università degli studi di catania, società di storia patria per la sicilia orientale
LastModificationDate *timestamp*:	2007-08-11 20:54:00
Size *int*:	107433500
FileType *varchar(10)*:	TIFF
ImageWidth *int*:	5100
ImageHeight *int*:	7020
CreatorTool *varchar(50)*:	Adobe Photoshop CS2 Windows
XResolution *int*:	600
YResolution *int*:	600
Title	la lupa

References

- **AMGA Web Site**

http://cern.ch/amga

- **AMGA Basic Tutorial**

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

- **AMGA Manual**

http://amga.web.cern.ch/amga/downloads/2.0/amga-manual_2_0_0.pdf

- **AMGA API Javadoc**

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

- **AMGA API PHPdoc and download**

http://amga.web.cern.ch/amga/php/amga-php/MDClient.html

http://jra1mw.cvs.cern.ch/cgi-bin/jra1mw.cgi/org.glite.amga.api-php/

- **gLibrary**

https://glibrary.ct.infn.it/glibrary_new/index.php

Thank you for your kind attention !

Any questions ?

Thanks to:
Tony Calanducci

Practical session

- Login on the UI cluster30.fis.utfsm.cl with an user valparaiso01..29
- Be sure that proxy is available or type voms-proxy-init –voms prod.vo.eu-eela.eu
- Use wget for the archive at: http://cluster30.fis.utfsm.cl/AMGA/AMGA-samples.tar.gz
- Type **tar –xvzf AMGA-sample.tar.gz** on your home directory inside UI
- Use voms-proxy-info to get the full path where your certificate is located
- Inside AMGA-samples dir, edit Makefile with your user path, choosing an AMGA working directory, and setting the cert path:

```
CERT_PATH = /tmp/x509up_uXXXX
WORKING_DIR = /schooldir/valparaiso/testXXXX
```

- Now you can try each target for each level of sample and also clean:

make load show load-php show-php load-java show-java clean

NOTE: load-java needs to set CLASSPATH **before** show-php is also available online

- Internet PHP access sample on:
  http://cluster30.fis.utfsm.cl/AMGA/show.php?path=/schooldir/valparaiso/test01

-> While db is loaded look at mdclient.config and mdclient shell

-> Executing each target, look also at source codes