GSAF

Grid Storage Access Framework

Salvatore Scifo

INFN sez. Catania

JRA1 All Hands meeting

Catania, 07-09.03.2007

Partnership

- **GSAF**
  - The GSAF Project is carried out by INFN - Catania with the cooperation of the IR&T engineering s.r.l. (an SME of Catania). The context of this work is the *TriGrid VL Project* and the *ADAT Project* (“Archivio Digitale Antichi Testi”).
  - *Trigrid VL Project* aims to port several Industrial Use Cases over the Grid Infrastructure and the *ADAT Project* wants to design and implement a Digital Archive for Cultural Heritage that adopts Grid as a Content Management System (CMS).
- **Resources**
  - INFN
    - S. Scifo (s.scifo@ct.infn.it)
    - Gilda Team
  - IR&T engineering (http://www.irt-engineering.it)
    - V. Milazzo (v.milazzo@irt-engineering.it)
    - A. Magrì (a.magri@irt-engineering.it)

Web integration with the Grid

- **Designing and developing Web Application on the Grid is not easy.**
- **There is no a simple system that allows user to manage dynamic content for generic applications (e.g. web portal, digital libraries, …).**
- **Main objectives of web application**
  - Infrastructure side
    - Organize and handle big amounts of information
    - Share documents among several organizations
    - Security: Manage Access Control Policies
  - Development side
    - Build and maintain dynamic web content
    - Build application without specific technical knowledge
  - User side
    - Manage Groups and Users
    - Manage Digital Resources.

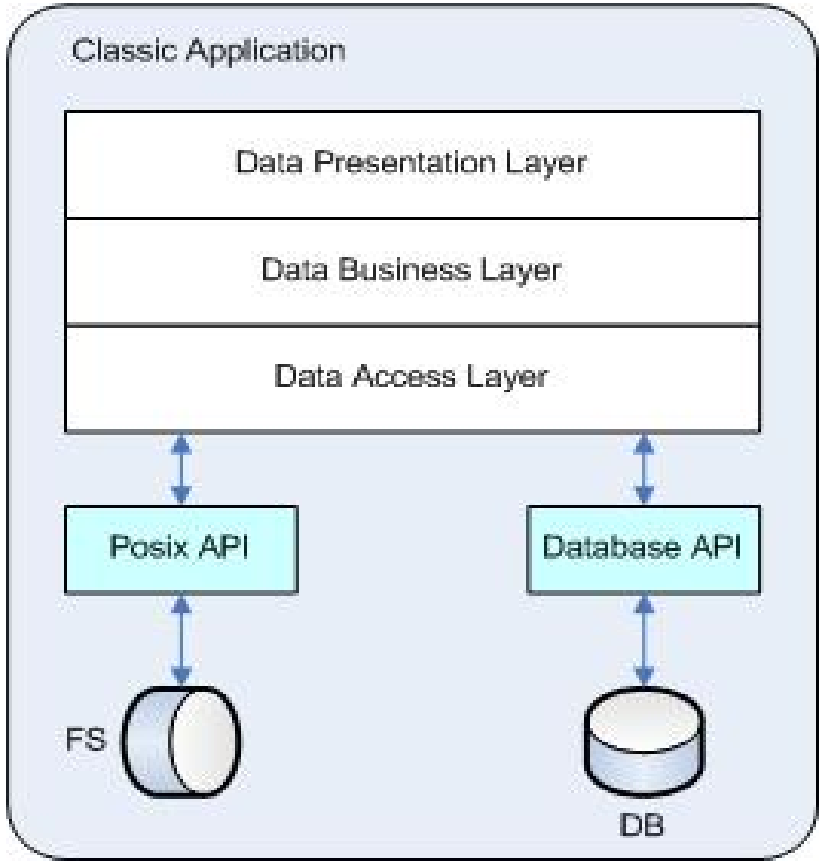
GRID Offer

- **Storage Virtualization**
  - Unique and uniform interface to manage DATA provided by the grid middleware
  - Unique and uniform interface to manage METADATA provided by the grid middleware
  - Large and numerous file handling capability also in a geographic distributed environment
  - Ubiquity: data access independently by their location.

- **Security capabilities**
  - Centralized access control mechanism based on x.509 certificates and user roles according to Virtual Organization policies that users belong to.

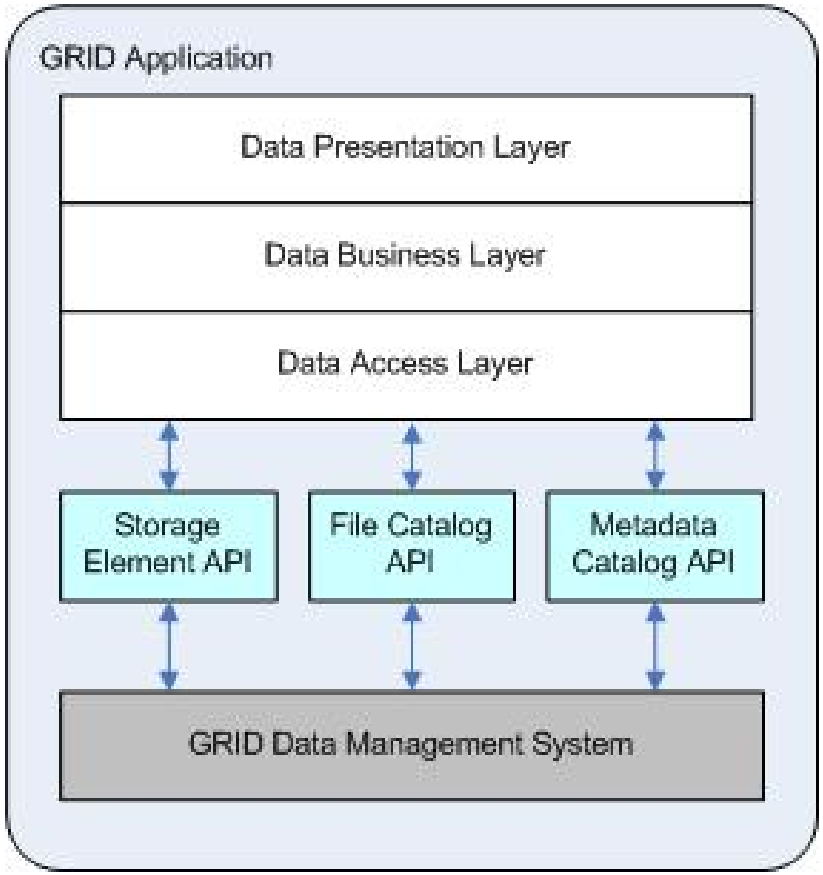
- **Availability, Scalability, Fault Tolerance.**

Classic Web Application



- Data Presentation Layer consists of all graphical interfaces that make user able to interact with application;
- Data Business Layer collects all software components that implement the behavior of the given application;
- Data Access Layer is made up by software components that allow application to manage data (ascii files, xml files, digital object, metadata, SQL data).
- Usually Data Access Layer components interact to several types of data sources (by means of proper APIs), and typical data source are file system (for data stored into files), or Relational Database Management System (for data organized into SQL tables).

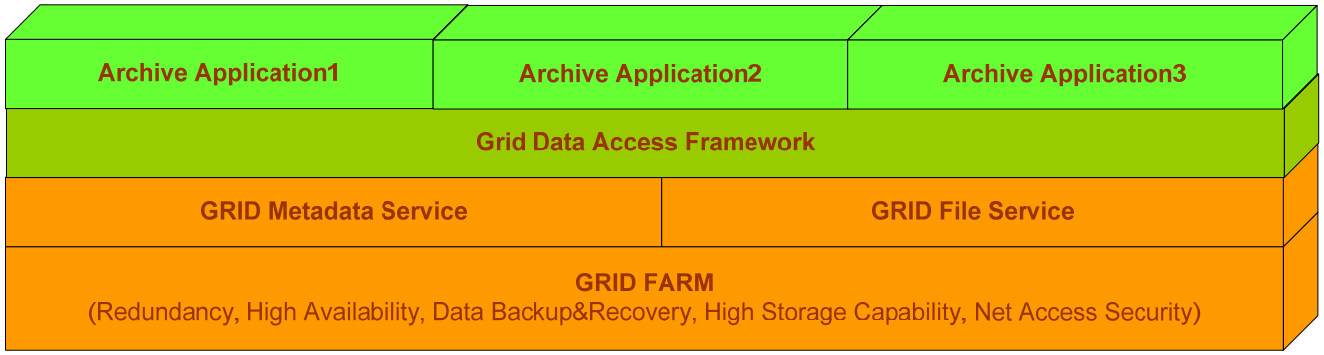
Grid Web Application



- Inside the Grid environment files are stored inside a Storage Element (SE);
- files can be replicated on several SEs for ubiquity, security and sharing needs; relationship among locations of files and replicas and theirs identifier are kept within a specific File Catalogue Service
- for each file is possible to associate descriptive metadata arranged by means a specific Metadata Catalogue Service.
- Developing applications for Grid means just substitute the traditional Data Access Layer with an appropriate interface that permits business components to manage data stored within the DMS and presentation objects to search and retrieve data from DMS.

GSAF: building blocks

- GSAF means Grid Storage Access Framework and it is a kind of Development Toolkit designed to help developers in building applications based on Grid Storage Services for managing files and data.
- The most important requirement of the GSAF is to hide the complexity and the fragmentation of the several APIs provided by the gLite 3.0 middleware in order to interface the main three Grid Data Services.

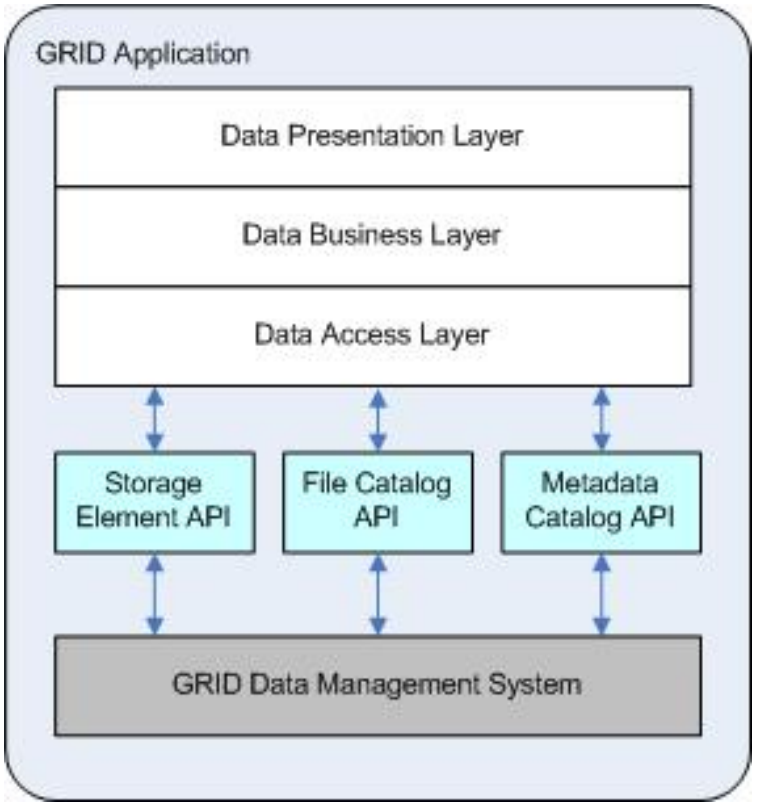


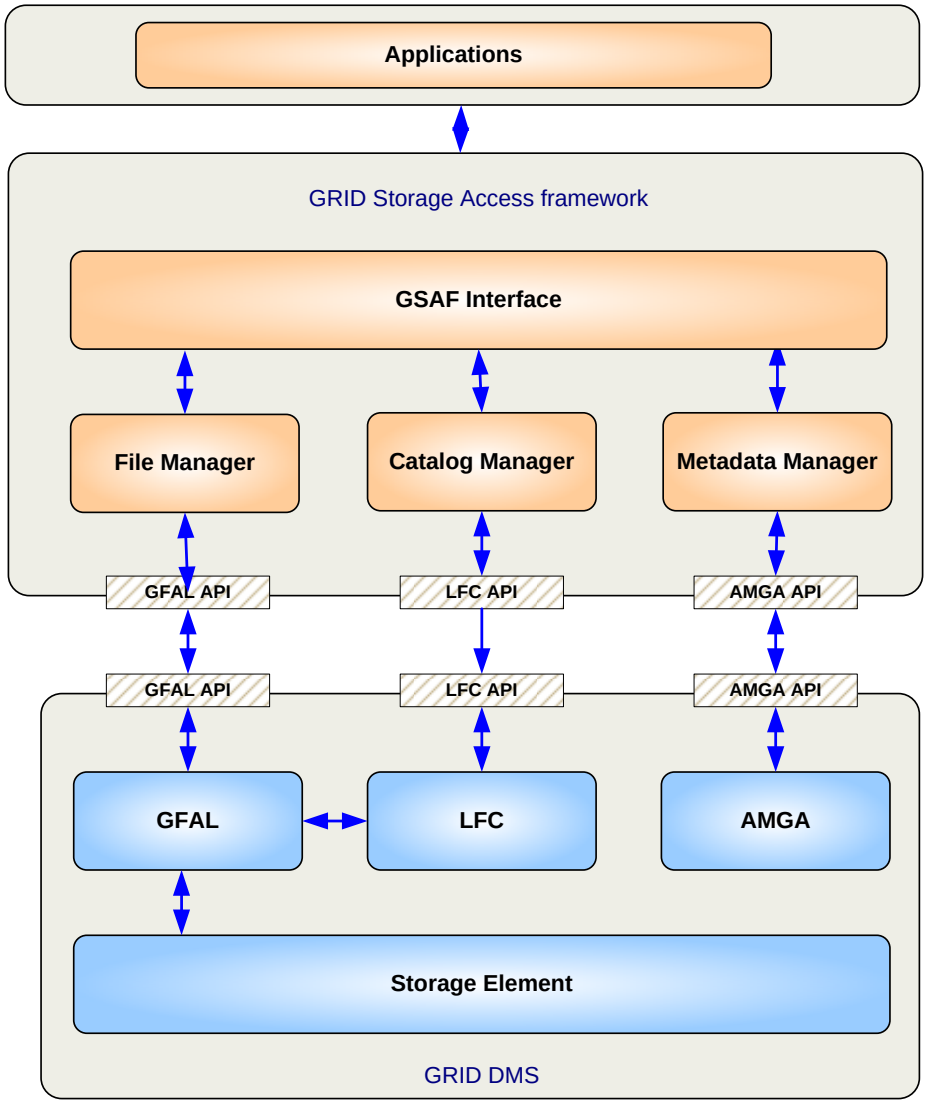
GSAF: goals

- **Implement the main framework capabilities:**
  - Managing Metadata Schemas for data collection
  - Managing Group and User to access metadata
  - Uploading file to the SE, registering LFN to the LFC and saving metadata into AMGA in a coherent and atomic mode.
  - Browsing Metadata Catalogue to download file and/or access to attributes schemas and values.
  - Search file by Metadata to download file and/or access to attributes schemas and values.
  - Deleting file in atomic mode from SE, LFC and AMGA

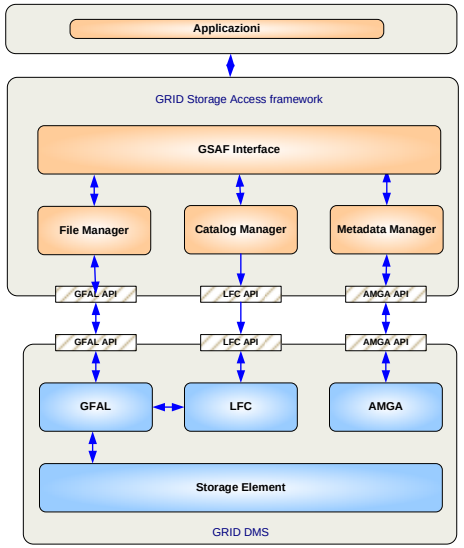
- **Develop a web application as a demonstrator**
  - The application demonstrates the framework behaviour allowing Grid User to manage file and metadata remotely towards a web user interface.

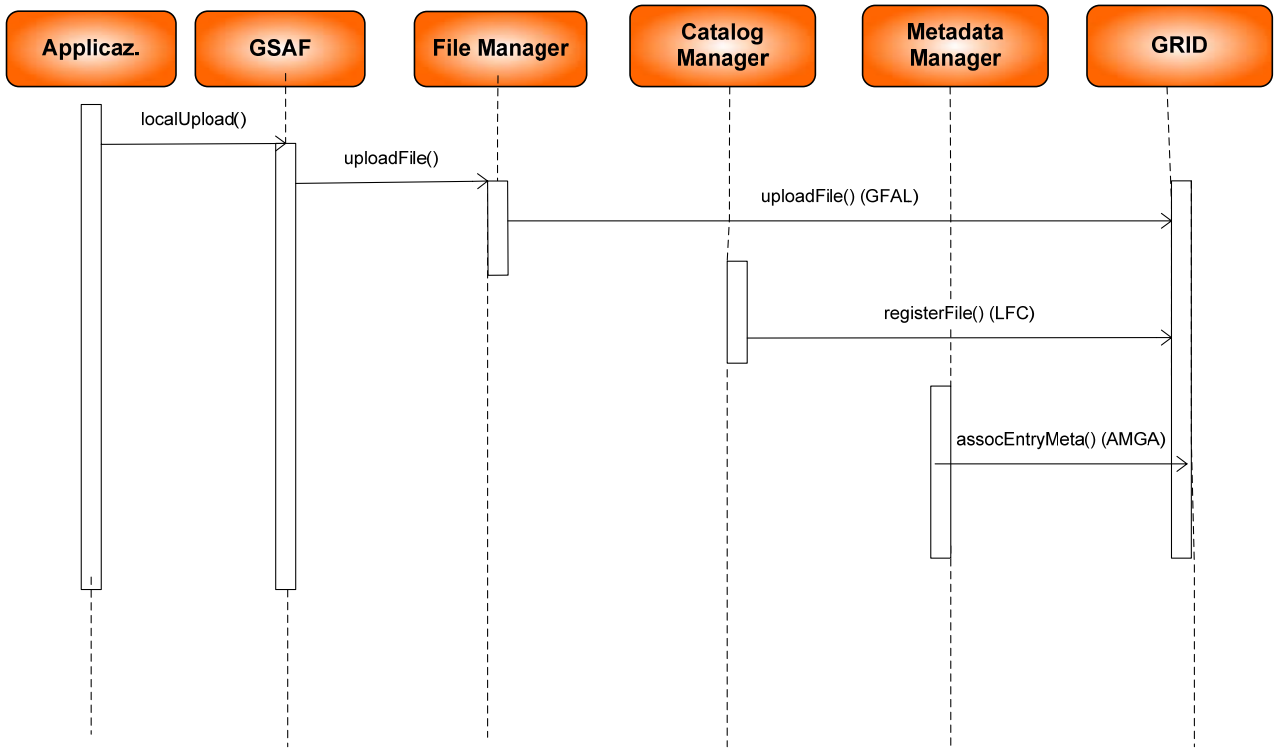
High Level Architecture



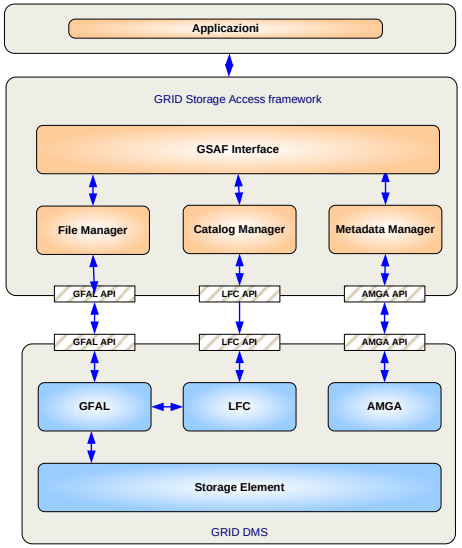


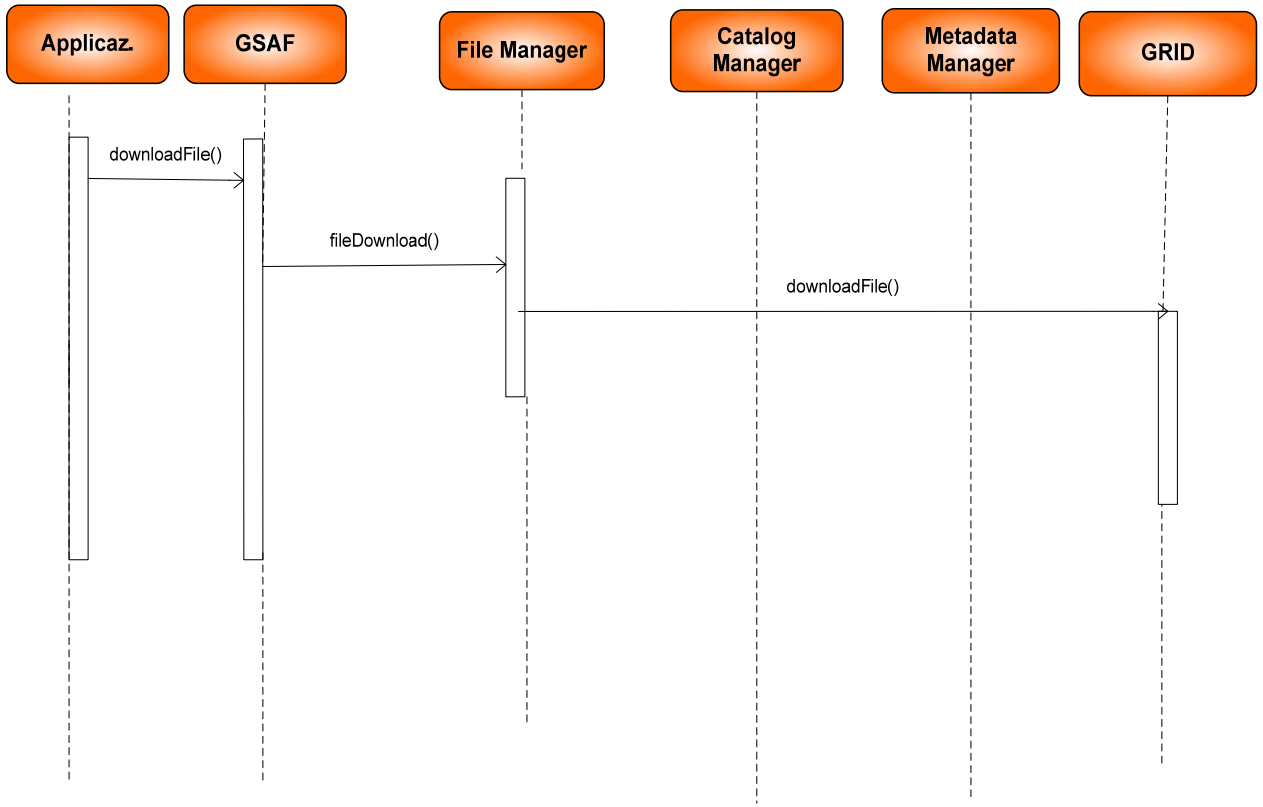
File Upload


Applicazioni
GRID Storage Access framework
GSAF Interface
File Manager
Catalog Manager
Metadata Manager
GFAL API
LFC API
AMGA API
GFAL API
LFC API
AMGA API
GFAL
LFC
AMGA
Storage Element
GRID DMS

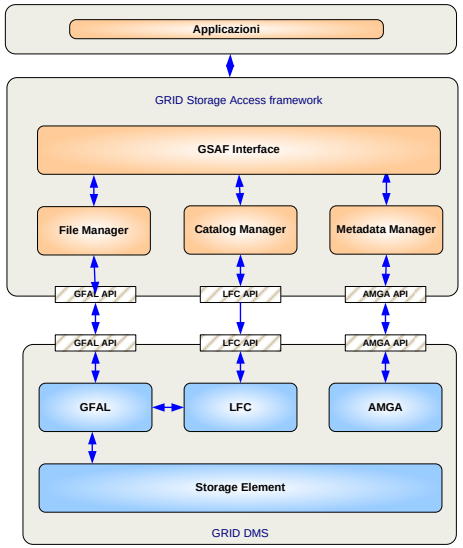

Applicaz.
GSAF
File Manager
Catalog Manager
Metadata Manager
GRID
localUpload()
uploadFile()
uploadFile() (GFAL)
registerFile() (LFC)
assocEntryMeta() (AMGA)

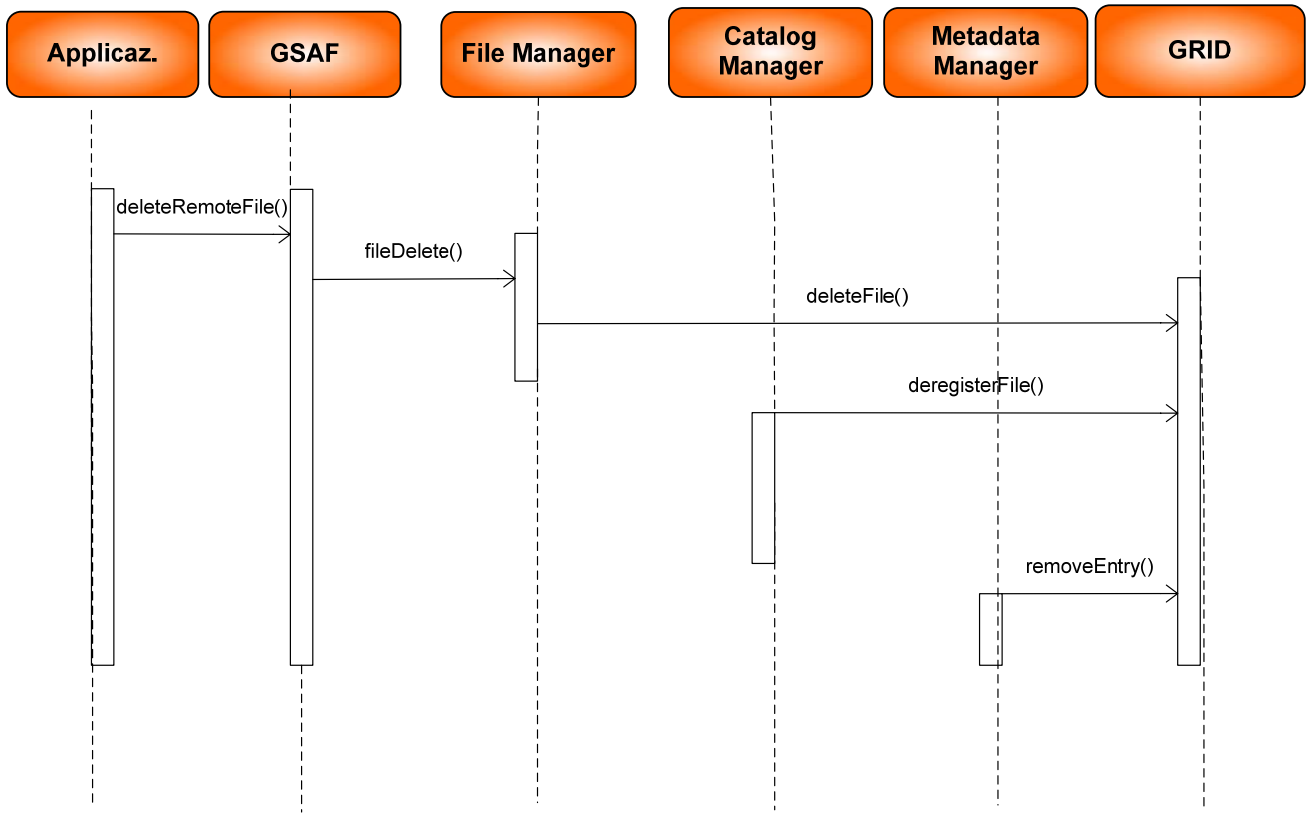
File Download



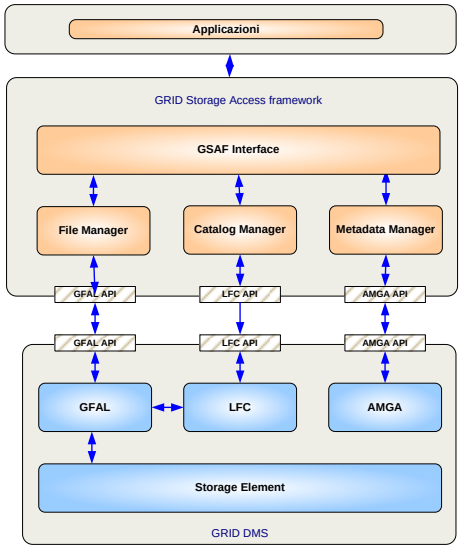


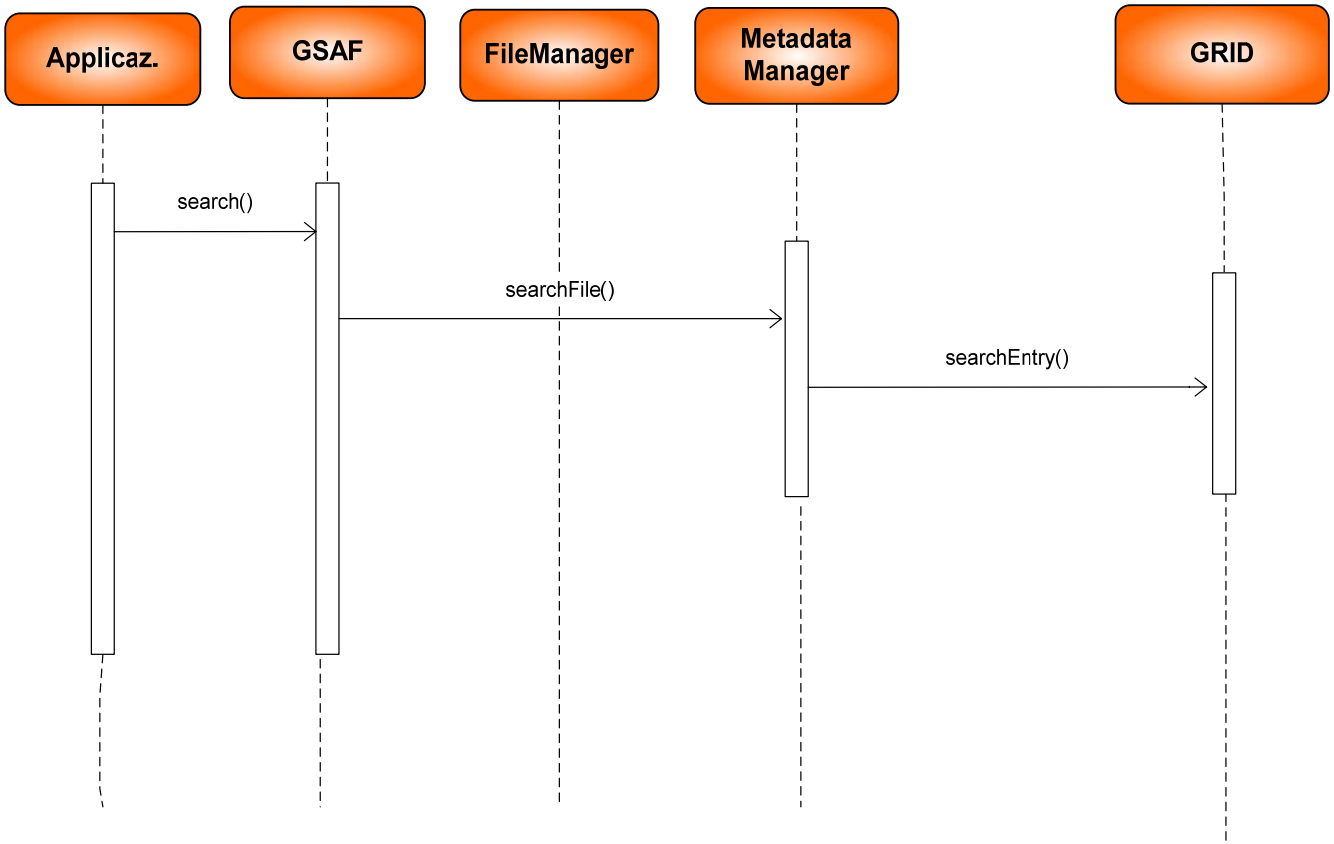
File Delete


Applicazioni
GRID Storage Access framework
GSAF Interface
File Manager
Catalog Manager
Metadata Manager
GFAL API
LFC API
AMGA API
GFAL API
LFC API
AMGA API
GFAL
LFC
AMGA
Storage Element
GRID DMS

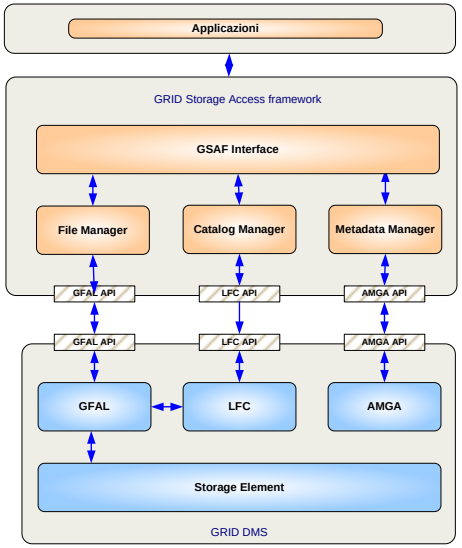

Applicaz.
GSAF
File Manager
Catalog Manager
Metadata Manager
GRID
deleteRemoteFile()
fileDelete()
deleteFile()
deregisterFile()
removeEntry()

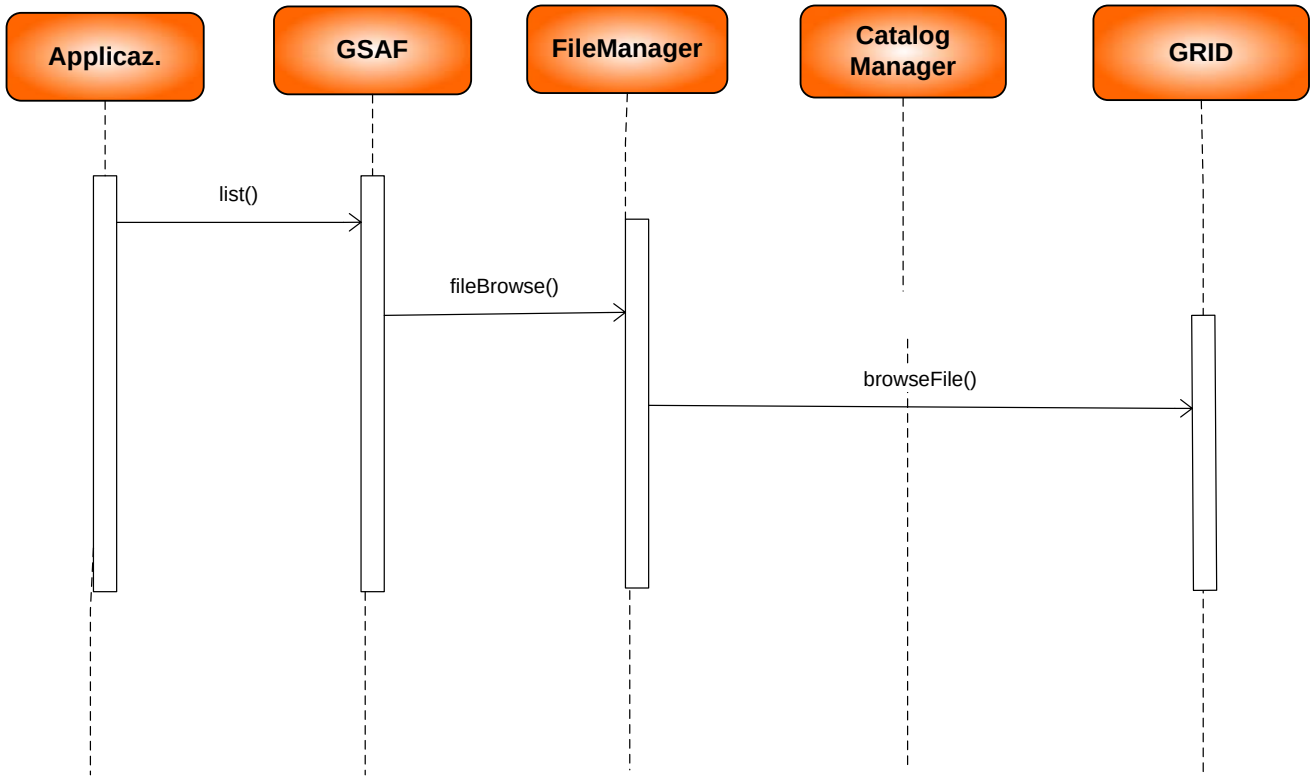
File Search


Applicazioni
GRID Storage Access framework
GSAF Interface
File Manager
Catalog Manager
Metadata Manager
GFAL API
LFC API
AMGA API
GFAL API
LFC API
AMGA API
GFAL
LFC
AMGA
Storage Element
GRID DMS

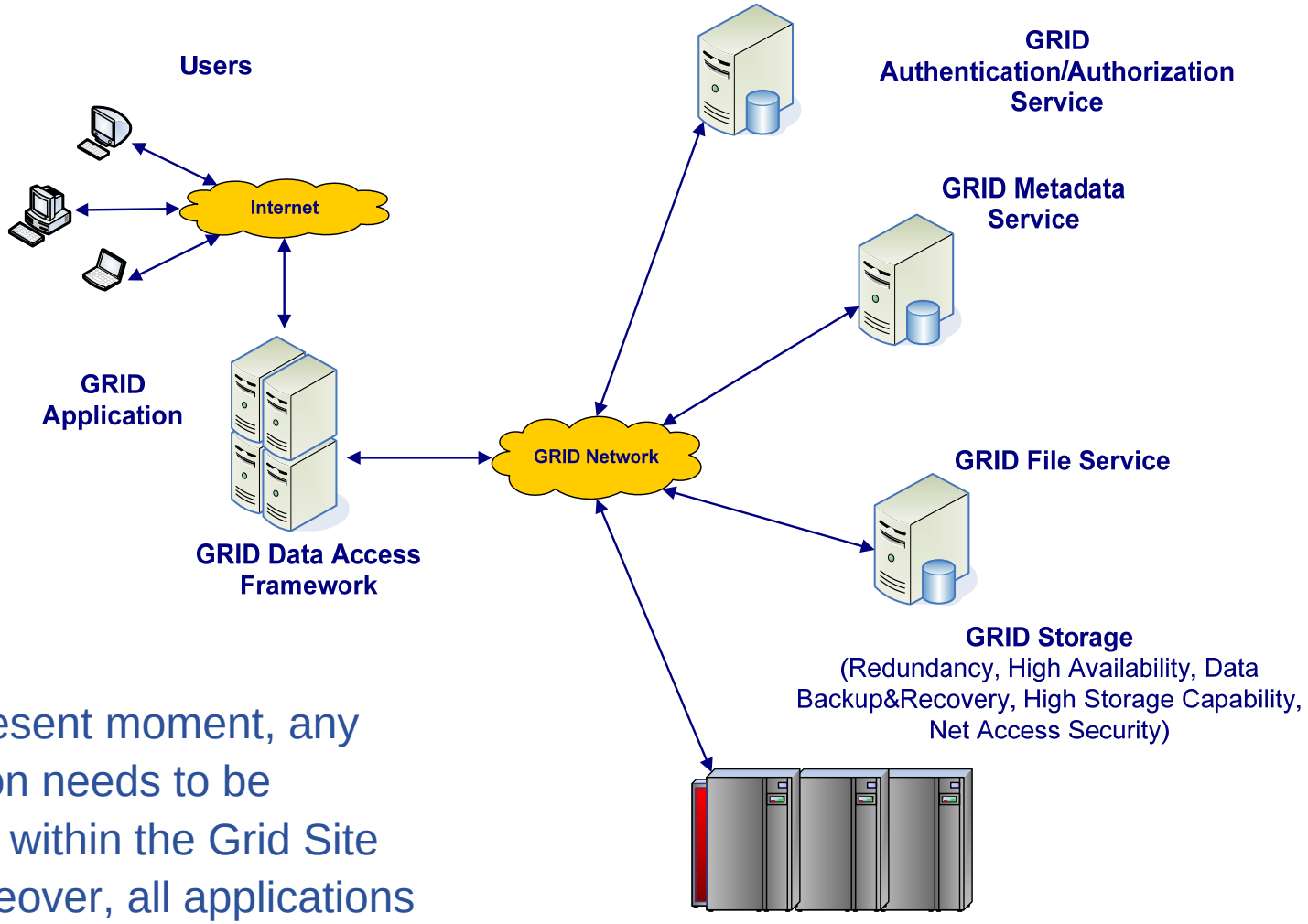

Applicaz.
GSAF
FileManager
Metadata Manager
GRID
search()
searchFile()
searchEntry()

File Browse


Applicazioni
GRID Storage Access framework
GSAF Interface
File Manager
Catalog Manager
Metadata Manager
GFAL API
LFC API
AMGA API
GFAL API
LFC API
AMGA API
GFAL
LFC
AMGA
Storage Element
GRID DMS


Applicaz.
GSAF
FileManager
Catalog Manager
GRID
list()
fileBrowse()
browseFile()

Deployment Plan



At the present moment, any application needs to be deployed within the Grid Site and, moreover, all applications could be developed in Java.


GSAF
GRID STORAGE ACCESS FRAMEWORK
Browse
Search
Upload
Login
BROWSE
Path
//grid/trigrid
folder's...
/ADAT
/gLibrary
/AmgaWiDemo
/ArcheoGrid
/Biomed
//grid/trigrid/ADAT
bolla_urbano_II_fronte_master
bolla_urbano_II_fronte_web
bolla_urbano_II_fronte_preview
bolla_urbano_II_retro_master
bolla_urbano_II_retro_web
bolla_urbano_II_retro_preview
bolla_urbano_II_indice_master
bolla_urbano_II_indice_web
bolla_urbano_II_indice_preview
ir&t
INFN
eGee
Enabling Grids
for E-sciencE

GSAF

GRID STORAGE ACCESS FRAMEWORK

Browse Search Upload Login

Upload

Path //grid/trigrid

Entry browse

FileID 1

Author Urbano II

Title Bolla Papale

CreationTime 1092-09-03

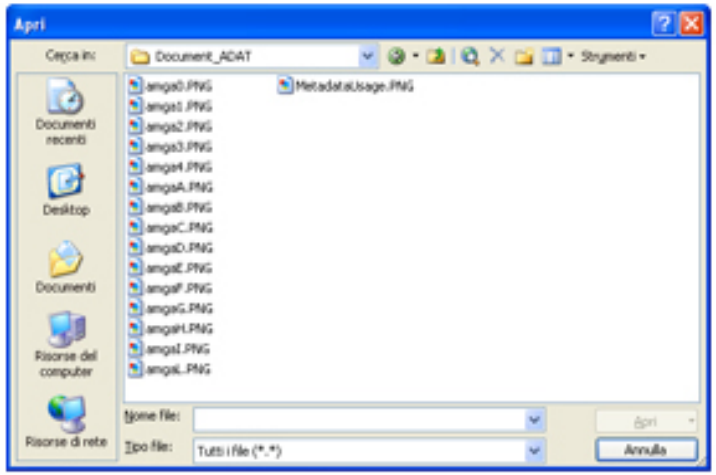
Category Manoscritto

Type

Submitter Salvatore Scifo

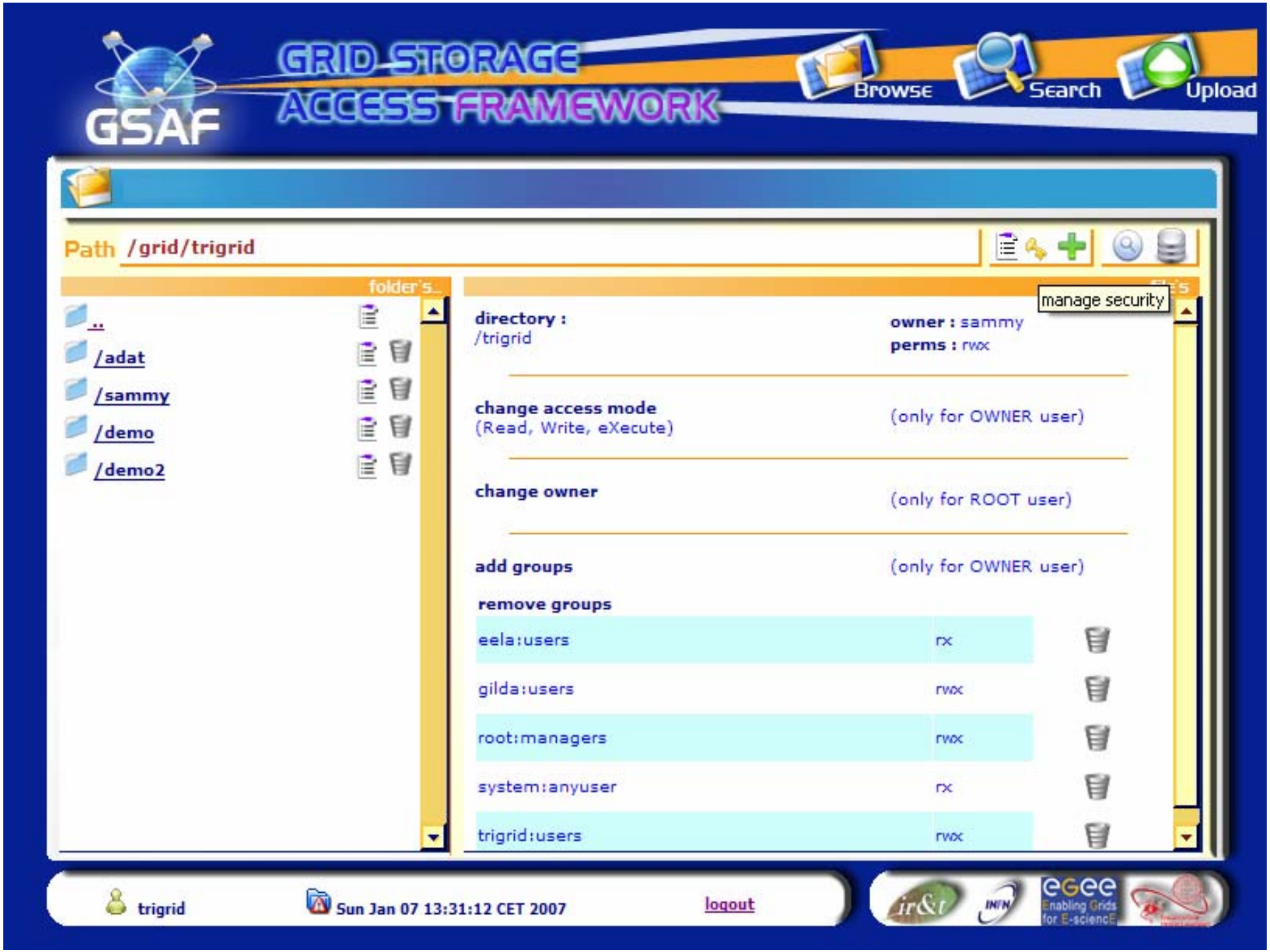
SubmissionTime 2006-11-30

upload



ir&t INFN eGee Enabling Grids for E-sciencE


GSAF
GRID STORAGE ACCESS FRAMEWORK
Browse
Search
Upload
Login
SEARCH
Path
//grid/trigrid
//grid/trigrid/ADAT
Search criteria
FileID
Author
Urbano II
Title
CreationTime
Category
Manoscritto
Type
Submitter
SubmissionTime
bolla_urbano_II_fronte_master
bolla_urbano_II_fronte_web
bolla_urbano_II_fronte_preview
bolla_urbano_II_retro_master
bolla_urbano_II_retro_web
bolla_urbano_II_retro_preview
bolla_urbano_II_indice_master
bolla_urbano_II_indice_web
bolla_urbano_II_indice_preview
search
close
ir&t
INFN
eGee
Enabling Grids for E-sciencE


GSAF
GRID STORAGE ACCESS FRAMEWORK
Browse
Search
Upload
Path /grid/trigrid
folder's
..
/adat
/sammy
/demo
/demo2
manage security
directory :
/trigrid
owner : sammy
perms : rwx
change access mode
(Read, Write, eXecute)
(only for OWNER user)
change owner
(only for ROOT user)
add groups
(only for OWNER user)
remove groups
eela:users rx
gilda:users rwx
root:managers rwx
system:anyuser rx
trigrid:users rwx
trigrid
Sun Jan 07 13:31:12 CET 2007
logout
ir&t
INFN
eGee
Enabling Grids for E-sciencE

ADAT

ARCHIVIO DIGITALE
ANTICHI TESTI

Browse | Search | Info

/grid/trigrid

- /ADAT
- /gLibrary
- /AmgaWiDemo
- /ArcheoGrid
- /Biomed

/grid/trigrid/ADAT

- bolla_urbano_II_fronte_master
- bolla_urbano_II_fronte_web
- bolla_urbano_II_fronte_preview
- bolla_urbano_II_retro_master
- bolla_urbano_II_retro_web
- bolla_urbano_II_retro_preview
- bolla_urbano_II_indice_master
- bolla_urbano_II_indice_web
- bolla_urbano_II_indice_preview

Conclusions

- **Sharing information belonging to different organizations in secure, scalable and efficient way is very frequent and actual in the ICT context.**

- **GRID offers**
  - **Reliable Resources Organization**
  - **Distributed storage virtualization**
  - **Uniform data access**
  - **Security and data Preservation**

- **GSAF means**
  - **Useful API to develop Storage based applications**
  - **Useful and simple web interface to access Data Management Services remotely**

References

- **GSAF wiki pages**
  - https://grid.ct.infn.it/twiki/bin/view/TRIGRID/GSAF
- **Amga Web Interface wiki pages**
  - https://grid.ct.infn.it/twiki/bin/view/TRIGRID/AMGAWI
- **AMGA Service and Java API**
  - http://project-arda-dev.web.cern.ch/project-arda-dev/metadata/index.html
- **GFAL Java API**
  - http://grid-deployment.web.cern.ch/grid-deployment/gis/GFAL/gfal.3.html
  - https://grid.ct.infn.it/twiki/bin/view/GILDA/APIGFAL
- **LFC Java API**
  - http://wiki.egee-see.org/index.php/SEE-GRID_File_Management_Java_API