gLibrary: Digital Asset Management System for the Grid

Friday 11 May 2007 09:00 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

In every Grid Infrastructure a huge amount of distributed storage resources is available to save user's data.
However, finding a specific file is not always easy. gLibrary challenge is to offer an extensible, secure and
easy-to-use system to handle digital assets stored as files on a Grid, providing an intuitive cataloging system that
allows to find assets in seconds. Target communities are both end-users and organizations that need a secure way
to save, share and organize their assets.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

A first prototype of gLibrary has been implemented as a web application available at https://glibrary.ct.infn.it
The
front-end was developed as a Web 2.0 application to offer a pleasant user experience. This implies the usage of
technologies like AJAX, Javascript and Java Applets to offer a desktop-like interface. The business logic of the
application is built as PHP 5 objects that interact with several gLite services, such as AMGA Metadata Catalogs, VOMS
servers, Information Systems, File Catalogs and Storage Elements. Since AMGA is the most used grid service by this
web application, to guarantee an immediate user interface response, it has been decided to develop a low level PHP
5 API whose role is to interact directly with the AMGA server daemons. All communication between clients and the
application server are sent over HTTPS connections. Authentication on grid services makes use of VOMS-enabled
proxy certificates and all file transfers are carried over GridFTP channels.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

An issue that we faced during gLibrary developing is the inconsistency of the gLite API and the non-availability of
multiplatform API: almost all of the gLite Data Management APIs are provided for Scientific Linux Cern as C
and Python/Perl libraries. Moreover. VOMS APIs are not fully completed, lacking in methods to handle attribute
chains and again available only to one platform. gLibrary goal is to be a multiplatform tool, allowing users to handle
their assets from any operating system.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

gLibrary is entirely based on the Grid services provided by the gLite middleware. This implies the exploitation of the

benefits that comes by grid technologies usage, such as its built-in security and the availability of huge storage resource for saving data. Users do not have to care about the complexity of the underlying systems and the geographical location of their data and they can consider the available Grid storage as a huge virtual disk. Permissions can be set on the users'assets in order to grant or deny access to given users, groups or even whole

organizations.All entries in gLibrary are organized according to their type, a list of specific attributes to describe each

kind of asset to be managed by the system. These are the same attributes that can be queried by users. Types can

be defined according to the assets users want to manage. The flexibility and extensibility offered by this type system

allow different communities to adopt gLibrary for any cataloging purpose

Primary authors: Dr CALANDUCCI, Antonio (INFN Catania); Dr CHERUBINO, Claudio (INFN Catania); Dr SCARDACI, Diego (INFN Catania); Dr NEUMANN CIUFFO, Leandro (INFN Catania)

Presenter: Dr CALANDUCCI, Antonio (INFN Catania)

Session Classification: Data Management

Track Classification: Data Management