

gLibrary/DRI: A grid-based platform to host multiple repositories for digital content

Wednesday, February 13, 2008 4:50 PM (25 minutes)

A gLibrary/DRI repository is made of large digital content (as image files, video, etc) and metadata associated with it (annotations, descriptions, etc). In a typical scenario, new repository providers could use the built in mechanisms to store repository items (e.g. studies made of textual data and multiple medical images) in a combined GRID and federated RDBMS by simply describing the structure of their data in a set of XML files following the gLibrary/DRI specification. In a more elaborated scenario, repository providers can implement specific data management policies and use custom viewers for their specific data structures, still relying on the platform for navigation and management of their repository.

As example, we present a repository based on mammograms, composed of both a repository and a viewer application, to manage patient's mammograms and diagnostics. This includes both the patient's data (stored as metadata) and the mammography digital content (large images stored in SE)

3. Impact

Repository providers describe the structure of the repository contents by following the DRI Data Model specification, indicating how the model is distributed into different relational entities (tables) and also marking what parts of it are to be stored in the federated database/metadata server and what parts are to be stored into Grid SEs.

The Storage DRI API Specification provides method definitions for loading and persisting model nodes. Through this API we isolate data management from its storage technology. (However we provide an implementation of this API using Grid SRM SEs and AMGA technologies.) These methods are transparent to the node complexity and content, and also to the storage system chosen for storing the data.

The GUI Navigation functions are used for providing to the user a quick and effective way of finding any node of the data model into the repository. The navigation system is based on categories trees and a set of filters that reduce the nodes search to the user.

If demonstration is requested please explain what visual or interactive aspects of the contribution necessitate a demonstration rather than a presentation or poster?

A live demonstration showing a working repository of Mammographies will be presented

4. Conclusions / Future plans

We have developed a platform that reduces the cost for developing new digital repositories. It provides a set of API and specifications that decouples the repository developing from the underlying platform. Multiple repositories can be hosted, just by providing the UI and Storage modules. The architecture is totally Grid based (VOMS authentication/authorization, data federation and distribution, usage of the computing power in the future). A mammograms repository has been also developed.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Digital Libraries, Metadata, Mammography, Medical Repositories, Data Management

1. Short overview

gLibrary/DRI (Digital Repositories Infrastructure) is a platform to host any kind of repository for digital content, providing a common infrastructure and a set of mechanisms (APIs and specifications) that repository

providers use to define the data model, the access to content (by viewers, navigation trees and filters) and the storage model. The main goal of the platform is to reduce the cost in terms of time and effort that a repository provider spends in order to get its repository deployed

Primary authors: Dr CALANDUCCI, Antonio (INFN Catania); Dr TCACI, Dorin (MAAT-G Knowledge); Dr GONZÁLEZ MARTÍN, Juan Manuel (MAAT-G Knowledge); Dr RUBIO DEL SOLAR, Manuel (CETA-CIEMAT); Dr RAMOS POLLÁN, Raúl (CETA-CIEMAT)

Presenter: Dr CALANDUCCI, Antonio (INFN Catania)

Session Classification: Data Management

Track Classification: Existing or Prospective Grid Services