Enabling Inter-repository Access Management between iRODS and Fedora

Bing Zhu,
Univ. of California: San Diego
Richard Marciano
Reagan Moore
University of North Carolina at Chapel Hill

May 18, 2009
Atlanta
An Environment with Heterogeneous Technologies

- DSpace
- Fedora
- iRODS
- dLibra
- Eprints
- Greenstone
- Handle System
Enabling inter-repository data management allows us to share data by connecting the repositories of:

- Different groups, projects
- Different institutions, locations
- Different disciplines
- Diverse types of data
- Diverse hardware, software infrastructure
Issues for inter-Repository data management

- Object Model
- Virtual Registration of Digital Objects from One Repository to Another
- Inter-repository Service Management
- Policy Enforcement across Repositories
iRODS - integrated Rule-Oriented Data System

A middleware providing functions to:

- manage distributed storages
- provide metadata support for digital preservation and search functions
- allow running distributed workflows to enforce system policies and harvest distributed computing power.

iRODS can be used for

- building datagrid
- building digital library
- building digital repositories
Why iRODS + Fedora?

User -> Application -> Fedora -> iRODS

Complex Object Modeling Layer
Digital Preservation Layer
iRODS for Digital Preservation

- Data replication service
- Periodic data integrity check
- Distributed storages for disaster recovery
- Metadata support for preservation description information
A replacement for Fedora’s local storage module
A standalone plug-in module (independent of Fedora release)
iRODS manages both Fedora objects (XML) and data streams for Fedora (for managed content)
Implemented based on SRB Storage Module for Fedora by DART project
iRODS Storage Module

- Manual Management of Distributed Stores - Admin Selects a Storage Resource for Storing Data Objects

- Auto Management of Distributed Storages - Use a Logical Storage Resource
Deployment of iRODS Storage Module in Fedora

- Download Jar files from iRODS web site
  
  https://www.irods.org/index.php/Fedora

- Copy the jar files into Fedora place under Tomcat
  
  $CATALINA_HOME/webapps/fedora/WEB-INF/lib

- **Edit the Fedora config file** $FEDORA_HOME/server/config/fedora.fcfg

```xml
<module role="fedorax.server.storage.lowlevel.ILowlevelStorage"
         class="fedorax.server.module.storage.lowlevel.irods.IrodsLowlevelStorageModule">
  <param name="file_system" :value="fedorax.server.module.storage.lowlevel.irods.IrodsIFileSystem"/>
  <param name="irods_host" value="irods.sdsc.edu"/>
  ...
```
iRODS Digital Object Model

- File Based
- Distributed Digital Objects
- System metadata
- User defined metadata

/dzone/home/nerual/eek_001.jpg

System metadata
User metadata
Fedora Digital Object Model

A Compound Object Model

Four Types of control group for Fedora datastream:

- Internal XML Metadata
- Managed Content
- External Reference Content
- Redirect
Referencing iRODS objects as Fedora External Datastreams

- An extension to the Fedora’s external content manager
- iRODS files can be virtually linked inside Fedora
- Syntax:
  - future: irods://…
Registering iRODS Objects into Fedora

- Create a Fedora object. The full path of the iRODS object becomes the label of the Fedora object.
- Create an external reference datastream for the iRODS file.
- iRODS system metadata is registered as an externally referenced datastream in Fedora.
- iRODS user-defined metadata is registered as an externally referenced datastream in Fedora.
**Example: Create a Fedora Object**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Active</td>
</tr>
<tr>
<td>Label</td>
<td>irods://srbrick15.sdsc.edu:7547/pzone/home/testuser/NeSC-Moore-intro.ppt</td>
</tr>
<tr>
<td>Created</td>
<td>2009-03-30T18:22:02.744Z</td>
</tr>
<tr>
<td>Modified</td>
<td>2009-03-30T18:45:31.238Z</td>
</tr>
<tr>
<td>Owner</td>
<td>fedoraAdmin</td>
</tr>
</tbody>
</table>
Example: Create a datastream for the iRODS file

- **ID**: NeSC-Moore-intro.ppt
- **Control Group**: External Reference
- **State**: Active
- **Versionable**: Updates will create new version
- **Created**: 2009-03-30T18:45:31.238Z
- **Label**: The PPT file for e-Science workshop
- **MIME Type**: application/ms-powerpoint
- **Format URI**:
- **Alternate IDs**:
- **Location**: http://srbrick15.sdsc.edu:7547/pzone/home/testuser/NeSC-Moore-intro.ppt?fs=irods
- **Fedora URL**: http://srbrick7.sdsc.edu:8089/fedora/get/irodsObj:2/NeSC-Moore-intro.ppt
- **Checksum**: DISABLED, none
Example: a dynamic reference to iRODS system metadata
Example: a dynamic reference to iRODS user metadata
Mapping Fedora Objects in iRODS

Tar File Format

User

Access compound tar file

iRODS Server

Manage the Tar file as a container

iRODS Rule Engine

Meta Data Catalog

Persistent ID
PID

Object Properties

Relations
RELS-EXT

Dublin Core
DC

Audit Trail
AUDIT

Datastream 1

Datastream ...

Datastream N

Manage the Tar file as a container
Invoking iRODS service in Fedora

User

Fedora

Invoke a rule

(Event, Condition, Action chains, Recovery chains)

iRODS Catalog

iRODS Rule Engine

iRODS Server

iRODS Rule Engine

iRODS Server

iRODS Rule Engine

iRODS Server

iRODS Rule Engine

iRODS Server
DICE Center

iRODS:
- http://www.irods.org
- http://www.dice.unc.edu
- http://www.diceresearch.org

Fedora-iRODS Integration:
- https://www.irods.org/index.php/Fedora