

Introduction to Federico 2.0 and Fedora Commons



<http://aforge.awi.de/gf/project/federico/>

Dr. Bernadette Fritsch
Bernadette.Fritsch@awi.de

Ing. José A. Mejía Villar M.Sc.
Jose.Mejia@awi.de

Computing Center of the Alfred Wegener
Institute for Polar and Marine Research

Contents

1. Introduction to Federico
 - 1.1 What is Federico?
 - 1.2 System Requirements
 - 1.3 Live Demo
2. Fedora Commons Repository
 - 2.1 What is Fedora Commons?
 - 2.2 Key Features vs Disadvantages
 - 2.3 Digital Object Model
 - 2.4 Content Model Architecture
 - 2.5 Web Service Interfaces
 - 2.6 Framework Services
 - 2.7 Security

1. Introduction to Federico



1.1. What is Federico?

1.2. System Requirements

1.3. Live Demo

1.1 What is Federico?

- **Fedora-Enabled Repository with Cocoon**
- **AJAX**-based frontend for a C3Grid local repository of metadata
- Transparent Integration of Fedora with the Framework Services GSearch and OAI Provider
- Developed in the scope of the work package #3, **Long-term Preservation of Digital Archives** of **Wissgrid**, sponsored by the **German Federal Ministry of Education and Research**



1.2 System Requirements [1/2]

Hardware

- PC with a 1 gigahertz (GHz) processor or faster and network card
- 2 GB RAM
- 800 MB free disk space for the installation

Software

- Linux Distribution with X Window System
- Java JDK 1.6
- 3 MySQL Databases for Fedora Commons, Fedora OAI Provider, and openID accounts

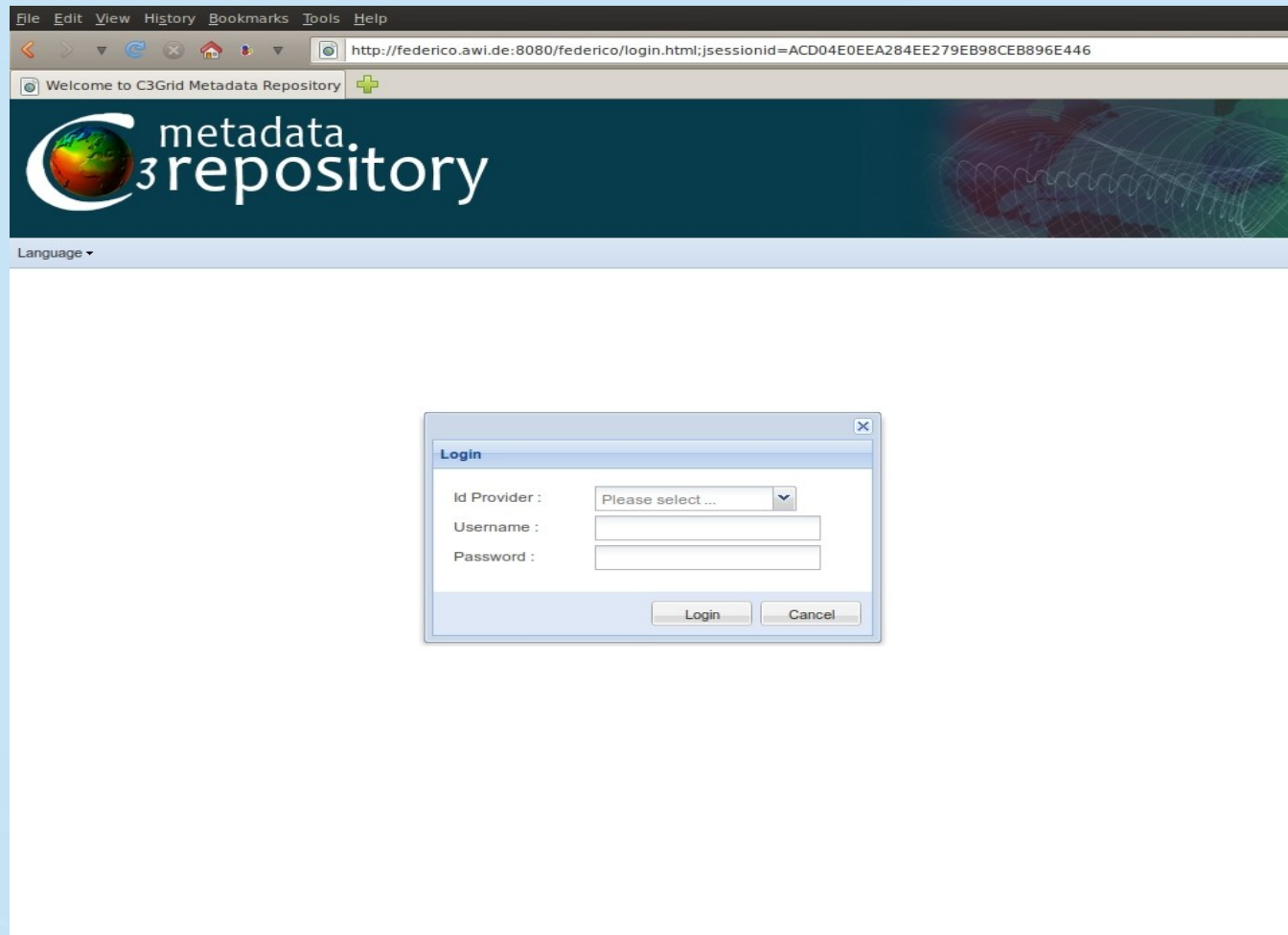
1.2 System Requirements [2/2]

User

- PC with graphical interface and network card
- Keyboard and mouse
- Browser (preferably Mozilla Firefox) with Javascript enabled

1.3 Federico's Live Demo: Screenshots

Login Form



1.3 Federico's Live Demo: Screenshots

Browse Collections Form

The screenshot shows a web browser window displaying the 'Browse Collections' page of the Federico metadata repository. The browser's address bar shows the URL `http://federico.awi.de:8080/federico/collections.html`. The page features a header with the 'metadata repository' logo and a navigation menu with options for 'General', 'Repository', 'Administration', and 'Language'. The 'Language' dropdown menu is open, showing options for 'Deutsch', 'Español', and 'English'. The main content area is titled 'Browse Collections' and includes a sub-header 'Collection demo:root [Root Set]'. Below this, there is a table with columns for 'Type', 'Id', 'Label', 'Owner', and 'Date'. The table is currently empty. To the right of the table, there are several informational panels: 'Nomenclature' (explaining that a folder represents a Set and a box represents a Collection), 'Login as' (showing 'fedoraAdmin' and a 'logout' button), 'What is a Set?' (defining a Set as a series of collections), and 'Download JANEME' (providing information about the J-NetCDF Metadata Extractor). The browser's status bar at the bottom shows 'Done'.

1.3 Federico's Live Demo: Screenshots

Metadata Upload

The screenshot displays a web browser window with the URL `http://federico.awi.de:8080/federico/upload_collection.html?parent=demo:1`. The browser's address bar shows the URL, and the page title is "Upload Collection Metadata".

The main content area features the "metadata repository" logo and a navigation menu with options: "General", "Repository", "Administration", and "Language". The "Administration" menu is currently selected.

The main content area is titled "Upload Collection Metadata" and contains a yellow dashed box with the following text: "Upload an XML file describing a collection for its online edition. Its content should conform to the community established metadata profile schema defined in Federico".

Below this text is an "Upload File [fill.xml]" button with a file selection icon. The status below the button reads: "Status: Successful upload of fill.xml, 35589 bytes". A "Submit" button is located below the status message.

The sidebar on the right contains several links and buttons:

- Nomenclature**: A lightbulb icon. Below it, a folder icon with the text "represents a Set" and a document icon with the text "represents a Collection".
- Login as**: A green circular arrow icon. Below it, the text "fedoraAdmin" and a "logout" button.
- What is a Set?**: A question mark icon. Below it, the text "A Set is a series of collections."
- What is a Collection?**: A question mark icon. Below it, the text "A Collection is a member".
- Download JANEME**: A green download icon. Below it, the text "The J-NetCDF Metadata Extractor is available for free download and use under the terms of a BSD License from [AWIForge](#)".

The browser's status bar at the bottom shows "Done".

1.3 Federico's Live Demo: Screenshots

Metadata Edition

File Edit View History Bookmarks Tools Help

http://federico.awi.de:8080/federico/590f087f0e4f6b8c77133a30331d707a7b3d5370.continue

Edit Collection Metadata

Edit the following fields and submit the form in order to ingest the collection.

MD_Metadata

1 fileIdentifier [?](#)
CharacterString :

2 language
CharacterString :

3 parentIdentifier [?](#)
CharacterString :

Repeater hierarchyLevel

4 hierarchyLevel

4.1 MD_ScopeCode
value :
@codeList :
@codeListValue :

[+](#) Add "hierarchyLevel" [-](#) Delete selected "hierarchyLevel"

Done

File Edit View History Bookmarks Tools Help

http://federico.awi.de:8080/federico/590f087f0e4f6b8c77133a30331d707a7b3d5370.continue

Edit Collection Metadata

[+](#) Add "hierarchyLevelName" [-](#) Delete selected "hierarchyLevelName"

Repeater contact

Error :

- This repeater must contain 1 to 2147483647 rows.

[+](#) Add "contact" [-](#) Delete selected "contact"

7 dateTime [?](#)
DateTime :

8 metadataStandardName
CharacterString :

9 metadataStandardVersion
CharacterString :

10 dataSetURI
CharacterString :

Repeater locale

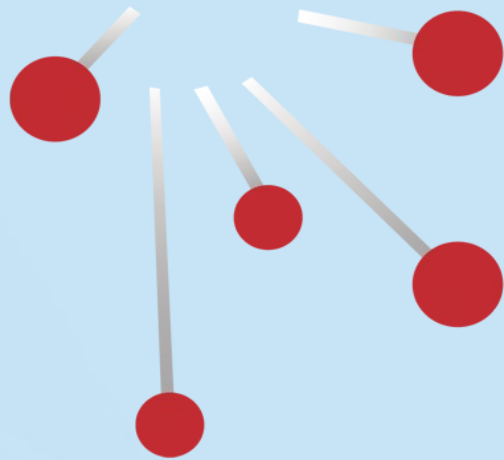
11 locale

11.1 PT_Locale
languageCode
LanguageCode

Done

2. Fedora Commons

Fedora
Commons[™]



- 2.1. What is Fedora Commons?
- 2.2. Key Features vs Disadvantages
- 2.3. Digital Object Model
- 2.4. Content Model Architecture
- 2.5. Web Service Interfaces
- 2.6. Framework Services
- 2.7. Security

2.1 What is Fedora Commons?

- **Fedora** stands for *Flexible Extensible Digital Object Repository*.
- Fedora is a **general-purpose**, **open-source** digital object repository system.
- Java based conceptual framework using a set of abstractions about digital information to provide the basis for software systems that can manage digital information.
- The Fedora software distributed by **Duraspace** (<http://www.duraspace.org>) is available from <http://fedora-commons.org> under the terms of the **Apache License, version 2.0**.

2.2 Key Features [1/3]

- Store all types of content and its metadata
- Scale to millions of objects
- Access to data via Web APIs (REST/SOAP)
- Provides RDF based Resource Index search
- Rebuilder Utility (for disaster recovery and data migration)
- The entire repository can be rebuilt from the digital object and content files.

2.2 Key Features [1/3]

- Store all types of content and its metadata
- Scale to millions of objects
- Access to data via Web APIs (REST/SOAP)
- Provides RDF based Resource Index search
- Rebuilder Utility (for disaster recovery and data migration)
- The entire repository can be rebuilt from the digital object and content files.

2.2 Key Features [2/3]

- Content Model Architecture (define "types" of objects by their content)
- Many storage options (database and file systems)
- JMS messaging provider (your apps can "listen" to repository events)
- OAI-PMH Provider Service

2.2 Disadvantages [3/3]

- Front-end Adaptation

<https://wiki.duraspace.org/display/DEV/Fedora+Tools>

- Object Store Scalability Strategy

<https://wiki.duraspace.org/display/AKUBRA/Akubra+Project>

2.3 Digital Object Model

- All content in Fedora is managed as **data objects**
- Data objects are made up of **datastreams** that store the content or metadata about it.
- Each datastream can be managed directly by the repository or left in an external, web-accessible location to be delivered through the repository as needed.
- A data object can consist of any number of data and metadata components, combining managed and external datastreams in **any desired pattern**.

2.3 Digital Object Model: FOXML

FOXML (Fedora Object XML) is a simple XML format that directly expresses the Fedora Digital Object Model.

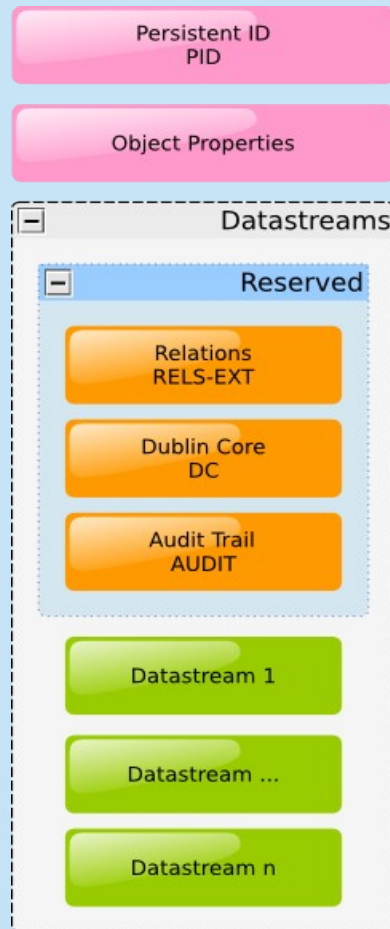
FOXML 1.1 XSD Schema on: <http://fedora-commons.org/definitions/1/0/foxml1-1.xsd>

```
<digitalObject PID="uniqueID">
  <!-- there are a set of core object properties -->
  <objectProperties>
    <property/>
    <property/>
    ...
  </objectProperties>

  <!-- there can be zero or more datastreams -->
  <datastream>
    <datastreamVersion/>
    <datastreamVersion/>
    ...
  </datastream>
</digitalObject>
```

2.3 Digital Object Model: Datastreams

Fedora reserves three datastreams for its use, namely “DC” (Dublin Core), “AUDIT”, and RELS-EXT.



Basic Datastream Properties

- Datastream Identifier
- State: Active, Inactive, or Deleted
- Created Date
- Modified Date
- Versionable: true/false
- Label
- MIME Type
- Format identifier (optional)
- Alternate Identifiers (Handlers or DOI)
- Checksum
- Bytestream Content
- Control Group
 - Internal XML Content
 - Managed Content
 - Externally Referenced Content
 - Redirect Referenced Content

2.4 Content Model Architecture

- The **Content Model Architecture** (CMA) describes an integrated structure for persisting and delivering the essential characteristics of digital objects in Fedora.
 - Structural, behavioral, and semantic information.
 - Description of the permitted, excluded, and required relationships to other digital objects or identifiable entities.
- The content model is expressed in a modeling language.

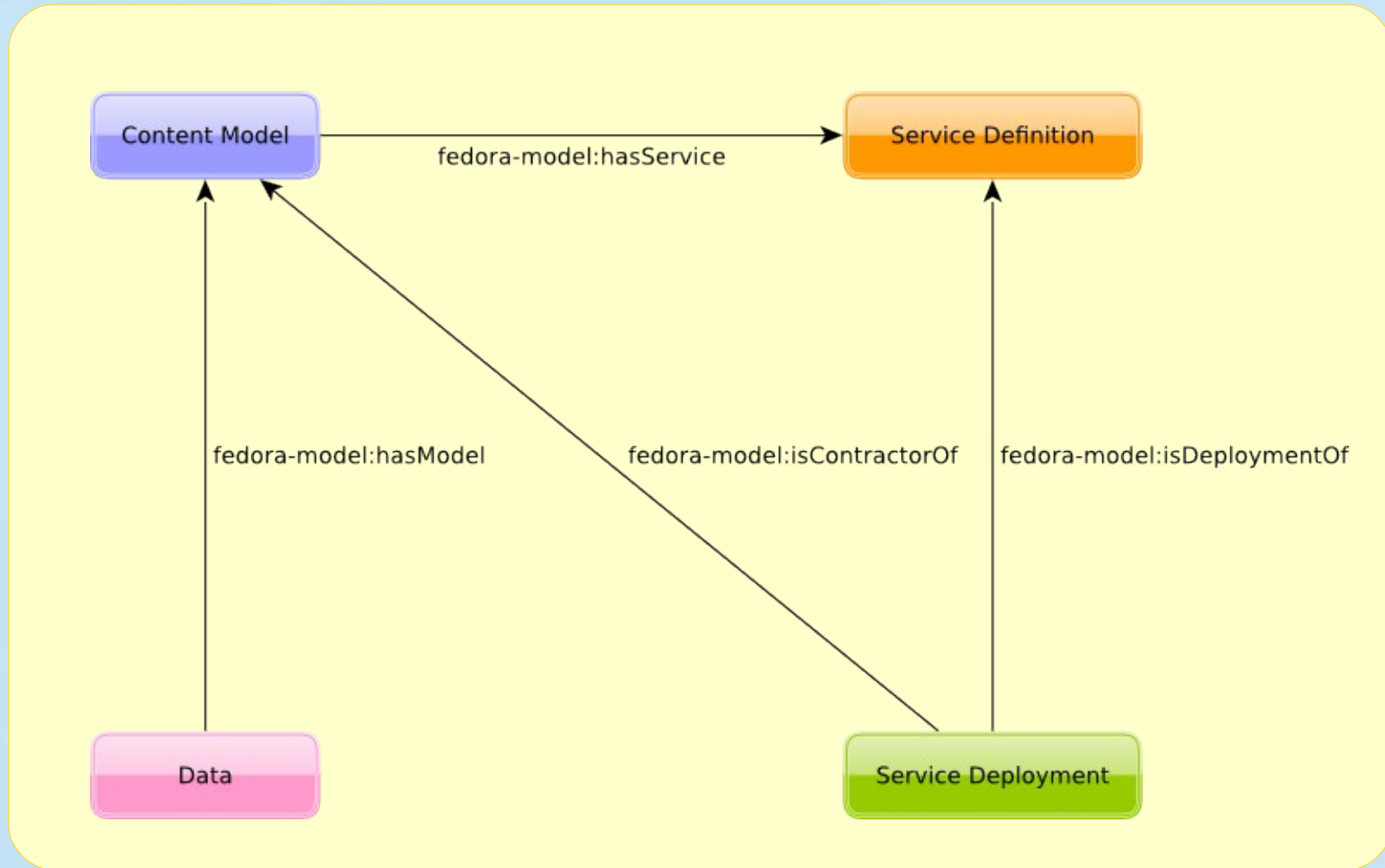
2.4 CMA: Object Types

Fundamental Fedora Object Types

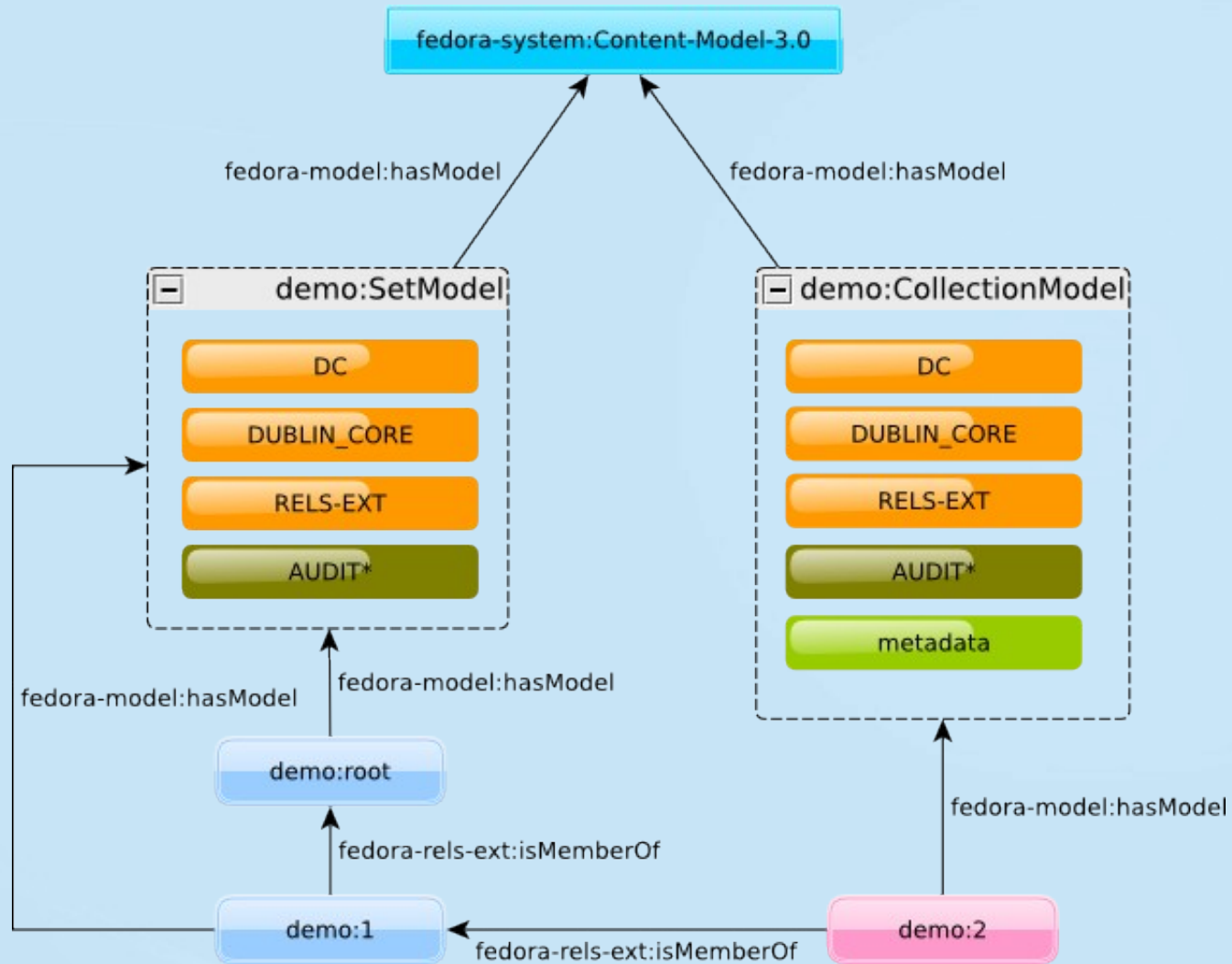
Object Type	Code	Description
Data	Data	A container for content
Service Definition	SDef	A container for the service definitions
Service Deployment	SDep	A container for service deployment bindings
Content Model	CModel	A container for content models

2.4 CMA: Object Types

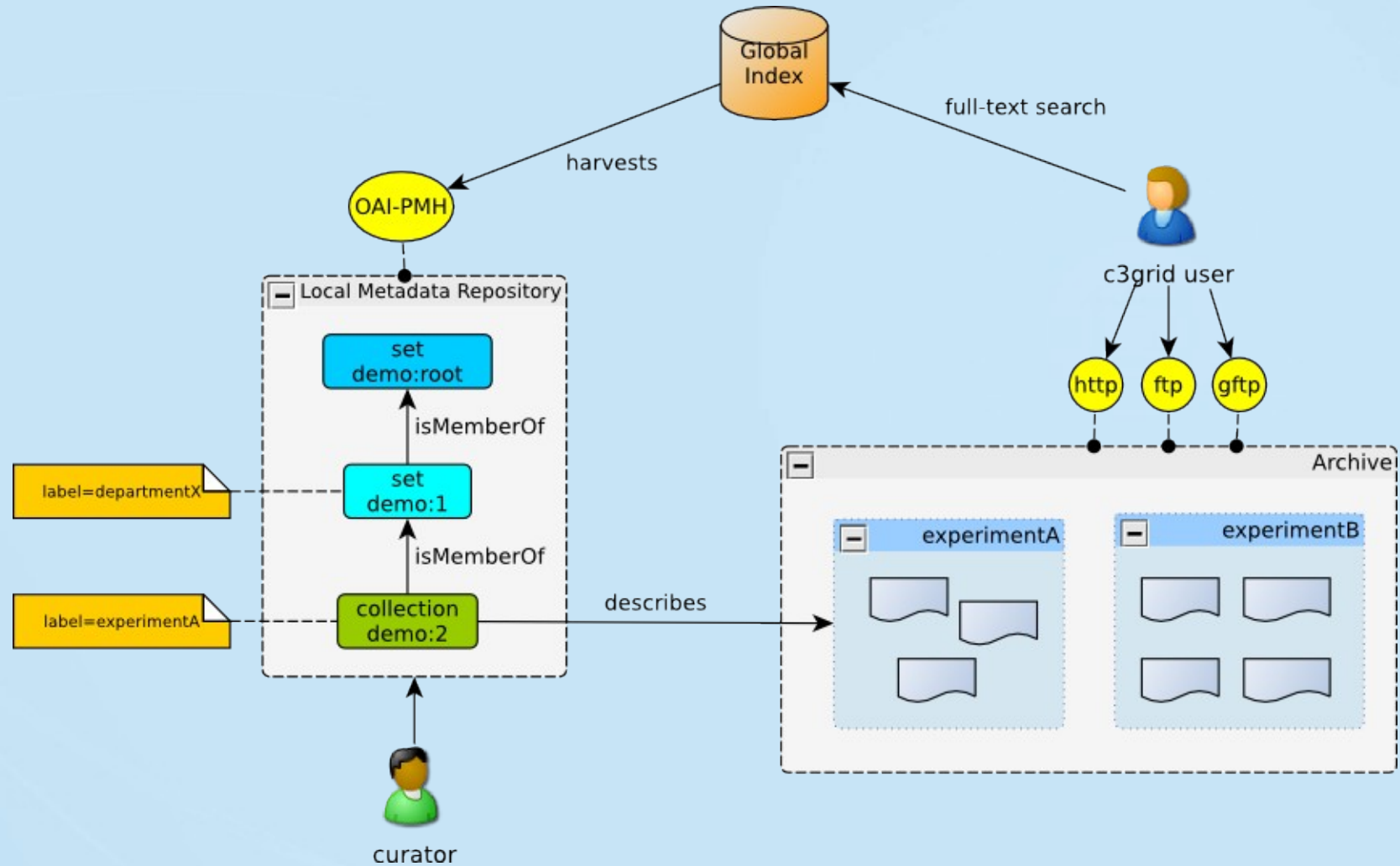
Fundamental CMA Relationships



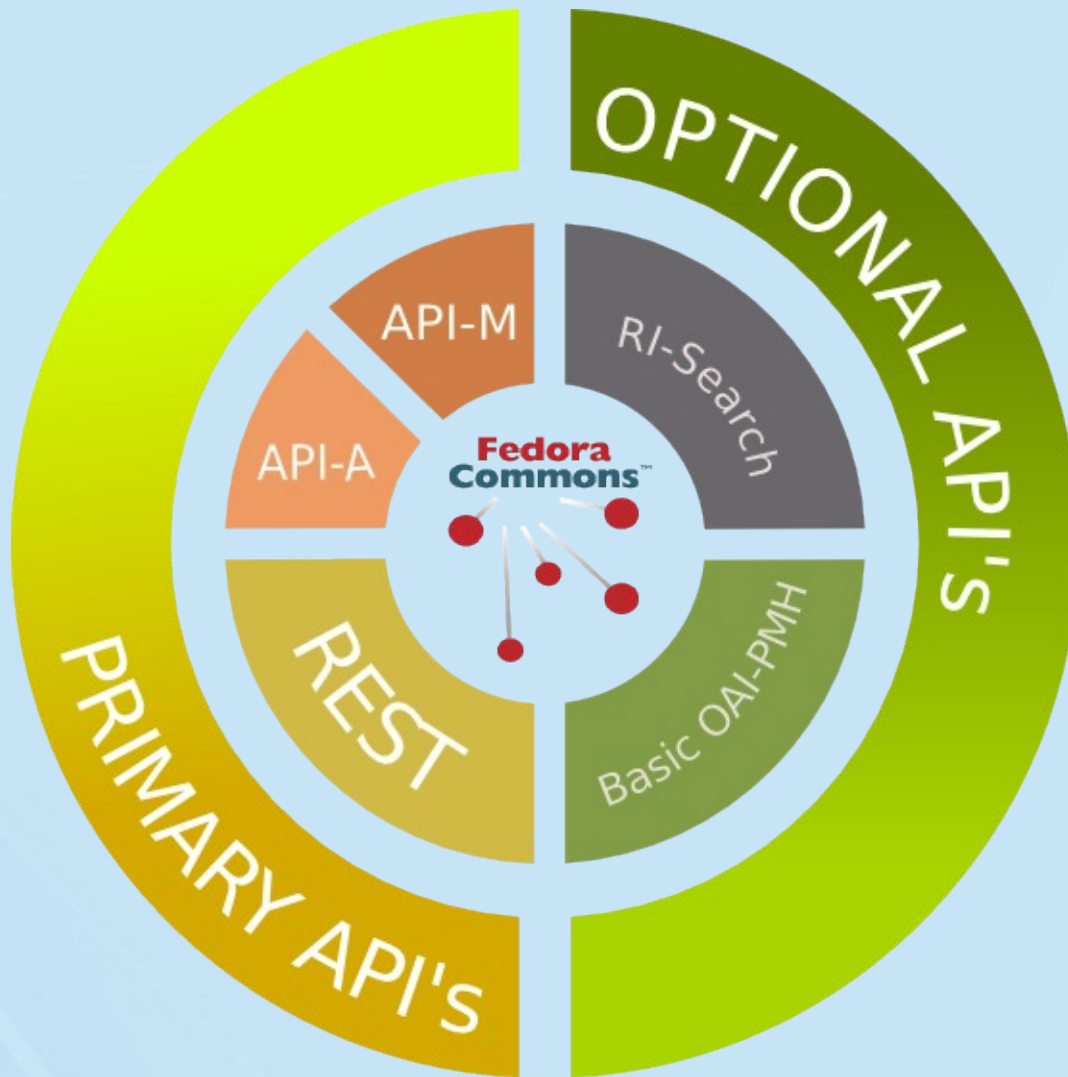
2.4.1 Federico: Content Model [1/2]



2.4.1 Federico: Content Model [2/2]



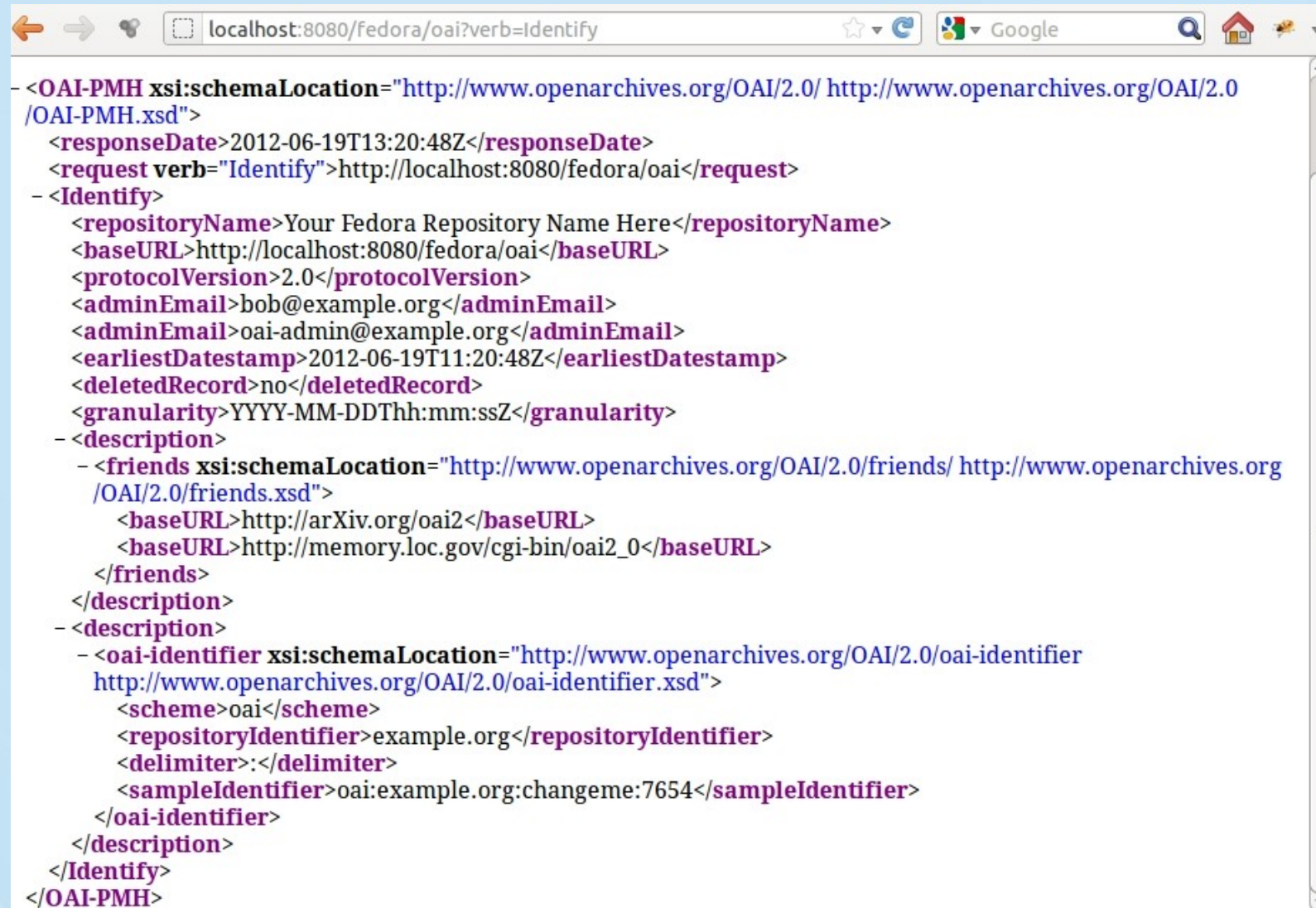
2.5 Web Service Interface



- **Primary API's**
Allow the creation, reading, modification, and deletion of Fedora digital objects.
- **Optional API's**
 - Basic OAI-PMH
 - RI-Search

2.5.1 Basic OAI

http://localhost:8080/fedora/oai?verb=Identify

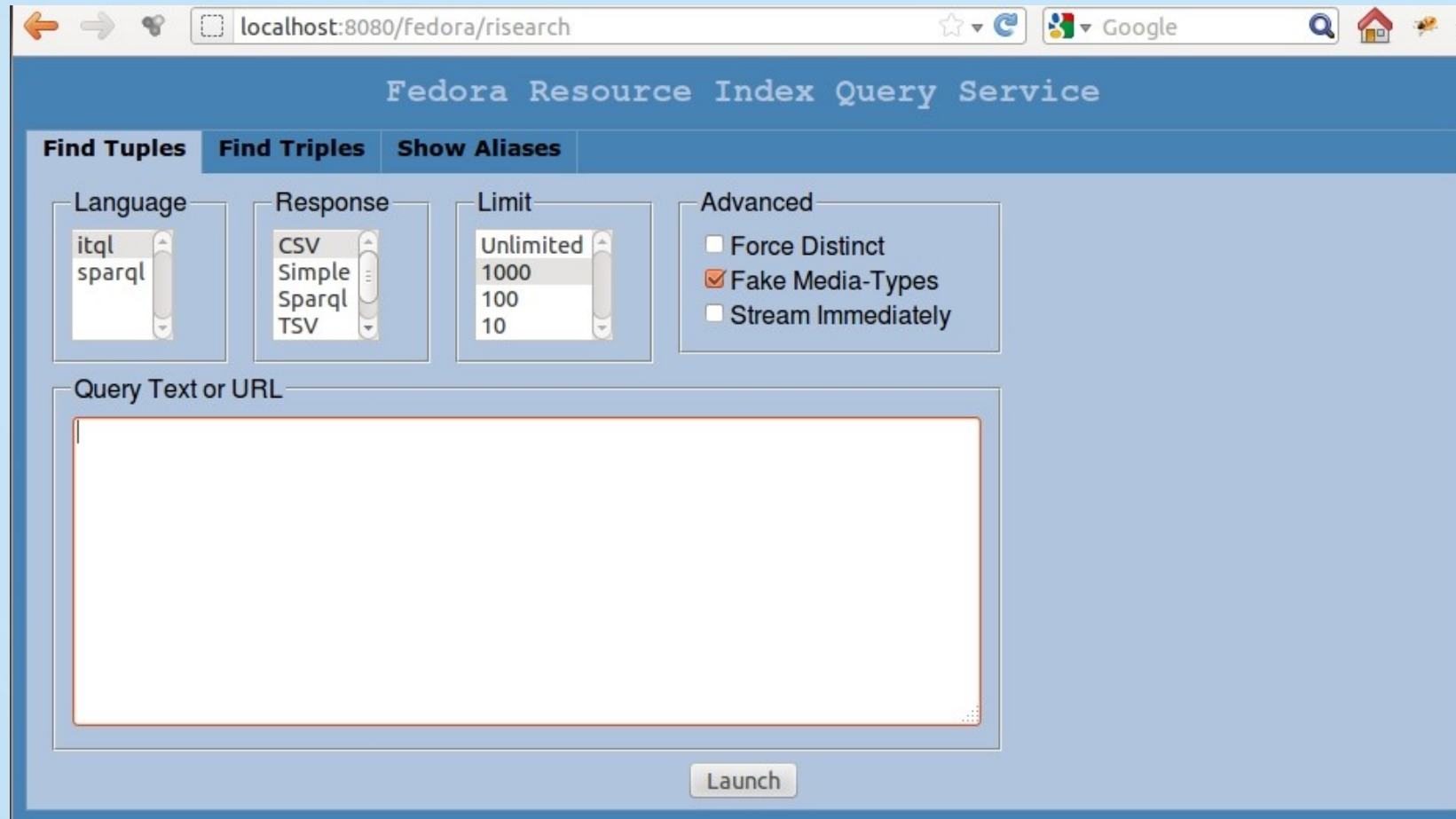


```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-06-19T13:20:48Z</responseDate>
  <request verb="Identify">http://localhost:8080/fedora/oai</request>
  - <Identify>
    <repositoryName>Your Fedora Repository Name Here</repositoryName>
    <baseURL>http://localhost:8080/fedora/oai</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>bob@example.org</adminEmail>
    <adminEmail>oai-admin@example.org</adminEmail>
    <earliestDatestamp>2012-06-19T11:20:48Z</earliestDatestamp>
    <deletedRecord>no</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
  - <description>
    - <friends xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/ http://www.openarchives.org/OAI/2.0/friends.xsd">
      <baseURL>http://arXiv.org/oai2</baseURL>
      <baseURL>http://memory.loc.gov/cgi-bin/oai2_0</baseURL>
    </friends>
  </description>
  - <description>
    - <oai-identifier xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
      <scheme>oai</scheme>
      <repositoryIdentifier>example.org</repositoryIdentifier>
      <delimiter>:</delimiter>
      <sampleIdentifier>oai:example.org:changeme:7654</sampleIdentifier>
    </oai-identifier>
  </description>
</Identify>
</OAI-PMH>
```

- Supports only DC (Dublin Core)

2.5.2 RI-Search

<http://localhost:8080/fedoragsearch/rest>



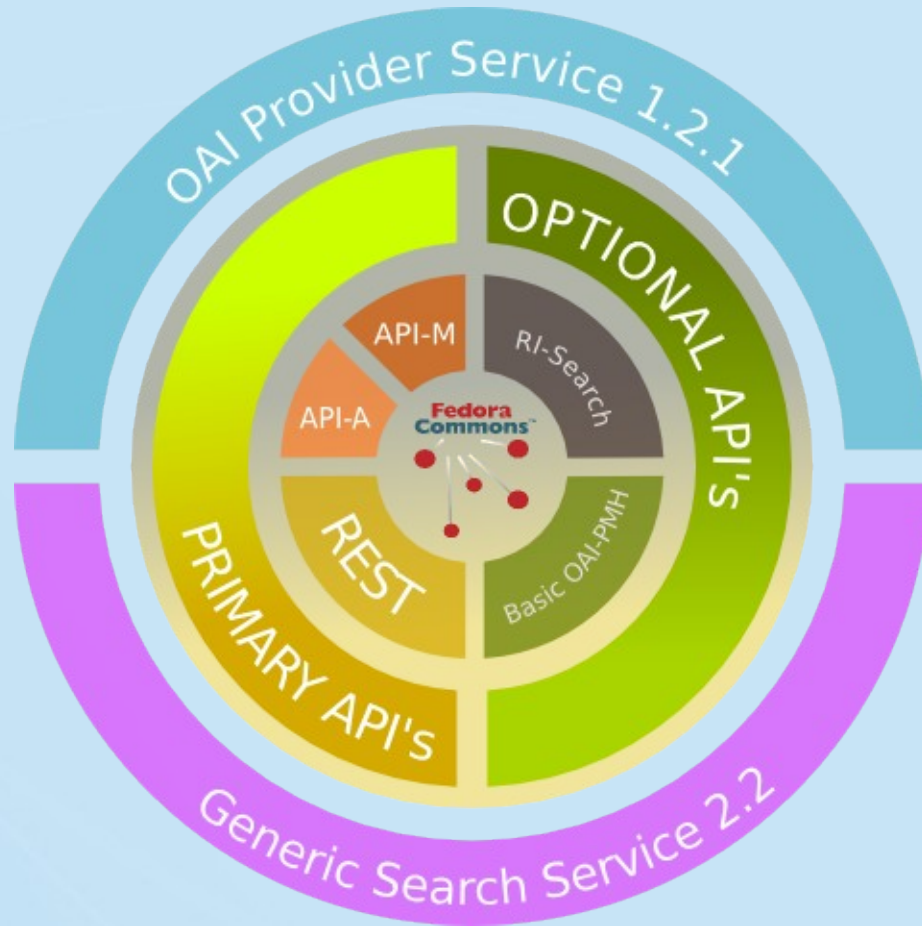
2.5.2 RI-Search

Example iTQL Query:

Find the children of the uppermost set demo:root with paging.

```
select $object $label $description $owner $date $type
from <#ri> where
$object <fedora-model:label> $label
and $object <dc:description> $description
and $object <fedora-model:ownerId> $owner
and $object <dc:date> $date
and $object <dc:type> $type
and $object <fedora-rels-ext:isMemberOf> <info:fedora/demo:root>
order by $date asc limit 12 offset 0
```


2.6 Framework Services



- Generic Search Service
- OAI Provider Service

2.6.1 Generic Search Service (GSearch)

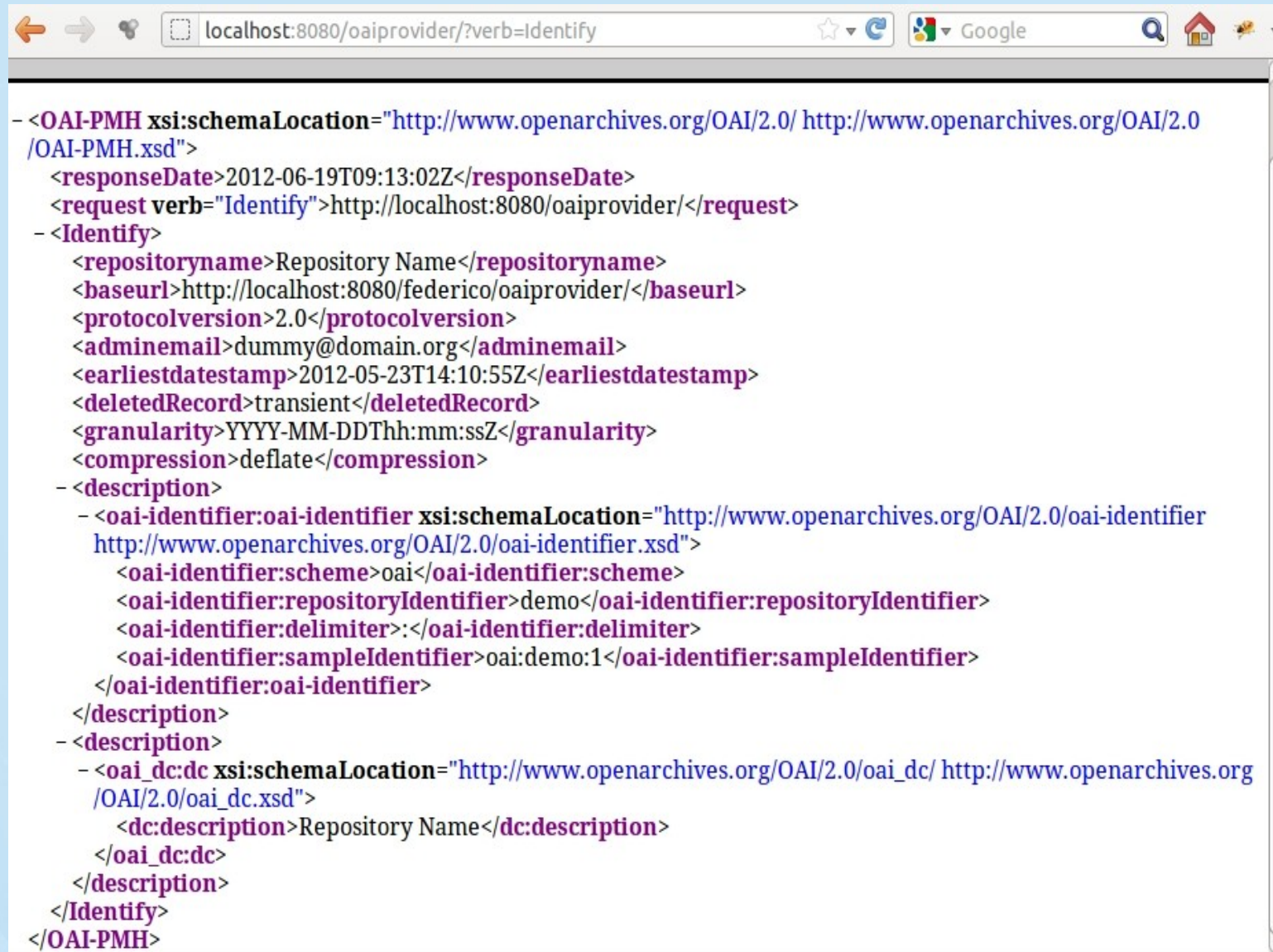
http://localhost:8080/fedoragsearch/rest



The screenshot shows a web browser window with the address bar containing the URL `localhost:8080/fedoragsearch/rest?operation=gfindObjects`. The page title is "REST Client Demo of Fedora Generic Search Service (configDemoOnSolr)". The page features a navigation menu with links for [updateIndex](#), [gfindObjects](#), [browseIndex](#), [getRepositoryInfo](#), and [getIndexInfo](#), followed by a response time of "(0 milliseconds)". Below the menu, the "gfindObjects" section contains a search form with a "Query:" input field, a "Hit page size:" input field, and a "Search" button. The footer includes the copyright notice "Copyright © 2008 Technical University of Denmark, Fedora Project" and the modification date "Last Modified 2006-10-13 by Gert Schmeltz Pedersen".

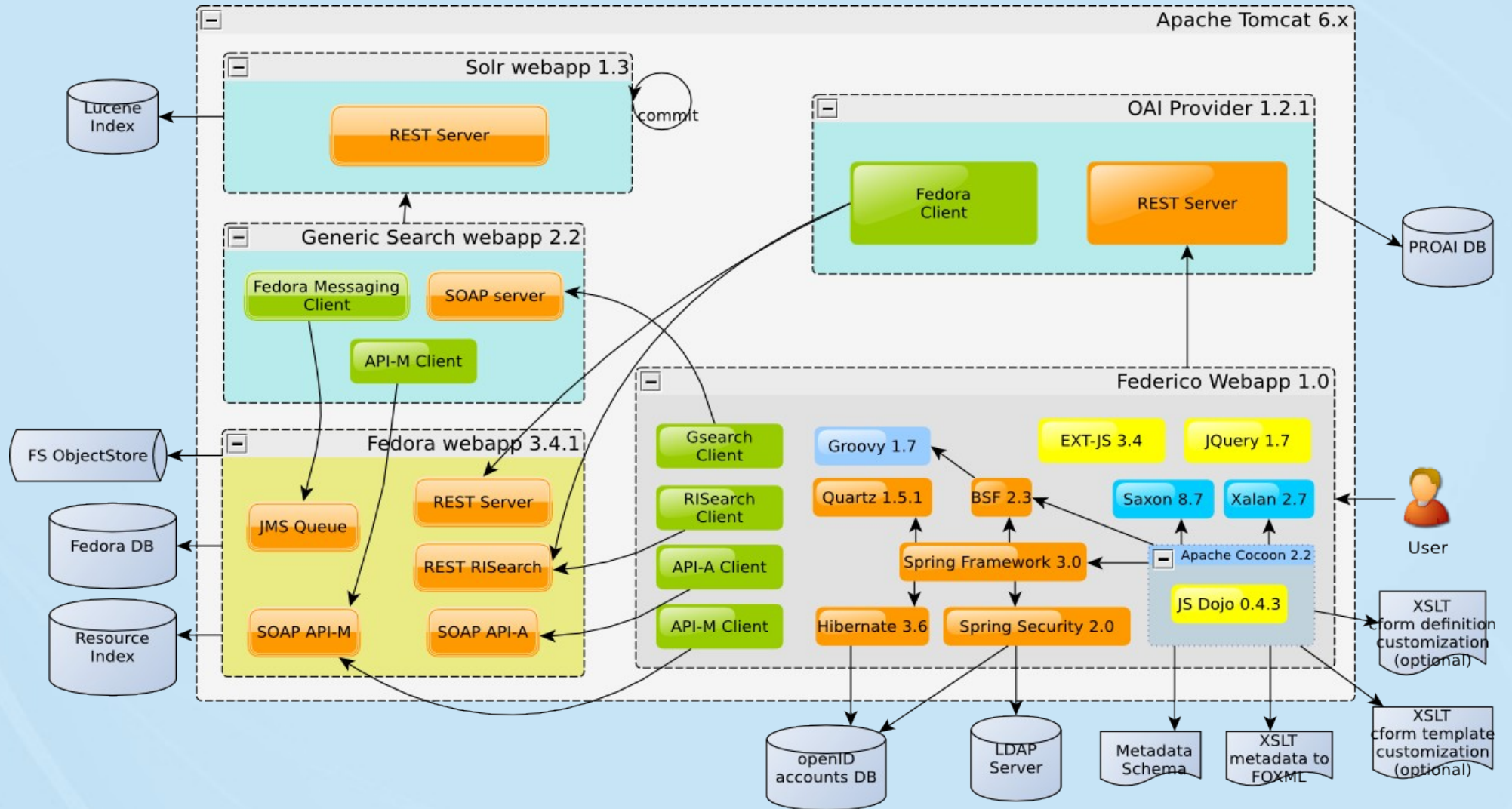
2.6.2 OAI Provider Service (PROAI)

http://localhost:8080/oaiprovider/?verb=Identify



```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-06-19T09:13:02Z</responseDate>
  <request verb="Identify">http://localhost:8080/oaiprovider/</request>
  -<Identify>
    <repositoryname>Repository Name</repositoryname>
    <baseurl>http://localhost:8080/federico/oaiprovider/</baseurl>
    <protocolversion>2.0</protocolversion>
    <adminemail>dummy@domain.org</adminemail>
    <earliestdatestamp>2012-05-23T14:10:55Z</earliestdatestamp>
    <deletedRecord>transient</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <compression>deflate</compression>
    -<description>
      -<oai-identifier:oai-identifier xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
        <oai-identifier:scheme>oai</oai-identifier:scheme>
        <oai-identifier:repositoryIdentifier>demo</oai-identifier:repositoryIdentifier>
        <oai-identifier:delimiter>:</oai-identifier:delimiter>
        <oai-identifier:sampleIdentifier>oai:demo:1</oai-identifier:sampleIdentifier>
      </oai-identifier:oai-identifier>
    </description>
    -<description>
      -<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
        <dc:description>Repository Name</dc:description>
      </oai_dc:dc>
    </description>
  </Identify>
</OAI-PMH>
```

2.6.3 Federico: Architecture



2.7 Security

Authentication

- Security Filters in fedora web application web.xml
 - XmlUserfileFilter (default)
`$FEDORA_HOME/server/config/fedora-users.xml`
 - LdapFilterforAttributes (optional)
 - LdapFilterforGroups (optional)
- Fedora Security Layer (FeSL)
 - New and experimental
 - Based on JAAS (Java Authentication and Authorization Service)
 - `$FEDORA_HOME/server/config/jaas.conf`

Authorization

- XACML Policy Enforcement
 - `$FEDORA_HOME/data/fedora-xacml-policies/repository-policies/default`
 - Definition of repository-wide policies and object-specific policies
 - Each XACML policy defines:
 - (1) a "target" describes what the policy applies to (by referring to attributes of users, operations, objects, datastreams, dates, and more) and
 - (2) one or more "rules" to permit or deny access
- Fedora Security Layer (FeSL)
 - Based on XAMCL
 - Save policies in datastreams of digital objects

See more on <https://wiki.duraspace.org/display/FEDORA34/Security>

Summary

- Fedora as repository for digital information in research environment
 - Well defined API's
 - Content Model Architecture for the definition of “types” of objects
 - Harvesting through OAI-PMH
- Knowledge of XML is crucial
- Complex UI implementation

