

# DESY Computing Seminar on Data Management in Climate Research



**federico**

Ing. José A. Mejía Villar M.Sc.  
[jmejia@awi.de](mailto:jmejia@awi.de)

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

Hamburg, 31. January 2011

1. Fedora Commons Repository
2. Federico
3. Federico's Live Demo



# 1. Fedora Commons



- 1.1. What is Fedora Commons?
- 1.2. Key Features
- 1.3. Digital Object Model
- 1.4. Content Model Architecture
- 1.5. Web Service Interfaces
- 1.6. Framework Services



# 1.1 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**.



# 1.2 Key Features [1/2]

- 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.



# 1.2 Key Features [2/2]

- 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



# 1.3 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**.



# 1.3 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>
```





# 1.3 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



# 1.4 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.



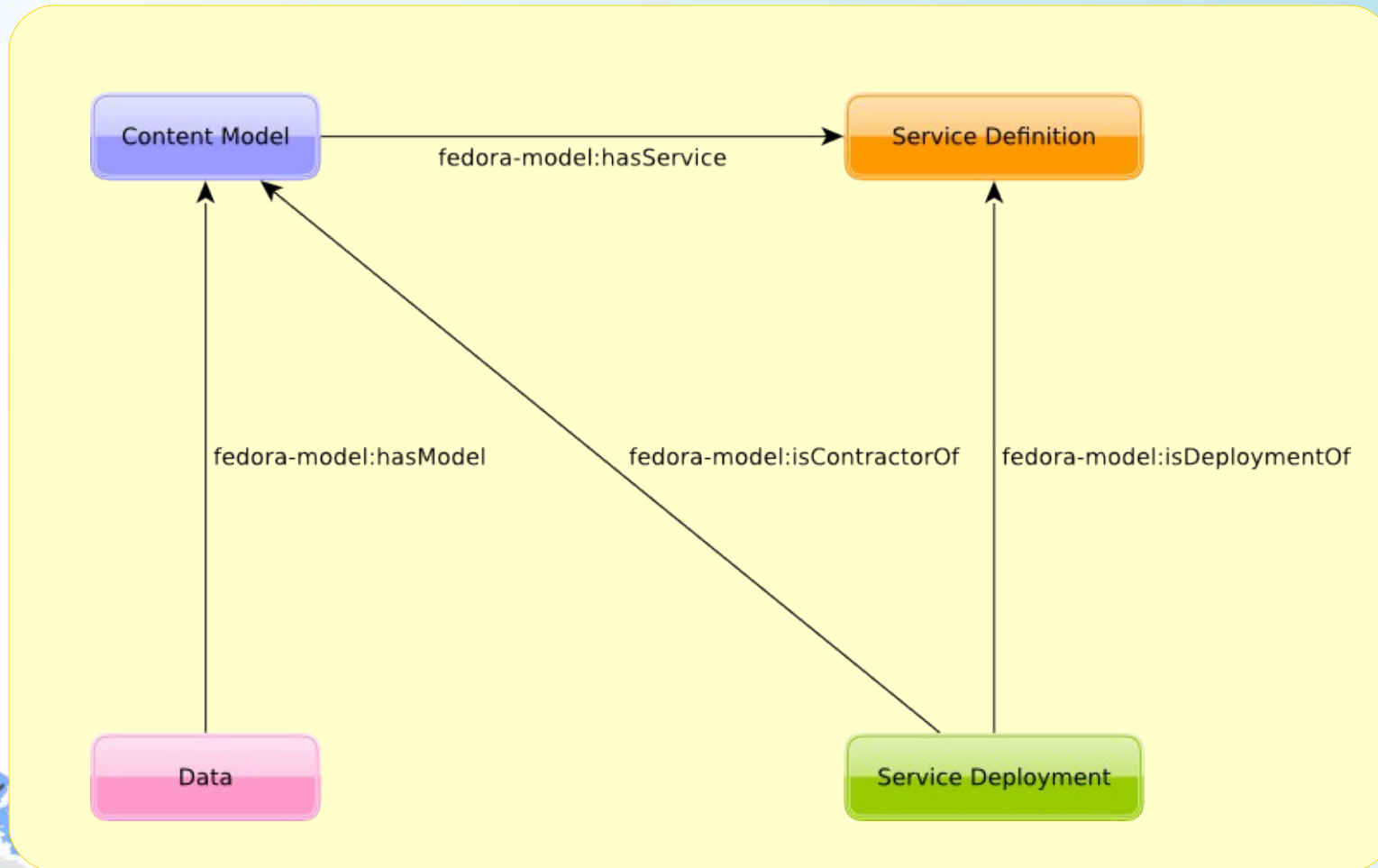
## Fundamental Fedora Object Types

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

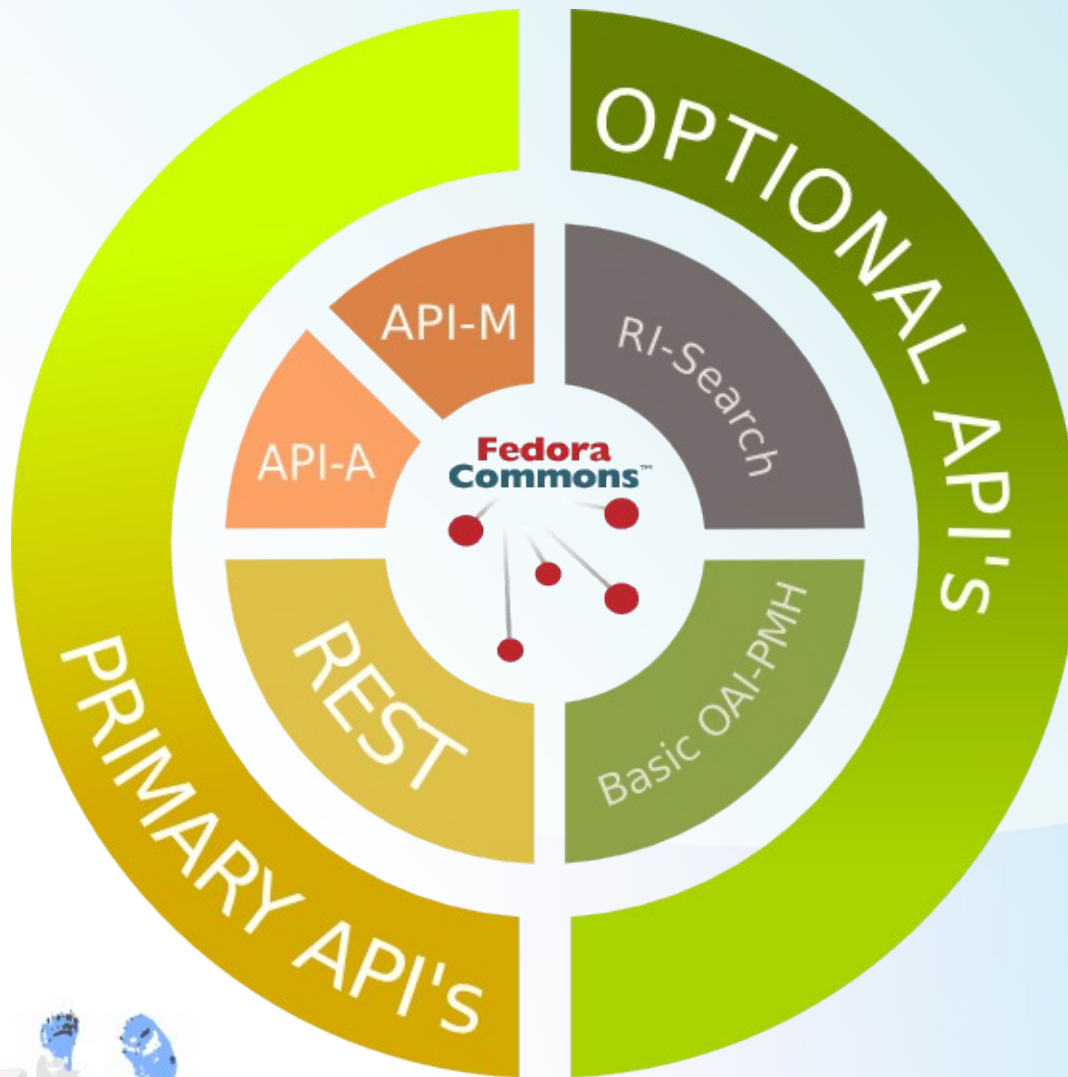


# 1.4 CMA: Object Types

## Fundamental CMA Relationships



# 1.5 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



## Fedora Access service methods

- Repository Access
  - DescribeRepository
- Object Access
  - findObjects
  - resumeFindObjects
  - getObjectHistory
  - getObjectProfile
- Datastream Access
  - getDatastreamDissemination
  - listDatastreams
- Dissemination Access
  - getDissemination
  - listMethods



## Fedora Management service methods

- Datastream Management
  - addDatastream
  - compareDatastreamChecksum
  - getDatastream
  - getDatastreamHistory
  - getDatastreams
  - modifyDatastreamByReference
  - modifyDatastreamByValue
  - setDatastreamState
  - setDatastreamVersionable
  - purgeDatastream
- Relationship Management
  - addRelationship
  - getRelationships
  - purgeRelationship
- Object Management
  - modifyObject
  - purgeObject
  - export
  - getNextPID
  - getObjectXML
  - ingest
  - validate



- The Resource Index Search Service (RISearch) is a web service that exposes the contents of a repository's Resource Index guide for outside use.

## RISearch Service Functionality

	Find Tuples	Find Triple
Query Language	SPARQL, iTQL	SPO
Response Type	CSV, Simple, Sparql, TSV, count	N-Triples, Notation 3 RDF/XML, Turtle, count



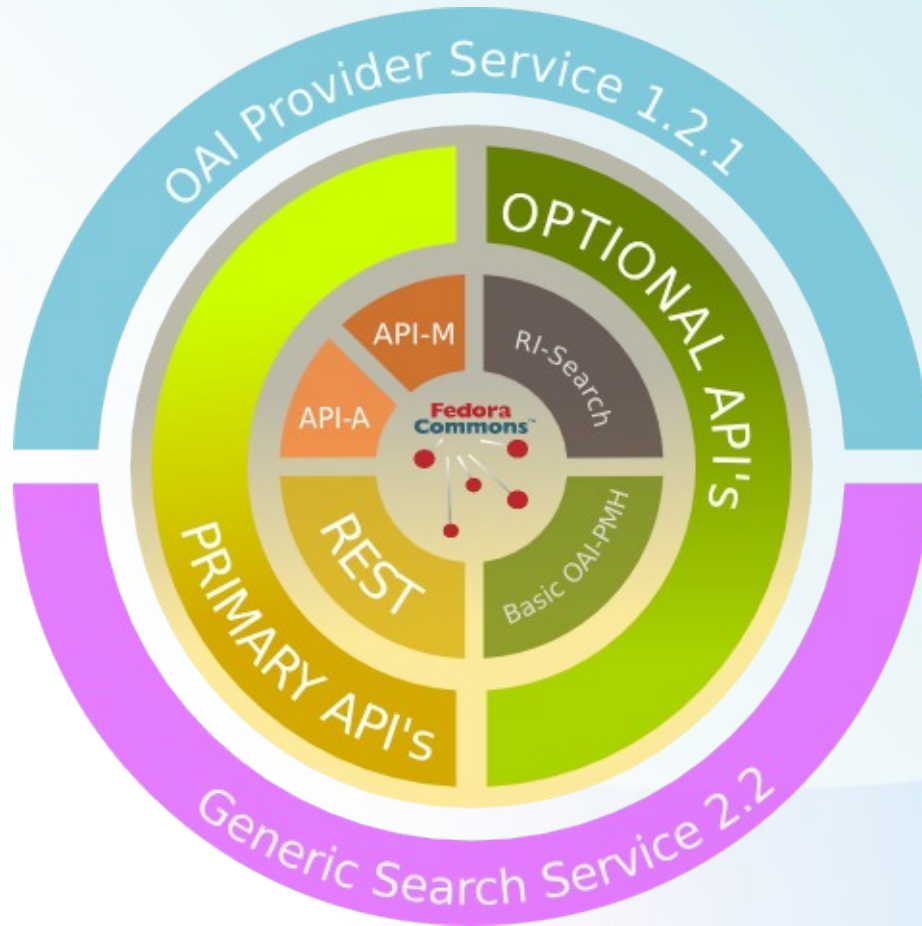


- Example iTQL Query:

```
select $object $label $description $owner $date $type from <#ri>  
where $object <fedora-model:label> $label  
and $object <fedora-model:ownerId> $owner  
and $object <dc:date> $date  
and $object <dc:description> $description  
and $object <dc:type> $type  
and $object <fedora-rels-ext:isMemberOf> <info:fedora/demo:1>  
order by $label asc limit 15 offset 12
```



# 1.6 Framework Services



- Generic Search Service
- OAI Provider Service





- 2.1. What is Federico?
- 2.2. System Requirements
- 2.3. Use Cases
- 2.4. Content Model
- 2.5. Architecture



# 2.1 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**



# 2.2 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



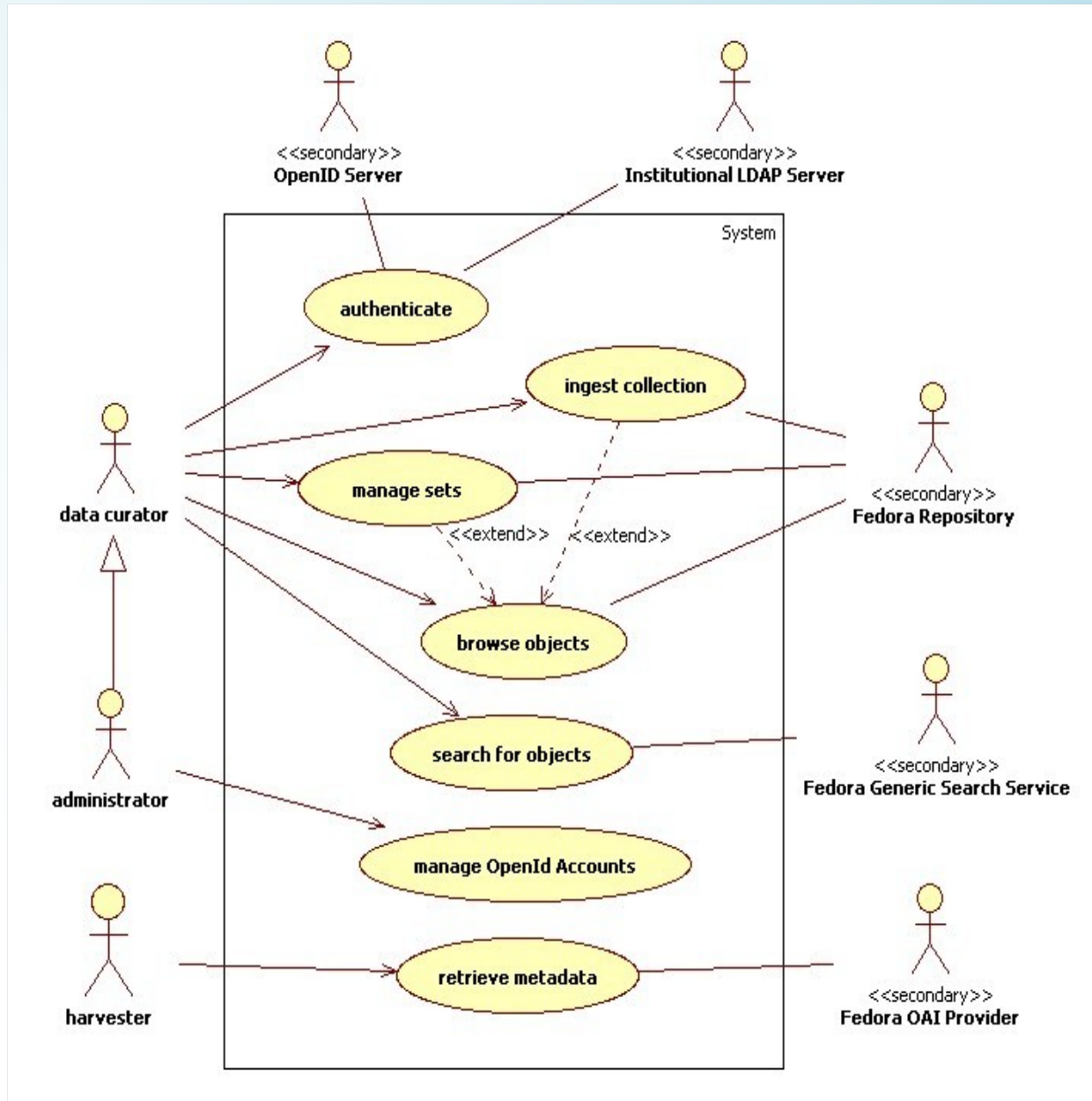
# 2.2 System Requirements [2/2]

## User

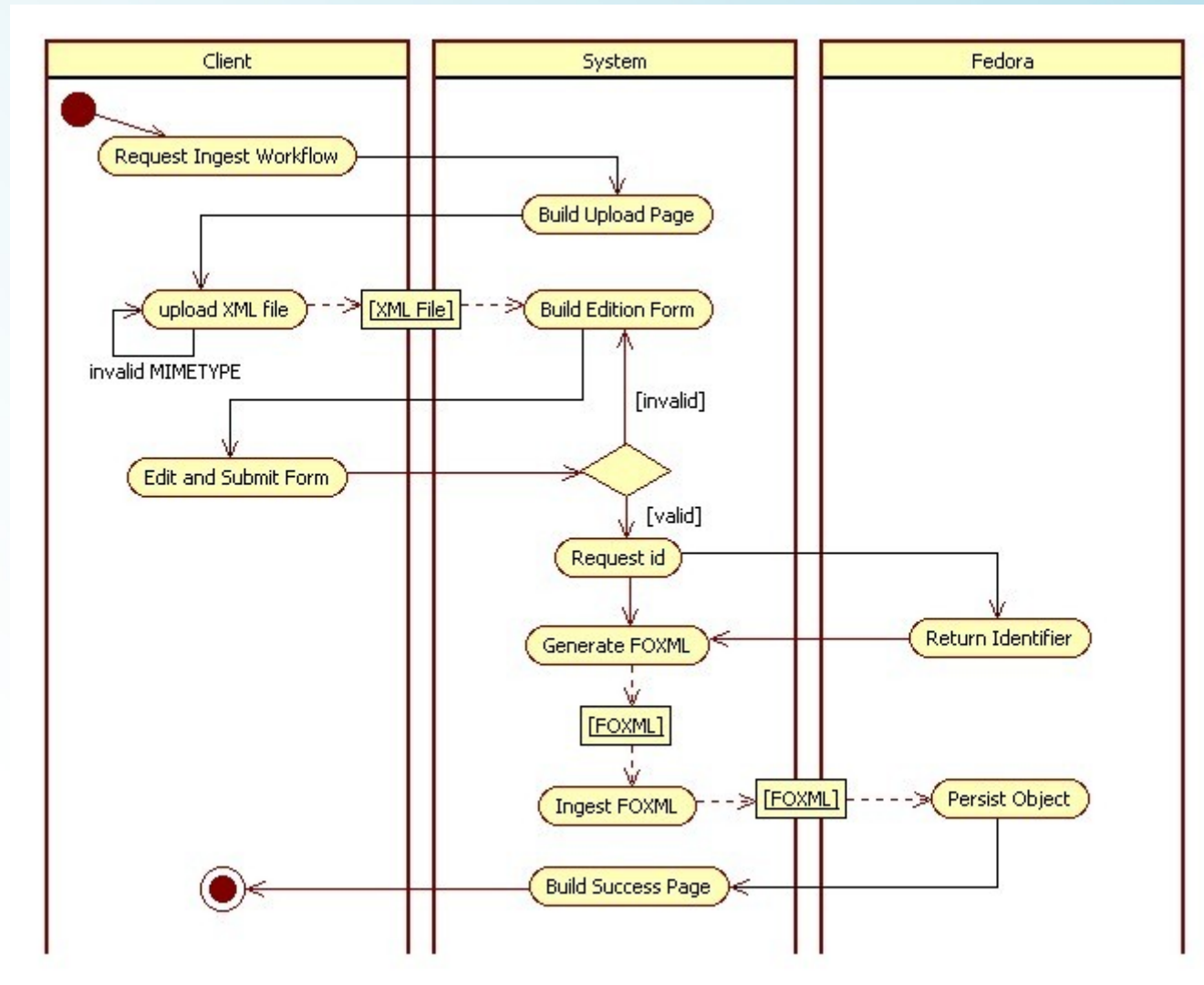
- PC with graphical interface and network card
- Keyboard and mouse
- Browser (preferably Mozilla Firefox) with Javascript enabled



# 2.3 Federico Use Case

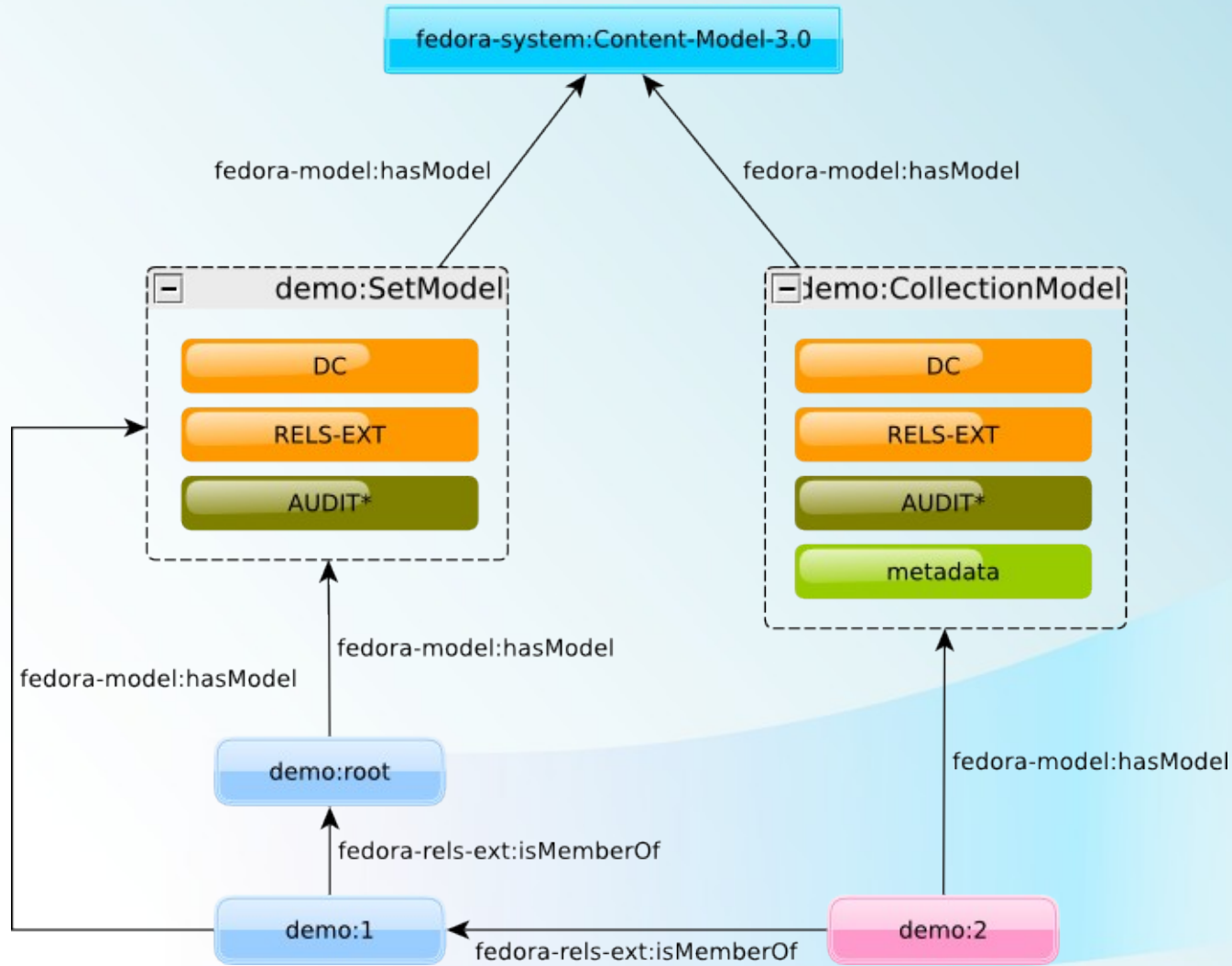


# 2.3 Activity Diagram – Ingest Collection

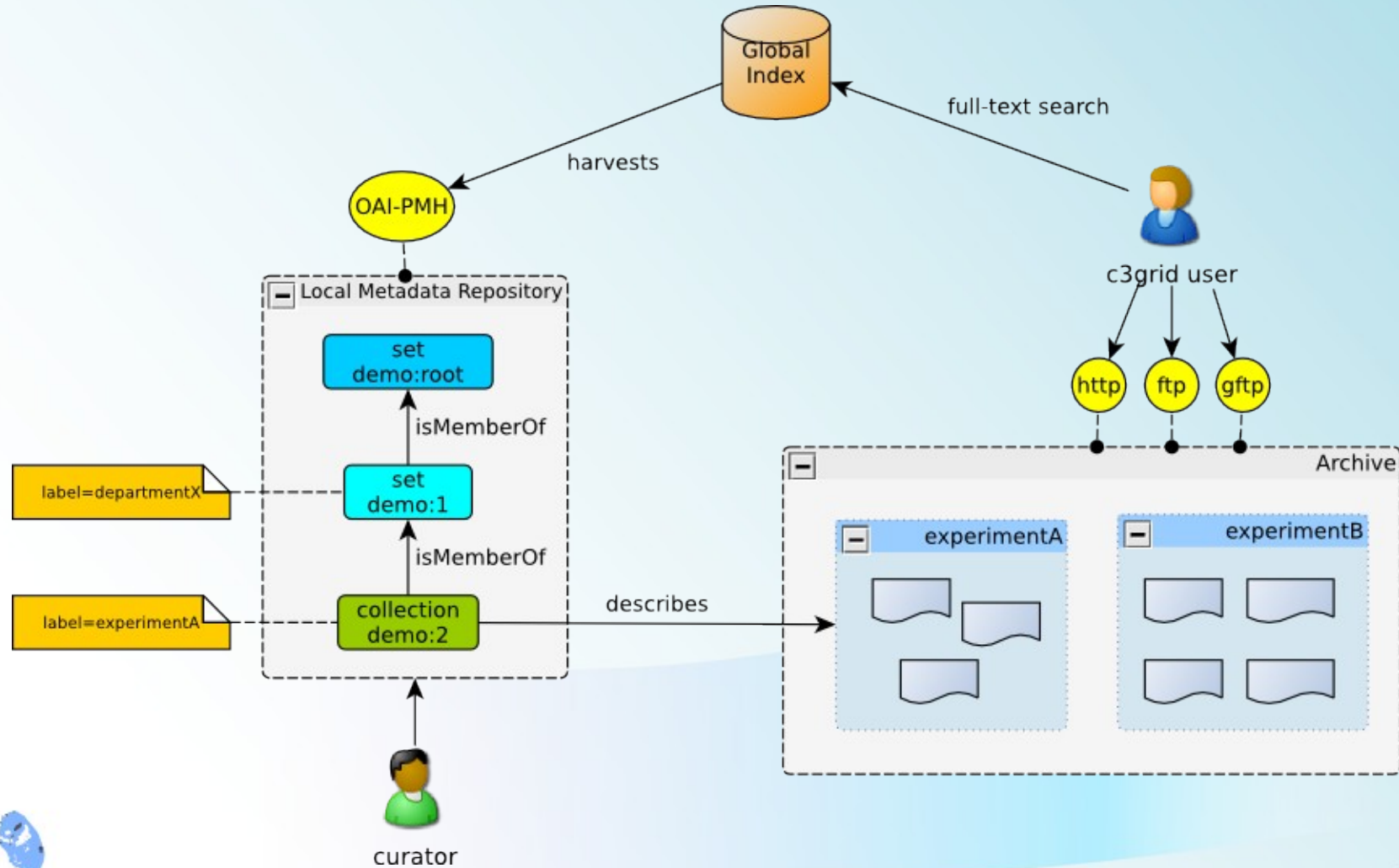




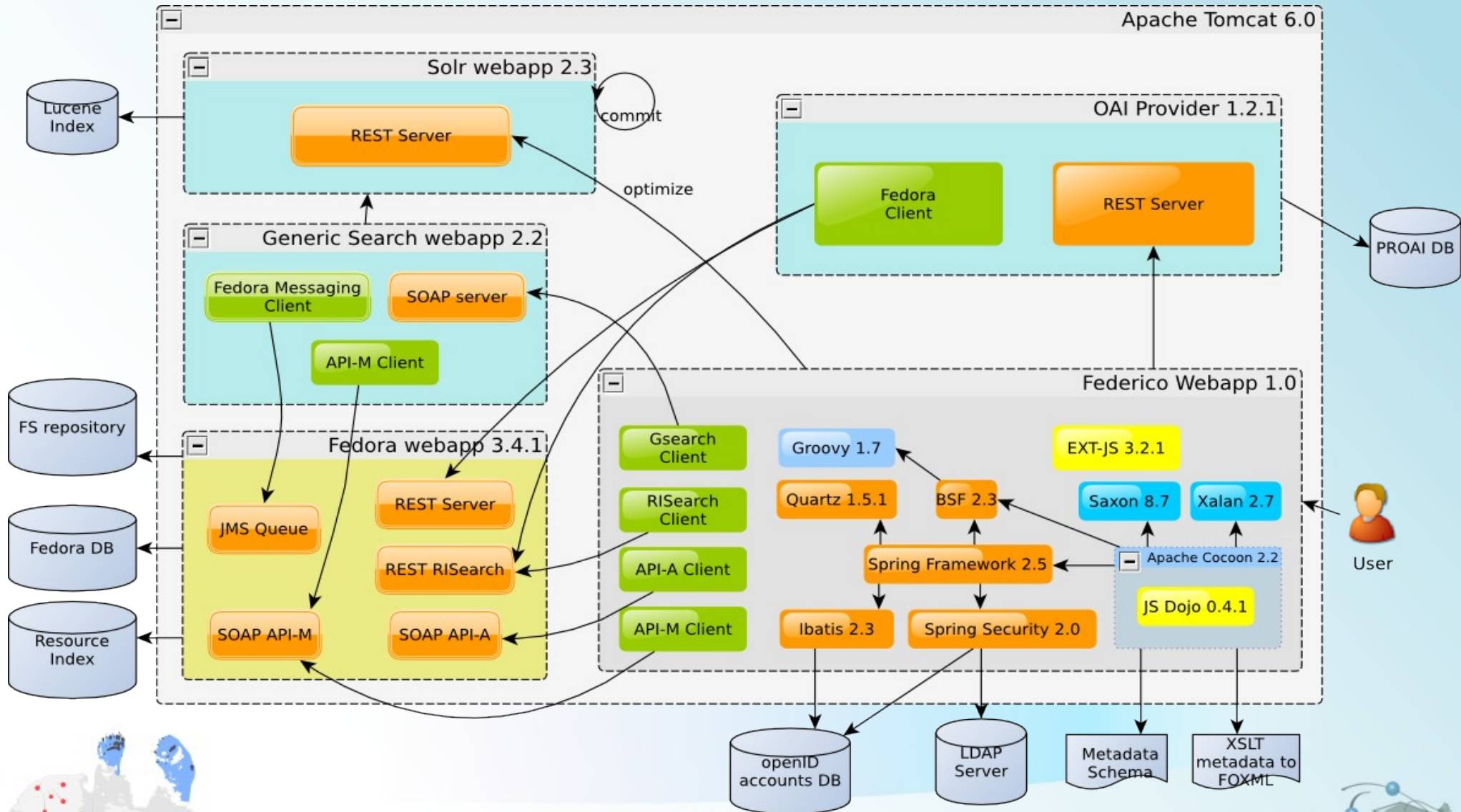
# 2.4 Content Model [1/2]



# 2.4 Content Model [2/2]



# 2.5 Architecture



# 3. Federico's Live Demo

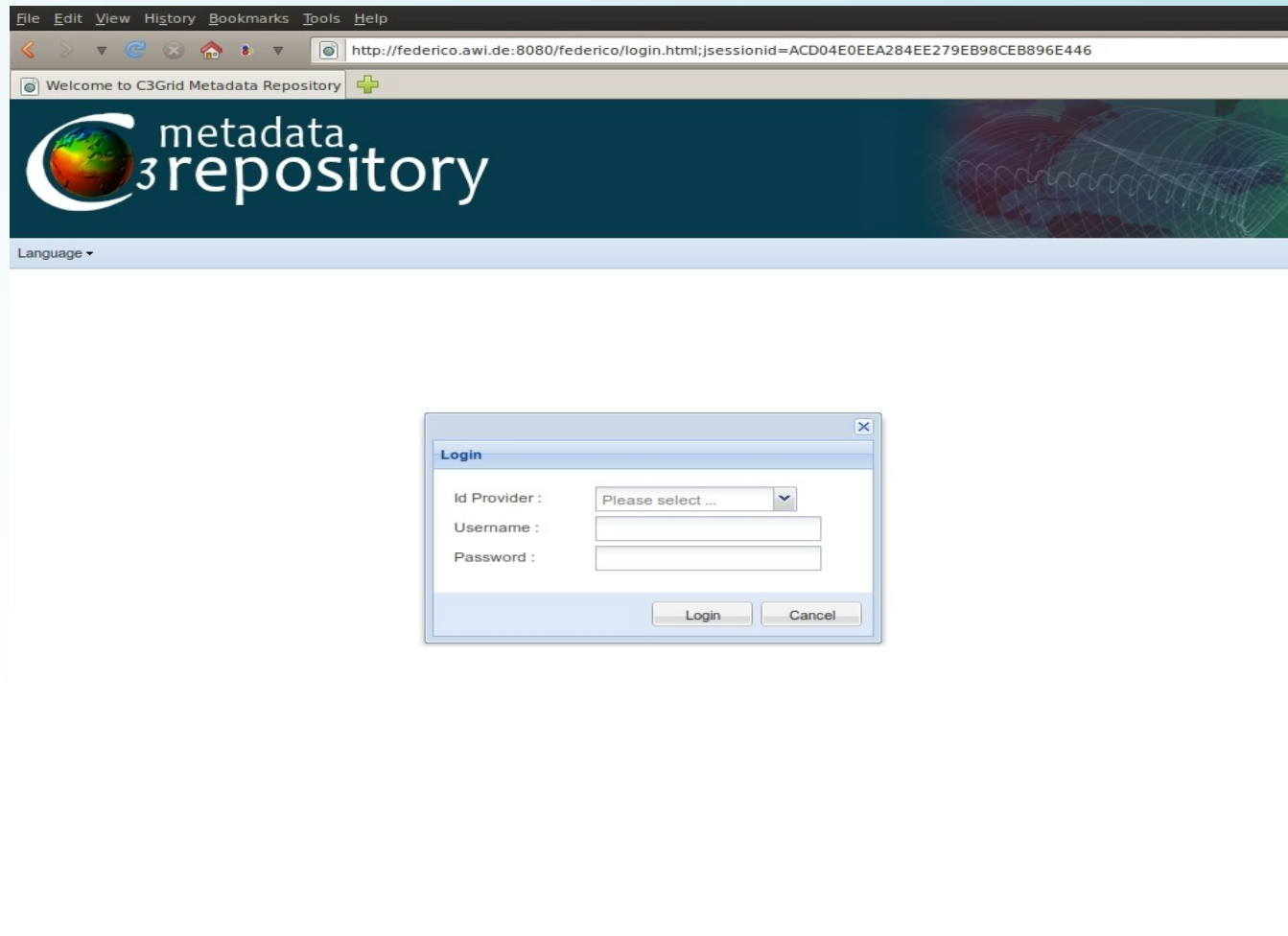


- 3.1. User Interface
- 3.2. Authentication
- 3.3. Ingest Collections
- 3.4. Full-text Search
- 3.5. OAI-PMH



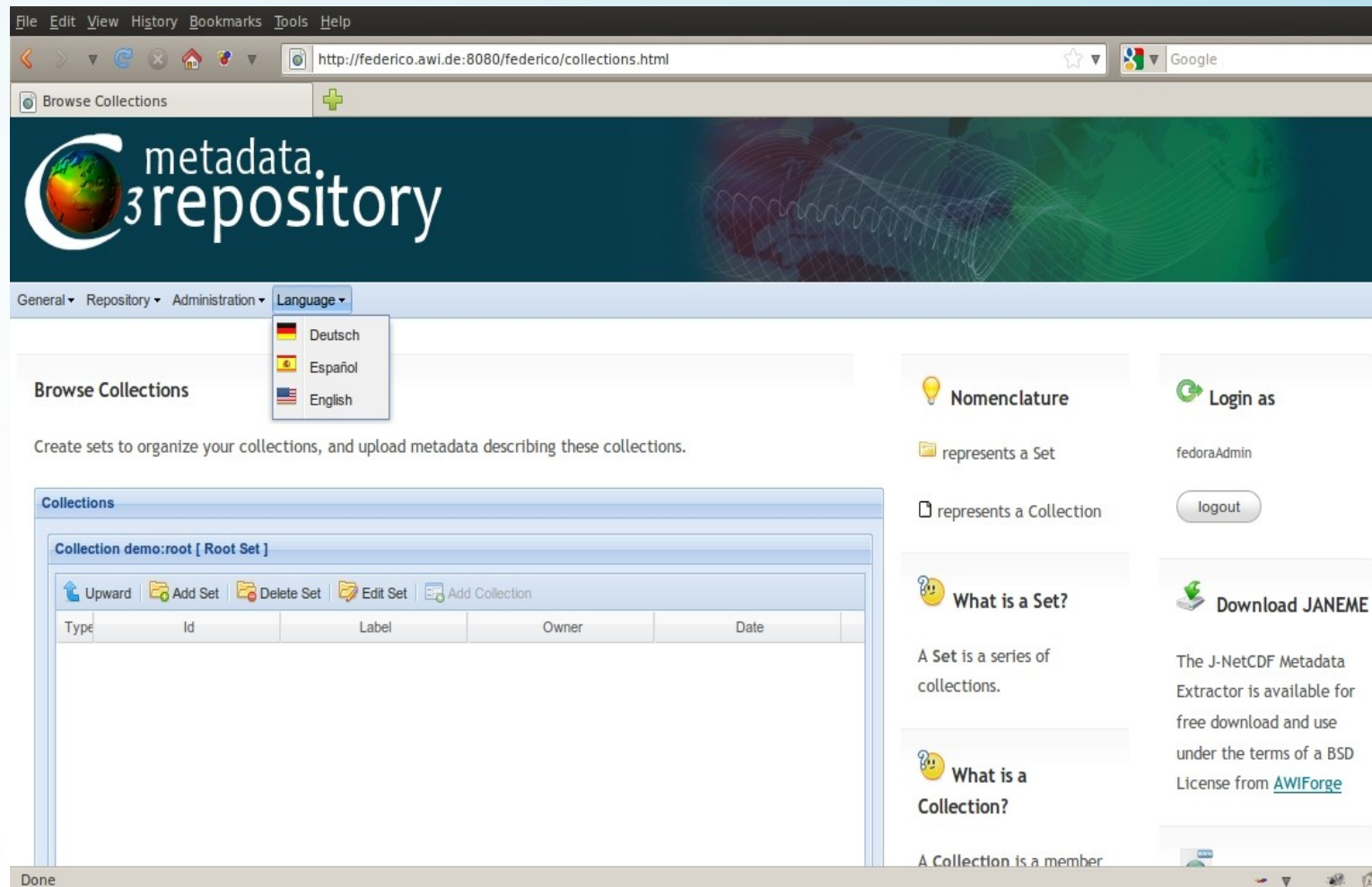
# 3. Federico's Live Demo: Screenshots

## Login Form



# 3. Federico's Live Demo: Screenshots

## Browse Collections Form



File Edit View History Bookmarks Tools Help

http://federico.awi.de:8080/federico/collections.html

Browse Collections

metadata repository

General Repository Administration Language

Deutsch  
Español  
English

**Browse Collections**

Create sets to organize your collections, and upload metadata describing these collections.

**Collections**

Collection demo:root [ Root Set ]

Upward Add Set Delete Set Edit Set Add Collection

Type	Id	Label	Owner	Date
------	----	-------	-------	------

Done

**Nomenclature**

represents a Set

represents a Collection

**What is a Set?**

A Set is a series of collections.

**What is a Collection?**

A Collection is a member

**Login as**

fedoraAdmin

logout

**Download JANEME**

The J-NetCDF Metadata Extractor is available for free download and use under the terms of a BSD License from [AWIForge](#)



# 3. Federico's Live Demo: Screenshots



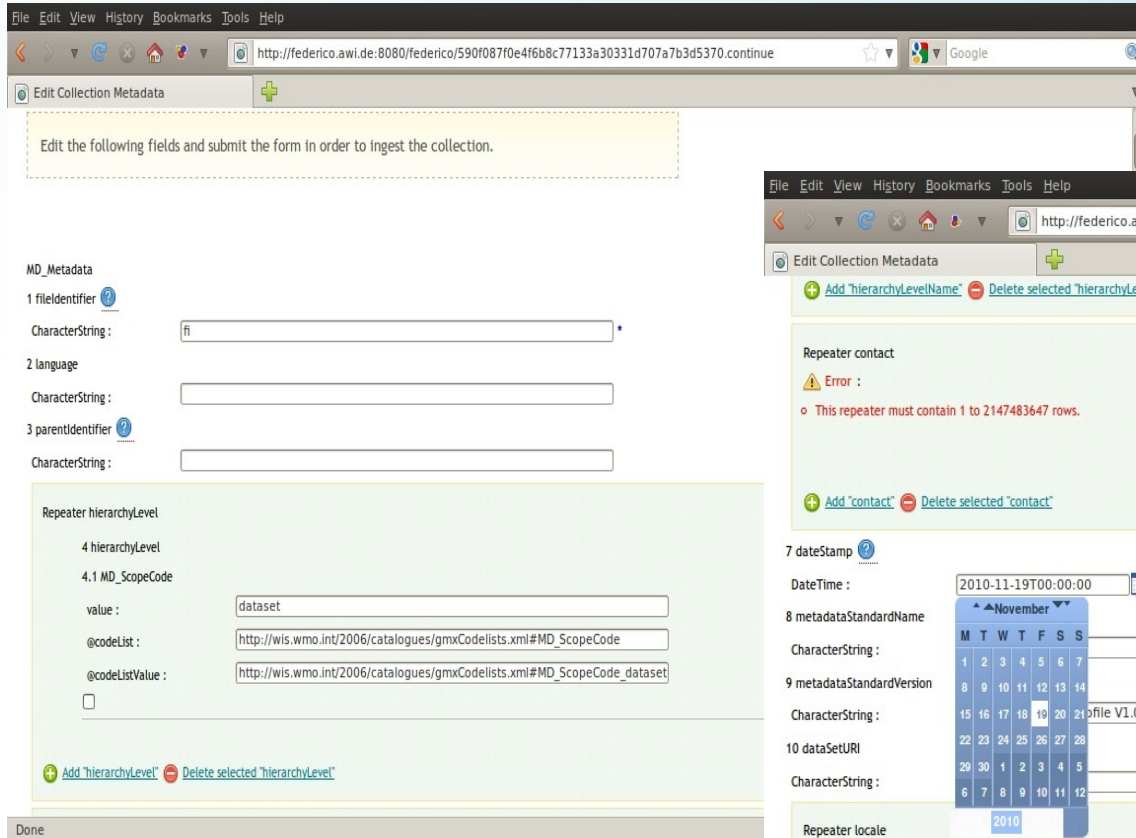
## Metadata Upload

The screenshot shows a web browser window displaying the 'Upload Collection Metadata' page of the Federico Metadata Repository. The browser's address bar shows the URL: `http://federico.awi.de:8080/federico/upload_collection.html?parent=demo:1`. The page header includes the repository logo and navigation menus for 'General', 'Repository', 'Administration', and 'Language'. The main content area is titled 'Upload Collection Metadata' and contains a yellow dashed box with instructions: '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, there is a file upload field showing 'Upload File [fill.xml] ...' and a status message: 'Successful upload of fill.xml, 35589 bytes'. A 'Submit' button is located at the bottom of the upload section. On the right side of the page, there are several informational panels: 'Nomenclature' (explaining that a folder represents a Set and a document represents a Collection), 'Login as' (showing the user is logged in as 'fedoraAdmin' with 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, available for free under a BSD License from AWIForge). A 'Done' status bar is visible at the bottom of the browser window.



# 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 : fi

2 language

CharacterString :

3 parentIdentifier

CharacterString :

Repeater hierarchyLevel

4 hierarchyLevel

4.1 MD\_ScopeCode

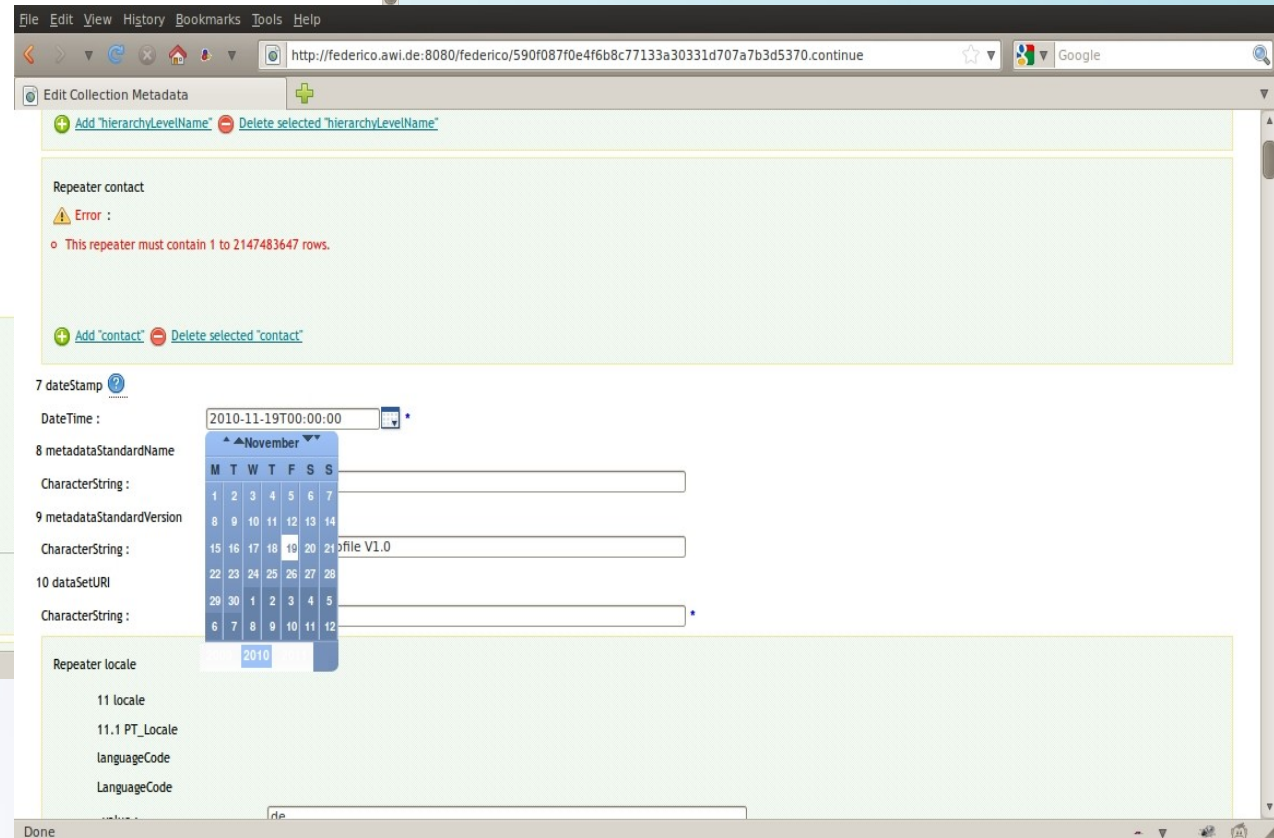
value : dataset

@codeList : http://wis.wmo.int/2006/catalogues/gmxCodeLists.xml#MD\_ScopeCode

@codeListValue : http://wis.wmo.int/2006/catalogues/gmxCodeLists.xml#MD\_ScopeCode\_dataset

+ 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 dateStamp

DateTime : 2010-11-19T00:00:00

8 metadataStandardName

CharacterString :

9 metadataStandardVersion

CharacterString : 2010file V1.0

10 dataSetURI

CharacterString :

Repeater locale

11 locale

11.1 PT\_Locale

languageCode

LanguageCode

Done





- 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
- Difficult UI implementation

