

Earth Science Data in Digital Object Repository Architecture (ESDORA)

Metadata Driven Display

Goal

An Active Archive System that supports better preservation, provenance, and access.

Architecture

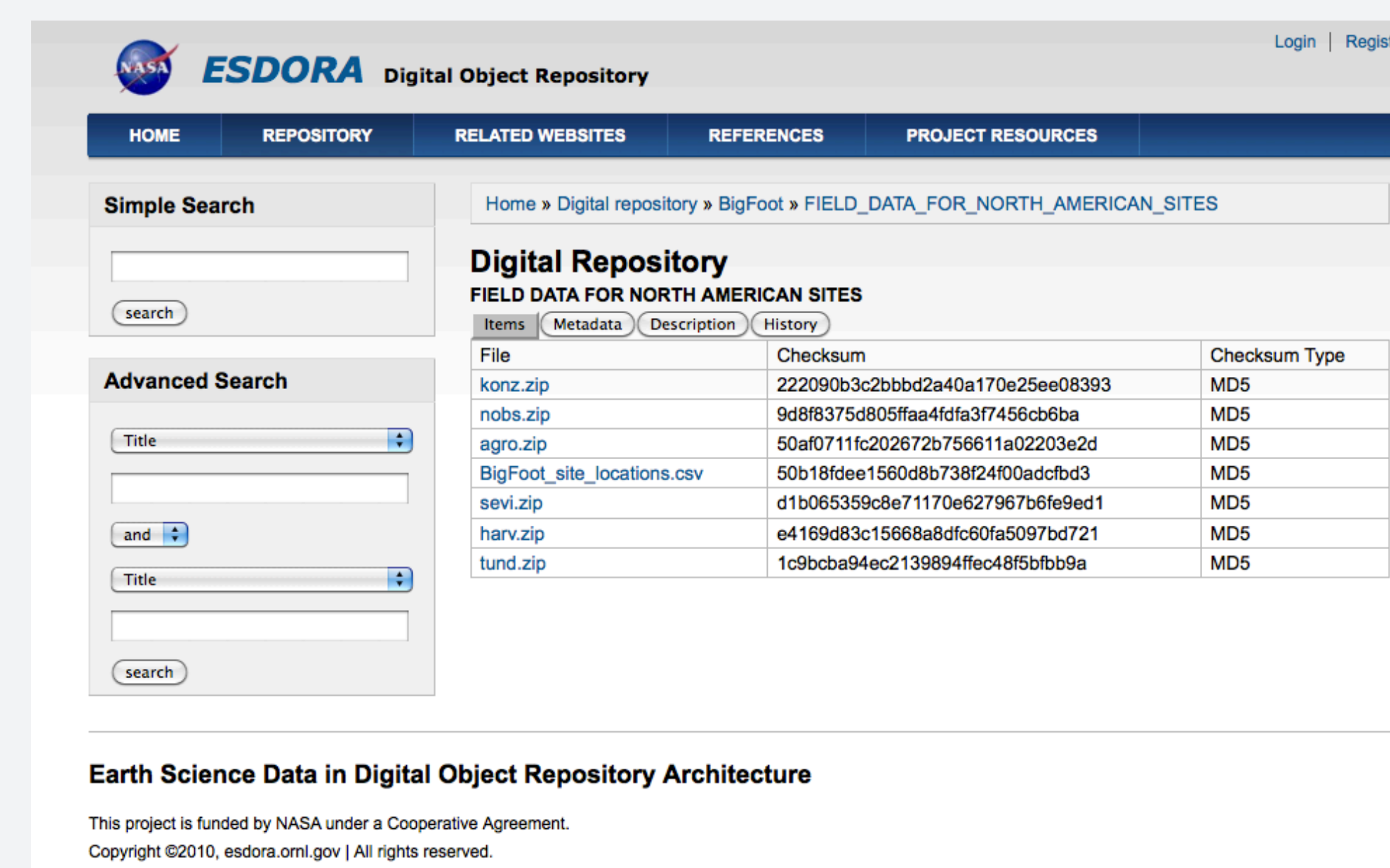
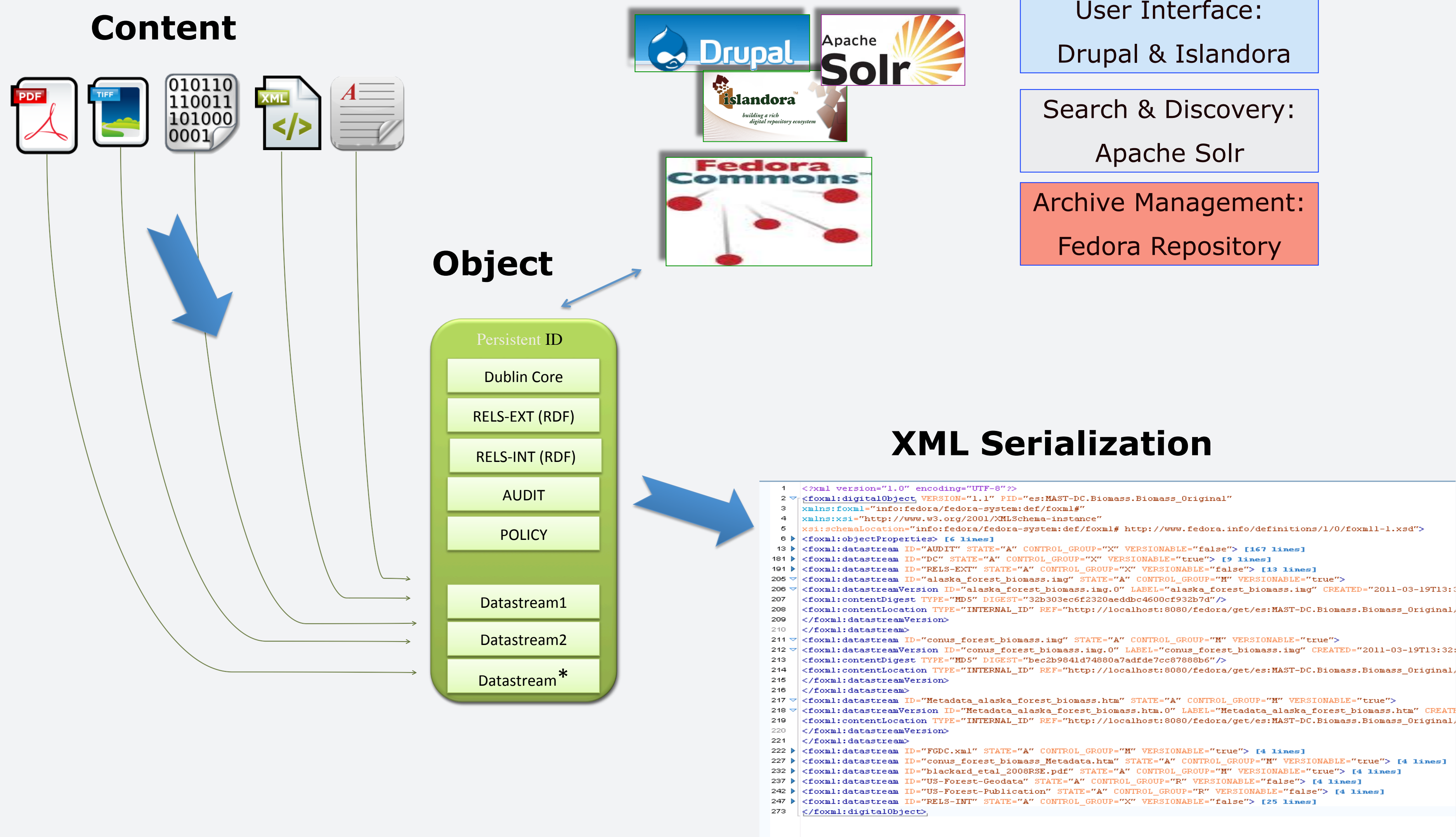
Data is modeled as digital objects – abstract containers for any digital artifacts, and a digital object is encoded in XML.

A generic RDF-based relationship model is used as a semantic store for inter- and intra object relationships.

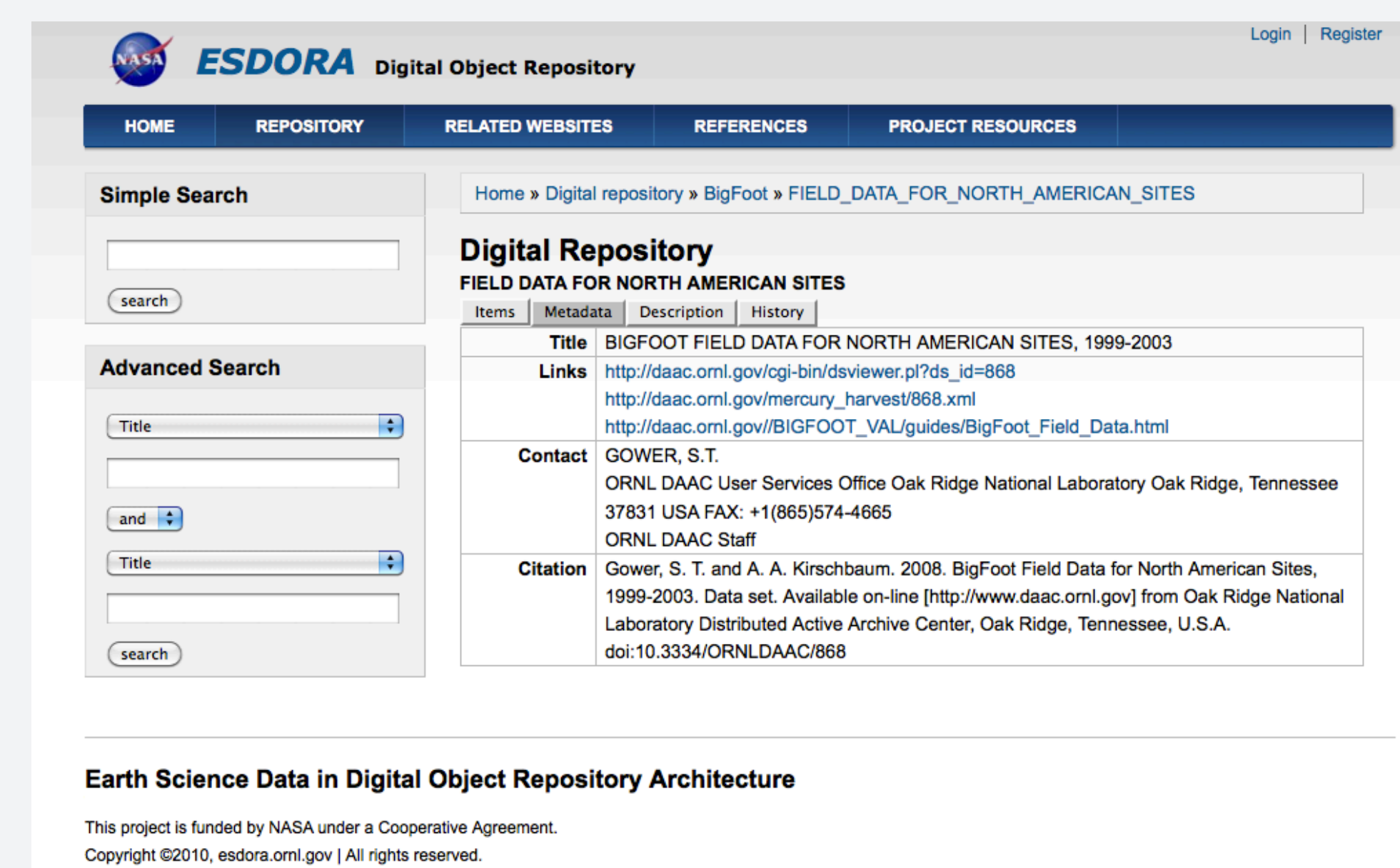
Open source frameworks are integrated together to store, manage, discover, and distribute object content.

Benefit

- Manage data, metadata consistently
- Preserve knowledge semantically
- Hide heterogeneity and promote interoperability
- Better provenance, stewardship, and preservation
- Ensure data integrity and quality in its life cycle in a repository

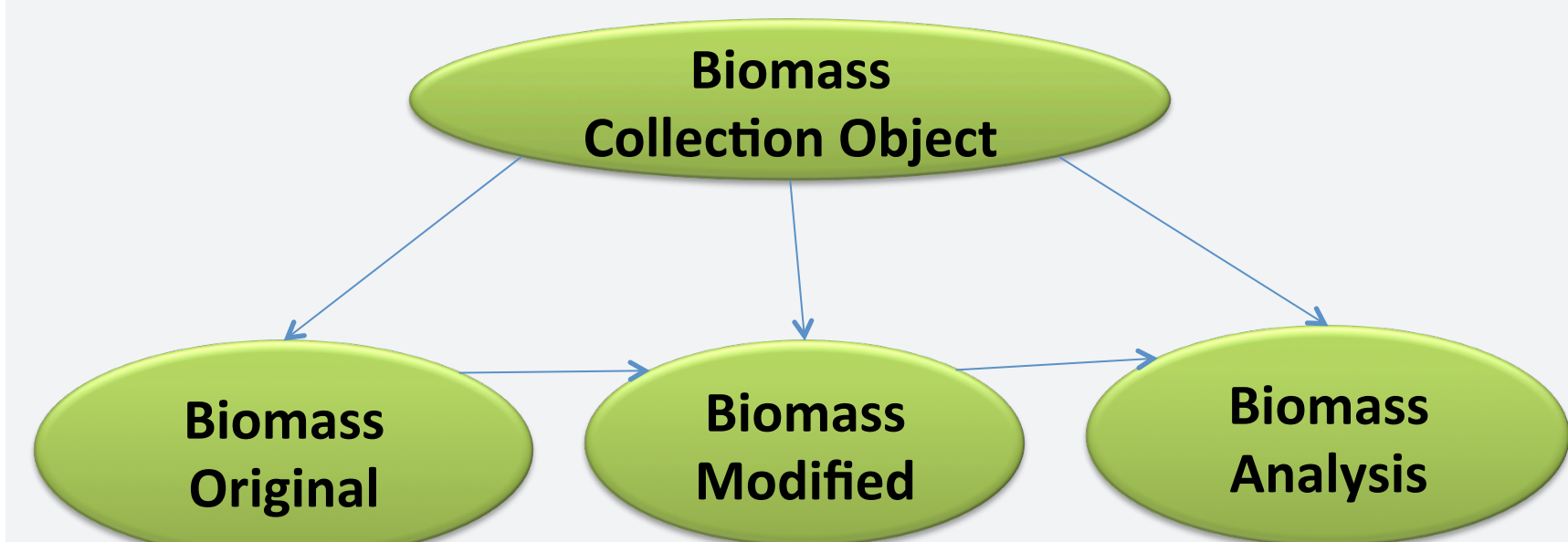


The Items Tab lists all data items, download links, and MD5 checksums



The Metadata Tab represents FGDC metadata for a data object, and the Description Tab lists non-structured companion files.

Data History



An example of inter-objects relations: "isMemberOf", "isDerivedFrom"

RDF Snippet

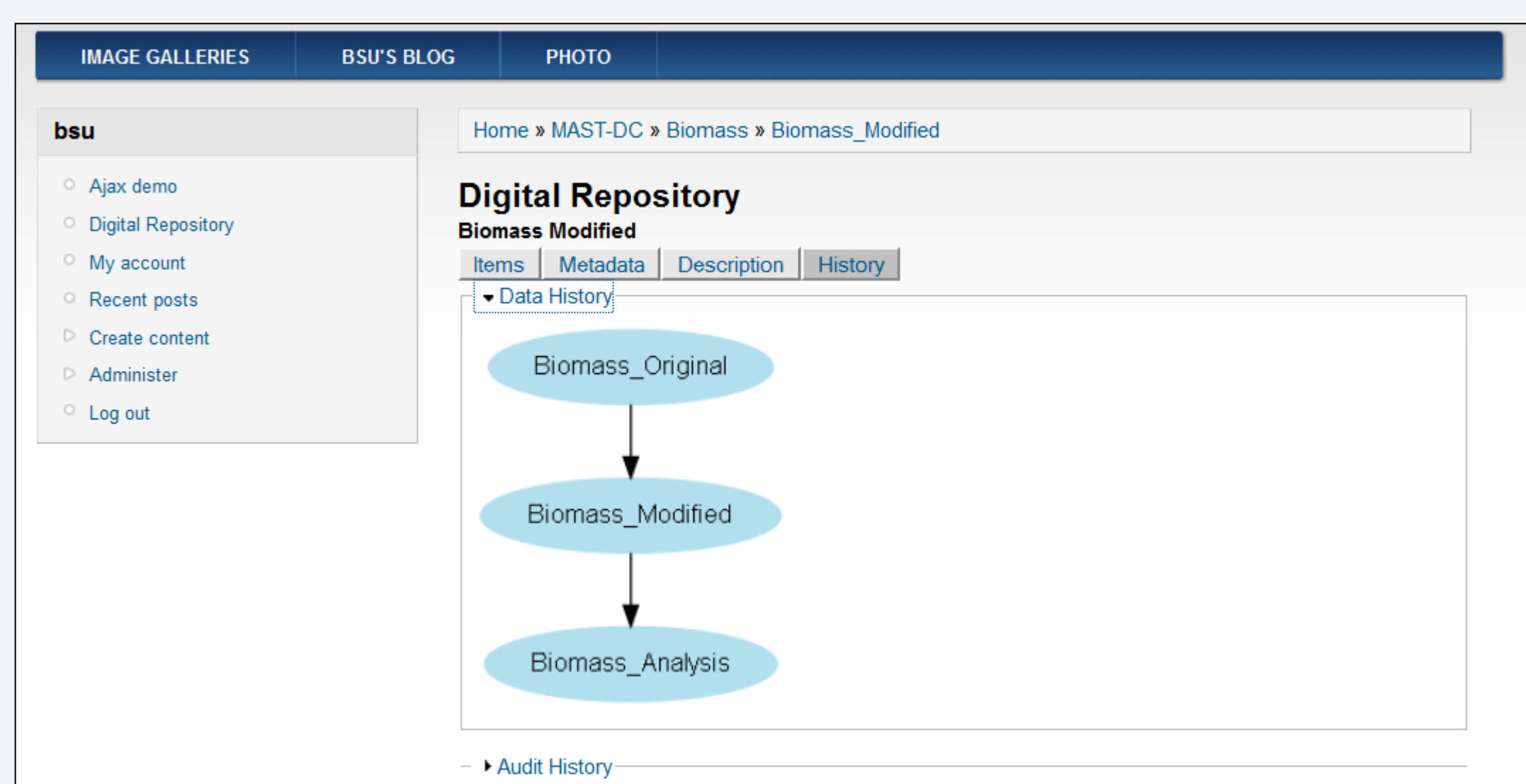
```
<rdf:RDF>
  <rdf:Description rdf:about="info:fedora/es:MAST-DC.Biomass.Biomass_Modified">
    <isMemberOf rdf:resource="info:fedora/es:MAST-DC.Biomass">
    <hasModel rdf:resource="info:fedora/es:defaultDataCModel"/>
    <isDerivedFrom rdf:resource="info:fedora/es:MAST-DC.Biomass.Biomass_Original"/>
  </rdf:Description>
</rdf:RDF>
```

Provenance information are recorded in RELS-EXT datastreams (RDF) and are stored in RDF semantic store.

The semantic store are indexed, and can be queried using SPARQL and iTQL

iTQL Query Sample

```
select $subject 'info:fedora/fedora-system:def:relations-external#isDerivedFrom' $object
from <?i> where walk(
$subject <info:fedora/fedora-system:def:relations-external#isDerivedFrom>
<info:fedora/es:MAST-DC.Biomass.Biomass_Modified>
and $subject <info:fedora/fedora-system:def:relations-external#isDerivedFrom> $object)
```



Data object derivation history

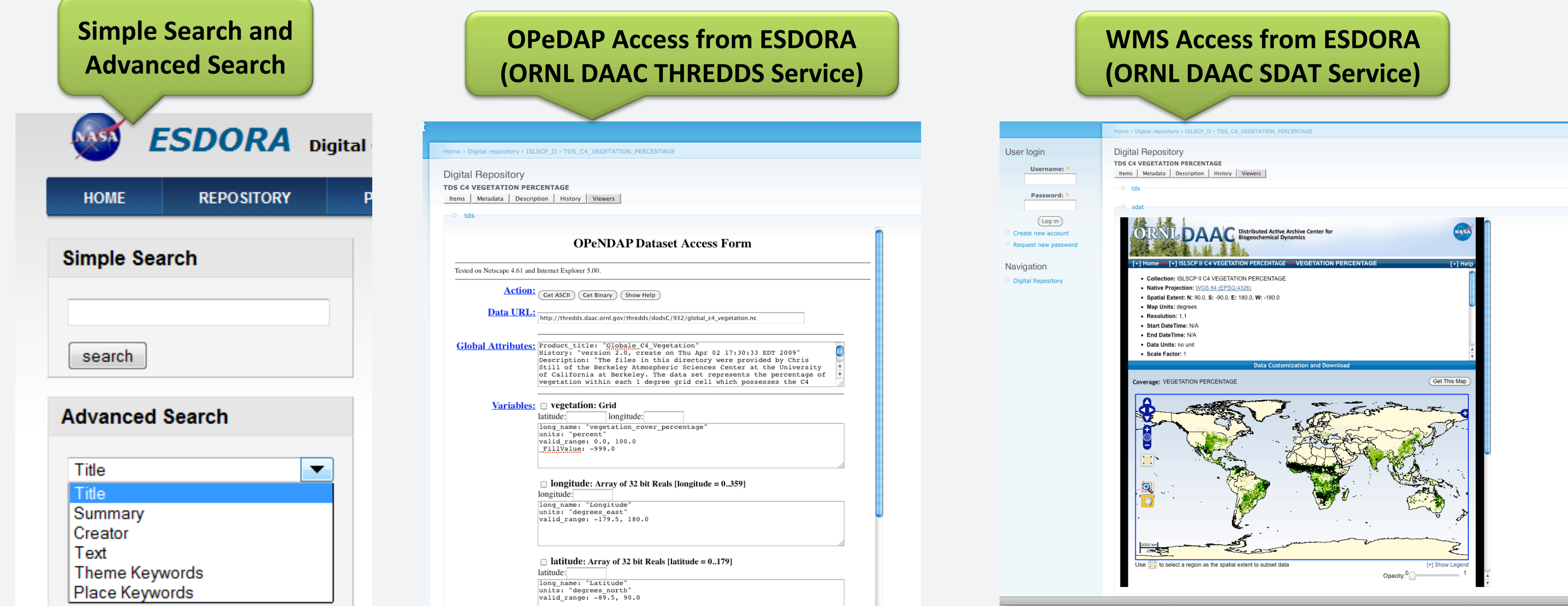
Versioning and Audit Trail

Content versioning and auditing trails are captured in the XML of the digital object model, and are presented on the GUI.

Count	Timestamp	Component	Action	Justification
1	2011-03-29T14:37:39.473Z		ingest	Initial object ingestion.
2	2011-03-29T14:42:32.495Z	us_forest_biomass.tif	addDatastream	Initial ingestion.
3	2011-03-29T14:44:34.207Z	FGDC.xml	addDatastream	Initial ingestion.
4	2011-03-29T14:44:34.415Z	MetadataLinks	addDatastream	Initial ingestion.
5	2011-03-29T14:44:34.869Z	RELS-INT	addDatastream	add RELS-INT for metadata
6	2011-04-07T21:04:48.452Z	RELS-EXT	modifyDatastreamByValue	

Audit Trails of all components of an object.

Discovery & Viewer Tools Integration



Summary

A repository system based on a digital object model, such as FEDORA used in this project, has a great potential to serve as a core technology to archive science data content. The digital object data model accommodates different content types and metadata standards in a uniform fashion, which is particularly applicable to Earth Science where the data formats and metadata standards are diverse and numerous. Provenance and descriptive metadata, structured or non-structured, are all accounted for and managed by the system.

A modern Web content management system such as Drupal is also a key component here, as many plugins and functions have been continuously developed for science applications, by a large talent pool, on a well-designed modular architecture.