Fedora Repositories
For Humans

Lauren Magnuson & Mary Wahl
California State University, Northridge
What is a Fedora Repository?

A Fedora 4 repository consists of a directed acyclic graph of resources where edges represent a parent-child relation.
“No Fedora object can ever be its own Grandpa.”
Resources, Containers, Binaries, Datastreams

Object Container

Resources

NonRDFSourceDescription

Binary

Describes

Image Source: Carbon Creek, 26 June 1960, CSUN Digital Library
Content Models > Compact Node Definitions + Mixins

https://github.com/projecthydra-labs/hydra-works/wiki/Lesson:-Build-models-with-Hydra-Works
Fedora in Context: Hydra Use Case

Adapted from https://wiki.duraspace.org/display/hydra/Hydra+Stack+-+The+Hierarchy+of+Promises
Differences between ContentDM content model and Fedora Commons

ContentDM: Flat

Metadata (desc.all)

- `<title>Carbon Creek</title>`
- `<find>8.jp2</find>`
- `<archiv>hah016.tif</archiv>`
- `<sourca>Homer Halverson Collection</sourca>`

Objects

- 8.jpg
- hah016.tif

Fedora Commons: RDF Graph

Carbon Creek

Has Title

Homer Halverson Collection

Is Member of Collection

Resource hah016.tif

(URI: http://myrepo.com/files/hah016.tif)

8.jpg

Has Derivative
# Migrating: Getting Metadata out of ContentDM

<table>
<thead>
<tr>
<th>Tab-delimited export</th>
<th>XML export (Dublin Core)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Where do I get it?</strong></td>
<td><strong>ContentDM Admin &gt; Collections &gt; Export &gt; Tab-delimited</strong></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Tab-delimited" /></td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>- Fields will differ based on collection and your ContentDM Setup</td>
</tr>
<tr>
<td></td>
<td>- May not contain administrative metadata</td>
</tr>
<tr>
<td><strong>What does it look like?</strong></td>
<td>- Plain text or csv</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Example" /></td>
</tr>
</tbody>
</table>

**Example:**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>Author/Photographer</td>
<td>Description</td>
<td>Date</td>
<td>Place</td>
<td>Subject</td>
<td>Keywords</td>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Carbon Creek, 2 Valherson, Non Looking B.S. Inc</td>
<td>6/26/1960</td>
<td>Orange County</td>
<td>Channels (Hydraulic engineering)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Carbon Creek, 2 Valherson, Non Construction at</td>
<td>6/26/1960</td>
<td>Orange County</td>
<td>Channels (Hydraulic engineering)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Caballero Creek Valherson, Non Construction at</td>
<td>10/7/1960</td>
<td>Calabasas</td>
<td>Calabasas Channels (Hydraulic engineering)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Caballero Creek Valherson, Non Caballero Creek</td>
<td>10/7/1960</td>
<td>Calabasas</td>
<td>Calabasas Channels (Hydraulic engineering)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Caballero Creek Valherson, Non Cedardale control</td>
<td>10/7/1960</td>
<td>Calabasas</td>
<td>Calabasas Channels (Hydraulic engineering)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Migrating: Getting Metadata out of ContentDM

## OAI-DC Harvesting (via MARCEdit)

### Where do I get it?

*MarcEdit > Add-ins > OAI Harvester Tools > Harvest OAI Data*

*Found in `\collection folder\index\description` on ContentDM Server*

### Notes

- Uses ContentDM's OAI Gateway to retrieve metadata record in OAI-DC format
- May not contain all administrative metadata

### What does it look like?

- Dublin Core or MODS in XML
- plain text or XML (no specific schema)
- completely flat

---

<table>
<thead>
<tr>
<th>OAI-DC Harvesting (via MARCEdit)</th>
<th>desc.all file</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where do I get it?</td>
<td></td>
</tr>
<tr>
<td><em>MarcEdit &gt; Add-ins &gt; OAI Harvester Tools &gt; Harvest OAI Data</em></td>
<td><em>Found in <code>\collection folder\index\description</code> on ContentDM Server</em></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>- Uses ContentDM's OAI Gateway to retrieve metadata record in OAI-DC format</td>
<td>- ContentDM master metadata file for each collection</td>
</tr>
<tr>
<td>- May not contain all administrative metadata</td>
<td>- Contains administrative, rights, and descriptive metadata</td>
</tr>
<tr>
<td>What does it look like?</td>
<td></td>
</tr>
<tr>
<td>- Dublin Core or MODS in XML</td>
<td>- plain text or XML (no specific schema)</td>
</tr>
<tr>
<td></td>
<td>- completely flat</td>
</tr>
</tbody>
</table>

```
1 <title>Carbon Creek, 26 June 1960</title>
2 <creator>McGovern, Mooney</creator>
3 <recip>Recip</recip>
4 <date>1980-06-26</date>
5 <description>"Looking D.S. from Merrill Ave. Bridge at Carbon Creek, 26 June 1960."</description>
6 <note>Notes</note>
7 <subject>Hydraulic engineering</subject>
8 <subject>Engineers; Orange County (Calif.)</subject>
9 <key>Key</key>
10 <type>Image</type>
```
**Metadata in Fedora 3 vs. Fedora 4**

**Fedora 3: FOXML + METS extension**

```xml
<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>
```

**Fedora 4: Native RDF**

Snippets:

```xml
<http://127.0.0.1:8983/fedora/rest/dev/test-1>
  <http://purl.org/dc/terms/title> "Anna Karenina";
  <http://fedora.info/definitions/v4/repository#hasParent>
    <http://127.0.0.1:8983/fedora/rest/dev>;
    <http://fedora.info/definitions/v4/repository#mixinTypes>
      "fedora:Container",
      "fedora:Resource";
    <http://fedora.info/definitions/v4/repository#primaryType> "nt:folder";
    <http://purl.org/dc/terms/creator> "Tolstoy, Leo";
    <info:fedora/fedora-system:def/model#hasModel> "Book".
</http://127.0.0.1:8983/fedora/rest/dev/test-1>
```


Metadata in Fedora 3 vs. Fedora 4

Fedora 3

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary Content</td>
</tr>
<tr>
<td>Rights</td>
</tr>
<tr>
<td>RELS-EXT</td>
</tr>
</tbody>
</table>

Fedora 4

- Description, Rights

Binary Content

(relationship)

Why RDF?

It facilitates **linked data**.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://127.0.0.1:8983/fedora/rest/dev/test-1">http://127.0.0.1:8983/fedora/rest/dev/test-1</a></td>
<td><a href="http://purl.org/dc/terms/creator">http://purl.org/dc/terms/creator</a></td>
<td>&quot;Leo Tolstoy&quot;</td>
</tr>
</tbody>
</table>

Anna Karenina

Has Creator

Leo Tolstoy

RDF properties can be string literal (e.g., “Leo Tolstoy”) or URIs, e.g. to the VIAF Authority: [http://viaf.org/viaf/96987389](http://viaf.org/viaf/96987389)
Why RDF?

It provides building blocks for the semantic web.

Image from: http://commonplace.net/tag/library-systems/
Sparql in Fedora 4

- SPARQL (SPARQL Protocol and RDF Query Language) provides a semantic mechanism for querying RDF data
- An **External Triplestore** is required to enable a queryable SPARQL endpoint for Fedora 4 data
  - **Fuseki**
  - **Sesame**

---

1c. Selecting Containers With PDF Binaries

```sparql
prefix fcrepo: <http://fedora.info/definitions/v4/repository#>
prefix relation: <http://www.iana.org/assignments/relation/>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix ebucore: <http://www.ebu.ch/metadata/ontologies/ebucore/ebucore#>
select ?container where {
    ?ds fcrepo:mixinTypes "fedora:Binary"^^xsd:string .
    ?ds fcrepo:hasParent ?container .
    ?ds ebucore:hasMimeType "application/pdf"^^xsd:string
}
```

https://wiki.duraspace.org/display/FEDORA44/SPARQL+Recipes
Hydra: CDM2BAG Batch Ingest (Oregon Digital Hydra)

- Designed for Fedora 3 for the Oregon Digital Hydra System.
- Translates desc.all file to RDF
- “Bags” together RDF and files, which are uploaded through the Hydra interface
- Requires some knowledge of Ruby as well as a Ruby/Rails-enabled environment
  - [https://github.com/OregonDigital/cdm2bag](https://github.com/OregonDigital/cdm2bag)

Islandora: MODS Batch Ingest

- Requires Islandora Batch Importer Model
- Convert ContentDM files to MODS
  - OAI-DC (can use MarcEdit to retrieve) must be crosswalked to MODS
  - Separate each record into its own file
  - Place metadata files into a folder with objects, using identical filenames
    - e.g., File1.xml → File1.JPG
    - Zip the folder and upload to Islandora Interface

More Info: https://wiki.duraspace.org/display/ISLANDORA711/How+to+Batch+Ingest+Files
Examples of Repository software Using Fedora

Oregon Digital (Hydra Fedora 3, moving to Fedora 4; ContentDM migration)  
https://github.com/OregonDigital/oregondigital

Sufia Institutional Repository (Hydra, Institutional Repository, Fedora 4)  
https://github.com/projecthydra/sufia

Avalon Media (Hydra, Fedora 3, moving to Fedora 4)  
https://github.com/avalonmediasystem/avalon

Islandora (Fedora 3, moving to Fedora 4)  
https://github.com/Islandora
MORE INFORMATION

Hydra

Hydra-Tech Google Group
https://groups.google.com/forum/#!forum/hydra-tech

Hydra RDF Working Group
https://wiki.duraspace.org/display/hydra/Hydra+RDF+Working+Group

Hydra MODS and RDF Subgroup
https://wiki.duraspace.org/display/hydra/MODS+and+RDF+Descriptive+Metadata+Subgroup

Dive into Hydra
https://github.com/projecthydra/hydra/wiki/Dive-into-Hydra

Hydra in a Box
https://wiki.duraspace.org/display/hydra/Hydra+in+a+Box
More Information

Islandora

Islandora Google Group
https://groups.google.com/forum/#!forum/islandora

Islandora Metadata Interest Group
https://github.com/islandora-interest-groups/Islandora-Metadata-Interest-Group

Metadata in Islandora
https://wiki.duraspace.org/display/ISLANDORA715/Metadata+in+Islandora
Special Thanks

- Trey Terrell @ Oregon State and Sarah Seymore @ University of Oregon for amazing help with CDM2Bag
- Mountain West Digital Library for this page about OAI queries: http://mwdl.org/getinvolved/oai_queries.php
- CSUN Sandbox Interest Group
Thank you!

Lauren Magnuson & Mary Wahl
California State University, Northridge