Breaking Up with CONTENTdm…
AlabamaMosaic: Using VuFind to Index a Distributed Digital Collection

Midge Coates
Digital Projects Librarian
Auburn University Libraries
AlabamaMosaic: Using VuFind to Index a Distributed Digital Collection

Midge Coates
Auburn University Libraries
What Does “Distributed” Mean?
Why Did We Leave MultiSiteServer?
• 2012: Page views for AlabamaMosaic Web site had been going down since 2010
• 65% of visits lasted 10 sec or less
• Traffic from AlabamaMosaic to Auburn’s digital collections was 0.1%
• 2009 Survey: Users wanted a better search
• Rumor: OCLC would discontinue support for MultiSite Server “soon”
• University of Alabama digital collections no longer in CONTENTdm: MSS couldn’t index
VuFind to the Rescue!
- Open source software => Local customization
- Local (Auburn) VuFind experience => Local expertise
- Intuitive public interface (“Google-like”); better searches; facets for narrowing results
- VuFind can index digital collections via OAI-PMH and XSL transforms
- VuFind can index from content management systems other than CONTENTdm (MSS software can only index CONTENTdm collections)
What Metadata Fields Should We Use?
• Hard-coded:
  – Collection name
  – Institution/publisher name

• Harvested:
  – Link to item thumbnail
  – Link to item in native collection
  – Title
  – Creator(s)
  – Description
  – Subject terms
How Does the Harvesting Process Work?
Step 1: Check OAI-PMH feed.
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<OAI-PMH xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <!-- xmlLeech schemaLocation="http://www.openarchives.org/OAI/2.0/oai.pmh.xsd" -->
    <!-- xmlLeech schemaLocation="http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd" -->
    <responseDate>2013-05-29T18:15:47Z</responseDate>
    <result>
        <ListRecords>
            <metadataPrefix>oai_dc</metadataPrefix>
            <setSpec>austest</setSpec>
            <request>
                <identifier>http://content.lib.auburn.edu/cgi-bin/oai.exe</request>
            </ListRecords>
            <record>
                <header>
                    <identifier>oai:content.lib.auburn.edu:austest/0</identifier>
                    <datestamp>2010-05-11</datestamp>
                    <setSpec>austest</setSpec>
                </header>
                <metadata>
                    <dc:title>County agents and other extension leaders</dc:title>
                    <dc:description>
                        Caption: "A group portrait of county agents and other Alabama Cooperative Extension Service leaders with Alabama Farm Bureau leaders. Place: Montgomery, Alabama. Date: April 15, 1922. Fred Stewart is the last man on the right on the front row." April 15, 1922.
                    </dc:description>
                    <dc:date>1922-04-15</dc:date>
                    <dc:coverage>Montgomery Co. (Ala.)</dc:coverage>
                    <dc:creator>Unknown</dc:creator>
                    <dc:publisher>Auburn University Libraries</dc:publisher>
                    <dc:type>Image</dc:type>
                    <dc:format>JPEG</dc:format>
                    <dc:identifier>
                        http://www.lib.auburn.edu/archive/find-aid/071p/, http://www.lib.auburn.edu/archive/find-aid/071.htm
                    </dc:identifier>
                    <dc:identifier>71.12.1.1</dc:identifier>
                </metadata>
            </record>
        </ListRecords>
    </result>
</OAI-PMH>
```
Step 2: Customize XSLT file.
Collection Information

<!-- AU hard coded values and most of the <doc></doc> fields are generated or called from here -->
<!-- CUSTOMIZE THESE FOR EACH SOURCE -->

<!-- Begin hard coded values -->
<field name="format">Electronic</field>
<field name="collection">Alabama Cooperative Extension Service Photographs Collection</field>
<field name="building">Auburn University Digital Library</field>
<field name="publisher">Auburn University Libraries</field>
<field name="topic">Alabama Cooperative Extension Service photos</field>
<field name="fulltopic">Alabama Cooperative Extension Service photos</field>

<!-- End hard coded values -->

<!-- Condense repetitive dc:description elements into one <field></field> -->
<field name="description">
    <xsl:for-each select="dc:description">
        <xsl:value-of select="."/>
    </xsl:for-each>
</field>

<!-- Condense all oai_dc:dc elements into one <field></field> -->
<!-- I assume this is for search purposes within VuFind jdg -->
<field name="allfields">
    <xsl:for-each select="*">
        <xsl:value-of select="."/>
    </xsl:for-each>
</field>

<!-- Call all other field templates matching all children under oai_dc:dc -->
<xsl:apply-templates select="*"/>

<!-- Define each field you want in the destination document below here as a xsl:template -->
List of possible Solr fields is available at:

<!-- Match the first dc:identifier to generate a unique ID for it -->
<xsl:template match="dc:identifier"></xsl:template>
Item Identification Information:
Item URL, Thumbnail URL, Collection Set Name
Step 3: Customize test.bat file. Run it.

```sh
java -jar AUOAItoVufind2.jar -b http://content.lib.auburn.edu:81/cgi-bin/oai.exe -f AUautest_oai_vufind_JG1.xsl -s autest -S -C
```
<xml version='1.0' encoding='UTF-8'>
<oai:PMH xmlns:oai='http://www.openarchives.org/OAI/2.0/'
    xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
    xsi:schemaLocation='http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd'>
    <responseDate>2013-05-29T18:36:31Z</responseDate>
    <request verb='ListRecords' metadataPrefix='oai_dc'
        set='autest'>http://content.lib.auburn.edu:81/cgi-bin/oai.exe</request>
    <ListRecords>
        <record>
            <header>
                <identifier>oai:content.lib.auburn.edu:autest/0</identifier>
                <datestamp>2010-05-11</datestamp>
                <setSpec>autest</setSpec>
            </header>
            <metadata>
                <oai_dc:dc
                    xmlns:oai_dc='http://www.openarchives.org/OAI/2.0/oai_dc/
                    xmlns:dc='http://purl.org/dc/elements/1.1/'
                    xsi:schemaLocation='http://www.openarchives.org/OAI/2.0/oai_dc/
                    http://www.openarchives.org/OAI/2.0/oai_dc.xsd'>
                    <dc:title>County agents and other extension leaders</dc:title>
                    <dc:description>Caption: "A group portrait of county agents and other Alabama Cooperative Extension Service leaders with Alabama Farm Bureau leaders. Place: Montgomery, Alabama. Date: April 15, 1922. Fred Stewart is the last man on the right on the front row." April 15, 1922.</dc:description>
                    <dc:date>1922-04-15</dc:date>
                    <dc:coverage>Montgomery Co. (Ala.)</dc:coverage>
                    <dc:creator>Unknown</dc:creator>
                    <dc:publisher>Auburn University Libraries</dc:publisher>
                    <dc:type>Image</dc:type>
                    <dc:format>JPEG</dc:format>
                    <dc:identifier>71.12.1.1</dc:identifier>
                    <dc:source>Auburn University. Special Collections and Archives</dc:source>
                    <dc:rights>This image is the property of the Auburn University Libraries and is intended for non-commercial use. Users..."</dc:rights>
                </oai_dc:dc>
            </metadata>
        </record>
    </ListRecords>
</oai:PMH>
</xml>
OAI Output, Transformed
Step 4: Customize import.bat file. Run it.

After All That …
What Does the Index Look Like?
MultiSite Server: Search Results for Alabama River
VuFind Index: Search Results for Alabama River
Alabama River at Selma, Alabama.

Format: Electronic
Link: http://digital.archives.alabama.gov/cdm/ref/collection/photo/id/7649
Published: Alabama Department of Archives and History
Subjects: Alabama Photographs and Pictures
          Bodies of water; Buildings; Rivers; Alabama River (Ala.); Selma (Ala.); Dallas County (Ala.)
Who Does What?
• Programmers:
  – Created VuFind instance
  – Created XSLT file templates
  – Created Java tools and bat file templates
  – Manage database at server level
  – Trouble-shoot as needed

• Digital Projects Librarian:
  – Monitor changes in native collections
  – Create XSLTs and bat files for new collections/update for existing collections
  – Harvest collections as needed
  – Trouble-shoot as needed
How Did It All Work Out?

Essentials for Success with Hairy Vetch and Austrian Winter Peas
1. Plant Early — September 15 to October 15.
2. Dwarf Inoculant.
3. Phosphate — Apply Not Less Than 400 lbs per acre. It is preferable to have done when seeding.
5. Drill the seed. Two can be done with a three-row drill, a one-row drill, or a distractor. Consult your county agent about proper method of planting these crops.

Why Grow Hairy Vetch or Austrian Winter Peas
1. A crop of Hairy Vetch or Austrian Winter Peas will easily produce 18 to 20 bushels per acre. Corn yield.
2. Hairy Vetch or Austrian Winter Peas will add as much nitrogen to the soil as 400 lbs of nitrate of soda and 1500 lbs. of manure. As five tons of stable manure.
3. By growing Hairy Vetch or Austrian Winter Peas to turn under, two weeks before planting corn, enough corn can be made to supply the needs of Alabama.
4. Hairy Vetch or Austrian Winter Peas means richer soils and prosperous farmers.

No Vetch
480 pounds seed cotton per acre

After Vetch
836 pounds seed cotton per acre

No Vetch
15 bushels corn per acre

Gom after Vetch
Made 37 bushels per acre
Increased 22 bushels per acre
• AlabamaMosaic page views tripled
• Number of visits lasting 10 sec or less went down to 49% (from 65%)
• Traffic from AlabamaMosaic to Auburn’s collections went up to 4.7% (from 0.1%)
• People tell us they like the new look and the new search function.
• When OCLC discontinued support for MSS, VuFind was already in place.
• We have indexed the UA Acumen collection and an Auburn dSpace collection, along with all the CONTENTdm collections
Thank you.

www.alabamamosaic.org