

TESTING VUFIND AS A COLLECTION INDEX FOR ALABAMAMOSAIC

Midge Coates, Clint Bellanger, and Jeremy Goslin
AUBURN UNIVERSITY LIBRARIES

ABSTRACT

In 2005, the Network of Alabama Academic Libraries launched AlabamaMosaic: a consortial collection of Alabama cultural heritage materials digitized by 21 member partners. At present, all partners use OCLC's CONTENTdm content management system (CMS) software for their AlabamaMosaic collections, and the project uses OCLC's Multi-Site Server (MSS) software to index these. Recently, partner institutions have been exploring alternative CMS software for their digital collections. But these non-CONTENTdm collections cannot be included in AlabamaMosaic if the project is using the MSS indexing software.

VuFind is an open-source next-generation catalog software created at Villanova University. Records for items in a digital collection can be created in VuFind using OAI metadata harvesting. XML output from OAI harvesting feeds is transformed into index records using a software tool and transform files customized for each collection. Like MSS, VuFind allows searching across all indexed collections. Facets then allow the search results to be browsed by institution, collection, author, and/or subject term. Collections in CONTENTdm, dSpace, and Acumen have been harvested into the AlabamaMosaic VuFind test index.

INTRODUCTION

AlabamaMosaic is a collection of digitized Alabama cultural heritage materials created by the Network of Alabama Academic Libraries consortium. The AlabamaMosaic Web site makes materials from 21 contributing partners searchable via a single search interface. When the project was launched in 2005, the ENCompass content management system (CMS) software was used. Several years later, the collections were converted to OCLC's CONTENTdm CMS, and the project began using OCLC's Multi-Site Server (MSS) software to index collections and provide a search interface.

Recently, partner institutions have been exploring alternative CMS software for their digital collections. The University of Alabama converted all its digital collections to its locally-created CMS Acumen. Auburn University began creating collections using the dSpace CMS. But MSS cannot index these non-CONTENTdm collections, and OCLC has no immediate plans for adding this capability. For this reason, non-CONTENTdm collections cannot be included in AlabamaMosaic as long as the project continues to use the MSS indexing software.

VuFind is an open-source next-generation catalog software created at Villanova University. Auburn University Libraries has been using VuFind as its default catalog since 2010. Auburn's experience indexing its own digital collections into VuFind using the OAI metadata harvesting protocol suggested that VuFind could be a suitable replacement for MSS as the indexing software for AlabamaMosaic. In 2011, workers at Auburn began creating a VuFind test index for AlabamaMosaic collections.

METHODOLOGY

Figure 1 shows a flow chart of the VuFind metadata harvesting process. For each collection to be imported, an XSL transform file is created using the collection's OAI base URL and set name. A Java tool is used to apply the transform to the collection's OAI output XML and import the resulting metadata records into the VuFind database.

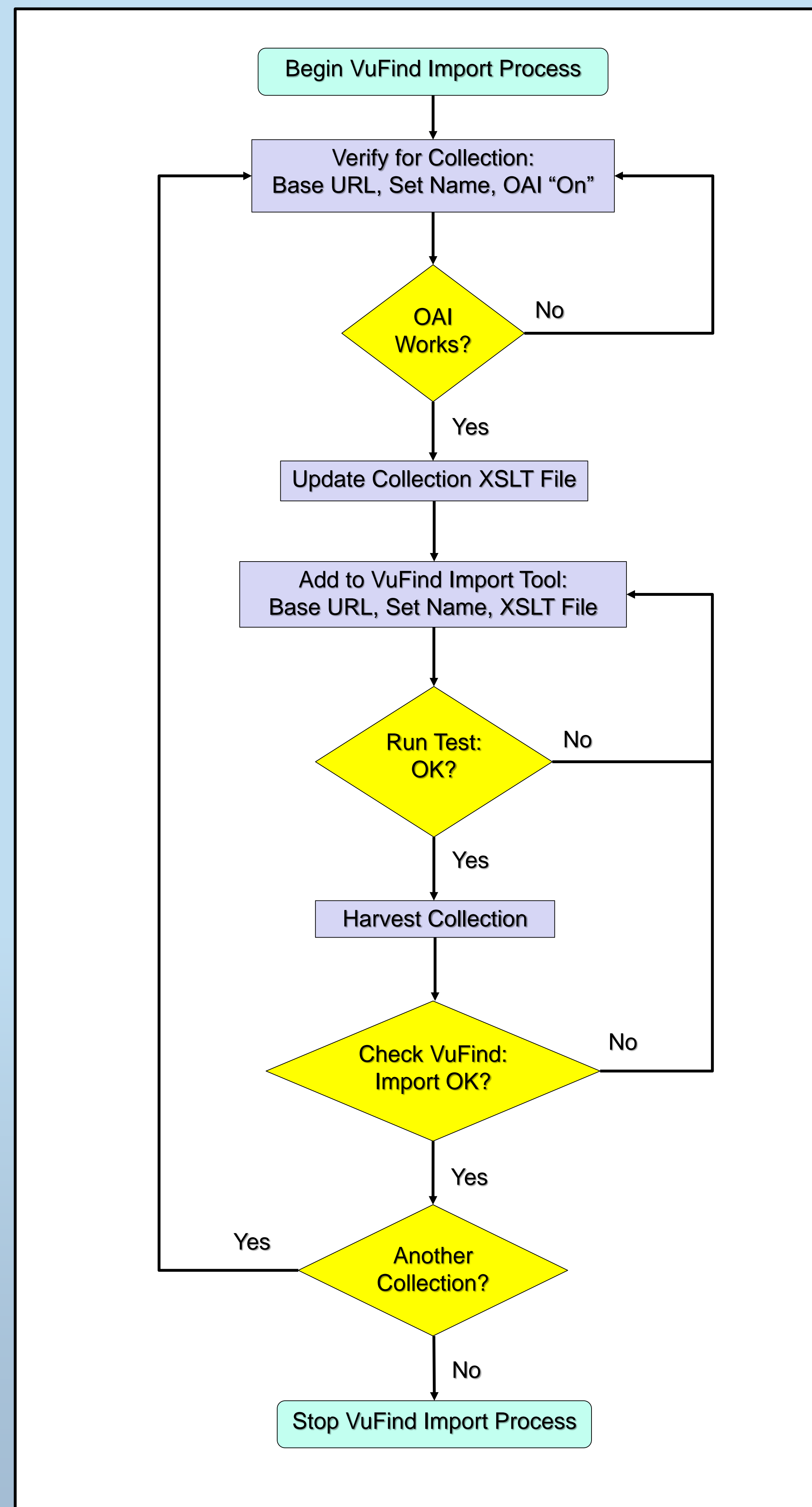


Figure 1. Flow Chart showing VuFind collection import process using OAI metadata harvesting protocol.

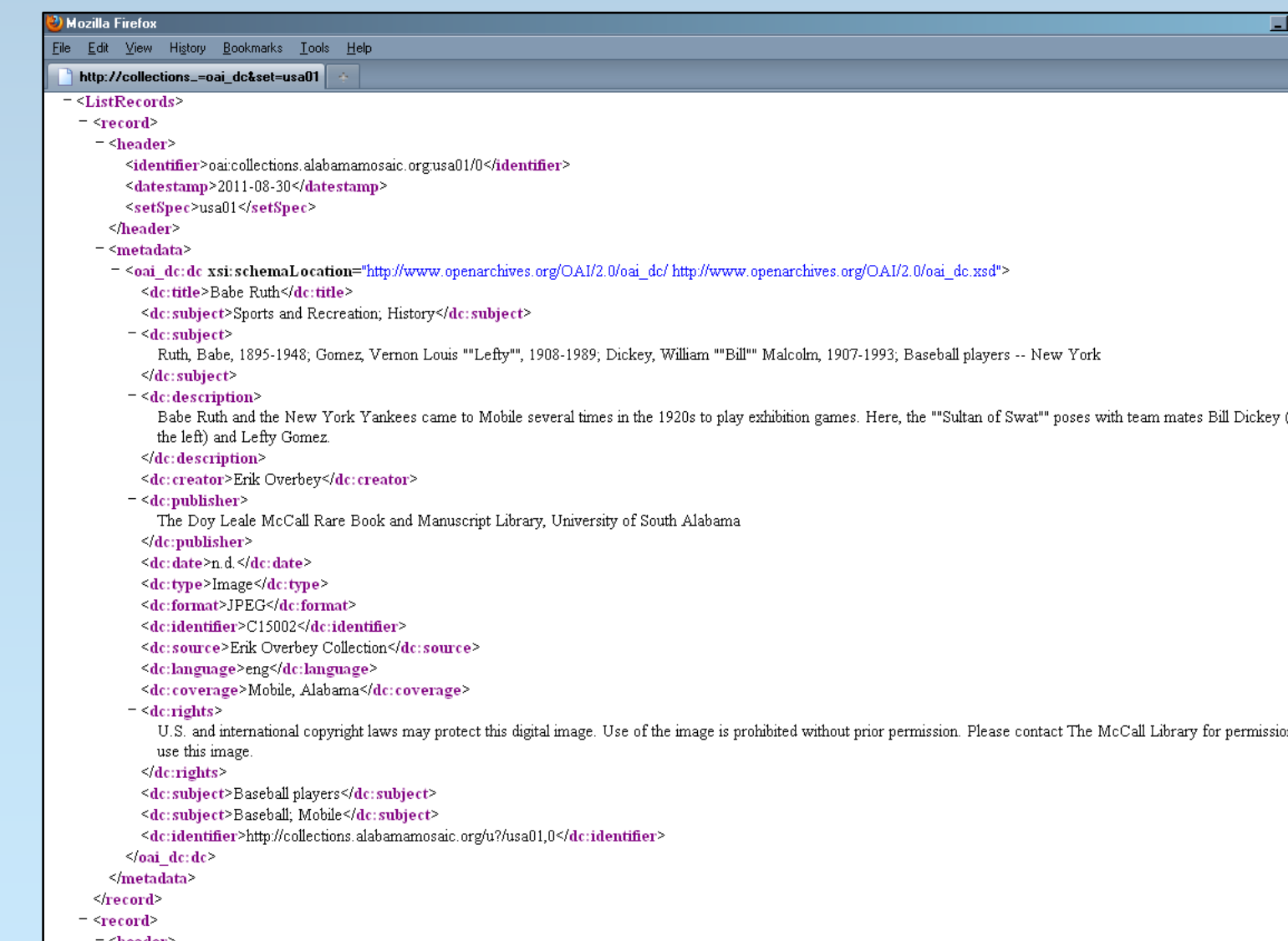


Figure 2. Portion of OAI output XML for University of South Alabama, McCall Library collection, showing record for Item Number "0".

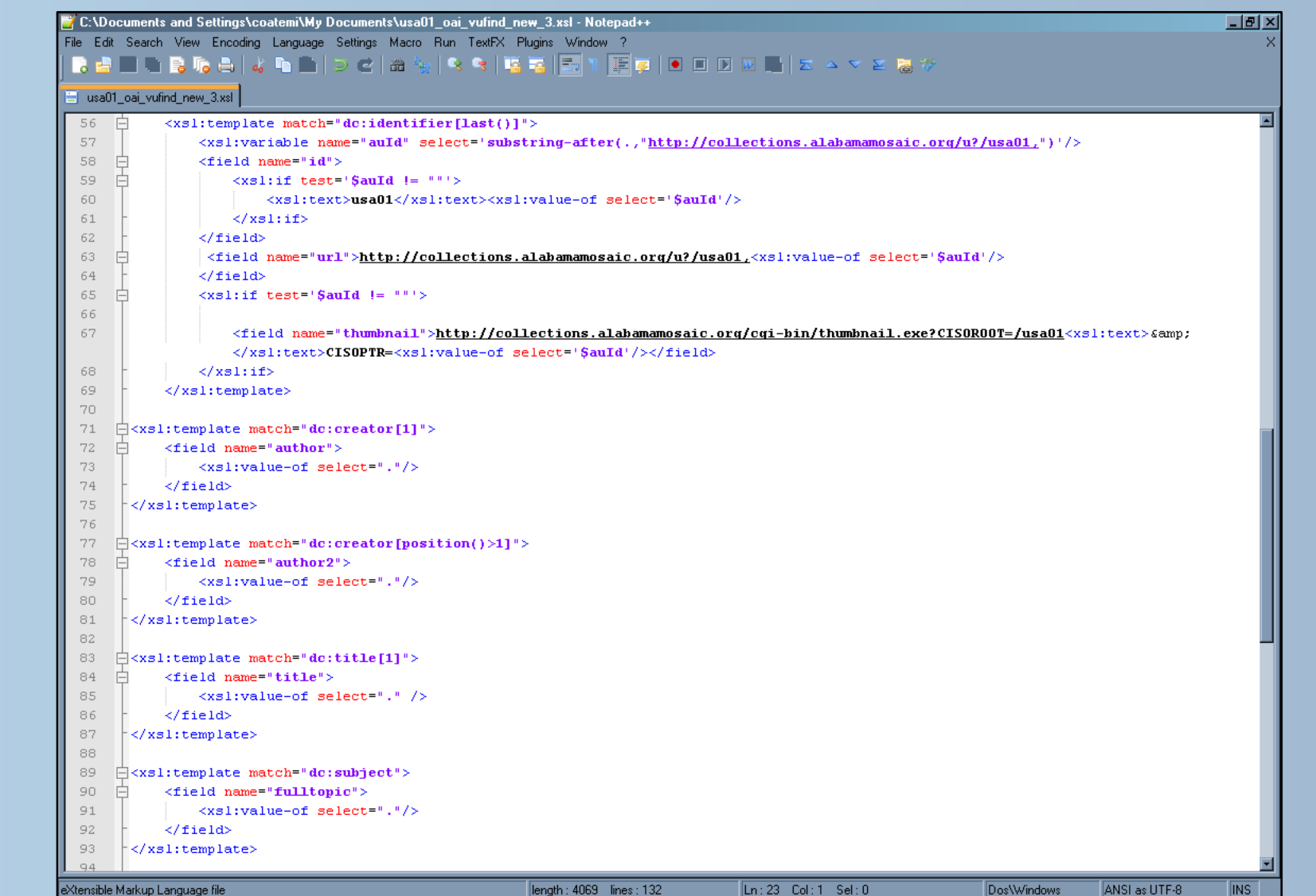


Figure 3. Portion of XSL transform file for University of South Alabama, McCall Library collection.

Metadata for collection items is obtained from the native collections using the OAI metadata harvesting protocol. OAI produces an XML document with elements corresponding to Dublin Core metadata elements. Figure 2 shows a portion of the OAI XML output corresponding to the first record in the collection of University of South Alabama, McCall Library.

The XML produced by OAI must be transformed in order to create the corresponding VuFind records. An XSLT file is created which will split the single large XML file into individual records and copy the contents of only specific Dublin Core elements into those records. Figure 3 shows a portion of the XSLT file created for the University of South Alabama, McCall Library collection.

A Java tool transforms the OAI XML output using the XSLT file and creates the item records in the VuFind database. Figure 4 shows the Java tool interface. The user enters the collection's base URL and set name, the XSLT file, and the type of collection (Acumen, CONTENTdm Version 6, Generic/CONTENTdm Version 5).

To be sure the harvesting process will work, the Test mode of the Java tool is used first. Figure 5 shows a portion of the output from a harvesting test for the University of South Alabama, McCall Library collection. If the test is successful, records are then input using the Harvest to VuFind mode of the Java tool.

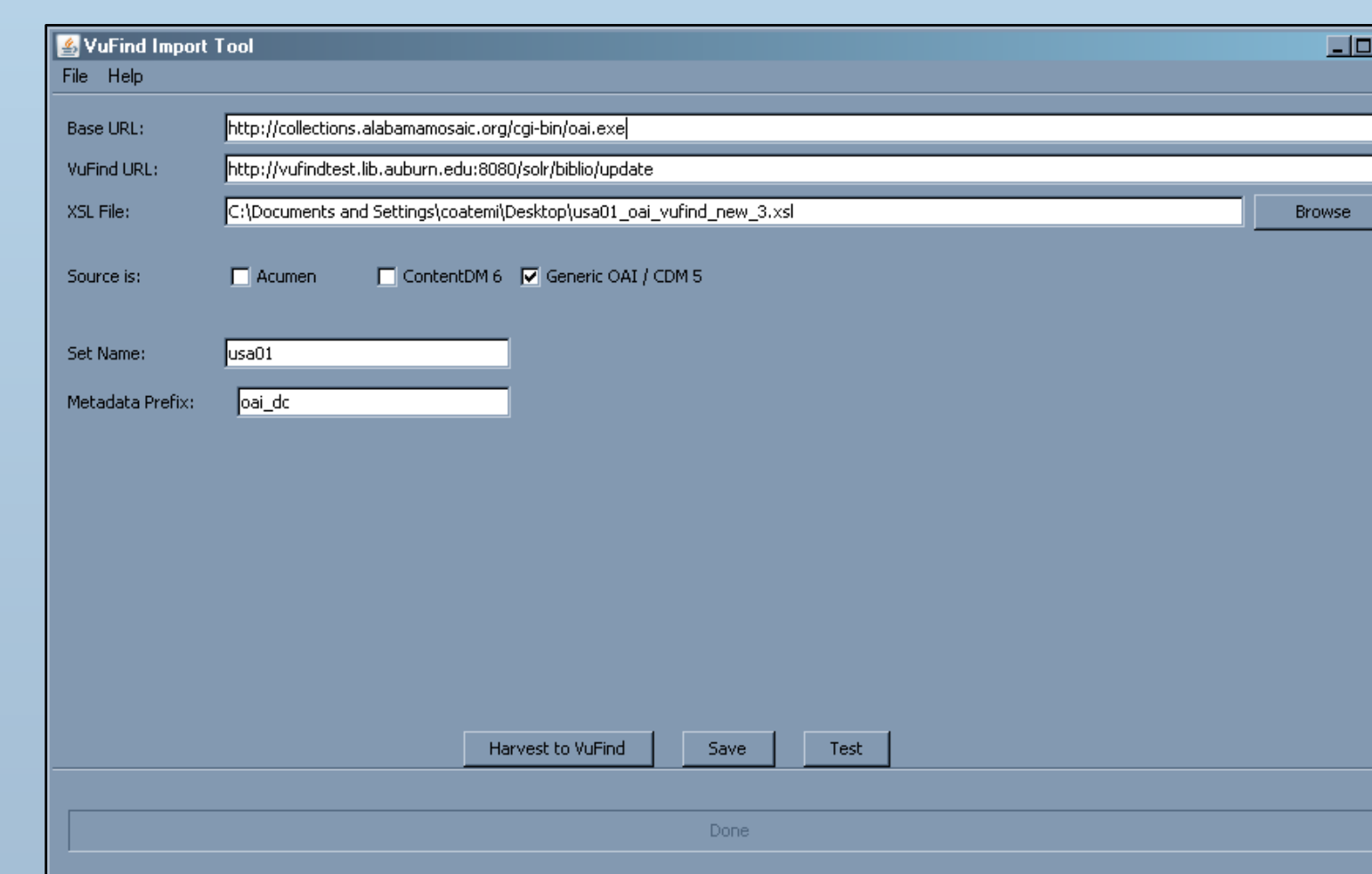


Figure 4. VuFind OAI import tool, set up to import University of South Alabama, McCall Library collection.

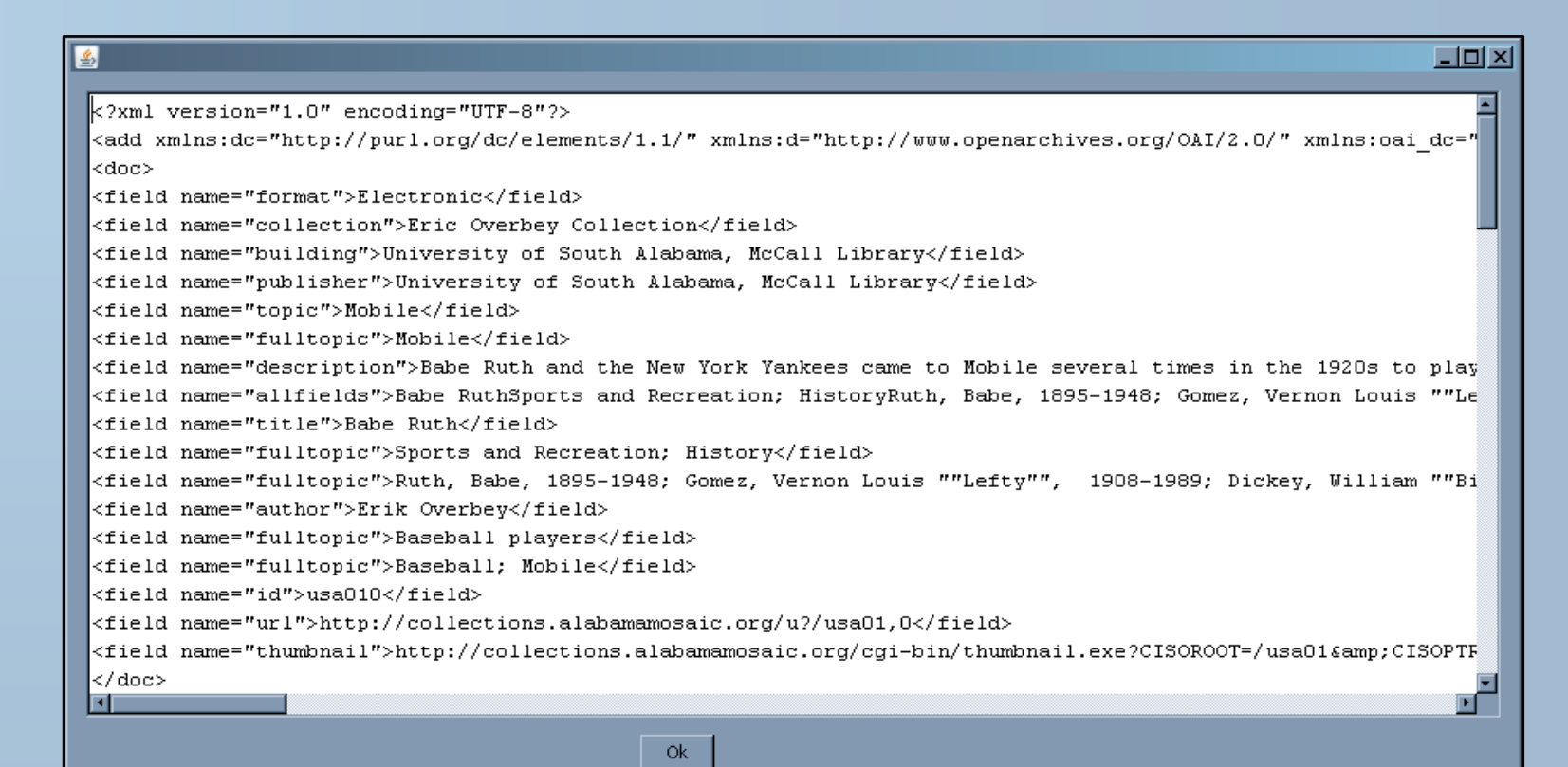


Figure 5. Test output: transformed metadata for University of South Alabama, McCall Library collection, showing record for Item Number "0".

RESULTS

AlabamaMosaic VuFind Test Web Site: <http://vufindtest.lib.auburn.edu/vufind/>

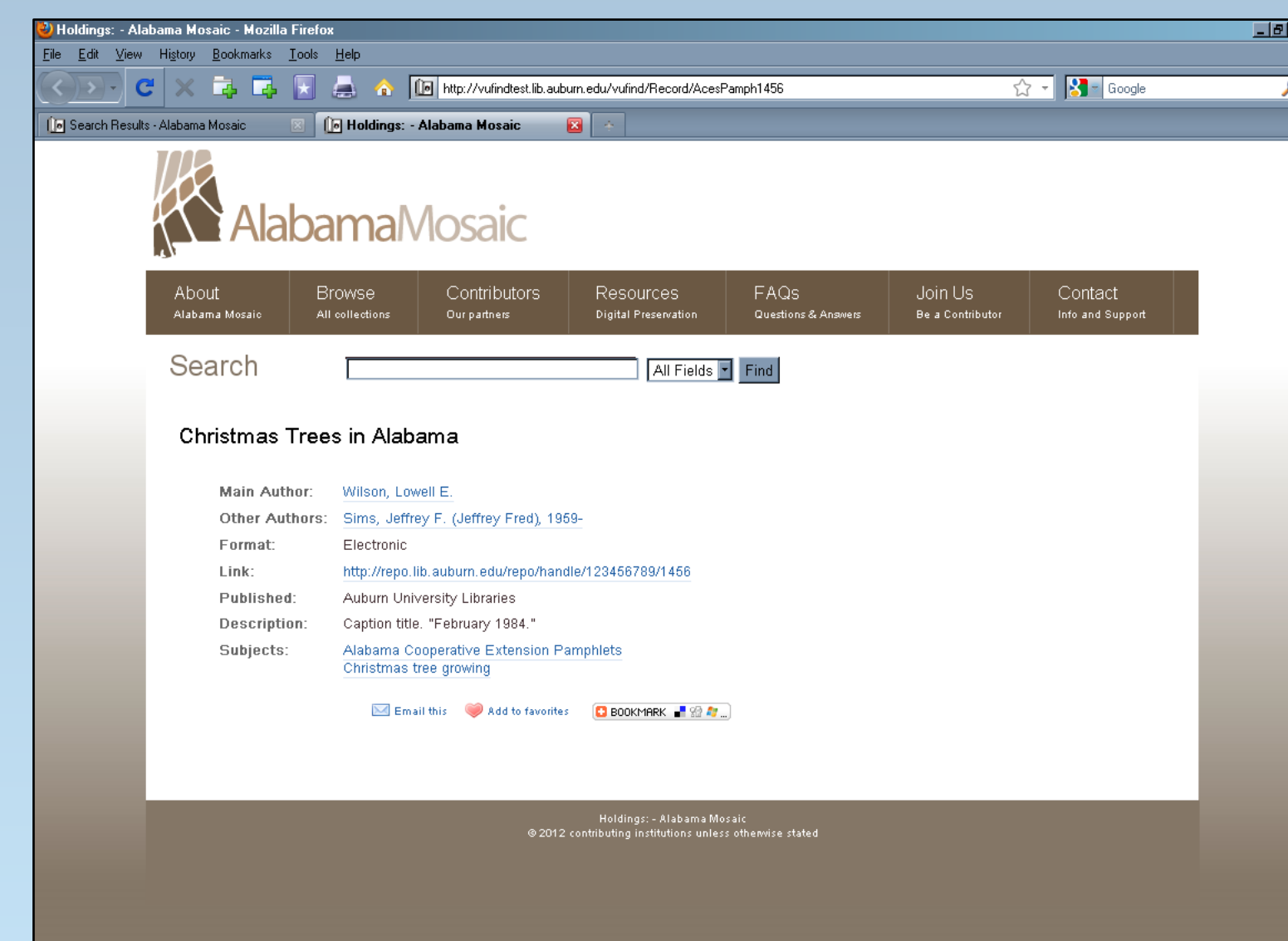
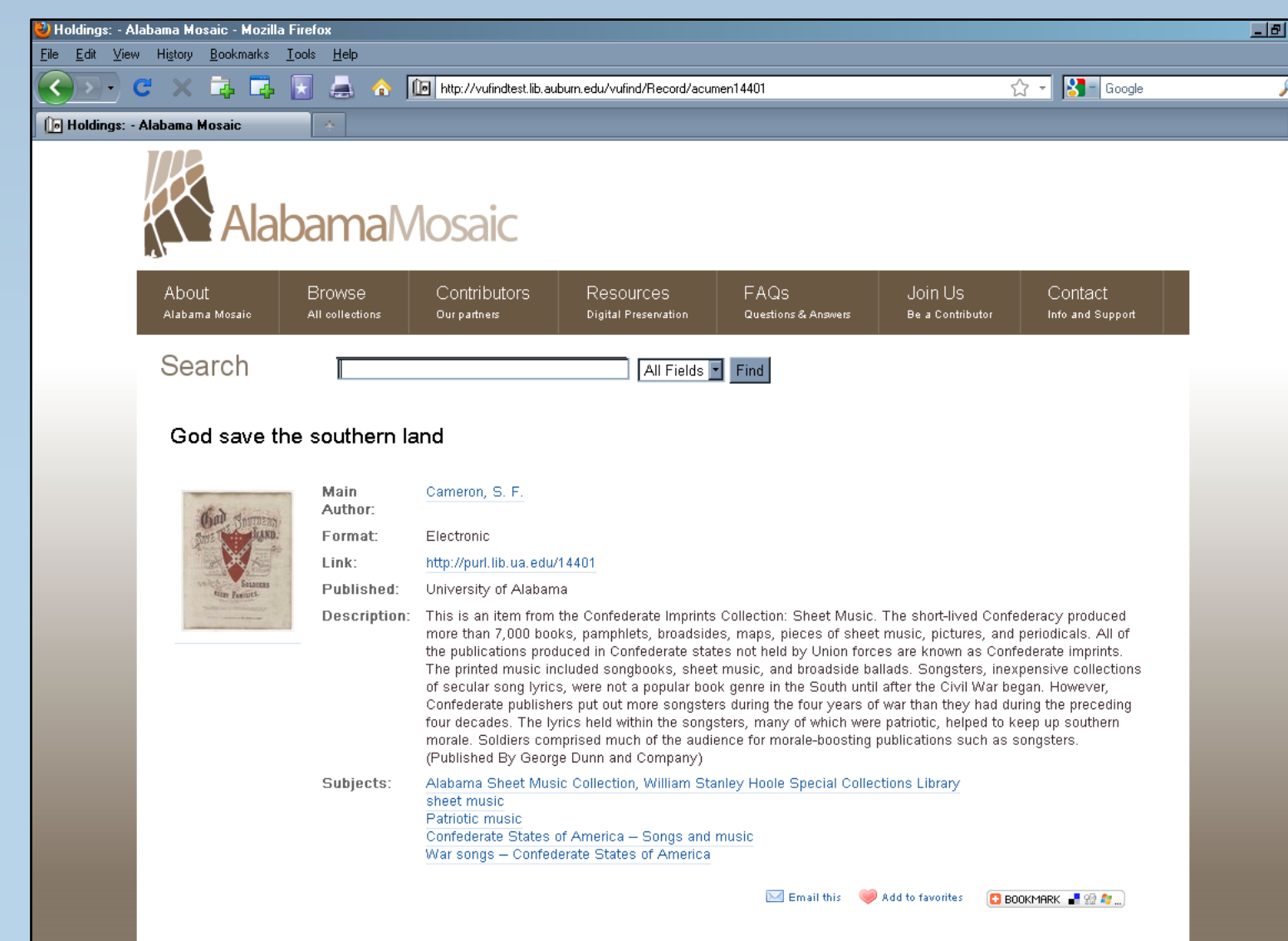
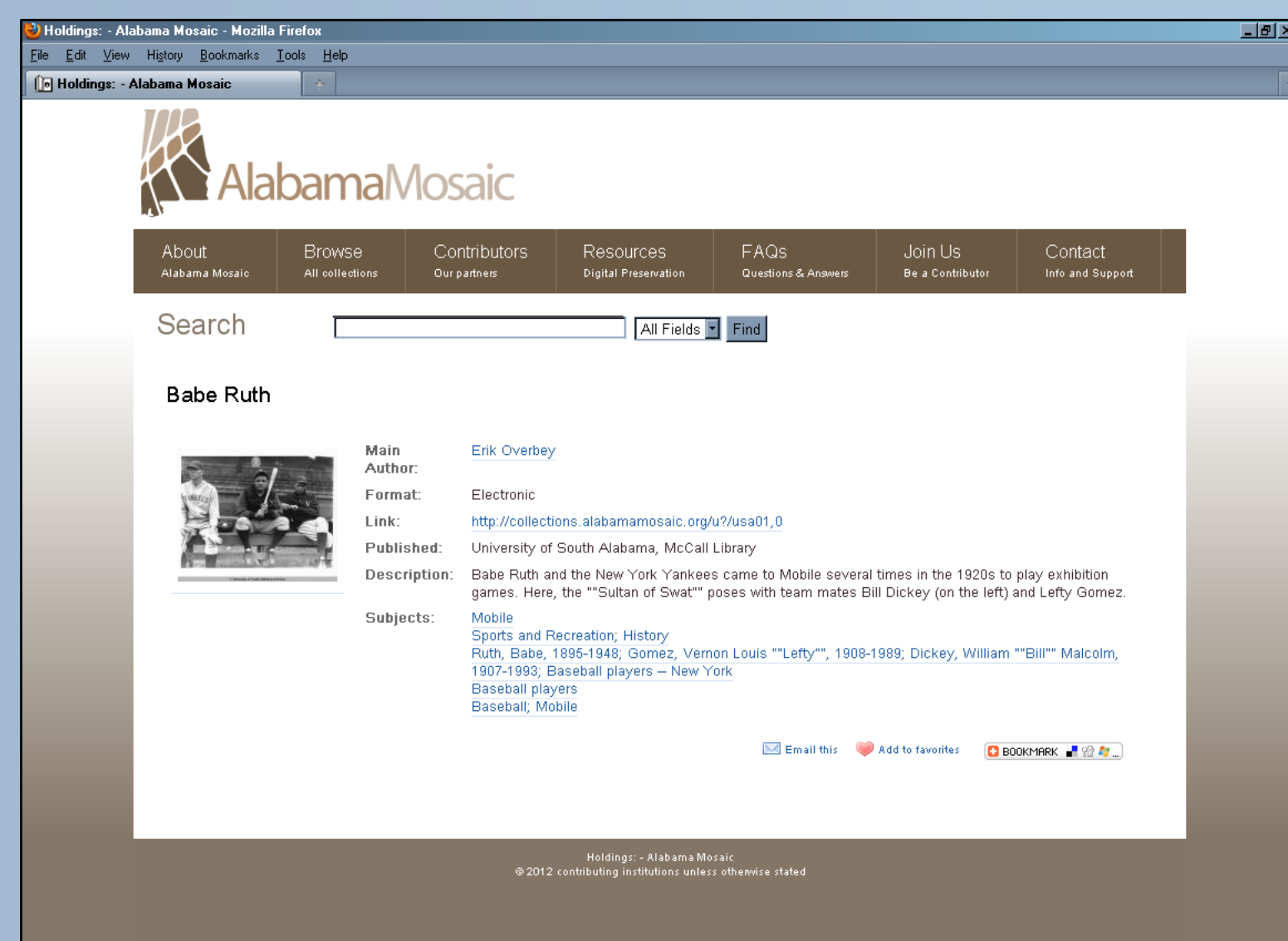


Figure 6. VuFind record for item in University of South Alabama, McCall Library collection (CONTENTdm).

Figure 7. VuFind record for item in University of Alabama Confederate Imprints collection (Acumen).

Figure 8. VuFind record for Auburn University Alabama Agricultural Experiment Station Pamphlets collection (dSpace).

Figures 6, 7, and 8 show VuFind records for items harvested from CONTENTdm, Acumen, and dSpace collections. Metadata imported includes title, creator, description, subject terms, thumbnail, and URL for the native item. Thumbnails cannot be harvested from dSpace collections at this time.

User interface Web pages are adapted from code used for Auburn's VuFind catalog interface. Graphics and colors have been changed to be consistent with project branding. The test site's home page (Figure 9) explains that we are trying out a new way to search the collections and lists some of the advantages of using VuFind. Its main feature, however, is the Google-like central search box.

Figure 10 is an example of a VuFind search results page. The right-hand side of the page shows the top results of a search for "Alabama River." The column on the left-hand side under the heading "Narrow Search" lists results by categories called "facets." Search results are grouped into facets by collection publisher, by individual collection, by creator, and by subject terms. Facets are also available when browsing the collections.

For comparison's sake, Figure 11 shows the MSS search results page for the same search for "Alabama River." The relevancy of the top results to the search terms is much lower than that of the top VuFind results. In addition, MSS doesn't have facets available to narrow the search results into conveniently browsable sub-groups.

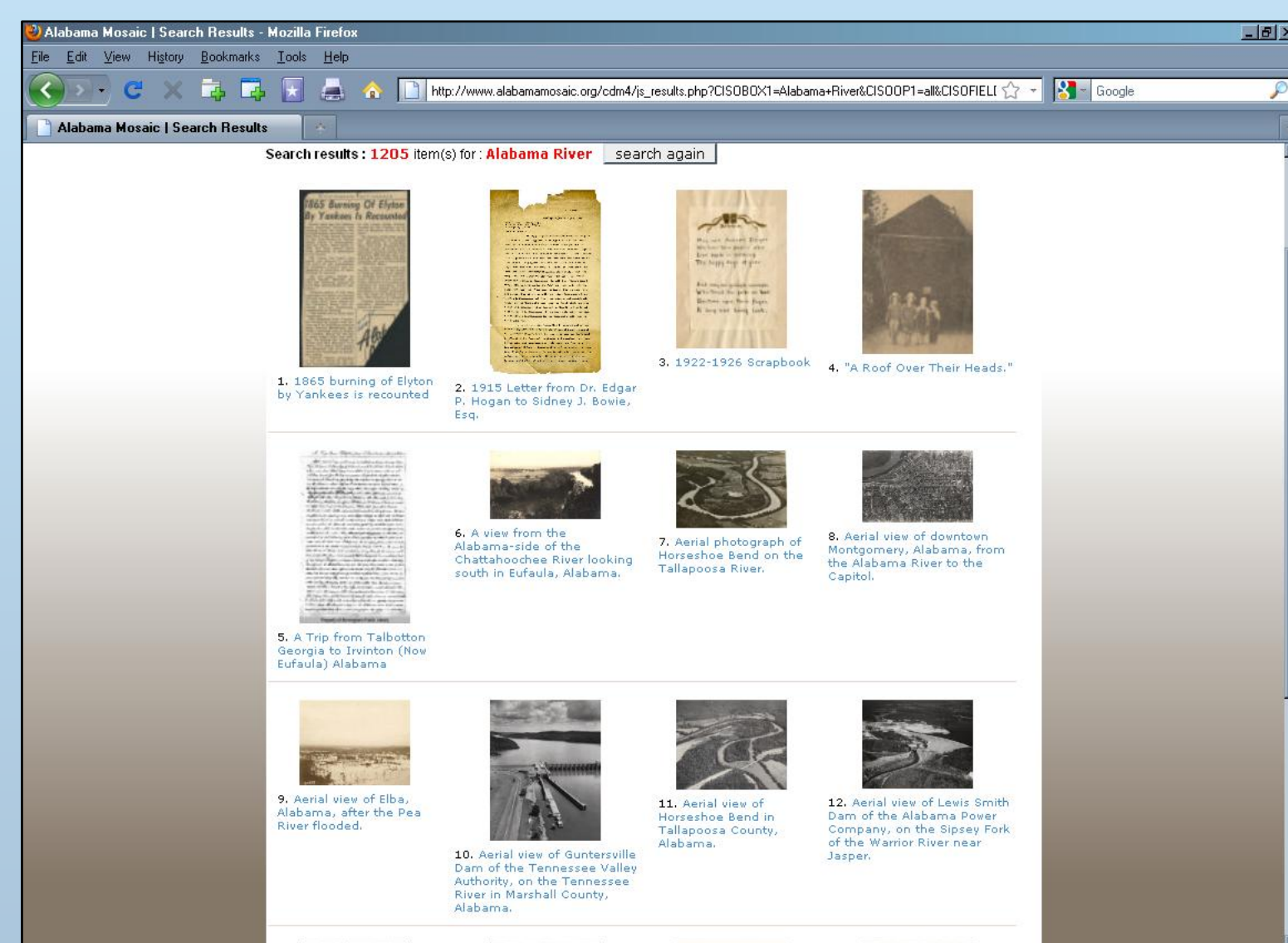
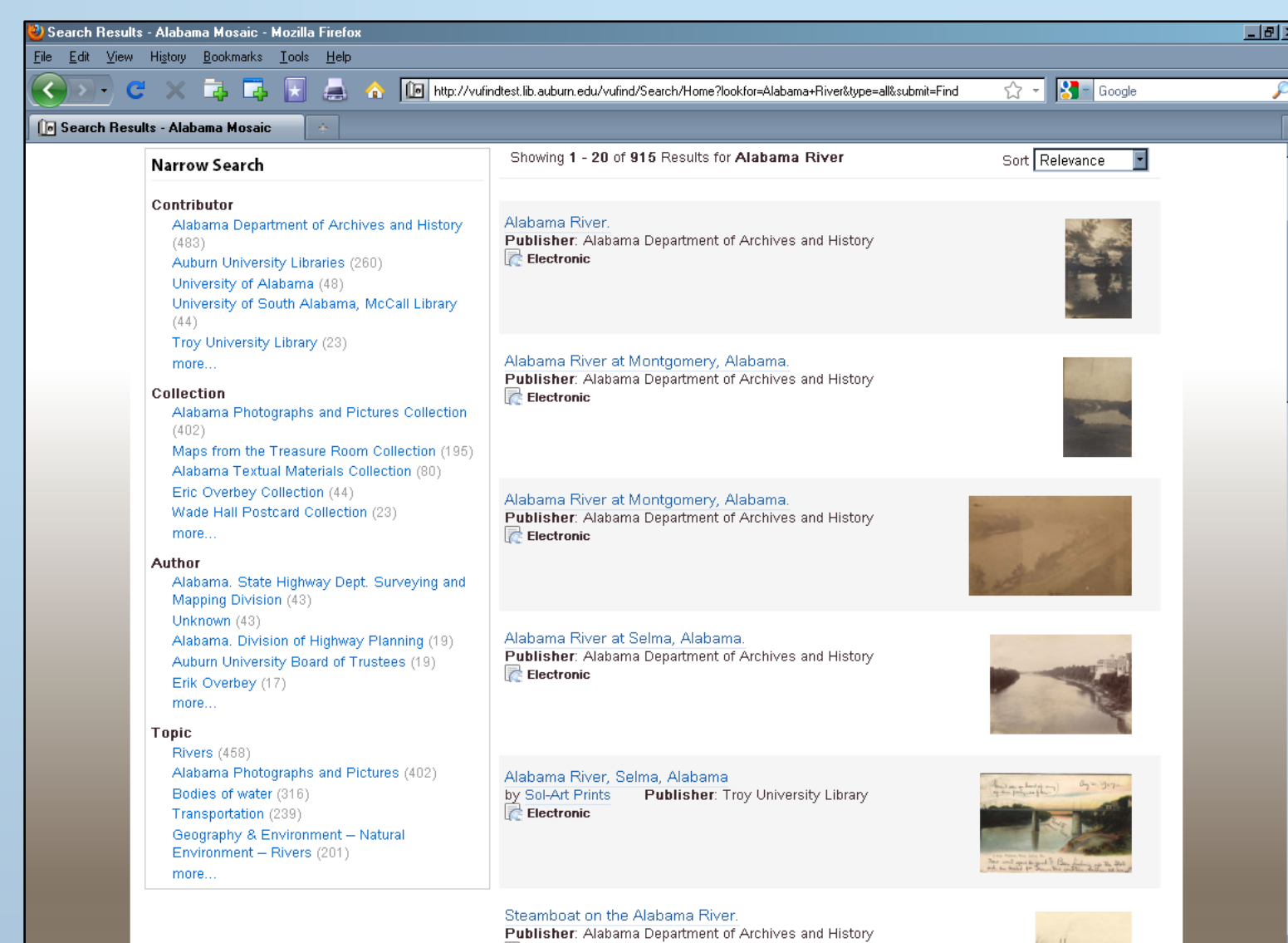
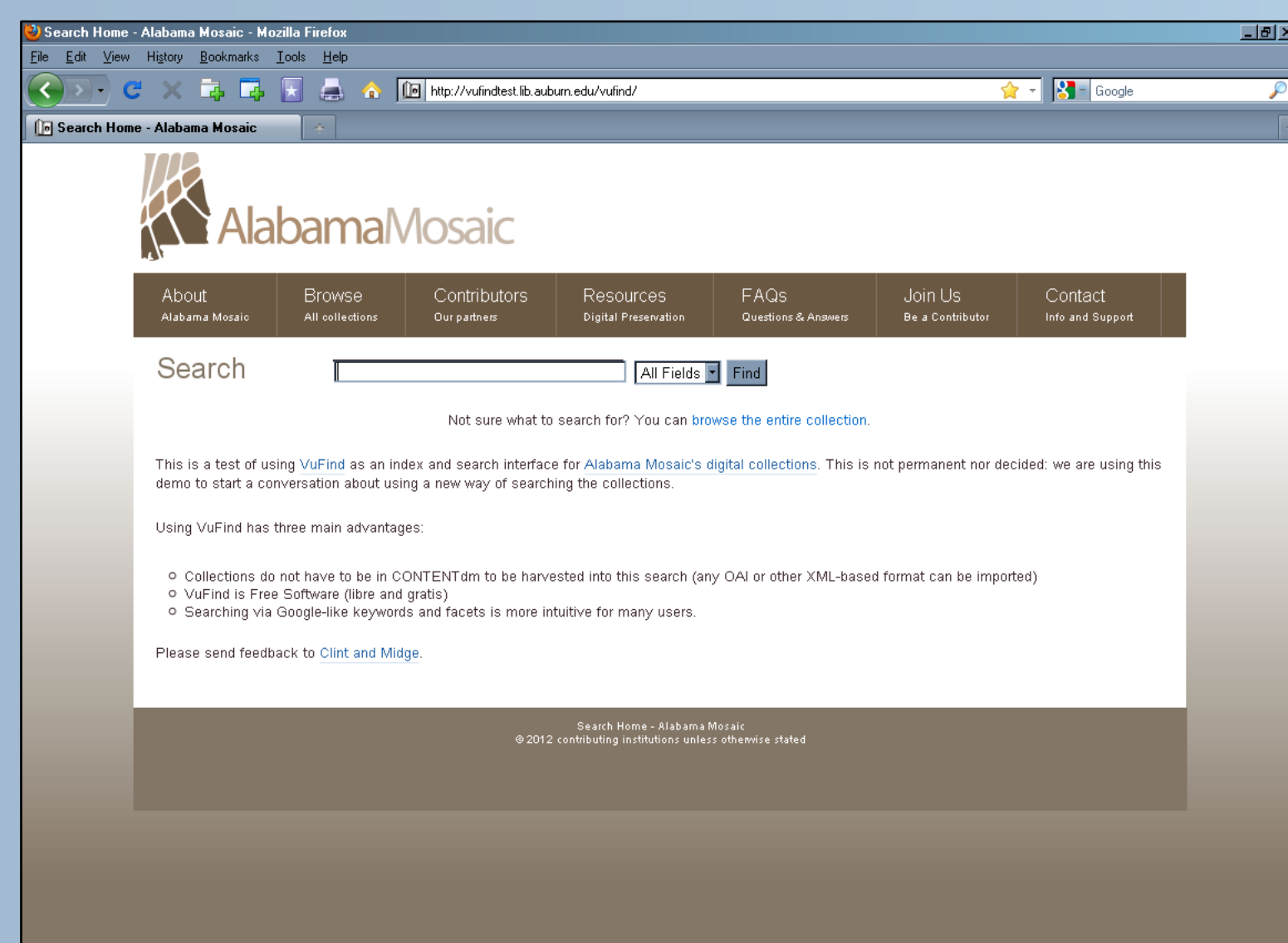


Figure 9. VuFind test home page, showing the search box.

Figure 10. VuFind search results page, showing the top results for a search for "Alabama River." Left-hand facets enable the user to narrow the search by contributing institution, by collection, by author, and by subject term.

Figure 11. MSS search results page, showing the top results for a search for "Alabama River."

CONCLUSIONS

Preliminary testing shows that the VuFind discovery tool software allows AlabamaMosaic to include digital collections from three CMS software packages: CONTENTdm, Acumen, and dSpace. It should also be possible to index collections from other systems provided the collections are OAI-compliant.

Search results obtained with VuFind appear more relevant when compared to MSS. Faceted navigation also allows users to narrow VuFind search results by contributor, collection, author, and topic. Browsing of the entire collection is also possible from the VuFind interface.

FUTURE WORK

Currently, harvesting of collection metadata into the AlabamaMosaic VuFind database is done manually, and collections must be checked periodically for new content. Future work will involve creating a more automated indexing process via cron jobs run at specified time intervals.

Thumbnails are not currently being displayed for collections harvested from dSpace. This is because dSpace handles thumbnails differently from CONTENTdm and Acumen. Future work will involve adjusting the harvesting process to obtain thumbnails for dSpace collections.

With CONTENTdm no longer required for participation in AlabamaMosaic, contributing partners are expected to continue exploring the use of alternative CMSs for their digital collections. Future plans for AlabamaMosaic VuFind include writing XSLT files to harvest collections from the Greenstone and Omeka CMSs.

BIBLIOGRAPHY

Arlitsch, Kenning, and Jeff Jonsson. "Aggregating Distributed Digital Collections in the Mountain West Digital Library with the CONTENTdm Multi-Site Server." *Library Hi Tech*. 23:2 (2005). 220-232.

Denton, William, and Sarah J. Coysh. "Usability Testing of VuFind at an Academic Library." *Library Hi Tech*. 29:2 (2011). 301-319.

Digby, Todd, and Stephen Elfstrand. "Open Source Discovery: Using VuFind to Create MnPALS Plus." *Computers in Libraries*. March 2011. 6-10.

Downer, Sherida, Sue Medina, Beth Nicol, and Aaron Trehub. "AlabamaMosaic: Sharing Alabama History Online." *Library Hi Tech*. 23:2 (2005). 233-251.

Emanuel, Jennifer. "Usability of the VuFind Next-Generation Online Catalog." *Information Technology and Libraries*. March 2011. 44-52.

Featherstone, Robin, and Lei Wang. "Enhancing Subject Access to Electronic Collections with VuFind." *Journal of Electronic Resources in Medical Libraries*. 6 (2009). 294-306.

Ho, Birong, Keith Kelley, and Scott Garrison. "Implementing VuFind as an Alternative to Voyager's WebVoyage Interface: One Library's Experience." *Library Hi Tech*. 27:1 (2009). 82-92.

Houser, John. "The VuFind Implementation at Villanova University." *Library Hi Tech*. 27:1 (2009). 93-105.

VuFind: The Library OPAC Meets Web 2.0. Villanova University. Falvey Memorial Library. <http://www.vufind.org>.