

10-1-2009

Preserving and Providing Access to Digital State Publications in Alabama: A Case of Cooperation

Rickey Best

Auburn University, rbest@aum.edu

Follow this and additional works at: <http://digitalcommons.kennesaw.edu/seln>

 Part of the [Library and Information Science Commons](#)

Recommended Citation

Best, Rickey (2009) "Preserving and Providing Access to Digital State Publications in Alabama: A Case of Cooperation," *The Southeastern Librarian*: Vol. 57: Iss. 3, Article 3.

Available at: <http://digitalcommons.kennesaw.edu/seln/vol57/iss3/3>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in The Southeastern Librarian by an authorized administrator of DigitalCommons@Kennesaw State University.

PRESERVING AND PROVIDING ACCESS TO DIGITAL STATE PUBLICATIONS IN ALABAMA: A CASE OF COOPERATION

Rickey Best

Rickey Best is Dean of the Auburn University Montgomery Library and can be reached at rbest@aum.edu.

The author wishes to express his appreciation to Dr. Sue Medina, Executive Director of the Network of Alabama Academic Libraries, along with the reviewers of this article, for their suggestions and recommendations to improve the work.

Introduction

As initiatives to achieve a more interactive web-format for government information expand, concerns continue to mount over data security and preservation. While efforts to address these issues are ongoing at the federal level, at the state and local levels there continues to be gaps between state and local agencies creating digital publications and services and ensuring the long-term preservation of those publications. The nature of government information on the Web also presents challenges. Some information is generated from databases (often federal) which pull together information and make it readily available. Such data, such as the Alabama's agricultural statistics, are actually derived from the United States Department of Agriculture's National Agricultural Statistics Service. In this case, the data resides in the "Deep Web." These statewide statistics cannot be harvested directly using web-crawling tools.

Among the challenges faced by libraries is a lack of clarity in legislation mandating electronic access to publications at the state level. To address issues relating to preservation and access to "born-digital" state publications the Network of Alabama Academic Libraries in 2005 created a Task Force to examine the issues relating to the preservation of "born-digital" state publications, along with the conversion of print publications to an electronic format. The Task Force examined various practices and developed a recommendation for a cooperative method of identifying and preserving state publications.

Born-Digital State Publications

The internet has created for libraries an uncertain environment for collection, preserving, and making accessible government publications. A number of issues relating to selecting Web publications for preservation and accessibility exist. Among these is the relatively short life-span of government information. Agency webmasters are more concerned with providing citizens updated and current information, rather than in providing long-term access to publications which may have been updated (Lin & Eschenfelder, 2008). While the United States Government Printing Office (GPO) has succeeded in implementing its 1996 plan (more than 1,300 titles were reported as being converted to electronic format only in the 2008 GPO annual report) (United States Government Printing Office [USGPO], 2008), Baldwin and Barnum (2008) report that the decisions on what to digitize and what to preserve are included in the Federal Depository Library Program's (FDLP) Superintendent of Documents Policy Statement 71. They note that "information content remains the primary selection criteria for inclusion in the FDLP, and that the GPO is legally mandated to retain FDLP publications permanently in either the regional depository library collections or in the FDLP electronic collection". In her article "From Double Fold to Double Bind," Sarah Thomas (2002) observed the tenuous nature of digital publications, noting that "digital technologies supplanted one another in very short cycles.... Librarians did not distinguish between the carrier and its content, and they wanted to provide appropriate stewardship for the artifact. Digital objects, by comparison, were often not 'owned' in a permanent sense, but were rather 'licensed' by libraries and consequently not even always accessible and hence not even physically in the library's custody".

In his article “Availability, access, authenticity, and persistence,” George Barnum (2002) reported on the history of the GPOs efforts to alter the FDLP to move towards a “more or mostly electronic program” in order to make the program more “relevant” in terms of user needs. Barnum noted the trend towards a “centralized approach to retention which reorients the responsibilities of libraries away from preservation of individual objects and shifts activity to preservation of access.” The issue of authenticity as described by Barnum identified the GPO electronic collection policy which points users to the origination agency server for as long as possible, while recognizing the necessity of technological developments to ensure the accuracy of digital publications. In describing these efforts, Barnum reported on efforts between GPO, OCLC and other libraries to develop “preservation metadata... that will assure a file migrating from one medium or operating environment to another has not been corrupted or degraded.”

While much of the literature has focused upon access to federal government publications (such as articles by Barbara J. Costello’s *“Moving in the Right Direction: Developments in on-line availability of full-text Congressional committee hearing transcripts”* in *Government Information Quarterly* 25 (2008) 104-117), Lin and Eschenfelder approached the topic from how to select items to preserve. While recognizing Barnum’s issue of preservation of access versus the preservation of the object, Lin and Eschenfelder (2008) focused upon the selection and capture practices used to populate electronic depositories of born-digital state government publications. Lin and Eschenfelder examined efforts underway in three states (Connecticut, New Mexico and Texas) to identify and select web-based state government publications, using different levels of collection building: selective collecting and bulk collection (California Digital Library and the Andrew Mellon Foundation [CDLAMF], 2003).

Selective collecting, as defined by the California Digital Library report, requires human selectors have some prior knowledge about the existence and location of certain specific Web resources, and the individuals select items according to pre-defined collection criteria. Bulk collection utilizes automated web crawlers to capture a target domain or set of sites (CDLAMF, 2003). The trade-off is the specificity with human intervention in terms of identifying and selecting resources, versus the huge amounts of data harvested using bulk harvesting that focuses upon a set of sites.

State Publications in Alabama

Academic librarians in the state of Alabama have long been concerned about access to state publications – the state does not have a funded and operational depository program and librarians must directly contact state agencies to acquire publications. In a digital environment, however, many publications are created, exist, and then disappear often without librarians or citizens becoming aware of their existence. Upon the inauguration of each new governor, many digital records are removed in preparation for transition to the new administration. At the same time, information about specific documents posted on state agency websites disappears as well because agencies are lax in complying with requirements to send web-pages to the Alabama Department of Archives and History for long-term preservation and archiving. Regretfully, citizens are as unlikely to find many of these publications as they a mis-shelved book.

Alabama’s academic librarians have long been concerned about the inadequate public access to state publications and the lack of system to collect and archive publications for long-term access. In September 2005, the Network of Alabama Academic Libraries (NAAL) established a task-force to address issues of access to and preservation of digital state publications. NAAL is “an unincorporated, self-governed and state-based consortium of Alabama’s academic institutions and the Alabama Commission on Higher Education.” Its purpose is to coordinate academic library resource sharing to enhance education and research. An Advisory Council governs NAAL and in approving the Network’s plan created a task force to address issues with state publications. The State Publications and Records Task Force (SPRTF) drew its charge the following activities in NAAL’s Plan (Alabama Commission on Higher Education, 2009):

- *Improve Access to State Publications:* NAAL will develop a plan to expand access to state publications and assure online publications are available electronically by coordinating efforts to create a distributed archival collection for state publications.
- *Digital Preservation:* NAAL will support the Library of Congress initiative encouraging each state to “work toward the effective preservation of information in digital formats” especially the long-term preservation of state publications and records.

The charge to the Task Force stated that “NAAL recognizes that implementation of these tasks can be done only in cooperation with the Alabama Department of Archives and History, the state agency legally responsible for state records and publications. However, there may [be] ways for the NAAL members to help facilitate a statewide effort to assure that this information is not lost to future generations. The charge to the Task Force is to recommend action by NAAL as a consortium and/or its members as individual institutions that will advance the identification and preservation of state records and publications distributed in digital formats” (Network of Alabama Academic Libraries Advisory Council, 2005).

At the initial meeting on the Task Force on 17 November, 2005, NAAL Director Sue Medina recounted the historical concerns of NAAL librarians over the lack of easy access to state publications. She recounted a past NAAL effort by librarians at Auburn University, the University of Alabama at Birmingham, and the University of Montevallo to catalog state agency web-sites into OCLC WorldCat as an initial step to use those web-sites as a means to identify state publications and to begin cataloging them to make the digital formats accessible to the public (Network of Alabama Academic Libraries State Publications and Records Task Force [NAALAC], 2005). The Chair of the Task Force, Diann Weatherly from the University of Alabama Birmingham, reviewed the national conference “Persistence of Memory: Stewardship of Digital Assets” and specifically a presentation by Richard Pearce-Moses. Pearce-Moses articulated the impossibility of building a preservation program based on acquiring digital assets item-by-item. Rather, he advocated strategies using the archival concept of provenance to acquire digital assets, and to use technical solutions to capture materials by categories (NAALAC, 2005). Provenance is the archival principle which states that records should be arranged according to their origins in an organic body or activity (Schellenberg, 1965). Next, Tracey Berezansky, Assistant Director for Government Records at the Alabama Department of Archives and History reported on an Archives project to test the ability of the Internet Archive to “spider” or web crawl the content of State Agency websites and to capture content loaded on those websites. A “spider” is a program that automatically fetches web pages and is used to feed web pages to search engines. It is called a “spider” because it crawls over the Web. Another term for these programs is *webcrawler* (Internet.com, 2009). Consensus of the Task Force members attending the meeting was 1) a need existed to identify digital state publications with the highest priority for public access (subsequently identified as the “Top 100”); and 2) further information on technical solutions was needed. Members immediately began discussions aimed at identifying the “Top 100” and to seek additional information on issues relating to harvesting and preserving state documents.

Defining the Top 100

Tracey Berezansky suggested that the Alabama Department of Archives and History might be able to target a well-defined set of publications for harvesting as part of a project with Internet Archives to archive government publications (<http://www.archive-it.org/>). Rickey Best, Dean of the Library for Auburn University at Montgomery (AUM), volunteered to visit state agency web-sites to determine which were posting publications online. His extensive list was later circulated among the Task Force members to select the Top 100 digital state publications that should be harvested and preserved.

State Documents in an Electronic Age Workshop

The Task Force decided that a workshop would be the best means to gather information about feasible solutions, and the NAAL Executive Council approved funding for the speakers' expenses. "State Documents in an Electronic Age" featuring presentations highlighting three initiatives to harvest, catalog, and archive state publications was held at AUM on 19 July, 2006. The speakers chosen included Joe Natalie, Executive Assistant to the Director for Planning and Outreach at the Illinois State Library; P. Toby Graham of the Digital Library of Georgia,; and Marci Smith, State Documents Librarian of the New Mexico State Library (State Publications and Records Task Force [SPRTF], 2006).

- Natalie described the FindItIllinois (<http://www.finditillinois.org>) program of the Illinois State Library. Natalie outlined an Illinois initiative to create a system to preserve state web-sites for permanent public access. Natalie also discussed the Electronic Depository Initiative (EDI) in Illinois, which requires all state agencies to provide an electronic copy of all publications, including those that are hardcopies as well as those born digitally. EDI serves as the depository for these electronic publications, and are assigned permanent URL's to facilitate access. The state agencies provide metadata for the documents, which are edited by the Illinois State Library's Technical Services Division to facilitate locating the documents using internet search engines.
- Graham reviewed the Georgia Government Publications (GGPO) initiative, a program of the Digital Library of Georgia (<http://www.libs.uga.edu/govdocs/collections/georgia.html>). The GGPO staff digitizes state historical documents and collect electronic publications submitted by state agencies. Graham discussed the digitization efforts for state publications and the access and preservation issues involved in making the documents available through GALILEO (Georgia Library Learning On-Line). In Georgia, approximately 80 state agencies were involved in the project, and Graham recounted issues relating to work-flow and other activities.
- Smith described participation by the New Mexico State Library in a pilot project to test OCLC's Digital Archives Toolbox to harvest and catalog electronic state and federal documents (SPRTF, 2006). The Digital Archives software reflects the description in Barnum's article "Availability, access, authenticity, and persistence" of the GPO / OCLC and partner libraries for developing what has become a digital content management system (Content dm®). Taylor Surface from OCLC assisted with a presentation detailing the Digital Archives Toolbox functionality. (<http://www.bpxchange.org/2006/presentations/msmith1.ppt>.)

The Pilot Project

As a result of the information gathered from the workshop and deliberations, the Task Force identified strategies it needed to accomplish which were approved by the NAAL Advisory Council. The approved tasks included (Network, 2006):

- Licensing the OCLC Digital Archive Toolbox
- Identifying electronic publications on state agency websites
- Harvesting (downloading) these publications
Storing the digital file for public access (presentation copy)
- Cataloging the electronic publications into the library's public access catalog with a link to the publication's digital file for public presentation

- Contributing the records with links to the electronic publication to OCLC [WorldCat](#)
- Sharing the OCLC record number(s) with other NAAL members
- Storing the digital file (archive copy) in an archive for long-term preservation (local archive, OCLC Digital Archive, ADPNet, or other archival storage facility)

Auburn University at Montgomery served as the pilot project site to test the OCLC Digital Archives Toolbox. Task Force members advocated testing the Digital Archives Toolbox to because a) it would allow for “cross-walking” the meta-data information into WorldCat, and b) the stability of OCLC as an organization gave confidence that digital images archived there would be safe.

The Task Force returned its attention to identifying the state publications selected for the pilot project. To facilitate discussion, the staff of the AUM Library provided access to a bibliography which listed the state agencies and the electronic publications found on their web-sites (Auburn University at Montgomery, 2009). In a meeting on 13 December 2006, the Task Force explored criteria it might use to select specific publications for harvesting. Suggestions included perceived demand by the public (for example, The Alabama Driver’s License exam study booklet), the perceived importance of the state agency (the Office of the Governor), and myriad other possibilities, such as electronic only continuations of a long published series. A breakthrough occurred with the suggestion by Lucy Farrow, Head of Public Services at AUM, to use the sixteen state publications reported in the Statistical Reference Index. The Statistical Reference Index (SRI), published by Lexis-Nexis, is the standard reference tool for identifying statistics appearing in reports by associations, state and local government agencies, research centers, and universities. The Index provides access to these statistics through microfiche publications containing the data. For the Alabama State Publications project, the following publications appearing in SRI were chosen (Network, 2006):

- Alabama Agricultural Statistics Annual Bulletins
- Annual Reports from the Alabama Department of Corrections
- Annual Report of the Alabama Unified Judicial System
- Alabama Vital Statistics (Alabama Department of Public Health)
- County Health Profiles (Alabama Department of Public Health)
- Alabama General Election Results (Secretary of State’s Office)
- Alabama Superintendent of Banks Annual Report
- Alabama Department of Education Annual Report
- State of Alabama Comprehensive Annual Financial Report (State Comptroller’s Office)
- Alabama Department of Public Health Annual Report
- Alabama Accident Summary: Statewide Abstracts (Alabama Department of Public Safety)
- Detailed Monthly Statistics (Alabama Department of Human Resources)
- Crime in Alabama (Alabama Criminal Justice Information Center)
- Alabama Public Libraries Statistics (Alabama Public Library Service)
- Annual Report of the Alabama Department of Insurance

Task Force members enthusiastically adopted this suggestion as it provided a manageable set of resources which were already discoverable through a standard reference tool. Task Force members from eight institutions (Alabama A&M University; Alabama Department of Archives & History; Alabama Public Library Service; Auburn University at Montgomery; Jacksonville State University; Samford University; Troy University; and the University of Alabama Birmingham) agreed to select one or more titles or state agencies from the list to investigate issues related to library support and workload involved in locating and identifying state publications; downloading them, and preparing a draft catalog record for the publication. Participants were also asked to identify issues with the consistency of formatting for the publication, once it had been downloaded (Network, 2006).

A public wiki was created (<http://sprtf.pbworks.com/>) to facilitate sharing information resulting from the investigations. Each person was asked to post a detailed report of their assessment of publications found online along with any problems arising during harvesting and cataloging of the documents. Participants identified a number of common difficulties they encountered in attempting to harvest and catalog state publications. These included “poor presentation of publications on the agency websites making them hard to find and/or hard to read; very large file sizes that took a long time to appear on the computer screen; files that were so large that long-term storage would become a major problem; ‘publications’ not structured as a publications that could be downloaded (multiple links to separate chapters of a single title or a mix of documents and spreadsheets); many postings that were ephemera (such as newsletters, brochures); a lack of bibliographic identity [for the publication]; links to a summary but not the entire document ; serial publications with individual URLs for each issue; inconsistent titles; incomplete content (some publication parts updated but other parts not updated); databases used instead of publications to provide information; poor quality or inappropriate orientation for scans of printed materials; publications posted in odd locations such as the ‘About’ section; no publications directory to assist with navigation; publications that could be found with a provided URL but not found by using the site’s own search engine ; and inconsistent records in [WorldCat](#) for the print equivalent (some cataloged as serials, some as monographs)” (Network, 2007). Following this test, it became abundantly clear that the harvesting of born-digital state publications would be a labor intensive activity.

In support of the State Publications project, the AUM Library had applied to NAAL for grant funding to license the OCLC Digital Archives Tool Box. The information provided at that point in time was that the Tool Box offered web-crawler software that could be profiled to automatically harvest publications. By licensing the Digital Archives services, AUM sought to test the feasibility of using the Digital Archives Toolbox to automatically harvest and preserve digital publications of the state agencies (Network, 2008). The license was signed for AUM to use the Digital Archives Tool Box, with Contentdm® as the preservation platform Content dm® (owned by OCLC) is a web-based software designed for the storage, management, and delivery of digital collections via the Web. The software integrates institutional metadata with WorldCat to promote discovery (Contentdm, 2009). The software can either be downloaded to an institutional computer system, or can be hosted by OCLC. At AUM, work began in February of 2008 to set-up the software. During the set-up period, we were informed that a software upgrade taking place in April would discontinue the automatic harvester feature. As was explained by OCLC representatives, the reasons for discontinuance related to the complexity of the software and the inability to get it to work consistently and reliably. With the release of the new software, we began using the Web Harvester software. Because we could not establish a protocol for the software to automatically check and harvest web based publications, we were manually identifying and harvesting those items. Unfortunately, the initial version of the software had problems in harvesting certain types of files and the harvester was withdrawn for further development. The Web Harvester software was re-released in June, and following this release, no further technical difficulties were encountered (Auburn University at Montgomery NAAL, 2009).

During the course of the project, the AUM Library staff harvested sixty-two state publications, cataloged them into WorldCat, and preserved the images in Contentdm®. An example of the OPAC display for the Department of Corrections Annual Report is provided in Image 1.



Image 1
AUM OPAC Record for Alabama Department of Corrections Annual Report

In conducting the cataloging for the records harvested, the AUM Library staff was responsible for creating meta-data for the publications in order to export them from the Web Harvester to Contentdm®. The Content dm® interface has forty-five fields for data input, including title (a required field), subject, description, creator, publisher, contributors, date, type, format, identifier, source, language, relation, coverage, rights, audience, alternative title, description (table of contents), description (abstract), date created, date (valid), date available, date issued, date modified, format (extent and medium), fourteen fields defining the relation of the item being harvested, coverage, audience, POI (Purl-based Object Identifier) title, POI date, and permissions (Woodley, Clement, & Winn, 2005). Once the fields were input, the metadata was transferred using the Dublin Core protocols and uploaded into WorldCat. The Dublin Core Metadata Initiatives project provides standardized descriptive metadata that will facilitate searching and retrieval of digital items. Specifically, “the Dublin Core standard includes two levels: Simple and Qualified. Simple Dublin Core comprises fifteen elements; Qualified Dublin Core includes three additional elements (Audience, Provenance and Rights Holder), as well as a group of element refinements (also called qualifiers) that refine the semantics of the elements in ways that may be useful in resource discovery. The semantics of Dublin Core have been established by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship and practice” (Hillman, 2005).

Image 2 includes the metadata converted into MARC format as it appears in the bibliographic record in the Voyager OPAC.



Image 2
Metadata in MARC Format AUM OPAC

Within the cataloging record for the item, a permanent URL is created and the document is housed in a

Content dm® server at OCLC's headquarters in Dublin, Ohio. Image 3 shows how the document appears in the public interface of Contentdm®. By clicking on the permanent URL in the Library's OPAC, the user is delivered to the screen in Image three.

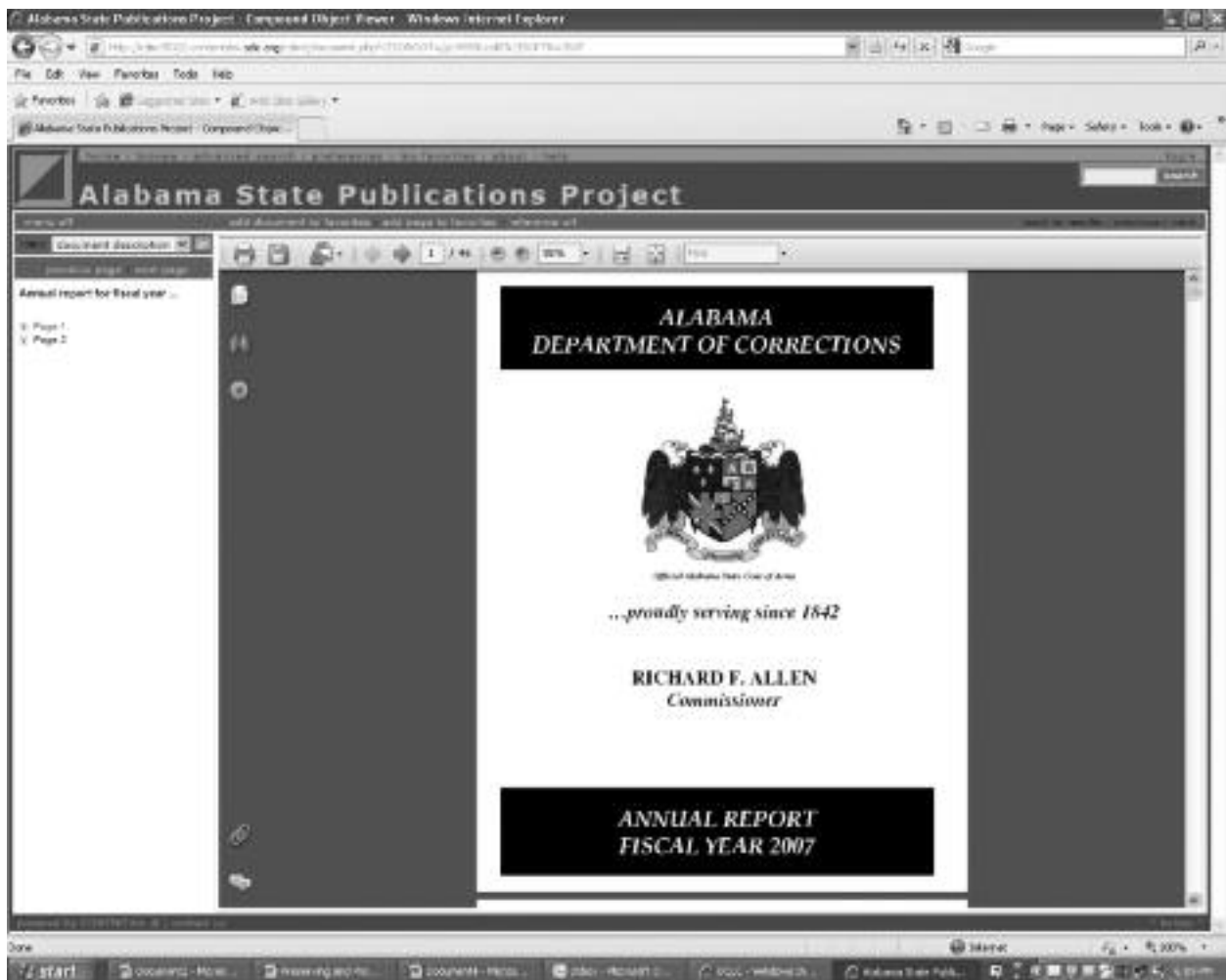


Image 3
Contentdm® View of Department of Corrections Annual Report

Project Results

For the period of the grant funding (February 2008-April 2009), a total of 62 state publications were harvested, cataloged, and archived into Contentdm®, with access to the digital publications made available through WorldCat. In conducting the harvesting, the AUM Library staff identified numerous problems / difficulties with the Digital Archives software.

- The lack of standards for digital publication by the state agencies created problems for harvesting the documents. The Web Harvester software was unable to harvest a number of documents because of incompatible file formats. As an example, the Task Force members had identified Accountability Reports for local schools from the Alabama State Department of Education as a document to harvest, catalog, and make available. Unfortunately, the reports are only accessible by using drop down menus or by retrieving by FTP from zipped files. The schools must be individually identified and harvested, and many of the reports lack URLs by which they can be retrieved.
- Some state agencies update their publications, but without doing a manual review of the publication, it was not possible to know a priori that the document had been updated.
- The lack of sophistication with the web crawling software and the fact that it was not seamlessly integrated into Contentdm® resulted in the AUM Library staff being required to identify and harvest titles individually.
- The project demonstrated that the Web Harvester software and Contentdm® were adequate to undertake the project of capturing and licensing state publications. However, the project also demonstrated the necessity of a significant amount of staff time being involved in identifying and manually harvesting the publications. It was felt that student workers could be trained and used to check state agency websites to determine whether changes had been made to previously harvested publications.

Coordinating with the Alabama Department of Archives & History and State Publications

The Alabama Department of Archives and History was represented in the Task Force membership by Tracey Berezansky. Berezansky informed the Task Force of the Department's experiment using Internet Archive to support the harvesting and digitizing of state agency websites. The Department, however, does not provide cataloging at the individual title level for publications, but rather catalogs at the collection level. Whereas in librarianship, the focus is upon cataloging a discrete item, archival cataloging focuses upon collections, consisting of "documents organically created, accumulated, and/or used by a person or organization in the course of the conduct of affairs and preserved because of their continuing value. They most often consist of aggregations of documents (largely unpublished) and are managed as such..." (Society of American Archives, 2004). The bibliographic records are not presently being loaded into WorldCat (Network, 2006). Berezansky noted that the Department was already capturing publications, and stated that if the Task Force could identify the top 100 publications, the Archives staff could ensure that those publications were captured.

As identified by the Task Force, the issue of lack of standardization in the digital formatting of state agency publications is a significant obstacle to harvesting and making those publications available, as well as to ensuring the long term digital preservation of the item. The staff of Archives and History has worked to provide the agencies with guidelines to ensure some standardization of formatting and design of the digital publications.

At the December, 2008 meeting Berezansky noted that the Archives had moved to Contentdm® as the platform for hosting the harvested websites and publications (Network, 2008). The harvested documents are accessible through the Archives website (<http://216.226.178.196/cdm4/statepub.php>) as well as in

Alabama Mosaic (<http://www.alabamamosaic.org/collections.php>), a digital project sponsored by NAAL and hosted by Auburn University. One of the issues the Task Force discovered relating to the cataloging of state publications was the need for authority work on the names of the state agencies. Berezansky reported that the staff of the Archives had done a great deal of work on authority control for the state agencies, and that these records would ultimately be added to WorldCat, but no time frame currently exists for that happening (Network, 2008). However, issues pertaining to access to the digital versions of state publications remains at the forefront. The Department of Archives and History utilizes the Ex-Libris Voyager system for an Integrated Library System. Items are cataloged into Voyager using the archival approach of provenance, i.e., at the collection level. Searches in the Archives ILS system for “Alabama Department of Corrections Annual Report” and various truncations on that title retrieved nothing. The Annual Reports, however, are available in Electronic State Publications Collections of Alabama, available at <http://216.226.178.196/cdm4/statepub.php>. Searching for specific reports are rather cumbersome, however, as each report is cataloged as a separate entity, and there is no serial record to provide title level access, with the individual reports nested beneath the title.

An Agreement for Cooperation

To address the issue of improved access, the Task Force recommended that volunteers be solicited to form a new working group consisting of NAAL Library representatives and representatives from the Department of Archives and History to address issues of creating title level MARC records to be loaded into WorldCat. The working group would consist of catalogers and digital collections librarians who would work to develop guidelines for cataloging state agency publications. The group would also work to provide a notification system to other libraries within the state (including public and special libraries) when new records became available in OCLC to allow those libraries to download the records into their local catalogs.

Conclusion

The NAAL Task Force was able to meet its charge for accomplishing the identification and preservation of state publications that were born “digitally.” The formation of a working group to participate with the Alabama Department of Archives and History to ensure the serial cataloging of state publications and loading them into OCLC WorldCat offers great hope of improving access to Alabama state publications, not only within the state but throughout the world. The Task Force verified the importance and value of standards in the design of digital publications. In addition, the Task Force reaffirmed that there is no simple, turn-key system that will automatically harvest all documents, catalog them, and distribute them to a preservation system without significant staff involvement.

The failure of software that can be profiled to automatically harvest publications is the greatest impediment in the Digital Archives suite of software. The average amount of time required for the staff of the AUM Library to harvest, create the metadata, catalog, and upload the item into Contentdm® and the bibliographic record into OCLC was 75 minutes. For the 62 documents harvested by AUM, a total of 77 and one-half hours were required. The actual time depended upon a number of factors: size of the document(s) and the number of issues involved in the harvesting being the most significant. Without the capability of the harvester software to automatically ingest published issues, for institutions to participate in harvesting digital content will require a significant investment of staff time to individually locate, harvest and catalog the publications.

In spite of the limitations to the Harvester software, the OCLC Digital Archives software has some significant advantages. By placing the images into Contentdm® and having Content dm® hosted at the OCLC Headquarters, we were able to ensure long-term preservation of the images. By importing the metadata into the MARC records and loading in WorldCat, we have increased the availability of the documents. Finally, as a group, we reaffirmed our feelings of the value and importance of state publications, not only for research needs but also for the average citizen. The levels of accountability

described by David Osborne and Ted Gaebler in their 1992 book, *Reinventing Government*, ‘argued for a “customer driven” government by pushing control of the bureaucracy and into the community”’ (Osborne & Gaebler, 1992). In Alabama, through the work of the State Publications and Records Task Force, we are working to push information out from governmental agencies to the public at large. In an age of accountability, we believe that these efforts are a significant benefit to the citizens of the state.

References

- Alabama Commission on Higher Education. Network of Alabama Academic Libraries Homepage. <http://www.ache.state.al.us/NAAL/> (accessed April 30, 2009).
- Auburn University at Montgomery. NAAL State Publications Project Report; prepared by Colleen Valente. Unpublished manuscript in possession of the author. March 24th, 2009.
- Auburn University at Montgomery. State Agencies Pub[lications] and Website Info. <http://aumnicat.aum.edu/departments/administration/StateAgenciesPubWebinfo.pdf>. (accessed May 7, 2009).
- Baldwin, Gail and Barnum, George (2001). On My Mind: Government Documents for the Ages. *American Libraries* 32 (11), 38.
- Barnum, George. (2002) Availability, access, authenticity, and persistence: Creating the environment for permanent public access to electronic government information. *Government Information Quarterly* 19, 37-43.
- California Digital Library. Project Report: Web-based Government Information: Evaluating Solutions for Capture, Curation and Preservation. Andrew Mellon Foundation. Project Report, 2003. http://www.cdlib.org/programs/Web-based_archiving_mellon_Final.pdf. (accessed October 5, 2009).
- Content dm® At A Glance. <http://www.oclc.org/contentdm/about/default.htm>. (accessed October 5, 2009).
- Hillman, Diane. Dublin Core Metadata Initiative. Using Dublin Core. November 7th, 2005. <http://dublincore.org/documents/usageguide/>. (accessed October 5, 2009).
- Internet.com’s Webopedia. <http://www.webopedia.com/TERM/s/spider.html> (accessed October 1, 2009).
- Lin, Chi-Shiou and Eschenfelder, Kristin R. (2008). Selection practices for Web-based government publications in state depository library programs: Comparing active and passive approaches. *Government Information Quarterly*, 25, 5-24.
- Network of Alabama Academic Libraries Advisory Council. FY 2006 Plan. Charge to the State Publications and Records Task Force. <http://sprtf.pbworks.com/Full+Charge+to+the+Task+Force>. (accessed April 30, 2009).

- Network of Alabama Academic Libraries. State Publications and Records Task Force. Minutes of Meeting, December 13th, 2006. <http://sprtf.pbworks.com/Minutes+12-13-06>. (accessed May 7, 2009).
- Network of Alabama Academic Libraries. State Publications and Records Task Force. Minutes of Meeting, March 7th, 2007. <http://sprtf.pbworks.com/Minutes+3-7-07>. (accessed May 7, 2009).
- Network of Alabama Academic Libraries. State Publications and Records Task Force. Minutes of Meeting, February 20th, 2008. <http://sprtf.pbworks.com/Minutes+3-7-07>. (accessed May 8, 2009).
- Osborne, David and Gaebler, Ted. (1992). *Reinventing Government*. Reading, Mass.: Addison-Wesley Pub. Co., p. 180-186.
- Schellenberg, T. R. (1965). *The Management of Archives*. New York: Columbia University Press, p. 41-45.
- Society of American Archives*. (2004). Describing Archives: A Content Standard. Chicago, IL: Society of American Archivists, p. xi.
- State Publications and Records Task Force. Workshop Agenda for "State Documents in an Electronic Age." Unpublished document in possession of the author. July 19, 2006.
- Thomas, Sarah. (2002) From Double Fold to Double Bind. *The Journal of Academic Librarianship*, 28(3). p. 104-108.
- United States. Government Printing Office. (2008). *Annual Report*, 2008. Washington, D.C.: GPO, 2008. p. vii. http://www.gpo.gov/pdfs/congressional/GPO_08AnnualReport_final.pdf. (accessed October 5, 2009).
- Woodley, Mary S.; Clement, Gail; and Winn, Pete. *Dublin Core Metadata Initiative. November 7th, 2005. DCMI Glossary* <http://dublincore.org/documents/usageguide/glossary.shtml#P>. (accessed October 5, 2009).