

DIGITAL GOVERNMENT INFORMATION INITIATIVES

The Arizona State Library and Archives has undertaken a number of initiatives to ensure that records and publications in digital formats are captured and preserved. As a result of its work, Arizona is seen as a leader in efforts to solve the complex problems that must be overcome to build a durable collection of digital government information.

Web SafetyNet Archives

The State Library and Archives, along with a number of other states, continued its work on a project led by the Illinois State Library to capture government publications on agency websites. The project was funded with federal funds from the Institute of Museum and Library Services. What was originally a test project is now moving into program status.

The Library has harvested documents from agency websites twelve times since beginning this project in July 2004. Currently, the SafetyNet software is harvesting publications from 141 websites on a monthly basis. The most recent harvest captured 261,623 files, an estimated 105,000 documents.

These files are being stored in compressed format. They are not immediately accessible through the web, but can be retrieved on demand.

Arizona Model for Curating a Repository of Web Publications

One of the biggest challenges of building a collection of web publications is knowing what to keep. Traditional approaches of selecting publications one at a time are impractical given the vast number of publications on the web. In response, many attempted to capture everything on the web, but that resulted in enormous duplication and capture of information with little value.

Arizona developed another model for selecting and describing state agency publications. That model has received significant attention over the last year, and staff have been invited to speak about the model at conferences hosted by the National Historical Records and Publications Research Fellows, the Library of Congress, American Library Association, the California State Library, and the Association of Research Libraries. The model was adopted as the basis for the Web Archives Workbench research project (see below).

A white paper describing the "Arizona model," included in last year's annual report, was published in *DttP: Documents to the People* 33:1 (Spring 2005).

Web Archives Workbench: Research at the University of Illinois

In 2004 the Library of Congress National Digital Information Infrastructure Preservation Program awarded a grant to the University of Illinois at Urbana-Champaign and OCLC to test the Arizona Model (see above) and to develop software to support the implementation of that model. Richard Pearce-Moses, Director of Digital Government Information, has been a key participant in this project.

The alpha version of the domain and entities tools was released for testing by project partners in May 2005. The alpha version of harvesting tools will be released in early 2006, and the remainder of the tools will be released before the end of 2006.

These tools make it possible to begin a more selective harvest of web publications than the SafetyNet software. More important, these tools will capture the web publications in a format that allows them to be made accessible through the Arizona Memory Project.

Web-at-Risk: Research at the California Digital Library

The Library and Archives is participating in another research project funded by the Library of Congress' National Digital Information Infrastructure and Preservation Program. Under leadership from the California Digital Library, the project will develop a set of tools for use by libraries to capture, curate, and preserve collections of web-based government and political information, including state agency publications and campaign literature.

Authoritative List of State Agency Websites

As a result of the work done on harvesting state agency publications from the web, the Library and Archives developed the most comprehensive list of state agency websites and web servers. That information was shared with the Government Information Technology Agency (GITA). It is currently available on the web at <http://dev.lib.az.us/safetynet/agencies.aspx>.

Specialized Search Engine for State Agency Publications on the Web

For a number of years, the Library and Archives has been experimenting with software to help the public find the information they need on state agency websites. The search engine is online at <http://find-it.lib.az.us/>.

The project has not been given significant attention or changed to program status due to concerns about software and questions about the need for such a program given the power of Google and other free search engines. Google is a powerful search engine that the public can use at no cost. However, Google is not ideal. Google does not index all web pages, nor does the state have the ability to know exactly what Google is indexing. Because Arizona agencies changed their domain naming conventions, it is now almost impossible to limit results to web pages on state websites. As a result, state information is often buried under many other results.

Because of these concerns about public search engines' ability to provide effective access to Arizona, especially the inability to limit search results to Arizona websites, the Library and Archives is putting more energy into this project.

Electronic Recordkeeping System Assurance Statement

The Library and Archives is very concerned that records created and maintained in digital form can be preserved as long as needed by the state. In particular, the Library and Archives is concerned that records of permanent value not be lost through hardware or software obsolescence. The Library and Archives is also very concerned that those systems can produce records that are trustworthy and admissible as evidence in court.

The Records Management Division developed a model agreement that allows agencies to certify to the Library and Archives that their systems can produce records that are acceptable as trustworthy evidence and that permanent records can be preserved. (A copy of the model assurance statement is attached.) The Arizona Office of Administrative Hearing was the first agency to provide the Library and Archives with a certificate of assurance.

Arizona Memory Project

Every library and archives needs stacks to store the collections. The Library and Archives has acquired hardware and software that will serve as the stacks for its digital collections. The system will be used to store web documents captured through the Web Archives Workbench, as well as reports received directly from agencies. The system was acquired using federal funds from the Institute for Museum and Library Services.

The Library and Archives is making the system available to other libraries, archives, museums, and historical societies in the state. Other institutions can now get their collections online for much less money than would be required if each institution installed their own system. This collaborative project will ensure that Arizonans will have access to a wealth of historical information for the Centennial.

National Archives and Records Administration: Electronic Records Archives

The Archivist of the United States, Dr. Allen Weinstein, ask Richard Pearce-Moses to serve on the National Archives and Records Administration's Advisory Committee for the Electronic Records Archives (ERA). The National Archives has issued a \$300 million contract to Lockheed Martin to develop the ERA over the next five years. The ERA system will be a comprehensive, systematic, and dynamic means for storing, preserving, and accessing virtually any kind of electronic record, free from dependence on any specific hardware or software. ERA, when operational, will make it easy for NARA customers to find the records they want and easy for the National Archives to deliver those records in formats suited to customers' needs. In a press release issued by the National Archives, the committee of twenty includes "recognized experts and leaders in their field." In part, Pearce-Moses' participation will ensure that states have sufficient input on the system, in hopes that the system can be adapted.

Other Presentations

Arizona State Library and Archives staff presented a number papers at conferences this past year.

- Society of Southwest Archivists, May 2005. "A Bridge to the Future: Committing Intentional Acts of Memory," Richard Pearce-Moses. Keynote speech. Online at <http://www.lib.az.us/diggovt/presentations/Bridge.pdf>.
- International Federation of Library Associations (IFLA), August 2005. "From Bibliographer to Curator: Archival Strategies for Capturing Web Publications, GladysAnn Wells and Richard Pearce-Moses. Online at http://www.lib.az.us/diggovt/presentations/ifla_2005.pdf.
- Society of American Archivists, August 2005. "Winds of Change: Blown to Bits," Richard Pearce-Moses. Closing plenary. Available online at http://www.lib.az.us/diggovt/presentations/saa_2005.pdf.