

Ingestion and Indigestion: Negotiating Between States in the Digital Cloud

Author Jerry Handfield; Washington State Archives; Olympia, Washington

Abstract

Begun in 2007 under the leadership of the Washington State Digital Archives, the Multi-State Preservation Consortium has brought together 10 states and dozens of different agencies to create a centralized digital archive based on the Washington State model. Though highly successful the process has not always been smooth, and the lessons we have learned can be applied to other multi-partner digitization efforts.

History of the Washington State Digital Archives

Established in 2004, The Washington State Digital Archives is the nation's first archives dedicated specifically to the preservation of electronic records from both State and Local agencies that have permanent legal, fiscal or historical value. Our records range from returns from Washington Territory's first election in 1854; to decades of laws and ordinances; to marriage and death records from the 1800s to the 1980s; to "born digital" records such as emails and audio files from the state legislature. The Digital Archives currently (March 2009) holds over 80 million records, making it one of the largest databases and the single largest digital repository of historic documents in the world.

The Digital Archives grew from rare alignment of vision, leadership and opportunity. In the late 1990s, Washington State was enjoying robust economic growth, fueled in part by Microsoft, Boeing, and other technology companies. The elected officials in Washington had resources to spend and an appreciation of technology. Strategic planning for the Digital Archives began in March of 2000, and planning for the physical design and technical infrastructure of the facility occurred during 2002. The state's 2001-2003 Capital Budget (SSB 6155) authorized the Secretary of State to enter into a financing contract for the construction. Early site work for the facility in Cheney, Washington began July 2002 with construction beginning January 2003.

Since its establishment the Digital Archives has attracted international interest, hosting delegations from many states and half-a-dozen countries, including China, Australia, Canada and the UK.

The Library of Congress and the Multi-State Preservation Consortium

The experiences of the Washington State Digital Archives are being shared with the nation through the Multi-State Preservation Consortium (MSPC). The MSPC is the most ambitious multi-agency digitization effort in the nation. Begun in 2007, the project has brought together 10 states under the leadership of the Washington State Digital Archives to preserve governmental



Figure 1: Washington State Digital Archives

records using the Washington State model. These states are Alaska, Colorado, Idaho, Indiana, Louisiana, Montana, Nevada, North Carolina, Oregon, and Washington.

The Multi-State Preservation Consortium grew out of the Library of Congress' National Digital Information and Infrastructure Program (NDIIPP). In 2000, Congress appropriated \$100 million for the Library of Congress to lead the effort to engage other stakeholders in efforts to capture, preserve and make available digital content. A special focus of NDIIPP is to preserve "born digital" records such as electronic publications. The Multi-State Preservation Consortium is a demonstration project funded by \$925,000 in grants from NDIIPP. This is a much smaller figure than initially conceived. The original plan was for a \$6 million project that would host servers in each partner state and develop a greater level of customization. In addition to the partner states listed above, non-state partners in the project include the Microsoft Corporation, Eastern Washington University and others.

The Multi-State Preservation Consortium was tasked with developing a centralized regional repository for state and local digital information. Sharing the Washington model with other states would create a cost-effective interstate technological archiving system and demonstrate a scalable approach to preserving and making available at-risk digital government information. The goals of the MSPC were to 1) identify varieties of digital records for preservation; 2) develop a broad-based workflow to transfer data to the archive; 3) to establish access procedures, including searchable public Web sites; and 4) to publicize best practices and lessons learned.

As of March 20, 2009, the project has digitally preserved 550,143 documents from 10 states and agencies in three different record series. The project is building "the other national archives."

The Consortium So Far

The project kickoff meeting was March 18, 2008. Working behind the scenes, the Washington State Digital Archives team had already created individual web sites for each of the initial four partner states and ingested state records into each of the sites. By doing this, the demonstration project was proven to be a success at the first meeting.

There was great enthusiasm and confidence in the project. Partner states left the kickoff meeting with a set of requirements for continued participation, such as choosing a URL for their digital archive websites, and identifying records for the project.

Dealing with 10 state agencies, complications quickly arose. Even something as simple as choosing a URL could lead to protracted negotiations and disagreements. The initial demonstration websites had URLs that were extensions of the Washington State Digital Archives site. For example, Idaho was www.idaho.digitalarchives.wa.gov. Obviously, it was important to create a URL for partners that did not reference the fact that their servers hosting the website were actually in Cheney, Washington. But what should the new URLs be? In Idaho, the state archives and state library had to come to agreement, which led to some complex negotiations. Should the URL reference the archives or the library, both of which were contributing material? And it was not possible to determine a URL without determining the name of the website. Would it be the Idaho State Digital Archive? The Idaho Digital Libraries? The Idaho Digital Archives—or Archives and Libraries? Eventually they did negotiate a solution. The banner on the website reads “The Idaho State Historical Society: Idaho’s Digital Archives,” and URL is www.digitalarchives.idaho.gov.



Figure 2: Banner for the Idaho Digital Archives

Determining website banners and URLs was the least of the issues that had to be resolved. Archivists are possessive of their records, and no less so of digital records. The Washington State Digital Archives had demonstrated at the kickoff meeting that the technological problems could be solved, but changing the mindsets of archivists, librarians, and government officials across the country would prove more complicated.

Some states had revenue concerns. In Louisiana, relevant and already digitized records were located. However, the agencies holding the records were used to selling access to the records and not willing to allow them to appear online free of charge.

Other states were concerned about the Intergovernmental Agreement. Early in the process, the WSDA developed a simple three-page agreement for partner states to sign. Several states, including Alaska and Oregon, refused outright to sign the document. In other states, the memoranda became an internal tug-of-war between state libraries, state archives, and other agencies.

Most difficult was Indiana. The Indiana Attorney General amended the 3-page agreement into a 19-page one. A timely retirement in Indiana paved the way for the original agreement to go through.

Partner Participation: Ingestion

With the negotiating complete, and the fears of some partners assuaged (though not eliminated), the work of the Multi-State Preservation Partnership could proceed.

Idaho was the first state to produce digital records for the archives on August 5, 2008. These records consisted of 73,000 searchable indexes, including Civil War indexes, Naturalization Records, and Old Age Pension Records.

The Oregon Library produced the first set of electronic publications (E-Pubs). These 1,715 records from 332 different agencies in Oregon included records from agencies as diverse as the Oregon Fireworks Task Force and the Oregon Genetics Program. These E-Pubs were the first set of records ingested in which the indices included the referenced digital objects.

A few days after Oregon's E-Pubs, the Idaho Library produced its own set of electronic publications. These came from four different agencies: the Department of Correction, the Criminal Justice Commission, the Idaho State Liquor Dispensary, and Bureau of Homeland Security, and consisted of 211 records.

In February of 2009, Montana produced the first audio files as part of the project. The audio files are of House, Senate, Legislative, and Joint session hearings from 2005 to present. These files will be keyword searchable thanks to software developed as part of a partnership between the WSDA and Microsoft.

In March of 2009, Colorado added its first set of indices to the project. They provided 469,258 records from 39 agencies. Colorado has identified additional digital records that they plan to participate with soon.

Partner Participation: Indigestion

This program is headed towards a successful conclusion—but not because all of the problems were solved. Some were worked around, others continue. As of March 2009 most partner states have signed the participation agreement, identified the record series they will contribute, and have either sent their digital files to the WSDA or are in the process of doing so. States are working towards creating successful workflows to suit their unique needs and bureaucracies.

Success creates its own problem. Publicity surrounding the Multi-State Consortium has generated interest in non-partner states eager to join the program. Nevada and North Carolina were invited to join the MSPC in the fall of 2008. Nevada responded promptly and returned a signed agreement. They are digitizing publication and preparing them for ingestion. North Carolina on the other hand has yet to produce a signed agreement.

The Future of the MSPC

In March of 2009 the Washington State Digital Archives was informed by the Library of Congress will provide an additional \$100,000 to extend the project another year, through 2010. One future focus of the program will be on sustainability. Developing a partner consortium with a schedule of maintenance fees is one possible avenue for continuing the work of the MSPC. The WSDA is also seeking additional sources of funding, both public and private, to expand “the other National Archives.”

Conclusion—Lessons Learned

The difficult part of digitization partnerships is never the technology. The Multi-State Preservation Consortium was successful because of strong leadership and extensive experience brought to the task by the Washington State Digital Archives. However, strong leadership is not enough.

A major lesson learned in the MSPC is the need to begin cooperation early on. Many partner states were not eager to sign an agreement they had not helped to design, or to adopt a workflow and standards not developed in-house.

Another lesson was that many organizations are not aware of the digital records they already hold. Many facets of state governments create digital records, but the state archiving communities are not always aware of these records or involved in their preservation.

The varying needs of partners also have to be considered from the beginning. We have discovered that many states were eager to participate but lacked personnel to digitize records or even to identify the digital records already in their possession.

Author Biography



Jerry Handfield was appointed Washington State Archivist by Secretary of State Sam Reed in 2001. In Indiana, he served as State Archivist for 14 years. Jerry holds a BA from Yankton College, an MA from the University of South Dakota, and a Ph.D. in history from Indiana University. His career includes many years of teaching college level history, oral history, and fieldwork for the Indiana State Library and the Indian Historical Society. Jerry was appointed Director of the Indiana Commission on Public Records by Governors Frank O'Bannon and Evan Bayh. His

professional experience includes leadership in several national organizations, special assignments in Indiana state agencies, and archives consultant to the government of Kazakhstan.