



STATE OF WASHINGTON DIGITAL ARCHIVES ¹

Background. The State of Washington is geographically diverse, with the Cascade Mountain Range separating the western half of the state from the east. In winter time, snowstorms and avalanches make the Cascades notoriously difficult to cross.

The State holds all Washingtonians' official records in repositories in and around the capital city of Olympia, located at the southern tip of Puget Sound. To reach Olympia can be difficult, especially for easterners who must travel distance and across elevation. State offices are open Monday through Friday, 9-5.

"What if you cannot travel?" asked Adam Jansen, the state's Digital Archivist and Deputy State Archivist. "These are the records of people's rights, the things that most directly affect them on a daily basis - marriage records, land records, powers of attorneys. We are holding them as custodians. But how do you gain *access* to your records?"

Sam Reed. In the mid 1990's Washington's state Auditor, Sam Reed, initiated an imaging program in the Olympia region, offering to scan and index records, transmitting the images rather than the paper to the state Archives for storage. But accustomed as they were to handling and storing microfilmed images, the Archives rebuffed his efforts to transmit digital images, as they lacked the capability or technical "know how" to acquire, process or store the digital images that the Auditor might transmit.

Upon his election to Secretary of State, Reed undertook to change that. Sensing that the time was right and the need great, he sought and obtained legislation to fund the state's digital archives, for the purpose of preserving the electronic records of long-term legal, historical, or fiscal value to the State of Washington.

"If Abraham Lincoln wrote the Gettysburg Address on a laptop," Reed declared, "we may not have it today. Electronic records have been disappearing at an alarming rate because we've had no means to preserve them. These are records we need to make public policy, to conduct day-to-day business, and to prepare for the future."

¹ This case was written by Zachary Tumin, Executive Director, Leadership for a Networked World Program, John F. Kennedy School of Government, Harvard University. Copyright @ 2007 by the President and Fellows of Harvard College. No part of this publication may be reproduced, revised, translated, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise) without the written permission of the John F. Kennedy School of Government.

Under the legislation, any agency that received any public funds in the State of Washington was now eligible to send its electronic records of permanent value to the State Archives. “We are talking somewhere in the tune of 3000 state and local agencies who are eligible to send those records to us for long-term preservation,” said Jansen.

“Our goal,” said Jansen, “is to preserve these records and to make them accessible in the future in perpetuity to the citizens of the State of Washington because the state does not own any records. It produces records in the course of business and preserves those records for the citizens. If you do not have the records, if you cannot show the records, if you cannot make them accessible, then we do not have a democracy.”

Jansen Runs the Project. Jansen was hired in 2003 to launch the digital archiving project. “I am the first digital archivist in the State of Washington,” said Jansen, “and as my boss likes to say, I am going to be the last digital archivist for the State of Washington.” From this point forward, Jansen observed, “all archivists need to be digital archivists.”

Jansen’s challenge was as clear as it was monumental: how shall the State of Washington best address the issue of preserving its electronic records? “The traditional archival methodology of preserving information does not really change with electronic record,” he observed. “Electronic records are just a different media and the media itself has some particular needs and wants that have to be satisfied. It is not that much different than say, microfilm or analog audio recordings on tape or photographs or maps. They each have their own special care and feeding that needs to be done.”

Scanning the horizon to see what was in use and discern best practice, Jansen found few working models of digital archives that focused, as he must, on long-term preservation of electronic records.

But he found that important progress had been made. There were some “intellectual” models, and good standards frameworks from the International Standards Organization (ISO). The Open Archival Information System (“OAIS”) ISO Standard, for example, was developed in a collaboration of NASA and the European, Russian, and Japanese space communities. Having used and preserved electronic records longer than most – at least as far back as the Apollo missions – Jansen found that the space agencies had “a very good framework” that identified the key issues he would have to address in preserving the State of Washington’s own electronic records.

In addition, the Department of Defense 5015.2 standard, which addressed short-term records management, offered good approaches to change control and security parameters. Further research uncovered international organizations that provided good guidance on issues to be addressed regarding authenticity and chain of custody.

Jansen's team arrived at a mission statement and requirements schema for storing its records in a database. It sought the flexibility to take in records regardless of their originating source, to save them, "serve them up", and, ultimately, "to be able to intelligently migrate them while preserving the authenticity and chain of custody."

We very much built on the core principles of archival preservation, which direct, "Do not do anything you cannot undo. Do not go down the path of no return." We wanted to make sure that the system itself was open enough that as new technologies came along, we could adapt, adopt, and convert into these new technologies because the one thing we knew is the one constant is change. Anything that we do today will need a major overhaul in two to four years. We knew that going in because file formats are continually evolving. Newer, better, faster tools are coming out. International standards on file formats are still continuing to evolve and those standards, I guarantee, will not be in effect in ten years. So, we wanted to make sure that any tool that we used was open sufficiently that we could migrate to the next tool that came along that suited our goals and our needs.

Tools Search. Scanning for current off-the-shelf approaches that might address the state's requirements, Jansen's team examined several content management systems. They deemed them powerful, good – but none meeting, perfectly, the State's requirements. "They were very good at categorizing, organizing, and scheduling records," he recalled, "but none really focused on the authenticity, chain of custody, long-term preservation aspects of what we needed." With much functionality that Jansen did not require, and some missing that he did, Jansen turned his attention to customizing a solution, which would involve using a toolset to create its own workflows, in order to take records in no matter what their original format, normalize the data, and put them into the state's database. "And," said Jansen, "most importantly for us, we could export them into an open standard or open file format method that we could easily exchange with, and import into any other system."

As a Microsoft "shop" already using Microsoft SQL Server, Jansen assessed that his agency had in effect standardized on that environment and had the skills sets available to operate there. "All of the internal agency expertise in development was all in SQL server and programming in .NET," he observed. "For us, there really was not any other option because any other options would have required us to acquire new skill sets."

In order to assess whether a system so configured could fit the state's requirements, Jansen's team worked with Microsoft to operate a proof of concept, to learn better whether an electronic archives system so configured could take records in any format and normalize them, or export them into an open file format.

“In order to do that,” Jansen said, “we actually wrote out every single record with all the metadata, both that which was sent to us and that which we assigned and assumed based on point of origin. All of that information including the actual digital object sent to us was written out into an XML file, and that XML stored in a raw uncompressed method as just a .XML file on a tape and sent offsite across the state.”

The Work Proceeds. In August, 2004, with the proof of concept behind it, the project team began work with county auditors. “One lesson that we learned in Katrina is that current disaster recovery model for records, both paper and electronic, were insufficient to stand up to a disaster of a regional scale,” Jansen stated. “How do you prove you own a house, if your copy of your house records was destroyed and the original copy of those records was stored in the courthouse 30, 40, 50 miles away? The courthouse is now six feet underwater and ravaged by the hurricane as well. We wanted to start doing it better.”

The team started with “the people’s records” – marriages, land records, powers of attorneys. By 2007, 19 of the states 39 auditors were online and accessible via the Internet.

Archivists’ technical process is now well-honed. As they receive original digital objects, they maintain them in perpetuity in their original formats, and also write-out self-tagging XML copy of every record they receive so that the records can be accessed and made available over the Internet. In some instances, the state maintains three copies of the same record: the native file, the preservation copy of that native file, and a presentation copy of that native file, which might be a compressed version of an extremely large file containing a map, for example.

The archives now include documents of historic importance from the Governor’s office, and legislative leaders leaving office, census records going back to the 1850’s, naturalization records, and now e-mails and web content from all the major state and local government agencies. “We just received all the papers of the Senate Minority Leader who just left office,” stated Jansen. “So for us, we are really bridging some very interesting gaps that have not been done before. We are really focused on convergence and collaboration of different agencies.”

The Changing Equation. It was not always thus. While the State Archives has always been open to all state agencies, not all state agencies utilized them. Some institutions were large enough that they pursued their own course. Higher education and state historical societies and museums, for example, were concerned for the unique nature of their own records and the special “care and feeding” they required, and pursued separate paths. Jansen explained:

In a paper-based environment there is one physical tangible copy of a record that either you own or I own, we cannot both own that piece of paper at the same time, and so historically, museums and historical societies and large agencies have wanted to maintain control of that paper because information is power. To maintain that copy is to maintain that information, and thereby receive the benefits of the power contained within that document. With electronic records, that paradigm no longer holds true. You and I can both have an exact same duplicate of the exact same record, bit-for-bit we can have identical copies of the same record. So that whole power struggle of who maintains the original of that record no longer holds true, and so, agencies are much more willing to give us a copy of that electronic record because it does not interfere or impinge on their ability to do something with their copy of the record.

In fact, few agencies now have either the budget or technical skills to undertake complex digital archiving. Some of the “hold outs” – museums, historical societies – “that in the hundred years that the Secretary of State has been preserving records have never collaborated with the State Archives” -- are moving online via the State Archives. “We are finding,” Jansen said, “that agencies are far more willing to work with us as we have become the dominant Internet presence for records e-commerce.” With over a million marriage records online that can be ordered directly from home, billed to a credit card, and shipped within 24-48 hours, “we are really enabling the people to have access to their records in a way that has never been done before.”

The Path Forward. “All archivists coming on board now need to understand digital archiving or they are obsolete dinosaurs,” Jansen asserted. In making his own choices – whether commercial-off-the-shelf products, or customized tool development – Jansen elected to write his own system. He chose a tool set knowing that his in-house developers were well-trained. But the tools are, in his view, interchangeable. “If you do not like that tool, go use another tool. There is not any one solution that will satisfy everyone’s needs,” he said.

“Our goal,” he continued, “was to make sure that we are not locked into any one vendor or any one technology because the field will change. As long as the file formats are open, as long as I have access to the source code, then I know my mission is potentially successful. That is, fifty or a hundred years from now, I am going to be able to go back and read these file formats, which I cannot guarantee you right now with proprietary file formats, such as Word 95 documents.”