

THE STATE OF WASHINGTON DIGITAL ARCHIVES



On October 4, 2004, Washington State opened the doors of the first government digital archives built from the ground up in the United States. Initiated by the Office of the Secretary of State Archives and Records Management

Division, the project was formed to transfer of the recorded history of the State into a single, widely-available digital repository beginning with documents from the establishment of Washington Territory in 1853 and continuing to the present.

Included among the collection are the papers of each governor, legislative records, court records, records from all state agencies, and all of the “official records” of the state, including governors’ proclamations, executive orders, election results, and the laws as passed and signed. To date, the archives store five terabytes of data and that amount could double for years to come. The system eventually could hold 800 terabytes which is about 200 billion pages of text.

When the project was initiated, the State of Washington discovered that there weren’t any states that had taken on an archiving project like this one. Other states backed up electronic files as part of disaster recovery efforts, but no other state had developed a large-scale, comprehensive plan for saving electronic documents and making them accessible for the future.

GlassHouse won a contract with Washington State for a consulting engagement that included a feasibility analysis of three key components of the Digital Archives project: storage area networks

(SANs), conversion of data from legacy software, and data archiving. GlassHouse consultants worked to assess the feasibility of a variety of storage, data conversion and archiving technologies for the project, as well as making recommendations, providing preliminary design parameters, documenting recommendations and best practices, and providing proof-of-concept testing.

Bill Peldzus, director of storage architecture for GlassHouse Technologies Inc., helped the state develop archiving options and aided in negotiations with vendors. “There was a lot of hype from vendors who wanted to be involved in the archiving project,” he says. “In choosing outside vendors for this archiving project, we focused on three components: the technical feasibility, data-conversion techniques, and content-management options.”

Unlike some other projects GlassHouse has handled, this one wasn’t high risk because it didn’t require new technology or infrastructure changes. As the project moved forward, Peldzus was surprised at how smoothly it went, considering all the variables. “One of the things we learned was that they didn’t need a top-of-the-line storage system. A mid-level storage system using off-the-shelf disk and tape drives and software would suffice.

DIGITAL ARCHIVING

Phase one of the project began when three counties and the Secretary of State’s office began contributing documents to the archive. In phase two, a handful of state agencies, including the Transportation Department, began sending data to the archive. Phase three, encompassed the collection

and storage of historic documents. The fourth and final phase scheduled for the first half of 2006, allows all state agencies and county governments to send electronic documents to the archive.

“While state and county agencies aren’t obligated to send documents to the archives, if they don’t, they are responsible for keeping them in perpetuity on their own,” says Adam Jansen, a digital archivist for the State of Washington. He hopes agencies will send documents to the archives because it will be easier, less costly, and more secure.

BEST PRACTICES

What makes up a record? Whether it’s a piece of paper, an e-mail, and electronic document or a digital image, it’s a record. “It’s the content that drives retention, not the media that it’s written on,” says Jansen. It’s the media that data is stored on and the software that it was created in that makes archiving it challenging. As files are uploaded via an FTP site and tapes and optical disks are amassed, it becomes more important than ever that they have a way to maintain that information and be able to read it when they need it.

At its core, the archiving technology is remarkably simple: electronic records are stored forever in their native file formats and a web-viewable version is created on the fly so researchers have nearly-instant access to newly-acquired records. “We’re focusing on capturing the data no matter what format it’s in, because if we wait until the state has a common standard, it will be too late,” Jansen says. “If we don’t capture it now, it will disappear.”

According to Jansen, the most important part of any long-term archival system is avoiding proprietary file formats for data where ever possible and centralizing the backup of data in order to be able to standardize the storage method. Most documents for the State of Washington are kept in a standard format – text documents are stored as PDF files and images are converted to TIFF files.

COST OF ARCHIVING

The archiving system cost \$2.5 million--about \$1 million for software and \$1.5 million for hardware. The two-story building cost \$12 million and houses paper documents on one floor and electronic documents on the other. Since the archives opened, the State of Washington has gotten queries from other states looking to build similar storage centers. Why aren’t other states doing this already? For some, the main reason is financial, but for others, perhaps officials don’t yet understand the importance of keeping electronic records.

While Secretary of State Sam Reed believes that state governments to do a good job of preserving paper documents, he estimates that his state may have already has lost more than half of its electronic records. In the early days of computers, state workers often printed crucial documents, so some of those have been retained. But as computers became more prevalent and reliable, most documents remained electronic, stored locally on PCs, and often were deleted when storage space became tight. “The prospect of keeping records for future generations isn’t only a legal and social obligation--it’s also very exciting,” Reed says.