Multi-State Demonstration Project for Preservation of State Government Digital Information

Final Report

Washington State Archives

June 30, 2011

Table of Contents

Final Report

The increasing use of digital records in government has created the need for new approaches to archiving. Every day, state governments lose irreplaceable records because of the lack of a strategy to address record preservation.

To ensure that records with long-term legal, historical, and/or fiscal significance continue to be available over the next century, the Library of Congress implemented a grant for a demonstration project to develop a digital archive that could be used as a blueprint to help states to implement a least-cost, efficient, and proven archival system. This paper describes the demonstration project and the lessons learned.

Project overview

The Library of Congress (LOC) project began in November 2007 with project planning. At the time of the project inception, Washington State already had a digital archive established in Cheney, WA. The archive stored approximately twenty million Washington State governmental records and made them available to the public through a web portal.

To contain costs for the project, much of the Washington State Archives code base was reused, with some modification, as a template for partner archives. While this approach helped reduce costs, it also created some issues with partners that will be addressed later in this document.

November, 2007 through March, 2008

During the startup phase of the project, activities consisted primarily of planning, setting up hardware and software, and engaging with

identified partners who wanted to participate. The four original partners were Alaska, Idaho, Montana, and Oregon.

The web portal for the state archives was originally created to support a single state, and no easy customization options were included. With the onboarding of new state partners, manual customization had to be completed to provide each state's portal with a custom banner. States were asked to obtain their own domain names. Idaho and Alaska were slow to respond to this request.

Databases were created for each partner state. The database behind the state archives required modification to make database names more generic. Some stored procedures required modification so they would work for state databases.

Ingestion code was modified to accommodate a multi-state environment. Configuration files were created to make modifications by state easier. As a test of the new setup, the Social Security Death Records were ingested for each partner state.

Computer hardware was obtained and set up in the same physical environment already occupied by the state archives. The initial plan called for clustering of servers to provide access, but it was later determined that mirroring provided better performance, and so the servers were reconfigured. Database instances were created on separate servers and designed to accommodate larger volumes and increased availability.

The formal project kick-off meeting was held in Cheney, WA on March 18-19, 2008. Representatives were present from the four partner states, the Washington State Archives, Eastern Washington University, Microsoft, and other individuals. Individual state archival websites were demonstrated for the partners.

Three categories of partners were created. States transferring governmental records were classified as Archives Partners. States transferring e-publications only were classified as Library Partners. States who were interested in continuing to receive information about the project but who would not participate were classified as Education Partners.

April through June, 2008

In the second quarter of the project, planning continued. States were asked to complete their Intergovernmental Agreements (IGAs). Those states that hadn't already provided domain names were asked again to obtain them. Plans for a demonstration were drawn up, and success criteria were defined.

Starting in late May, 2008, Emiley Jensen of Montana's Historical Society and Justin Jaffe from the state archives worked to find a free or inexpensive audio conversion utility to convert Montana's legislative audio files from Real Audio to .WAV format for compatibility with the state archives requirements. Montana's primary goal was to take advantage of the audio search software that the state archives beta tested for Microsoft.

The Library of Congress awarded additional funds to be used by Washington State to host educational conferences and inform other interested parties about the project.

At this point in the project, partners were expected to provide a domain name and website page header and footer information. No additional participation criteria were described. For Washington, the success criteria consisted of holding educational meetings to increase awareness of the need for and ease of creation of digital archives.

July through September, 2008

The third quarter proved to be particularly busy. Indiana, Louisiana, and Colorado all asked to become partners in the project. While the system included sufficient hardware capacity to take on the partners, each partner required a customized web portal and setup of a database. More partners signed and returned their IGAs.

Discussions of the record series that each state intended to use began. Again, because this was a demonstration, the metadata choices available to partners to describe their records were limited to those metadata types already available in the Washington state archives.

Idaho became the first state to send records. Steve Walker, the State Archivist of Idaho, provided 73,000 new searchable records from ten different record series. Steve showcased Idaho's digital archives site at the Idaho Association of Counties, where county officials learned how they could participate.

To communicate with the partners, a SharePoint website was created. However, problems with the SharePoint site required support calls to Microsoft. Issues were not resolved until the following quarter.

Library partners requested the ability to upload e-publications, and development of an e-pub submission tool and workflow began. New metadata types were also created to support the e-pubs.

October through December, 2008

The first annual partners' meeting was held in October in Seattle, Washington. The meeting was attended by representatives from the original partner states, members from the Library of Congress, and guests who were recipients of other National Digital Information Infrastructure and Preservation Program (NDIIPP) grants. These guests had partnered with other states, and they presented their project goals and direction. They also provided reports about what they planned to submit and any successes or trials they had encountered.

North Carolina was added as an archive partner during this quarter, and Nevada was added as a library partner. Alaska was asked to submit their IGA form again because the form hadn't been completed even though Alaska had expressed interest in participating.

The partner SharePoint site continued to be an issue. Eventually, the site was moved to new hardware and became a primary method to communicate with partners.

During this quarter, Oregon and Idaho Libraries added e-pubs to their archived content.

January through March, 2009

In this quarter, the e-pub submission portal came online for the first time. Nevada sent their first e-pubs to the archive.

North Carolina and Nevada both signed their IGAs. Because of lack of participation, Alaska was changed to an educational partner rather than an archive partner. This allowed the recycling of assets set aside for Alaska to be used by other partners.

Oregon experienced budget issues in the new year and asked to become an educational partner only. However, new budget was eventually found, and Oregon was able to continue as a library partner, contributing e-pubs to the archive.

Oregon and Idaho each provided e-publications on hard drives, but the publications for each state were not separated by agency in the submission. To allow for better search capabilities, a separate file was required for each agency. To save labor by the submitting states, Adam Miller at the state archives wrote a script that searched the submitted spreadsheet and separated it into spreadsheets by agency. Oregon's submission included 1,715 records across 332 agencies. Idaho provided 216 records from four agencies.

During February and March, Justin Jaffe worked with Montana's Mike Allen to format the metadata for Montana's legislative audio. The first data transfer was received on February 17. Of the 2,322 audio files converted to .WAV format and uploaded to the state archives, nine files were found to be corrupt. These files had to be recreated from the original audio tapes of the sessions.

On February 29, the state archives received Colorado's first shipment of records on a hard drive. The data consisted of 469,258 records from 39 record series. None had digital objects associated with them. Two of the record series had issues with the metadata where dates were incorrectly mapped.

At this stage, the project was ready to receive increased submissions from partners, but processes were not currently documented. New documentation was created for:

- Data preparation
- Data transport options to help ensure security and chain of custody
- Record series metadata requirements

April through June, 2009

Now that states had begun shipping increased volumes of records to the archive, it was time to better promote the project. The state archives modified their brochure to include information about the project, and the state archives website added information about the project.

During this quarter, new hardware was added to the project to accommodate the addition of North Carolina's e-pubs.

On April 3, Colorado provided their second data shipment by sending an additional 210,775 records from four record series, making them the first partner state to send multiple shipments of data.

With more data arriving at the archive, a pattern began to emerge with incorrect metadata shipped to the archive. Ingestion was frequently delayed or repeated because of metadata errors, such as the use of disallowed characters in some fields, incorrectly mapped metadata, and data missing in required fields.

In April, the Library of Congress granted an extension of the project through December 31, 2010. The extension included an additional $100,000 in funding.

July through September, 2009

The second annual partners' meeting was held in Seattle, Washington. Participants included current partners, agencies interested in joining the partnership, Library of Congress representatives, and others interested in learning more about the project. Partners provided updates about progress toward goals.

Indiana shipped images to the archive, making it the first state to have searchable digital objects available online. The records were

received on September 4, and consisted of 357,288 records spread over 31 record series. Of these records, 400 had digital images.

Nevada, Oregon, and Idaho also sent additional records. Talks began with Minnesota to include them as a partner.

Work on the state archives website partner page was completed, and the new page was launched.

At the mid-point of the project, work began on developing a cost-recovery model as a replacement for funding when the project ended. Work on the model has continued, and issues of funding after the end of the project remain open.

October through December, 2009

With the project now well-established, Montana, Idaho, Oregon, Indiana, Colorado, and North Carolina all shipped records. Metadata errors continued to be problematic. Development of the Archive This! tool that transferred the responsibility for the validation of metadata and transfer of records onto the partners was begun, to cut down on errors and on the time from submission until records could be successfully ingested.

The records shipped by Indiana in the previous quarter were ingested into the state archives in two batches; one on October 2, and a second on October 6.

On October 2, Nevada Library made their first transfer of records by using the e-pubs web portal to upload eight publications. Prior to this transfer, a hard drive had been used to transfer their records.

On November 16, Arlene Weible sent the first partner news article to be posted on Oregon's Digital Archives homepage. The article consisted of a brief summary about a new publication that Oregon released.

January through March, 2010

The state archives accepts only a limited number of file formats for conversion to display on the web portal. During this quarter,

Montana shipped a large number of Real Audio files that required conversion to .WAV format. The conversion took two months to complete.

Nevada also submitted audio files that proved problematic. The Nevada files had been stored in smaller, incomplete pieces, and these needed to be merged so that they created a continuous .WAV format file. The files were successfully merged, but they lacked metadata needed for ingestion.

Metadata continued to be the primary source of issues. Indiana shipped files but not the accompanying metadata. North Carolina had issues with metadata errors for their e-pub files.

Tennessee was added as an archive partner after Minnesota decided they would be unable to participate. Tennessee submitted audio files for their test.

April through June, 2010

In this quarter, all partners were migrated to new hardware and software. The new Tennessee site was tested and proved ready for additional content.

Indiana conducted launch activities for their archive site by sending out press releases. Their promotions proved so successful that for the promotional days, their search numbers tripled as users accessed their records.

As part of the ongoing effort to document the project, "at risk" issues were defined for electronic records and digital preservation.

July through September, 2010

Idaho, Montana, and Tennessee continued to send records to the archive. Metadata errors also continued to be an issue for partners.

October through December, 2010

The original end date for the project was December, 2010. Considerable time and effort was put into completing applications for

an extension of the project through June, 2011. That extension was granted.

A regional symposium was held in Maryland to begin planning for the next phase following the end of the demonstration project, and to demonstrate the project to interested parties. Delaware, New Jersey, Pennsylvania, and Virginia all sent representatives to learn about the project.

A conference for partners' IT professionals was held in this quarter. Attendees toured the Microsoft campus in Redmond, met with Scott Guthrie, Microsoft Corporate Vice President, and participated in a technology exchange. From there, the attendees toured Microsoft's data center in Quincy, Washington, and then went on to see a presentation at the state archives in Cheney, Washington. The tour provided a behind-the-scenes look at the technical aspects and process flows used at the state archives.

As the project neared its end, the following additional documentation was created:

- *Necessary Requirements to Apply the Washington State Archives, Digital Archives' Framework in Other States* – A list of requirements for a state that wants to create their own stand-alone state archive based on the state archives template.
- *Administrative, Descriptive, and Technical Metadata for Government Information* – How state archives structured their metadata, with consideration to why this structure was chosen.
- *Proposed Digital Preservation Service Level Agreement* – A service-level agreement (SLA) for partners who wish to join in an ongoing consortium for a digital archive.
- *Educational Programs and Products for Collaborative Catalyst* – A brief description of how to use educational programs to promote collaboration between agencies.
- *Recommendations for Changes to the Washington State Digital Archives' Framework* – Recommendations for how to

improve the utility of the framework and the compatibility of incoming records.

The state archives partner web page was updated.

January through March, 2011

New records were received from Idaho, Indiana, Nevada, and Oregon.

Another regional meeting was set up to evangelize digital archiving. The meeting was held in Indiana. Representatives from Iowa, Illinois, Kentucky, Michigan, Minnesota, New York, Ohio, and Wisconsin attended.

To help alleviate ongoing issue with faulty metadata, training on the Archive This! tool was provided to representatives from both Nevada and Tennessee.

Lessons learned

Washington already had a successful digital archives, one that could be relatively easily modified for use by other states (with some limitations to meeting states' needs for metadata schemas.) Because of reduced funding, Washington also used budget-priced hardware for the demonstration, not the high-end systems that they use for their own repository. Were partner states fully engaged in the creation of a shared digital archives, they might expect to pay for hardware similar to that which Washington uses, and for software development to upgrade the Washington system to one specifically designed for multi-state use.

Washington saw this as a project where functioning digital archives would be created for each participating state. Once states saw the ease with which they could preserve their records and the low cost of becoming partners, Washington expected that states would participate by sending large volumes of records.

What neither entity expected were the many roadblocks to success that had little do with funding. These issues can be separated into two classifications:

- Political/bureaucratic issues
- Technical issues

Political/bureaucratic issues proved harder to resolve and more time-consuming than technical issues.

Political/bureaucratic issues

States were asked to provide ten thousand hours of labor over the life of the project as their price for inclusion. Some states who initially expressed interest decided they were unable to meet this requirement.

In some cases, states were reluctant to send their records out of state for preservation. In other cases, states lacked the political leadership or will to participate. Louisiana is an example of where the political leadership initially expressed interest in participating, but at the next election cycle a new group of politicians decided against participation.

Indiana required a large legal addendum to the contract before they would sign. In other states, state contract and accounting offices all created roadblocks that had to be overcome before partners joined. Some states had laws against sending governmental records out of state for storage.

Many states have simply failed to realize that they are entering a digital age. They remain mired in the idea that record preservation means keeping large volumes of paper or microfiche records, and they are reluctant to embrace electronic media.

States that are successful tackling the challenges of digital records are those that have involvement at the top levels of government or have appointed officials who champion digital archiving.

State IT departments are not part of the archival divisions. Those states that were most successful in getting records uploaded to their

archives were the states that also had the closest and most successful working relationships between their archivist divisions and their IT divisions. Close cooperation between these two areas of expertise can be considered a requirement for success.

Technical issues

Metadata issues

Metadata incompatibility has proved to be a difficult obstacle. Other states' metadata does not map well to the metadata types created specifically for Washington State agencies' data. This resulted in data ending up essentially re-categorized to the Miscellaneous Historical Records series, a catch-all with few searchable fields.

Metadata mapping issues, incomplete metadata, and data sets arriving with no metadata were all issues over the life of the project. While the Archive This! tool helped reduce the issues, the tool was not generally used effectively unless additional training was provided.

The ideal solution to metadata issues is the creation of a standard set of metadata for government data that can be applied by governments and libraries alike. Currently this metadata standard does not exist. Because of limited funding, it was not possible during the demonstration program to customize metadata for individual states.

File format issues

There is no standardization or requirements for the file formats used by government agencies. This creates issues for storage and migration, particularly if the file format is proprietary. Open formats are easy to work with and migrate, and they are not vendor-dependent. Standardizing file formats would help archives and libraries best preserve their digital records.

Website issues

The state archives' website was built for archiving and retrieval of records from a single state. Much of the framework is static and not

easily customized. This adds to development, maintenance, and deployment costs as partners are added. Examples include Louisiana using parishes instead of counties, and libraries that want to use publications instead of record series.

An updated website template should be developed if joint partner states are to continue to use a regional archive. The new website design should rely on a configuration file and dynamic fields that can be easily customized and updated.

Unexpected results

The state archives had a well-established archiving program in place by the time the project began. The project had minimal influence on the organizational practice of the state archives except in the following two areas:

- The creation of an archiving solution for e-publications and the addition of metadata to describe e-publications.
- The addition of search functionality for audio files. Audio files turned out to be a popular type of content for states participating in the project to archive.

Ongoing plans

Following the end of the Library of Congress grant, the project has two available options:

- A consortium model – Partners will pay a fee for services, and the state archives will continue to archive and provide access to partner records. Partners can withdraw at any time with 90 days' notice. Partners who fail to pay fees in a timely manner will have their data removed from the archive and returned to them on a portable digital storage device.
- A grant model – The state archives have applied for grants to continue to fund the project.

Content plans

If partners decide not to participate in a consortium model and no grants are awarded for ongoing work on the project, states will receive their records back on a portable digital storage device.

Appendix 1: Data Transportation Plan

Document plan for transporting (including describing, packaging, etc.) data from Multi-State Partners

When a partner is ready to send us data, they request a hard drive. We package an external USB hard drive (currently 750GB) and place it in hard shock absorbing anti-static foam. This foam is incased in a hard plastic case secured by two metal clamps. The enclosure is zip-tied with a numbered tag that we track to verify no one has tampered with the case's contents.

Library partners have access to our E-Publication submission portal. This portal is a gateway to upload E-Pubs to our system using a web interface. This interface has required fields that are specific for every E-Pub. After uploading the E-Pub, it waits in an approval queue until a user with approval permissions verifies and approves the record. At that time, the record is ready for extraction. See "Process Flow for EPublication Data_03-09-2009.vsd" for the E-Pub process flow.

Identify costs involved with transporting data

We ship the hard drive via UPS. The package is about 9 pounds with dimensions of 18x15x8. The cost to ship a drive to a partner is about $10 on average. We also supply a return shipping label for the partners to send the drive back. This label is about $10 on average. For UPS to send the hard drive round trip it costs about $20.

Additional shipping costs include single layer cardboard to put the previously mentioned zip-tied enclosure in, and two numbered security zip-ties to verify that there has not been any tampering of the data on the hard drive during shipping. The cardboard can

withstand an average of three to five round trips, and costs $300 per 100. The zip-ties are a onetime use and cost $14 for a pack of 50.

The average cost to send a hard drive is:

- $20.00 for UPS shipping
- $0.56 for two zip-ties
- $1.00 for cardboard

Total = $21.56 to send a hard drive round trip to a partner.

Document storage requirements for identified datasets

There are several places that data is stored for our partners.

- Backup Server
- Inbox Server
- Databases (one for metadata and another for digital files)

Determine data format and transfer media

For the metadata file, we support the following formats:

- Excel
- CSV (Comma Separated Values)
- Pipe Delimited
- Tab Delimited
- Semicolon Delimited

For the digital objects, we support the following formats:

- TIFF (must be flattened, no multi-layer support)
- WAV – Audio only
- Office (Word 2003 and prior, PowerPoint 2003 and prior, Excel 2003 and prior)
- PDF

Identify processes for content tracking, packaging transfer, and ingest

UPS supplies tracking of packages. We also have an internal Excel file to track which partner has a hard drive and the date it was sent to them. When we get the hard drive back, we update that spreadsheet with the return date. Then after ingestion, the accession numbers for the data provided on that hard drive, and the ingest date, are also added to the spreadsheet.

Identify transfer tools

- The tools are the external USB hard drive for archives and library partners.
- The E-Pubs submission portal for library partners only.

Appendix 2: Search Criteria for Record Series

Marriage Records:

- Reference Number (Unique Number)
- GroomLastName
- GroomFirstName
- GroomMiddleName
- BrideLastName
- BrideFirstName
- BrideMiddleName
- MarriageYear
- RecordingYear

Birth Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- BirthYear
- FatherFirstName
- FatherMiddleName

- FatherLastName
- MotherFirstName
- MotherMiddleName
- MotherLastName
- Gender

Census Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- BirthPlace
- PageNumber
- LineNumber
- VolumeNumber

Death Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- DeathYear

Executive Orders Records:

- Reference Number (Unique Number)
- GovernorFirstName
- GovernorMiddleName
- GovernorLastName
- DocNumber
- DocStatus
- DocStatusCode
- Year
- Title
- DocType
- Supercede
- Obsolete

Frontier Justice Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- CaseNumber
- Cause
- CaseType
- CasePartyType
- CaseYear
- Keywords

Institution Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- ReceiptYear
- WhereConvicted

Land Records:

- Reference Number (Unique Number)
- FirstName
- MiddleName
- LastName
- PartyType
- Year
- LegalKeywords
- GenericPartyType
- DocTypeDesc

Military Records:

- Reference Number (Unique Number)
- LastName
- FirstName

- MiddleName

Minutes Records:

- Reference Number (Unique Number)
- Recording Date
- Document Type
- Entity (Department)
- Committee

Ordinance Records:

- Reference Number (Unique Number)
- Recording Date
- Document Type
- Entity (Department)
- Committee
- Subject (Document Number / Ordinance Number)

Resolution Records:

- Reference Number (Unique Number)
- Recording Date
- Document Type
- Entity (Department)
- Committee
- Subject (Document Number / Ordinance Number)

Misc Historic Family Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName (Optional)

Naturalization Records:

- Reference Number (Unique Number)
- LastName
- FirstName

- MiddleName
- DocumentYear

Oaths Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- Office
- OathYear

Photographs Records:

- Reference Number (Unique Number)
- Photographer
- People
- Title
- SubjectStartYear
- SubjectEndYear
- Keywords
- Date

Plat / Survey Records:

- Reference Number (Unique Number)
- FirstName
- MiddleName
- LastName
- PartyType
- Year
- LegalKeywords
- GenericPartyType
- DocTypeDesc

Power Of Attorney Records:

- Reference Number (Unique Number)
- FirstName

- MiddleName
- LastName
- PartyType
- Year
- GenericPartyType
- DocTypeDesc

Professional License Records:

- Reference Number (Unique Number)
- LastName
- FirstName
- MiddleName
- Year

Uniform Commercial Code Records:

- Reference Number (Unique Number)
- FirstName
- MiddleName
- LastName
- PartyType
- Year
- GenericPartyType
- DocTypeDesc

Agreement Records:

- Reference Number (Unique Number)
- FirstName
- MiddleName
- LastName
- PartyType
- Year
- GenericPartyType
- DocTypeDesc

Audio Records:

- Reference Number (Unique Number)
- Date
- Description
- Duration
- Source
- Notes (Optional)

Appendix 3: Ingestion Criteria for Record Series

Auditor Records (Includes: Land Records, Misc Auditor Records, Map Records, Military Records, Frontier Justice Records, Minute Records, Ordinance Records, Plat and Survey Records, Power of Attorney Records, Professional Licenses Records, Resolution Records, and Uniform Commercial Code Records) require the following or ingestion will fail:

- Reference Number (Unique Number)
- RecordingDate (Must be a valid date)
- Grantor and Grantee
- DocTypeCode

Audio Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Audio record must have one and only one audio file and must have one audio file

Birth Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name
- Birthdate (Must be a valid date)

Building Permit Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- ParcelNumber
- StreetAddress

- IssueDate - (Must be a valid date) Use 'nd' if the date is not known.
- LUApprovalDate - (Must be a valid date) Use 'nd' if the date is not known.

Census Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name
- LineNumber
- PageNumber

Clerk Of Courts Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Date
- Page Count

Corporation Records require the following or ingestion will fail:

- Corporation Name
- UBI
- ReceivedDate
- Reference Number (Unique Number)

County Assessor Records require the following or ingestion will fail:

- Reference Number (Unique Number)

Note that more fields will be defined later for this record series.

Death Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name
- DeathDate - (Must be a valid date)

Executive Order Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Governor

- Title
- Status
- Recording Date - (Must be a valid date)

Frontier Justice Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Parties element - (cannot contain a null value)
- Recording Date - (Must be a valid date)
- DocTypeCode

Historic Record Collection Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name

Insitution Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name
- ReceiptDate - (Must be a valid date)
- DocTypeCode

Marriage Record Collection Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- GroomName
- BrideName
- Either MarriageDate or RecordingDate - (Must be a valid date)

Military Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Parties - (cannot contain a null value)
- Name

Minute Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Recording Date - (Must be a valid date) Use 'nd' if the date is not known.
- Parties - (cannot contain a null value)
- Entity

Naturalization Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name

Oath Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Name
- OathDate - (Must be a valid date)
- Office

Ordinance Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Recording Date - (Must be a valid date) Use 'nd' if the date is not known.
- Parties - (cannot contain a null value)
- Entity

Photo Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- A photo record must have at least one file attached
- Extent
- Title
- SubjectDate
- SubjectEndDate - (Must be a valid date)
- SubjectStartDate - (Must be a valid date)

Professional Licenses Records require the following or ingestion will fail:

- Reference Number (Unique Number)

- Recording Date (LicenseDate)- (Must be a valid date)
- Parties - (cannot contain a null value)
- Names

Resolution Records require the following or ingestion will fail:

- Reference Number (Unique Number)
- Recording Date - (Must be a valid date) Use 'nd' if the date is not known.
- Parties - (cannot contain a null value)
- Entity

Appendix 4: At Risk Issues

Define "at-risk" issues for electronic records and digital preservation.

There are several key areas that we have identified which need to be addressed to maintain electronic records. All electronic records are at risk of data corruption from failed hardware, data integrity from malicious code or unauthorized access, record inaccessibility from some failure, or quality assurance not being implemented or enforced. Anticipating current and future needs will minimize the risks associated with electronic records.

Technology

- Maintain computer hardware for data integrity.
- Update computer software to minimize bugs, maximize security, and maintain system stability.
- Network infrastructure to insure communication between dependant critical systems.
- Active backups for disaster recovery.
- Fall back plans for power outages such as UPS backups, backup cooling for servers, and emergency auto-shutdown for non-critical servers.
- Physical and network security to prevent unauthorized access to systems.

- Thorough documentation of infrastructure, workflows, source code, and all other aspects of the agencies operation so all duties can be performed in the absence of key staff.

Training

- Electronic records management.
- Custodial issues and responsibilities.
- Policy creators' understanding of technological implications for their decisions.
- Developers' use of the latest best practices for data security, availability, and migration capability.
- Application development planning must include all stake holders to maximize effectiveness of the application and minimize costs.
- Digital object standards (image DPI, audio bit rate, etc).

Funding

- As digital repositories grow, additional storage for record preservation, tweaking of networks to maximize throughput, and additional staff to support the operation become critical.

Appendix 5: Administrative, Descriptive, and Technical Metadata for Government Information

November 3, 2010

Created by the Washington State Archives

Preliminary discussions on metadata profile for administrative, descriptive and technical metadata appropriate for government information

How to select searchable metadata

Washington State Archives has broken record series collections up into titles. Each title will be associated with a parent record series. Currently the state archives supports 35 record series online (see Appendix A).

Common pieces of metadata are identified for each record series which then become required for a collection to be part of that record series (see Appendix A). This makes it easier to set searchable fields for the website frontend. Required metadata also makes data validation easier; if a required piece of data is missing, then the record is rejected and will not be ingested into the database. This prevents orphaned records in the database; orphaned data will impact data usage and backup performance.

Every record that is ingested must have a unique reference number within that title. This is so that the record can be identified, retrieved, and updated.

How to account for additional metadata

If a record has metadata that is in addition to what is required by the Record Series, we still process that data and add it to the record. When a user retrieves the record online, the additional metadata is presented along with the required metadata. Since the extra metadata is not necessarily expected, we cannot make the extra metadata searchable. To address this limitation, we have added a keyword field. Record submitters can add descriptive metadata for users to search against. This feature is not available for every record series.

How to update, add, or change metadata

As requirements, needs, and technology change, so does the Washington State Archives. The staff constantly reviews emerging technologies and archiving techniques to identify ways in which they can improve. Regular meetings are held to discuss benefits and implications of incorporating new technologies or business rules. It is necessary to include all the stake holders in these discussions since one change could impact other areas, or even require an entire system redesign. With everyone involved, project requirements can be discussed and expectations can be outlined.

Restricting metadata online

The state archives' system is designed so that restrictions can be added or removed at any time. Restrictions can be a result of changes in laws, privacy concerns, user requests, etc. It is possible to create or remove restrictions on one piece of metadata, an entire record, a collection at the title level, or even an entire record series. This design allows for collections, records, or metadata to be preserved in our database while limiting access as needed to the web.

Appendix 6: Proposed Digital Preservation Service Level Agreement 2010 – 2013

Note: Formation of a consortium is predicated on receiving funds from grants.

Consortium with the Washington State Archives, Office of the Secretary of State

Revised on December 7, 2010

Consortium Overview

- A consortium will be formed to include the Washington State Archives and other state archives and state libraries which request services provided by the state archives for the storage, preservation, and retrieval of electronic records.
- The purpose of the consortium will be to provide an efficient, low cost alternative to host public records from other state archives and state libraries.
- Membership will require completion of all required forms and the ability of the Washington State Archives to administer the members' digital collection.
- Members of the consortium will pay an annual fee for services to the Washington State Archives, Office of the Secretary of State. The Office of the Secretary of State will bill monthly for all services.

- Members who are in arrears for more than 90 days will be removed from the consortium and their data returned to them on a hard drive.
- Any member of the consortium can remove themselves from the consortium with a 90 day written notice to other members. Upon leaving the consortium, members will receive a copy of their data in a standard format within 90 days.
- For the first year, prices are set at $35,000 a year for an archives and $10,000 a year for a library. These prices will be reviewed on a yearly basis, and are subject to change based on realized costs and participation levels.

Support Time

- Generally, support time will be Monday through Friday from 8:00am to 5:00pm PST.
- Upon enrollment into the consortium, each member will receive 20 hours of support time without charge. Once the 20 hours are used, or one year passes, members will begin to receive 5 hours of support time per month at no charge. Support time does not accumulate month to month.
- Upon exhaustion of a member's support time, the Washington State Archives will notify the member that their allocated support time has been exhausted. Any further support time will need to be purchased or wait until the following month.
- Additional Washington State staff support time will cost $150 an hour. At a minimum, two development staff will be allocated to support an issue. The average rate of salary and benefits for two staff members is $150 an hour.
- Members will not be charged support time for resolving glitches, bugs, or website outages that are the responsibility of the Washington State Archives.
- If the Washington State Archives' development staff determines that the support issue is a result of a bug, then no support hours will be charged to the member.

- If a member is charged for support time and it is later determined that they should not have been, the support time will be credited to that member's account.
- There will be an 800 number available for critical system support on weekdays and weekends.
- All non-secure information about bugs and glitches will be posted on our secure SharePoint site.

Storage

- Archives will receive 5TB of formatted storage with their annual fee to be broken out as follows:
  - 2TB formatted production storage.
  - 3TB formatted backup storage.
  - Backup storage includes:
    - Full Backups
    - Differential Backups
    - Transaction Logs
- Libraries will receive 1TB of formatted storage with their annual fee to be broken out as follows:
  - 500GB formatted production storage.
  - 500GB formatted backup storage.
  - Backup storage includes:
    - Full Backups
    - Differential Backups
    - Transaction Logs
- 1TB of additional formatted storage will cost $4,000. This will include:
  - 1TB formatted production storage.
  - 2TB formatted backup storage.
  - Installation and management of new data.
- A maintenance fee of $500 a year will be charged for each TB of formatted production storage to perform regular maintenance, cover refresh replacements, and cover any additional maintenance costs.

Services Provided

- The Washington State Archives has a 99.9% system uptime which does not include scheduled maintenance windows and unexpected Internet Service Provider outages.
- Members will be notified a week in advance of any scheduled downtimes for maintenance and software deployments.
- A checksum is created and recorded before and after records are uploaded to the Washington State Archives. Comparisons of the checksums are made prior to ingesting the records and kept permanent with the record for future comparisons.
- Microsoft Windows Server for standardization and regular updates to security vulnerabilities.
- Member access to servers hosted at the Washington State Archives will be limited to applications provided by the Washington State Archives. Limiting access will increase security, maximize system performance, and maintain the chain of custody of records in the repository.
- Permissions for members' systems are set by the Washington State Archives' Network Administrator and Database Administrator.
- Members will maintain ownership of the records preserved at, and backups made by, the Washington State Archives.
- All requests from non-members will be forwarded to the intended member. The Washington State Archives will work with members to produce records hosted in Washington at the member's request.
- A dedicated high speed Internet Connection - 10 Mbps, with sustained burst capacity to 100 Mbps.
- Active network monitoring Tipping Point - Intrusion Detection System.
- Enterprise level security PIX / ASA – Firewall.
- Redundant, Coyote hardware load balancers, with redundant Web Servers and Search Engines for a faster user experience, and reliable uptimes.
- Maintain Domain Name Service registration for one URL will be provided for each partner. Additional URL's can be

associated with a member's website at that member's expense.

- For the first year, ecommerce will not be available to members.
- High speed Ethernet switching for optimal network communication between servers.
- Critical service patching preventing known computer exploits and including up-to-date feature capabilities.
- Virus scanning protection and updates provides data integrity. Virus scanning is performed using two separate virus scanners prior to record ingestion.
- A high performance RAID Disk storage system that minimizes system downtimes due to hard drive failures, and increases data access speeds for a faster user experience.
- Scheduled backups are performed regularly for disaster recovery. Full backups are done weekly, differential backups are done nightly, and transaction logs are recorded hourly. The backup schedule has been set to maximize system performance and recoverability while minimizing the impact of the user site experience.
- Backups are stored for two weeks for onsite copies and one quarter for offsite copies.
- Disaster Recovery backups will be stored onsite and at our offsite location which is 400 miles away way from the state archives facility. We will also coordinate with your preferred offsite storage facility.
- It is the system architecture of the Washington State Archives to never delete a digital object once it has been ingested into the system. If there is an error with a record in the database, the corrected record can be ingested as a replacement.
- The emergency generator is tested weekly for one hour to maintain website uptime during a power outage. Generator power can be supplied for up to three days before a refill is needed. Refills can be made while the generator is running, so it is possible to provide continuous power.

- Redundant emergency UPS conditioned power to increase server life expectancy, and provide power to essential servers when the power is interrupted.
- Emergency backup cooling system keeps the datacenter at the appropriate temperatures when primary cooling fails.
- Physical security includes motion detectors and a redundant burglar alarm systems with three different technologies to transport the alarm to emergency responders. A card access entry system provides restricted access to different parts of the building.
- Staffing includes an Ingestion Coordinator, Sr. Database Administrator, .NET Application Developer, Web Developer, Network Administrator, Applications Architect, General Support, and Administrative Staff.
- Monthly and quarterly reports to keep members updated on site activity, resource usage, current issues, and solutions to any past issues.
- Hardware and software support to include bug fixes and enhancements.

Member Responsibility

- A detailed service level agreement will be signed by each member of the consortium.
- All communication and coordination will be done through the project coordinator.
- Members will log all bug and enhancement requests though a centralized high availability SharePoint site.
- Each member will have the ability to vote on enhancements and new features established by a Consortium Agreement.
- Members will be responsible to transfer their electronic records in the specified format using tools provided by the Washington State Archives.
- Members will be responsible for correcting errors associated with their electronic records metadata and file naming.
- For phase 1, the Washington State Archives will **not** take **confidential** records from members.

Additional Questions:

- Are all web design services done through WA or can state's design their web sites remotely?
  All website design will be done at the state archives.
- If the web design services are only available through WA, does that fall under the same hourly services rates as tech support?
  Any time spent customizing a member's website will be charged to that member.
- Support documentation: WA should plan on offering documentation on what members need to do in order to use the service. The best practices documents could inform this documentation.
  Support documents will be produced as part of the NDIIPP grant requirements.

Appendix 7: Necessary Requirements to Apply the Washington State Archives' Framework in Other States

November 3, 2010

Created by the Washington State Archives

Determine requirements necessary to apply Washington State Archives Framework in other, stand-alone state archives.

Internal Partners (stakeholders)

It is important to identify who the stakeholders are. Once the stakeholders have been identified, it will be necessary to build strong relationships with them. The stakeholders will need to be supportive of your digital archives vision for the project to have a chance. If possible, demonstrate how the digital archives will benefit each stakeholder to help them identify with the mission, and feel connected to its success.

Revenue Source

The Washington State Archives was able to add a recording fee of $1 per page collected by county auditors on all recorded documents. This fee helps pay for the digital archives facility and on-going operation costs. This independent revenue sources has buffered the state archives from some of the state's budget cuts during the down economy.

Infrastructure

All of our servers run Microsoft technology. This keeps consistent with the industry standard and has made administering, updating, and migrating server software much easier. The large Microsoft community has been a great resource when issues arise and quick resolutions are critical.

Staffing

The state archives has several specialized developers, a network administrator, and an archivist on staff. Regular staff meetings are scheduled to work out plans for new projects, and provide input for current projects. Staff members usually attend yearly technology conferences to keep current on industry trends and best practices. The staff often conducts independent research on new technologies or techniques and shares relevant findings with other staff members.

Maintenance

Computer technology needs to be replaced as it fails and as business needs change. Common practice is to refresh the computer hardware and software in regular intervals, about every four years. Other key maintenance items are UPS batteries, backup generator, network switches and routers, intrusion prevention software, and firewall systems. Regular database server maintenance is needed to optimize the database and keep backups current for disaster recovery.

Disaster Recovery

Regular database backups, both full and incremental, are necessary for disaster recovery. Periodic backups should be sent offsite to minimize risk in the case of disaster. Backups should be periodically tested to practice the restore procedure, and to verify that backups are useable. Procedures for restoration should be documented in detail in case someone unfamiliar with the project needs to restore the data and administer the systems.

Appendix 8: Educational Programs and Products for Collaborative Catalyst

November 3, 2010

Created by the Washington State Archives

Deliverable: Create educational programs and products to act as catalysts for collaborations

In this electronic age riddled with budget cuts and limited resources, it is becoming more difficult to fulfill core missions, build required infrastructure, and serve customers needs. Advancements of technology now allow people to work from anywhere and with anyone they need to. Now is the time for agencies to take the next step and find ways to collaborate with each other. Resources should be pooled to build communal infrastructure, create best practices for electronic records, help each other meet core missions, and serve all customer needs.

Identify Groups

Find other agencies that have the same goals and needs as you. Make a list of the agencies which will be the best possible partners. The best candidates are ones that can make decisions, have political flexibility, and like to take on new challenges. Once the other agencies have been identified, find out what motivates them. Some of the most common motivators included cost savings, efficiency in government, and a desire to serve patrons online 24/7 from anywhere.

Communicate With Each Group

Develop a plan that would be mutually beneficial to each of the identified candidate agencies. Open lines of communication with the other agencies and present the plan. Focus the communication on each agencies motivator. Be prepared to support the agencies if they need help sharing their vision with other stakeholders within their states. Work together and define the collaborative efforts. Create objectives that everyone involved would be happy with, and support each other in executing the plan.

Methods of Communication

Agencies need to get their message out there. Some low cost methods off communication include online videos, ephemera, and presentations at conferences. Whatever the medium, statistics and graphs will get important data across quickly and effectively. Too much information and messages that lack interest will frustrate people receiving your message. t is better to keep communication short, interesting, and to the point. End with contact number or email address and list other resources for people to get more information if they are interested.

Patrons

To have a successful educational program, the program creators must identify who the stakeholders and target audiences are. As the Washington State Archives developed the state archives, the stakeholders which were identified were the record creators, record managers, lawmakers, attorney general, and elected officials. The target audience was genealogists. As the target audience became more aware of the resources the state archives made available, the more support the state archives received from taxpayers and lawmakers. This will help to make the collaboration sustainable.

Summary

Agencies will be able to fulfill their core missions in these difficult financial and technologically overwhelming times if they work together. Identify your goals, and find other agencies with like goals

that are able to act. Partner with those agencies by proposing a partnership that solves both agencies' needs. Present your idea by focusing on each agency's motivators. Support each other, and gain the support of stakeholders of each agency.

Appendix 9: Recommendations for Changes to the Washington State Archives' Framework

November 3, 2010

Created by the Washington State Archives

Deliverable: Recommendations for changes to the framework

The Library of Congress' NDIIPP demonstration project has given the Washington State Archives a different perspective on how to view electronic records. In the initial development of the digital archives, there were certain things that were not considered. The NDIIPP project has given the Washington State Archives an opportunity to view how other agencies would use the digital archiving solution. We have identified three primary areas which could make the digital archiving solution more compatible with other state archives and libraries needs.

Metadata

The Washington State Archives has developed required types of metadata for each record series to be classified as a particular Record Series. These requirements have prevented several partners' datasets to be classified as anything other than Miscellaneous Historical Records which only requires a name and unique reference number. This is partially due to how partners have indexed their digital records, and partially due to how Washington has defined each Record Series metadata requirements.

Creating a standard set of required metadata for each record series would guide archives and libraries when indexing historical records. Standardized metadata would allow for archives and libraries to be compatible with the state archives or any other digital archiving solutions developed with the standards in mind.

File Formats

Each agency has digital records in a variety of file formats for many different reasons. Whatever the reason, the wide range of file formats makes compatibility with preservation solutions difficult and forward migration a constant concern.

Some file formats are classified as open file format standards and others are proprietary closed formats. Open formats are usually easy to work with, migratable, and not vender dependant. Identifying file formats that are open is important to digital archiving. Standardizing archival file formats would guide archives and libraries to the best preservation solution for their digital records.

Website

Much of the state archives website is coded with static content. In some cases, this does not work for many partners for several reasons. Examples include Louisiana calling regions in their state parishes instead of counties, and libraries preserving publications instead of record series. From our current website design, each partner would have to have their site customized to conform to all their needs. This would take much more developer time for coding, maintenance, and deployment of each website.

It is possible to code the website to use a configuration file. Each website would have customizable fields that would dynamically populate depending on the values in the configuration file. If settings needed to be altered, only the configuration file would need updating, and the results would propagate throughout the agency's entire site.

Appendix 10: Links to Online Resources

The following additional presentations and documents further describe the state archives and Library of Congress demonstration project:

- Digital Archives Feasibility Study

- Digital Archives Investment Plan
- ISB Presentation Sept 2004
- Digital Archives Overview
- Quality Assurance Report Q1
- Quality Assurance Report Q3
- Library of Congress project on the state archives site
- Library of Congress NDIIPP Project Site

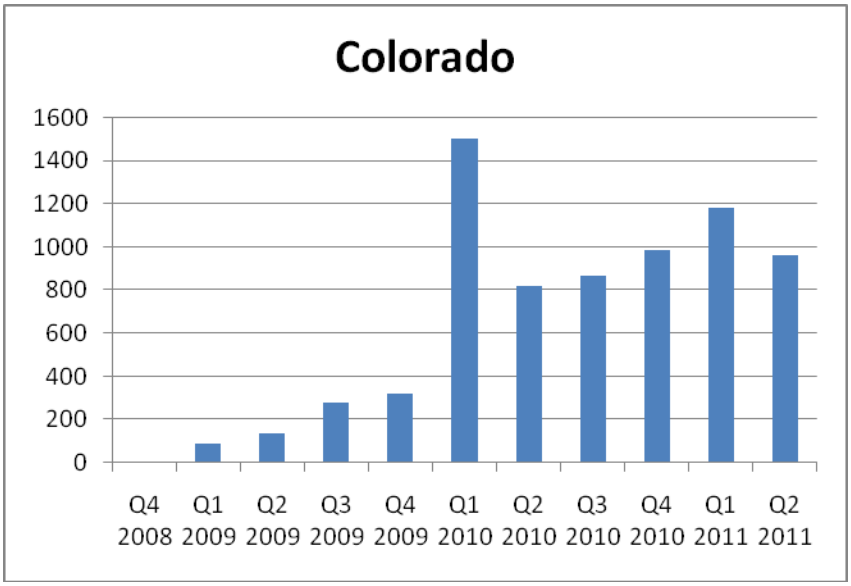
Appendix 11: Search Growth

Over the course of the Library of Congress demonstration project, website search query data was captured by the state archives. The following graphs represent the number of state-by-state queries by quarter.

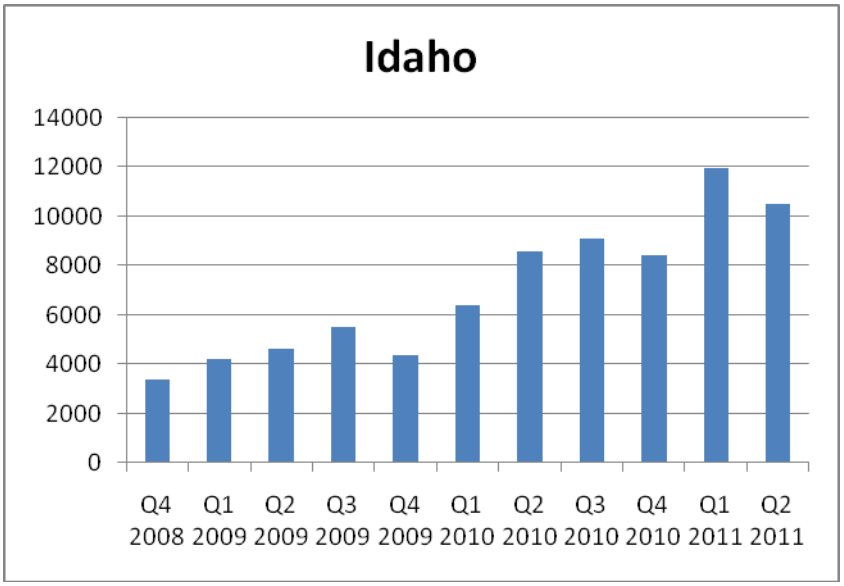
Alaska

After one quarter of participation, Alaska became an education partner. Only the Social Security Death Index was available on the Alaska site.

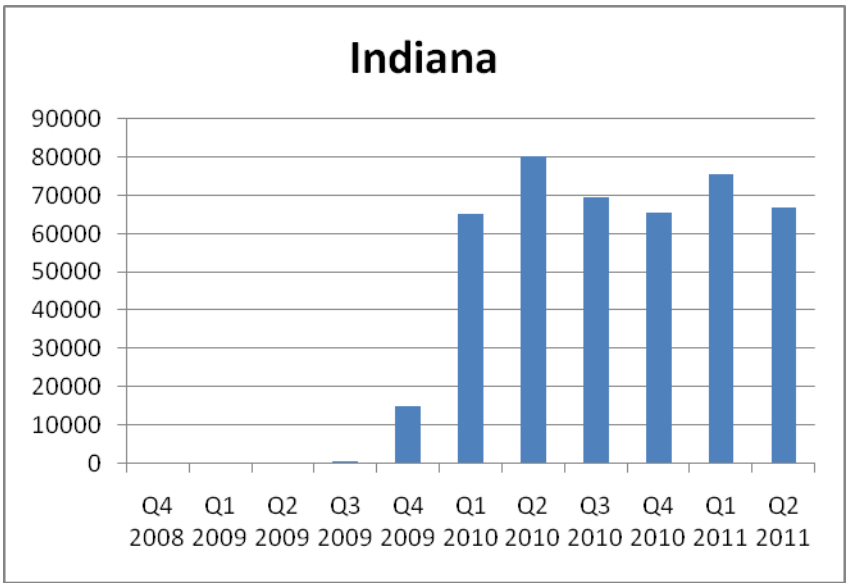
Colorado



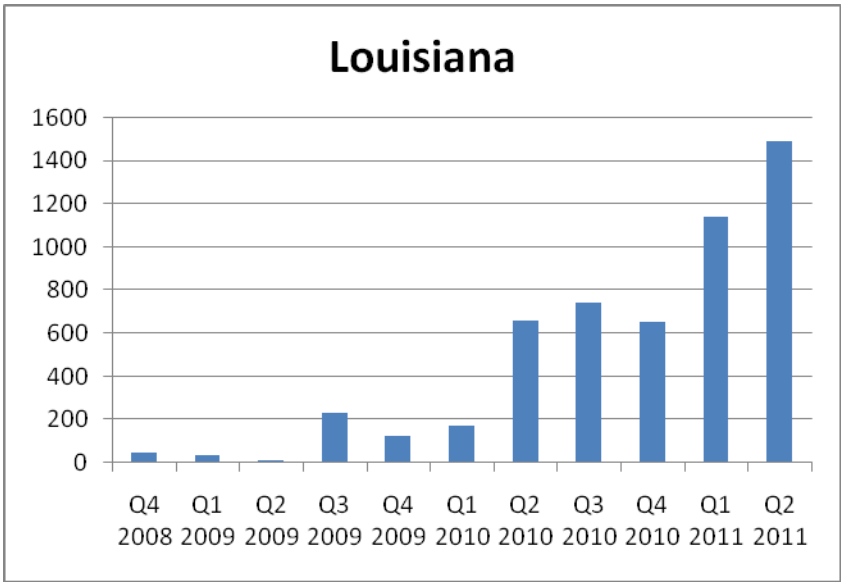
Idaho



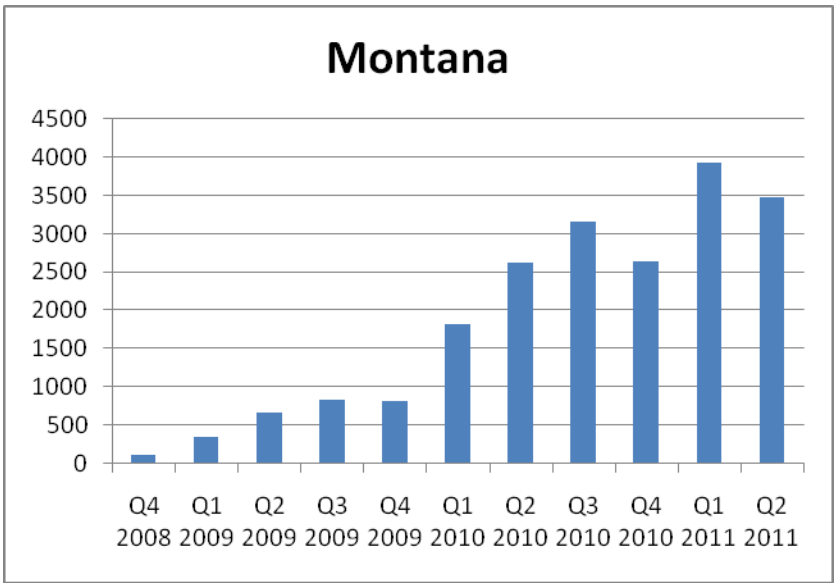
Indiana



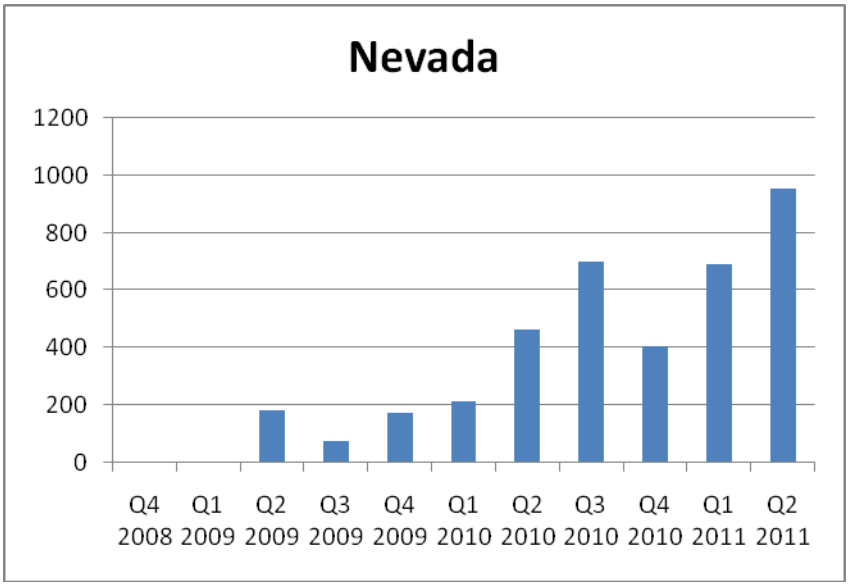
Louisiana



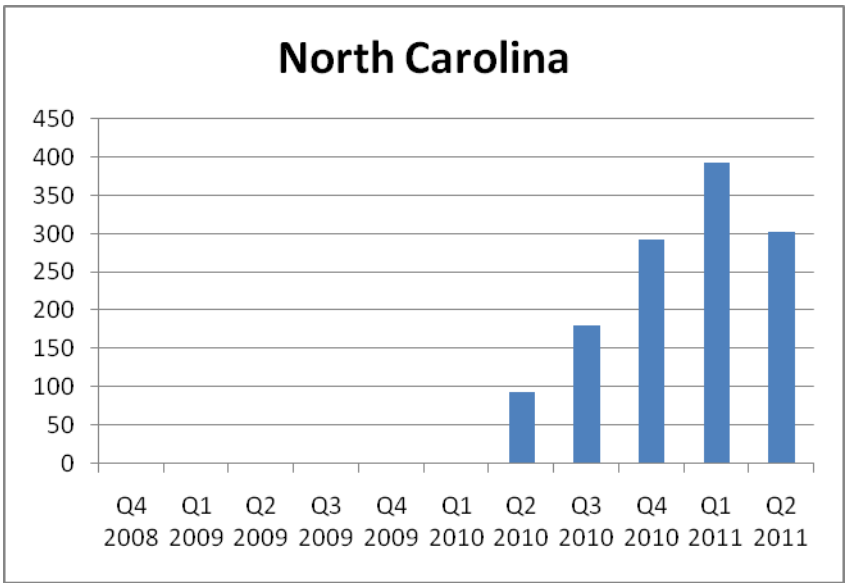
Montana



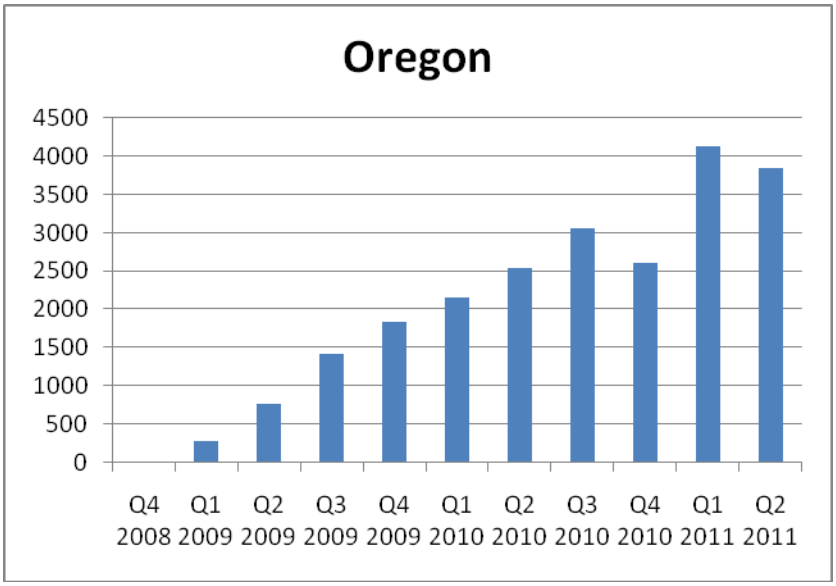
Nevada



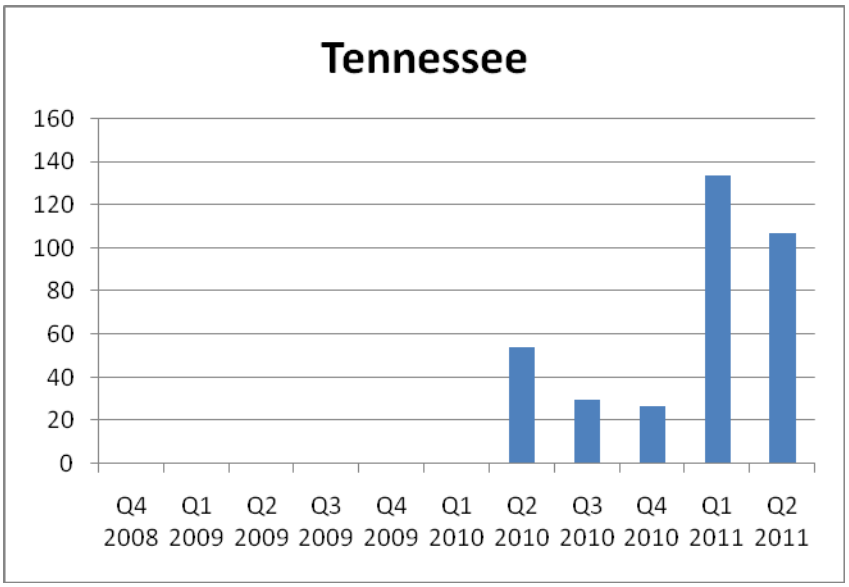
North Carolina



Oregon



Tennessee



Appendix 12: Record Totals by Partner

The following table lists the total number of records submitted to the state archives from each state's time of joining the project until the end of June, 2011.

State	**Records submitted** (Does not include SSDI)
Colorado	722,508
Idaho	162,544
Indiana	806,660
Louisiana	0
Montana	3,592
Nevada	25,782
North Carolina	887
Oregon	16,778
Tennessee	60