Digital Preservation 101: State Archives and Agencies Putting Digital Preservation into Action
Part 2

February 28, 2017
Welcome!

**PDP Briefings**
- Protecting and Preserving Long-Term Digital Information January 24
  - For IT Professionals & practitioners
- The Governance of Long-Term Digital Information May 23
  - For Senior Managers & Budget Administrators

**PDP “Hot Topic” Webinars**
- Tuesdays 2-3pm Eastern
  - Preserving and Protecting Audio-visual Files April 11
  - Preserving Digitized State Government Records May 9

**PDP Online Workshops - Digital Preservation 101:**
‘State Archives and Agencies Putting Digital Preservation into Action’
- Part 1: Practical Training in the Key Concepts February 14
- Part 2: Practical Training in the Key Concepts February 28
- Part 3: Case Studies March 14

**Sign up today** on the CoSA website – PERTTS Portal > Education -Training
Today’s Presenters

David Portman
Preservica

Lori Ashley
Tournesol Consulting

Tim Hodge
Preservica

Bonnie Weddle
New York State Archives
Council of State Archivists (CoSA)

- 56 state and territorial archives
  - Preservation of and access to records of enduring value
  - Efficient management of government records

https://www.statearchivists.org/
Your State Archives

- Records scheduling
- Advice and training
- Transfer of records of enduring value
- State-specific procedures and best practices

https://www.statearchivists.org/connect/resources-state/
Agenda

• Introduction

• Retention and Preservation Strategies and Tools

• Taking Action to Protect Long-term Records: Fostering Collaboration with IT

• Summary & Preview of Part 3

• Q&A
Retention and Preservation Strategies and Tools

Lori Ashley & Tim Hodge
Recap Part 1: Long-Term Record Requirements

• Long-term electronic records (10+ year retention) must be actively managed

• Start as early in the lifecycle as possible, ideally when planning to create records and/or configuring systems

• Decisions include formats and essential documentation about the systems and information objects (metadata)

• Integrate into IT infrastructure and architecture wherever possible

• Specialized expertise required for addressing the scope, scale and diversity of long-term records and metadata
Long-Term Digital Archiving and Preservation

Provenance
- Audit Trails and Metadata

Authenticity
- Fixity

Durability
- Self-healing Storage

Preservation
- Active File Format Migration

Disposition
- Retention Policies & Actions

Privacy & Confidentiality
- User Roles & Permissions

Access
- Search, Browse & Download
Retention Schedule Examples: Permanent Records

**STATE AGENCY RECORDS RETENTION SCHEDULE**

<table>
<thead>
<tr>
<th>Records Title</th>
<th>Function and Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>U0243 Capital Construction Disbursements and Receipts File</td>
<td>This series documents the disbursement of undertaking which is to be financed or funded by federal funds; (b) any undertaking which is construction item, or combination of items equipment estimated to cost $200,000 or an item of movable equipment if the total cost upgrade, or replacement of an information the Capital Planning Advisory Board of the as to state spending for capital projects.</td>
</tr>
</tbody>
</table>

**Access Restrictions**

None

**Contents**

Series may contain: Date of progress payment if also may contain supporting documentation

**Retention and Disposition**

Retain permanently in creating unit

---

**04550 Governor's Executive Order File - Miscellaneous (Includes employment contracts, court orders, non-merit appointments, paroles, re-organizations, special notaries) (V)**

This series documents the written action or order of the Governor, which has the force of law, and which is based upon his authority as the chief executive of the Commonwealth. The state's Constitution requires that the Secretary of State receive the original order signed by the Governor. After signature by the Secretary of State, the seal is affixed. This series incorporates the Governor's orders that are not related to the Governor's office and is the creation of the Governor. Much of the reference activity to Executive Orders is by courts requiring certification of the official acts of the state. This series reflects the Governor's orders, including special and temporary orders.

**Access Restrictions**

None

**Contents**

Series contains: Original order; supporting documentation

**Retention and Disposition**

Retain in Agency four (4) years after the end of the gubernatorial term; transfer to State Archives Center for permanent retention

---

Retain permanently in creating unit

Retain in Agency until disposition trigger; transfer to State Archives
Permanent Records: From Governor’s Office to State Archives

We as the Texas State Library & Archives Commission have a statutory responsibility to collect, preserve and make records accessible for the future.

https://www.tsl.texas.gov/texasdigitalarchive

Jelain Chubb
Texas State Archivist

“We as the Texas State Library & Archives Commission have a statutory responsibility to collect, preserve and make records accessible for the future”
Four Preservation Strategies You Can Use

1. Preferred File Formats
2. Metadata Management
3. Content Integrity Checks
4. File Format Migration
Strategy 1: Preferred File Formats

• Characteristics of good preservation formats include:
  – Public and open documentation
  – Non-proprietary
  – Widespread adoption
  – Can be opened, read and accessed using readily-available tools

• Examples of Open Standard/Neutral Formats:
  – HTML, Plain Text, XML, ODF and PDF/A for text
  – PDF/A, PNG and TIFF for scanned images
  – CSV for spreadsheets
  – SVG for vector graphics
Strategy 2: Managed Metadata

Unique Identifiers can also store descriptive data
EX: MFM121587

Journal of Information Science
March 18, 2013
doi: 10.1177/0165551513480107

Major Land Resource Area coverages
- Continental United States
- Alaska
- Hawaii
- Puerto Rico

Patent No.: US 6,539,396 B1

Stable URL: http://www.jstor.org/stable/3051911

Description
Major Land Resource Area (MLRA) boundaries for Puerto Rico and the U.S. Virgin Islands.

Purpose
This coverage is intended as a base map for US national level natural resource analysis.

Supplemental Information
This coverage contains the Major Land Resource Area (MLRA) boundaries for the
U.S. Virgin Islands and Puerto Rico. MLRA boundaries were revised in the
mid-1990s from earlier versions established in the 1970s and 1980s by the U.S.
Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS),
formerly the Soil Conservation Service (SCS). The revised MLRA boundary
coverage was digitized at the Soil Survey Center in Lincoln, Nebraska, the
Strategy 3: Integrity Checks

• A checksum is a mathematical algorithm run on/against a file and its resulting value

• The resulting “fixity” value uniquely identifies the file based on its content down to the bit level

• The resulting value is called:
  – Fingerprint
  – Hash Value
  – Cryptographic Hash Function

• A checksum can be recalculated and compared at any time to determine if there have been changes
Strategy 4: File Format Migration

- File format is characterized and captured in digital preservation systems
- Original record in native format is retained as the preservation master
- Records are transformed from one file format to another as needed and the metadata captured
- Guards against file format obsolescence
DEMO: Format Migration Workflow
## Long-Term Records Management Partners

### Archives
- Appraisal support
- Expertise in digital preservation techniques
- File format monitoring
- Access rights
- Transfer processes and procedures
- Search and retrieval
- Best practices & guidance
- Education & training

### Information Technology
- Apply conversion and migration techniques
- Document records destruction and transfer
- Metadata management
- Interoperability protocols
- Integrity checks and chain of custody
- Redundancy in storage architecture
- Data security
Taking Action to Protect Long-Term Records: Collaboration with IT

Bonnie Weddle

New York State Archives
Records Management Is Collaborative

- We build relationships with
  - Records creators
  - Senior management
  - Legal counsel
  - State archives or records management agency
  - Information technology
“But I’m not tech-savvy!”

• You don’t have to become an IT expert

• Include records creators in the conversation

• You can always ask questions

• The Internet is your friend
## IT Speaks a (Slightly) Different Language

<table>
<thead>
<tr>
<th>Term</th>
<th>Information Technology</th>
<th>Records Officers and Archivists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record</td>
<td>A complete item in a database</td>
<td>Any item, “regardless of physical form or characteristics,” that meets the definition outlined in relevant statute or regulation</td>
</tr>
<tr>
<td>Archive</td>
<td>Move to less frequently used storage medium (e.g., tape)</td>
<td>Transfer to state archives -or- Active preservation care</td>
</tr>
<tr>
<td>Permanent</td>
<td>About 10 years</td>
<td>Forever</td>
</tr>
</tbody>
</table>
Need to Emphasize Benefits to IT

• Increased efficiency
  – Information readily located and retrieved

• Reduced costs
  – Records transferred / deleted at the appropriate time
  – Obsolete and duplicate information destroyed
  – Appropriate storage choices made

• Reduced risk
  – Protection of mission-critical information
  – Protection of information of long-term value
  – Legal and regulatory compliance
  – Litigation, audit, and investigation
2017 State CIO Priorities

1. Security and Risk Management
2. Consolidation/Optimization
3. Cloud Services
4. Budget, Cost Control, Fiscal Management
5. Legacy Modernization
6. Enterprise IT Governance
7. Data Management and Analytics
8. Enterprise Vision and Roadmap for IT
9. Agile and Incremental Software Delivery
10. Broadband/Wireless Connectivity

http://www.nascio.org/topten
“We’ve Got Storage and Backup”

• **Storage and backup characteristics**
  – Wide range of systems and practices
  – Multiple copies
  – Bit-level storage

• **Challenges**
  – Locating needed information
  – Ensuring appropriate disposition
  – Records that must be kept for >10 years
“We Have an ECMS”

• Electronic Content Management System characteristics
  – Widely accepted strategies, tools, and practices
    • Content capture and management
    • Access mechanisms and controls
    • Version control
  – Bit-level storage
  – May offer records management functionality
    • DoD 5015.2 compliance

• Challenges
  – Records management modules add to cost
  – Records that must be kept for >10 years
“What’s a Digital Preservation System?”

• Key characteristics
  – Long-term availability, usability, and trustworthiness
  – Protection against file format obsolescence
  – Integrity safeguards
  – Preservation metadata
  – ISO 14721 (Open Archival Information System Reference Model) compliance
  – Optimal environment for records that must be kept for >10 years
  – Moving toward integration with ECMS

• Challenges
  – Making the case to IT and senior management
Enterprise Content Lifecycle

Content Management System

Digital Preservation System

Digital Content Contributor

Creation

Use/Maintain

Short Term Retention

Long-Term Retention

Permanent Retention

Digital Content Consumer

Information Manager

CoSA
Council of State Archives

Preservica
Digital Preservation
Getting Your Foot in the Door

• Keep emphasizing
  – Increased efficiency
  – Risk and cost reduction
  – Storage/backup and ECMS don’t support long-term preservation

• Keep looking for potential allies
  – Information security officer
  – Chief data officer
  – Past partnership with records management
  – Chaotic freedom of information or e-discovery experience
  – Senior management
  – Legal counsel

• Contact your state archives
Potential Collaborations

• Develop standards and guidance
• Identify information with high, long-term value
• Update retention schedules
• Develop disposition workflows
• Support preservation systems design and procurement
• Raise awareness of electronic records risks and benefits
Summary

• Effective records management through time, systems and custodians

• Preservation Strategies to use for long-term electronic records:
  – File integrity checks
  – Preferred file formats
  – Robust metadata management
  – Periodic transformation of electronic records

• Minimizing barriers to cross-agency records transfer will improve collaboration and reduce duplication of effort

• Capabilities for long-term digital preservation that are integrated into public sector IT infrastructure and architecture should result in significant economic savings.
PERTTSS Portal

Welcome to the Program for Electronic Records Training, Tools, and Standards - the PERTTSS Portal.

Education - Training

Webinar and training opportunities available through CoSA. Information on and links to opportunities from other organizations are also included.

SERP Framework

The State Electronic Records Preservation (SERP) Framework provides information and guidance on 15 different areas on how to move forward in areas associated with digital preservation. The SERP Framework is based on the Digital Preservation Capability Maturity Model which states are using in the CoSA requested Self-Assessments.

Electronic Records Resources

Searchable database of resources including tools, standards, policies/guidelines, and other informational guides related to digital preservation and records management topics can be found in the CoSA Resource Center.

* Add resources! Login to submit resources that you find most helpful. When ready to add a resource, click the Add Resource button on the Resource Center Home page.

CoSA thanks the National Historical Publications and Records Commission and Preservica for their substantial support of the PERTTSS Portal.

https://www.statearchivists.org/pertts/
Next Steps & Resources


- Preservica Resources Center http://www.preservica.com/resources/

- Preservica Webinars
  - Live Demo: March 2 @ 10am Eastern
    Register at www.preservica.com

- Upcoming Events
  - CoSA: Boise, ID (10-13 July)
  - NAGARA: Boise, ID (12-15 July)
  - SAA: Portland, OR (23-29 July)
Preview of Part 3 – Case Studies

*Tuesday March 14 @ 2pm Eastern*

- Introductions & Recap
- State Archive Use Cases
  - *Examples from Arkansas & Alabama State Archives*
- **Resources**
  - *CoSA Digitization Guidance*
  - *Tools & Resources for your agency*
- Next Steps
- Summary
- Q&A
Questions?