Long-term Digital Preservation and the Texas Digital Archive

E-Records Conference

Nov. 4, 2018
WHAT IS A RECORD?

Texas Code 441.108(11) defines a “state record” as:

• “means any written, photographic, machine-readable, or other recorded information created or received by or on behalf of a state agency or an elected state official that documents activities in the conduct of state business or use of public resources. The term includes any recorded information created or received by a Texas government official in the conduct of official business, including officials from periods in which Texas was a province, colony, republic, or state.”
What is an electronic record?

Sec. 441.189. ELECTRONIC STATE RECORDS.

• (a) Any state record may be created or stored electronically . . .
Characteristics of an Electronic Record

• Four essential characteristics:
  – Authenticity - A record must be what it purports to be.
  – Reliability - A record must be a full and accurate representation of the transactions, activities, or facts to which it attests.
  – Integrity - A record must be complete and unaltered.
  – Usability - A record must be able to be located, retrieved, presented, and interpreted.

• Digital preservation is the challenge of maintaining all of these characteristics over time

International Standards Organization (ISO) 15489-1:2001
Authenticity = TRUST

My son as Abe

Portrait of Abraham Lincoln
Reliability = completeness
Integrity = Protecting the bits

Example of Bit Rot
Usability
It’s all about the metadata!
(Sung to the tune of Meghan Trainor’s – “All About the Bass”)
Authentic Electronic Records

• It’s all about the Metadata: documents the activities of creation and use
  – Administrative metadata:
    • Access information
    • Audit trails
    • Retention Schedules
  – Technical metadata:
    • File format information
    • Checksums
    • System information/requirements
  – Descriptive Metadata
    • Bibliographic data (Who, what, when, where, why)
    • Indexes/Finding aids
Records Life Cycle

• **Birth** - Creation or receipt

• **Working life**
  – Active records – accessed more than 1/month
  – Inactive records – less than 1/month

• **Death** - Ultimate disposition
  (destruction or transfer to an archive)
What do we mean by “Long-term”?

- **Short-term (0-5 years)**
  - Highly active – on-line storage
  - “Normal” management issues
  - Possible application version upgrade

- **Mid-term (5-10 years)**
  - Less active – Near-line/Off-line storage
  - Multiple version control issues
  - System upgrades
  - Possible hardware/software migration

- **Long-term (10+ years)**
  - Least active – Off-line storage
  - Migration/conversion likely

Eventually all electronic records must migrate
Preservation Challenges
Unlike some paper records, digital records do not survive without constant attention.

Biggest Challenge facing electronic records - **CHANGE**

**Preservation Challenges**

- **Technical obsolescence**
  - Hardware, software, media

- **Technological Dependency**
  - Hardware, software, etc.

- **Media Deterioration**
  - Magnetic, Optical,
When Records Are at Risk

• Software is obsolete or getting to end of life
• Dependency on specific hardware or system
• Non-standard/specialty formats used
• Systems being replaced
• Records no longer in use
• Records are on removable media
1. Identify - Know What You Have

- Inventory your records
  - File types
    - Digital Photos
    - Text files (Word, Open Office, PDF)
    - Digital Audio/Video
  - Where are your files located
    - On your computer
    - On device (camera, phone, iPad)
    - On removable media (CD, DVD, Flash Drive)
    - Online (Google drive, iCloud, Facebook, Instagram)
Get control over your records!

• Starts with Records Management
  – Records Retention Schedules
    • Identify what records you have and what purpose they serve
    • Establish appropriate retention periods to manage volume
      – Dispose of records at the right time
      – Protect records that need to be retained
  – Recordkeeping systems
    • Organize and Categorize Records
    • Centralized control
    • Improved Access and storage
  – Policies and Procedures
    • People know how to use the system
    • Train people on the culture of management
What are you trying to preserve?

• **Information** - The raw data or information contained in the record

• **Functionality/Appearance** - Look & Feel

• **Context** — How the records relate to other records
Information versus Appearance

2. Active Records

Records management can have the greatest positive impact when applied to active records. These are records that are in use and need to be maintained, and are essential to the government because they are used most frequently, cost the most to maintain, and are essential to the current management of the government.

Activities Eligible for Support

a) File Management supports projects to reorganize paper or electronic files and develop and implement file classification systems, to develop written policies and procedures, and to train staff. Eligible expenditures include file shelving (including that with locking covers), side-tab file folders and associated supplies. Fire-resistant filing cabinets are eligible only if their need is sufficiently justified, but filing cabinets (storage devices with drawers) and top-tab file folders are not eligible.

b) Disaster and Business Recovery Planning supports projects to develop and test disaster and business recovery plans.

c) Imaging and Document Management supports projects to index or improve access to any active records, including minutes, vital records, or student records. Methods used to improve access might include traditional indexing, implementing full-text searching software, scanning and converting printed text to electronic text, or some combination of these.

d) Imaging and Document Management supports projects to index or improve access to any active records, including minutes, vital records, or student records. Methods used to improve access might include traditional indexing, implementing full-text searching software, scanning and converting printed text to electronic text, or some combination of these.

e) Geographic Information Systems (GIS) supports GIS needs assessment and implementation projects. The State Archives provides seed money for the initial implementation of GIS in local governments, but does not fund continued improvements to GIS beyond this point.

f) Electronic Records Systems covers the development of needs assessments or the implementation of records systems not covered under another category. Such recordkeeping systems could include database management systems (such as fire incident reporting software), computer output to laser disc (COLD) applications, and many others.

2) Business Process Analysis (BPA) supports the analysis and improvement of business processes that create or maintain records. BPA projects are a good choice for a government that has identified a problem with the way it conducts a specific recordkeeping activity but does not have a specific technological solution to that problem.

Category Requirements

General Technology Project Requirements:

Records Management Focus. The State Archives does not fund technology projects; it funds records management projects that sometimes have a technology focus. To be eligible for funding, a
Information versus Functionality
The value of context

This is a GIS shapefile – one layer out of GIS system. By itself it’s just a bunch of dots. (Stockyards in KY)

All of the dots have data attached to them

But without more map layers behind it, it’s just a bunch of dots.
Who are the users of your records?

• With electronic records, different users may require:
  – Different levels of access to data
    • Log-ins and permissions
    • Redaction
  – Different formats
    • “live data” (spreadsheet or database) rather than static printout or PDF
    • High resolution image/recording (for commercial use) versus a low resolution copy for casual printing
Long-Term Preservation Strategies: Standard (Sustainable) Formats

Formats expected not to change, or change slowly

• Widely supported & used
• Easily Transferable
  – Compatible with other applications
  – Forward/Backward Compatibility
  – Version Control
• Non-proprietary or open/published standards

Types of Standards
- State/Enterprise Standards
- National/International Standards (ANSI, ISO)
- Industry Standards
- “De facto” Standards
Sustainability Factors for Digital Formats (Library of Congress)

1. Disclosure. Degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. A spectrum of disclosure levels can be observed for digital formats. What is most significant is not approval by a recognized standards body, but the existence of complete documentation.

2. Adoption. Degree to which the format is already used by the primary creators, disseminators, or users of information resources.

3. Transparency. Degree to which the digital representation is open to direct analysis with basic tools, such as human readability using a text-only editor.

4. Self-documentation. Self-documenting digital objects contain basic descriptive, technical, and other administrative metadata.

5. External Dependencies. Degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments.

6. Impact of Patents. Degree to which the ability of archival institutions to sustain content in a format will be inhibited by patents.

7. Technical Protection Mechanisms. Implementation of mechanisms such as encryption that prevent the preservation of content by a trusted repository.

http://www.digitalpreservation.gov/formats/sustain/sustain.shtml
File FormatNormalization

• Goal: Sustainable formats
  – Some archives have policies about which formats they will accept
  – Archives should identify preferences for major categories of ER
    • TIFF, JPEG, PDF, XML, ASCII (flat files)
• Reduce the number of formats that need to be monitored for migration
• Repeatable (understood) migration procedures
• Preferred file formats
  – Link to the webpage
Integrity = Protecting the bits

Example of Bit Rot
Checksums

- Sometimes called a “hash”
- Uses algorithms to calculate a “message digest” value about the bit order in a file
  - Essentially the digital “fingerprint” of a file
- Doesn’t tell you what changed, just that something has
  - Intentional change or bit-rot?
- Types:
  - MD5
  - SHA-1, SHA-2, ...
  - Each type creates a different hash value (the md5 for a file will not be the same as its SHA-256)
Tools for Managing Integrity

ExactFile

- Checksum-only program
- What it does
  - Runs a wide variety of checksums
  - Creates a checksum file
  - Can run the checksum file at any time
- Does require installation
- Runs quickly
Biggest Preservation Challenge - Cost

• **Storage**
  – Multiple copies
    • Access Copies
    • Preservation Versions
  – Data Volume
  – Continuous increase without purging

• **Staff Resources**
  – Description, preservation, access
  – Backlog size

• **IT Support and Maintenance**
  – Unending and on-going
  – Backup/Replication Practices
Building the Texas Digital Archive
What is an Archival state record?

Texas Government Code 441.180(2)

"Archival state record" means a state record of enduring value that will be preserved on a continuing basis by the Texas State Library and Archives Commission or another state agency until the state archivist indicates that based on a reappraisal of the record it no longer merits further retention.
TSLAC’s Mission

• Preserve the record of government for public scrutiny,

• Secure and make accessible historically significant records and other valuable resources, both for print and electronic documents,

Been having issues with the electronic part. Now looking to change that.
Benefits of transferring records to the TDA

• Relieves the preservation burden from the agency
• Saving taxpayer money – Reducing agency storage and migration costs
• Reducing agency staff time for PIA requests
• Increases transparency by making more records accessible
• Preserves and protects electronic state records a secure trusted repository
Want to transfer electronic records to the TDA?

• **First step – Contact us!!**
• Complete our electronic records survey
• Identify the series to transfer
  – “A” & “R”
    – “R” series need to be appraised by archives staff
• Negotiate the timing and method transfer
- Laura Saegert – Assistant Director for Archives
  - lsaegert@tsl.texas.gov
  - 512-463-5500

- Mark Myers – Senior Electronic Records Specialist
  - mmyers@tsl.texas.gov
  - 512-463-5434

-If you have worked with an appraisal archivist you can contact them as well.
Questions?

Thank You!

Mark Myers
Senior Electronic Records Specialist
E-mail:
mmyers@tsl.texas.gov
https://www.tsl.texas.gov/texasdigitalarchive

Current Lobby Exhibit at TSLAC:
Greetings from Texas: A History of Postcards from the Lone Star State