

Library of Virginia  
Presents:  
Electronic Records  
Town Hall  
April 30, 2024


# Highs and Lows of Digital Preservation

1


## My adventures

- Alabama Department of Archives and History – 1998-2001
  - Electronic Records Management
- Kentucky Department for Libraries and Archives – 2001- May 2014
  - Electronic Records Management
  - Digital Preservation
- Texas State Library and Archives Commission – June 2014 – Present
  - Digital Preservation

2



Center aisle of the Texas State Records Center




<https://tsl.texas.gov/texasdigitalarchive>

3

## What is a Record?

"A record created, received, and maintained as evidence and information by an organization, in pursuance of legal obligations or in the transaction of business."


- International Standards Organization (ISO) 15489-1:2001 – Information and Documentation – Records Management – Part 1 – General.



4

## What is an Electronic Record?

- Data or information that has been captured and fixed for storage and manipulation in an automated system and that requires the use of the system to render it intelligible by a person.
- From "A Glossary of Archival and Records Terminology" (online edition at: <http://www2.archivists.org/glossary/terms/e/electronic-record>)



5



6

## Records are Records



Account book 1835-1856

Factory Details Store				Sales Ledger			
Date	Account Debit	Invoice No.	Part No.	Date	Part No.	QTY	Amount
1835	1	1000	100	1835	100	1	1,000.00
1835	2	1000	200	1835	200	1	2,000.00
1835	3	1000	300	1835	300	1	3,000.00
1835	4	1000	400	1835	400	1	4,000.00
1835	5	1000	500	1835	500	1	5,000.00
1835	6	1000	600	1835	600	1	6,000.00
1835	7	1000	700	1835	700	1	7,000.00
1835	8	1000	800	1835	800	1	8,000.00
1835	9	1000	900	1835	900	1	9,000.00
1835	10	1000	1000	1835	1000	1	10,000.00

Accounting software

7

## 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

8

## Authenticity

- An **authentic** record is one that can be proven to be what it professes to be, to have been created or sent by the person claiming to have created or sent it, and to have been created or sent at that time.

International Standards Organization (ISO) 15489-1:2001 – Information and Documentation – Records Management – Part 1 – General, section 7.2.2 – 7.2.5.

9

## Reliability

- A **reliable** record is one whose contents can be trusted as a full and accurate representation of the transactions, activities or facts to which they attest.

International Standards Organization (ISO) 15489-1:2001 – Information and Documentation – Records Management – Part 1 – General, section 7.2.2 – 7.2.5.

10

## Integrity

- “When we say that a digital object has ‘integrity,’ we mean that it has not been corrupted over time or in transit; in other words, that we have in hand the same set of sequences of bits that came into existence when the object was created.”

Clifford Lynch,  
*Authenticity in a Digital Environment*  
Council on Library and Information Resources, 2000

11

## Usability

“Information which is not communicated is valueless, and information that cannot be found is similarly worthless.”

Robek, Brown & Stephens,  
*Information and Records Management, 4<sup>th</sup> ed*

12

### It's all about the Metadata

Information about the record

- 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

13


### Access = Preservation

- The more a record is used, the more likely you are to keep it in a usable format and migrate it forward through time
- As use goes down:
  - Formats don't get updated
  - Removable media become obsolete
  - Media isn't monitored for deterioration

14

### Preservation Challenge

- ▶ 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

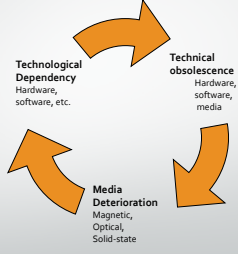


Eventually all electronic records must migrate

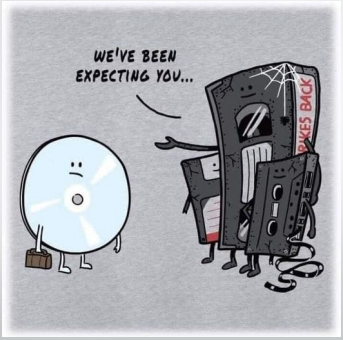
15

### Preservation Challenges

- Unlike some paper records, digital records do not survive without constant attention
- Biggest Challenge facing electronic records - **CHANGE**



16



17

### 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

18

### Starts with Records Management

- Identify what records you have and what purpose they serve
- Identify where they are located
- Records Retention Schedules
  - Establish appropriate retention periods to manage volume
  - Dispose of records at the right time
  - Protect records that need to be retained
- Policies and Procedures
  - Spell out the rules and processes
  - Train people on how to use tools

19

### 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

20

### Information versus Appearance

21

### Preservation Strategies

- Conversion to hardcopy (paper or microfilm)
- Sustainable formats (Normalization)
- Migration
- LOCKSS (Lots of Copies Keep Stuff Safe)

22

### Electronic vs Paper Records

<p><u>Advantages</u></p> <ul style="list-style-type: none"> <li>Increased Storage Capacity</li> <li>Improved Searching</li> <li>Increased Access</li> <li>Improved Workflow</li> </ul>	<p><u>Disadvantages</u></p> <ul style="list-style-type: none"> <li>Hardware/Software dependent</li> <li>Access/Security issues</li> <li>Maintenance Costs</li> <li>Long-term Retention</li> </ul>
--	---

23

### Long-Term Preservation Strategies: Convert to Hardcopy (Eye-readable)

**Best for text-based data:**  
Text files  
Imaged paper  
Digital Photos (mainly B&W)

- Solution when**
  - All necessary metadata is captured
  - No need to maintain functionality of records
  - Frequency of use goes down as time goes on
  - Color is not a major concern
- Examples**
  - Print electronic data to paper
  - Digital to Microfilm

24

### Long-Term Preservation Strategies: Standard (Sustainable) Formats

**Types of Standards**

- State/Enterprise
- National/International Standards (ANSI, ISO)
- Industry Standards
- "De facto" Standards

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

25

### Problems with changing formats

26

### LOCKSS – Lots of Copies Keeps Stuff Safe!

- Multiple copies
- Multiple media
- Multiple locations
- Separate from disaster backups

27

### Evaluate Preservation Strategies

- Cost effectiveness
- Accessibility
- Necessary functionality
- Newness of strategy

28

### Evaluate Resources and Cost

- Storage
  - In-house – Online (Cloud)
- Data Volume
  - Continuous increase without purging
  - Multiple copies
- IT Support and Maintenance
  - In-house vs Vendor support
- Staff Resources
  - New education and skills
  - New personnel


29

**Volume Issues: Multiple versions**

30

## Word about compression


- Digital media files come in 3 types:
  - Uncompressed – “raw” files in actual size
    - Audio – Wave (.wav)
    - Image - Tiff (.tff)
  - Lossless – Compressed but without loss of quality
    - Audio – Windows Media Audio (.wma)
    - Image – PNG (.png)
  - Lossy – Compressed but with loss of quality
    - Audio – Mp3
    - Image – JPEG (.jpg)



31

## Thoughts about the “cloud”

- Cost – Cheaper than stand-alone system - maybe
- Geographic dispersal of copies
- No on-site maintenance
  - No software to install
  - Accessed through web interface
- Level of In-house IT support
- Have a good contract



32

## Collaborating with IT

- IT maintains the technology
  - IT is the *physical custodian*
- IT does not own the data
  - Creator is the *legal custodian*
- IT focuses on maintaining the storage devices
  - IT does not consider long-term access/ preservation
- Data owner must define its need for retention and communicate with IT
  - Data owners often assume IT is already addressing this need

33

## “ARCHIVE”



THAT WORD

I do not think it means what you think it means

34

## Best Advice

- Figure out what you can do and work on it.
  - Exploit your strengths
- Keep it simple
  - Don't reinvent the wheel
  - Look for practical/realistic solutions
- Don't be afraid to get dirty!
  - Doing something is better than doing nothing

35

## Questions?

### Thank You!



Mark Myers

E-mail:  
mark.myers1031@gmail.com

36