STORAGE FOR ELECTRONIC RECORDS

A Discussion
LEAD PRESENTERS

Veronica Martzahl
Massachusetts Archives
Electronic Records Archivist

Tibaut Houzanme
Indiana Commission on Public Records
Electronic Records Specialist
ACKNOWLEDGEMENTS

This webinar is made possible by a grant from the National Historical Publications & Records Commission (NHPRC)
OVERVIEW

- Survey Results
- Definitions
- Presentation: A short review of Storage and Content Allocation
- Panel Discussion
SURVEY RESULTS

- 3 Questions plus demographics
- 18 Responses
  - Just under 1/3 of possible respondents

THANK YOU!
STORAGE OPTIONS BEING UTILIZED

- Nothing
- Floppies, CD, DVD, thumb drives
- External hard drives
- SAN
- NAS
- Unified Storage
- Tape
- M-disc
- Cloud
- Whatever the data center is using
COST OF STORAGE

- Absorbed by IT
- Rolled into other budget lines
- Per GB per month
  - Range from $.02 to $2.50 depending on type
  - May be additional maintenance charges or backup fees
- Cloud
  - Tends to be TB per year
  - $1,000-$2,000
QUESTIONS TO BE ADDRESSED

- Lots of them!
- BP&T Subcommittee will be setting up a Writeboard in Basecamp where people can comment on them
- Use them as start of questions for Forum/Interactive capabilities that will be added to the PERTTS Portal
DEFINITIONS

DAS – Direct-attached Storage

- Attached directly to a computer or a server
- Not part of a network
- Can be just the internal hard drive in a computer
- Often means dedicated storage arrays attached directly to servers
- Can include tape libraries or RAID hardware (redundant array of independent disks)
DEFINITIONS

SAN – Storage Area Network

- High-speed network of storage devices connected with services
- Can also include tape libraries and RAID hardware
- Removes storage from the server and consolidates it where it can be accessed by any application attached to the network
DEFINITIONS

NAS – Network Attached Storage

- Server that is dedicated to file sharing
- Storage-centric design that is drawn on by other servers
DEFINITIONS

Unified Storage (Network Unified Storage)

- Shared storage device
- Provides consolidated block and file services
- Basically – combination of NAS and SAN
CHOOSING, IMPLEMENTING & ALLOCATING STORAGE FOR DIGITAL PRESERVATION

A Short Review

Tibaut Houzanme, Electronic Records Specialist
Indiana Commission on Public Records – ICPR
thouzanme@icpr.in.gov
1. Storage, Repository & Preservation
2. Storage to Mitigate Preservation Risks
3. Content Characteristics
4. Storage Properties, Options, Locality, Likelihood & Control
5. Interoperability
6. Cost of Ownership (TCO)
7. Content Allocation
8. Storage Trends
9. Recap
So what?: Placing a storage sub-function within its meaningful context, is everything.
**ADDRESSING DIGITAL PRESERVATION THREATS & VULNERABILITIES**

---

**Techniques**

<table>
<thead>
<tr>
<th>Threats and vulnerabilities</th>
<th>Redundancy</th>
<th>Migration</th>
<th>Emulation</th>
<th>Refreshing</th>
<th>Diversity</th>
<th>Inertia</th>
<th>Metadata</th>
<th>Auditing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media faults</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>Media obsolescence</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td>Hardware faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td>Hardware obsolescence</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td>Communication faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td>Network service failures</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td>Software obsolescence</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td><strong>Disasters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural disasters</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Human operational errors</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td><strong>Attacks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal attack</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>External attacks</td>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td><strong>Management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic failures</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Organizational failures</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>-</td>
<td>-</td>
<td>R</td>
</tr>
<tr>
<td><strong>Legislation</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>r</td>
<td>-</td>
<td>r</td>
<td>-</td>
</tr>
</tbody>
</table>

r=: reduces risk/threat/vulnerabilities.  R=: required for recovery; -= does not fit.

LONG-TERM CONTENT CHARACTERISTICS

Primary characteristics of digital content(s) should drive storage choice

<table>
<thead>
<tr>
<th>Nature</th>
<th>Records</th>
<th>SIPs</th>
<th>AIPs</th>
<th>DIPs</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth means</td>
<td>Reformatted digital</td>
<td></td>
<td></td>
<td></td>
<td>Born digital</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>Digitized-Surrogates (Original is kept)</td>
<td>Digitized-Original (Original is destroyed)</td>
<td>Digital-Original (no other format available)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>High Traffic</td>
<td>Medium Traffic</td>
<td>Low Traffic</td>
<td>Very Low Traffic</td>
<td>Time-Bound Traffic</td>
</tr>
</tbody>
</table>
## PROPERTIES, OPTIONS, LOCALITY & CONTROL

### Medium Properties*  
**Storage Options, Locality, Likelihood and Records Control**

<table>
<thead>
<tr>
<th>Media</th>
<th>Longevity</th>
<th>Capacity</th>
<th>Viability</th>
<th>Obsolescence</th>
<th>Cost</th>
<th>Susceptibility</th>
<th>Total</th>
</tr>
</thead>
</table>

### STORAGE OPTIONS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DIY / Purchase Solution</td>
<td>Total Control</td>
<td>Good Control</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Seriously, check your laws</td>
</tr>
<tr>
<td>State IT</td>
<td>Good Control</td>
<td>Some Control</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Seriously, check your laws</td>
</tr>
<tr>
<td>Collaborative (Local)</td>
<td>Good Control</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Little Control</td>
<td>Seriously, check your laws</td>
</tr>
<tr>
<td>Collaborative (cross-state/Country)</td>
<td>Good Control</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Very Little Control</td>
<td>Seriously, check your laws</td>
</tr>
<tr>
<td>Commercial (Cloud)</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Little Control</td>
<td>Very Little Control</td>
<td>Seriously, check your laws</td>
</tr>
<tr>
<td>Commercial [Storage as a Service]</td>
<td>Some Control</td>
<td>Little Control</td>
<td>Very Little Control</td>
<td>Serious, check your laws</td>
<td>Seriously, check your laws</td>
</tr>
</tbody>
</table>

* Control refers to the ability of the State Archive to directly and independently make decisions about storage implementation and daily management.

INTEROPERABILITY FOR EFFECTIVENESS/ECONOMY

TIPR: Towards Interoperable Preservation (for object transfer and full repository ingest/migration)

VERS/VEO (encapsulation & system interoperability)

OASIS/CMIS: Content Management Interoperability Standard (for systems and integration)

OAI-PMH & OAI-ORE (cross-search/xfer)

Requirements \ Formats
1. Platform independent
2. Flexible packaging
3. Supports update packages
4. Standardised
5. Open
6. Easily understandable
7. Widely used in bit repositories
8. Tools available
9. Include files unchanged
10. Identifiers for files

SNIA/SIRF (open encapsulation format for objects)

SNIA/SMI-S (Storage Management Initiative)

EDRM: Electronic Document Reference Model (for cost saving on integrations, evidence production and preservation)
TOTAL COST OF OWNERSHIP

Storage TCO
• Acquisition cost
• Subscription cost
• Integration cost
• Increased storage
• Expertise & Mgmt
• Access/Dwnld cost
• Refresh cost
• Redundancy cost
• Recovery cost
• Bandwidth cost
• Computing cost
• Handling cost
• Facilities cost
• Security cost
• Fatigue/frustration

International Conference on Massive Storage Systems and Technology / Graph on left is borrowed from a presentation by David Merrill, Hitachi Data System: http://storageconference.us/2014/Presentations/Panel1.Wood.pdf
WHICH CONTENT TYPE(S) FOR WHICH STORAGE(S)?

Storage suggestion based on budget, expertise, content use and three data sizes*

<table>
<thead>
<tr>
<th>Size of data collection</th>
<th>MB</th>
<th>GB</th>
<th>TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommended storage</td>
<td>HDD, LTO tape, cloud</td>
<td>HDD, LTO tape, SSD</td>
<td>HDD, LTO tape, SSD</td>
</tr>
<tr>
<td>Size of budget</td>
<td>Small</td>
<td>Medium</td>
<td>Large</td>
</tr>
<tr>
<td>Recommended storage</td>
<td>HDD, LTO tape, cloud</td>
<td>HDD, LTO tape</td>
<td>HDD, LTO tape, SSD</td>
</tr>
<tr>
<td>Internal IT staff</td>
<td>0</td>
<td>1-5</td>
<td>5&lt;</td>
</tr>
<tr>
<td>Recommended storage</td>
<td>Cloud and remote backup on HDD by external vendor</td>
<td>LTO tape or HDD server systems</td>
<td>LTO tape, HDD and SSD server systems</td>
</tr>
<tr>
<td>Use of data</td>
<td>Daily access</td>
<td>Backup (refreshed)</td>
<td>Preservation archive</td>
</tr>
<tr>
<td>Recommended storage</td>
<td>HDD, SDD, cloud</td>
<td>LTO tape, HDD, SDD, cloud</td>
<td>LTO tape, HDD</td>
</tr>
</tbody>
</table>

STORAGE TRENDS*  
(DENSITY INCREASE, OPTICAL RE-FOCUS, FB USE CASE, AND AWS CLOUD GROWTH)

*Trends from 2 presentation events below (in addition, each image has its original hyperlink)
RECAP : THINGS THAT MATTER

Business Use Case and Interoperability Frameworks

Storage TCO
- Acquisition cost
- Subscription cost
- Integration cost
- Increased storage
- Expertise & Mgmt
- Access/Dwnld cost
- Refresh cost
- Redundancy cost
- Recovery cost
- Bandwidth cost
- Computing cost
- Handling cost
- Facilities cost
- Security cost
- Fatigue/frustration

Implementation
- Architectural options,
- Locality,
- Likelihood,
- Records Control

Content allocation
- Appraisal / Use of data
- Nature (S/A/M/D-IP)
- Birth means
- Uniqueness
- Access Status (Class.)
- Frequency of request
- Size of data collection
- Size of budget
PANEL DISCUSSION
DISCUSSION PANELISTS

- Glen McAninch
  - Manager of Technology Analysis and Support Branch (TAS)
  - Kentucky Department for Libraries and Archives
- Allen Ramsey
  - Assistant State Archivist
  - Connecticut State Library
- Mike Strom
  - State Archivist
  - Wyoming State Archives
- Matt Veatch and Matt Powell
  - State Archivist and Application Developer II
  - Kansas Historical Society
QUESTION #1

Describe your state’s storage situation
QUESTION #2

How many copies are you currently storing and how do you maintain and manage them?
QUESTION #3

Do access considerations impact your storage decisions, and if so, how?
QUESTION #4

Where do you see storage trending in your state?
QUESTIONS & COMMENTS

It's QUESTION TIME!!
CONTACT INFORMATION

Tibaut Houzanme
- THouzanme@icpr.IN.gov

Veronica Martzahl
- veronica.martzahl@sec.state.ma.us

Glen McAninch
- Glen.McAninch@ky.gov

Allen Ramsey
- Allen.Ramsey@ct.gov

Mike Strom
- michael.strom@wyo.gov

Matt Veatch
- mveatch@kshs.org
WRAP-UP

- Post-webinar evaluation will automatically open in your web browser when you exit the session.

- Complete webinar schedule is available on CoSA’s website: http://www.statearchivists.org/CoSA_Webinars.htm

- All webinar slides available from the SERI webinar page: http://www.statearchivists.org/seri/STEP/SERI_Educational_Webinars.htm