

CoSA & Preservica Practical Digital Preservation 2016/17



Digital Preservation Storage Choices

Benefits and Considerations for State Archives

Tuesday December 13, 2016

Practical Digital Preservation 2016/17

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)

- **Digital Preservation Storage Choices**
- Connecting Digital Preservation with Catalog Systems [CANCELLED]
- Preserving and Protecting Audio-visual Files
- Preserving Digitized State Government Records
- Best Practices in Digital Preservation: International Perspective

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



Sarah Grimm
Wisconsin Historical Society

December 13

January 10
April 11
May 9
June 13

Sign up today on the CoSA website – PERTTS Portal > Education -Training

Agenda

- Introduction
- Core requirements of a Digital Preservation System
- Checklist for selecting Cloud Storage for Digital Preservation
- Speaking the Language of IT: Veronica Martzahl
- Selecting and Negotiating Storage Solutions: Jim Corridan
- Storage and IT Analysis: Elizabeth Perkes
- Summary and Questions



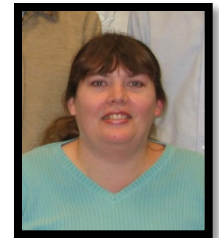
David Portman
Preservica



Veronica Martzahl
Massachusetts Archives

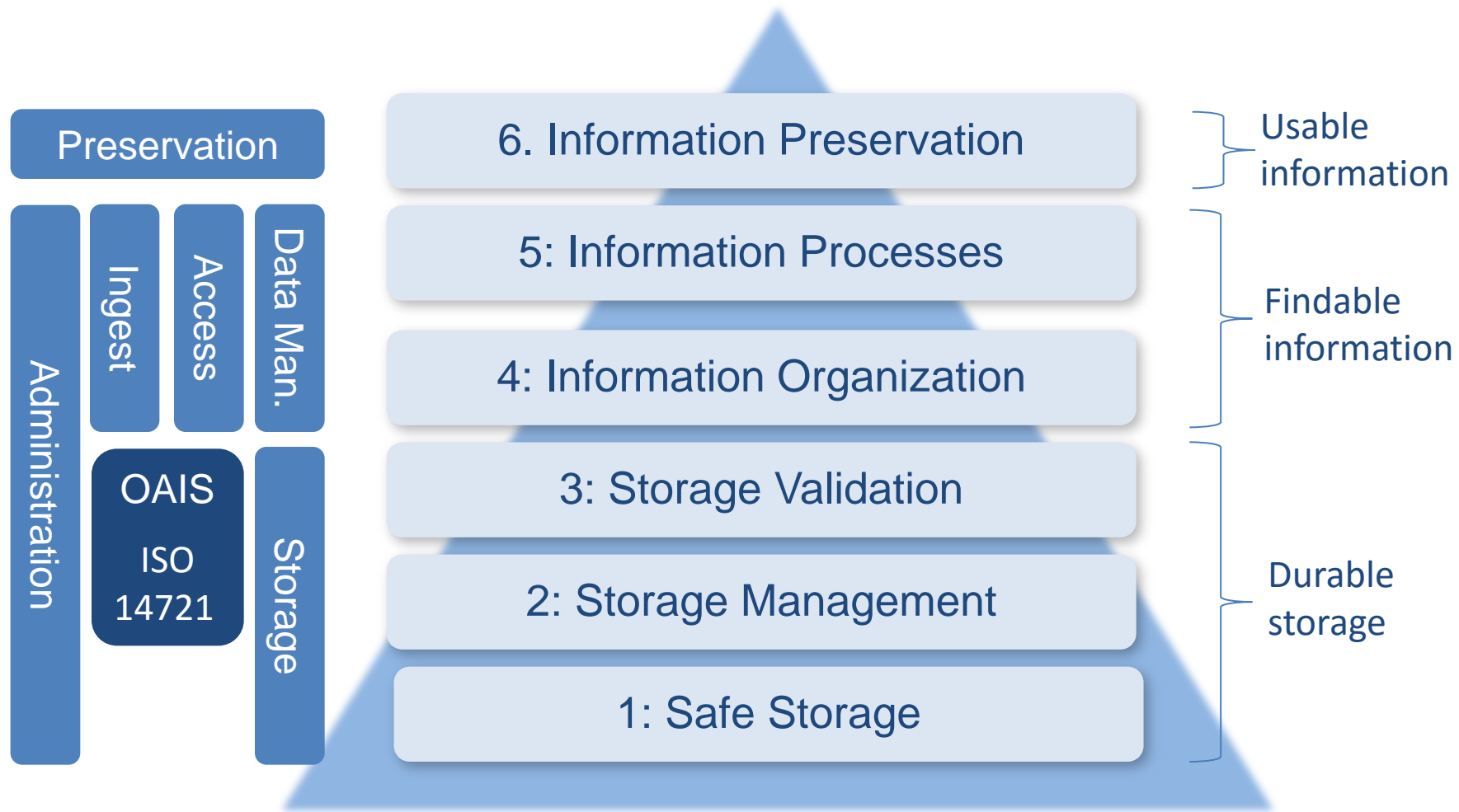


Jim Corridan
Indiana State Archives

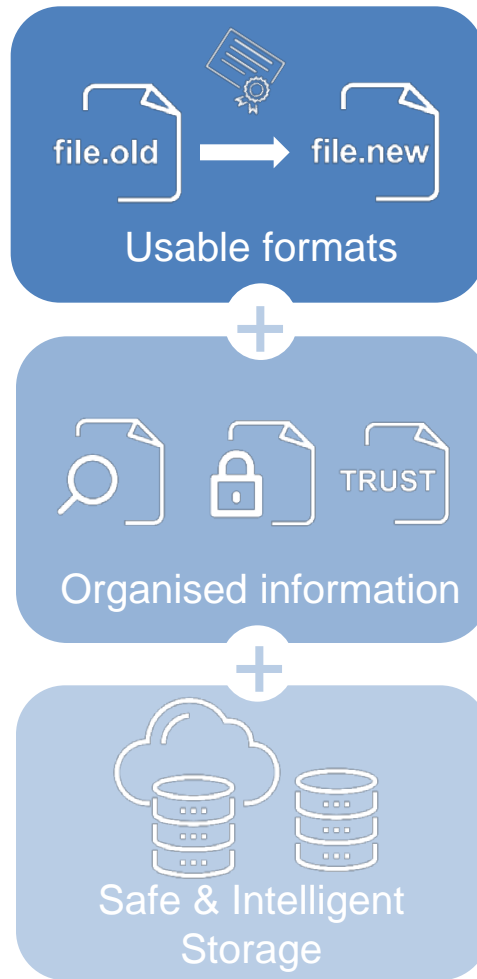


Elizabeth Perkes
Utah State Archives

Digital Preservation System : Requirements



Digital Preservation Technology



Active and automated migration to newer file formats and technologies

Flexibility to re-arrange your archive or collection post ingest

*Storage flexibility, intelligence and choice = optimized costs

*The big cloud suppliers – such as AWS, Microsoft, Google and IBM – are locked in a battle to the bottom on price, which is unlikely to let up any time soon. **Research Report by RBC Capital Markets January 2016**

SOFTWARE

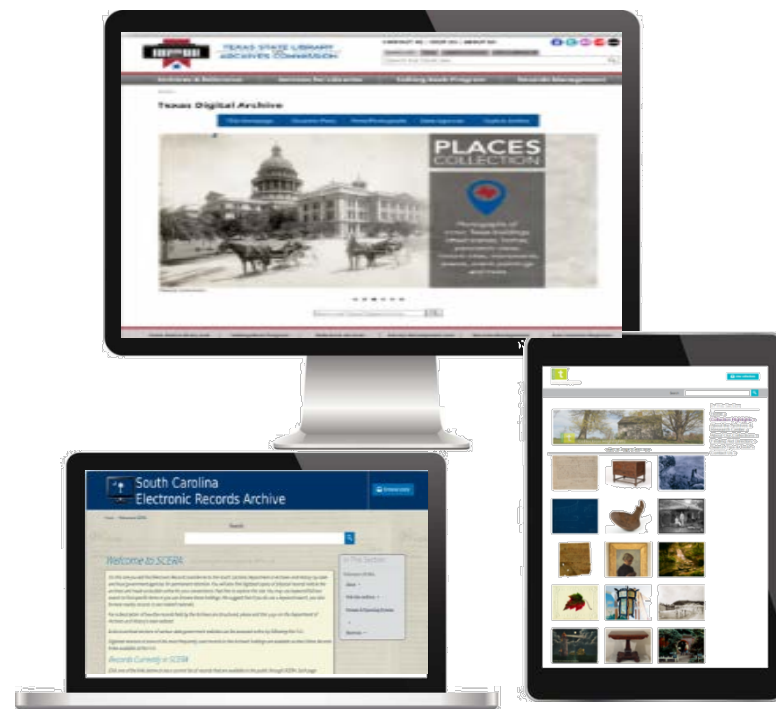
ADDED
VALUE

STORAGE

Delivering Value with Public Access

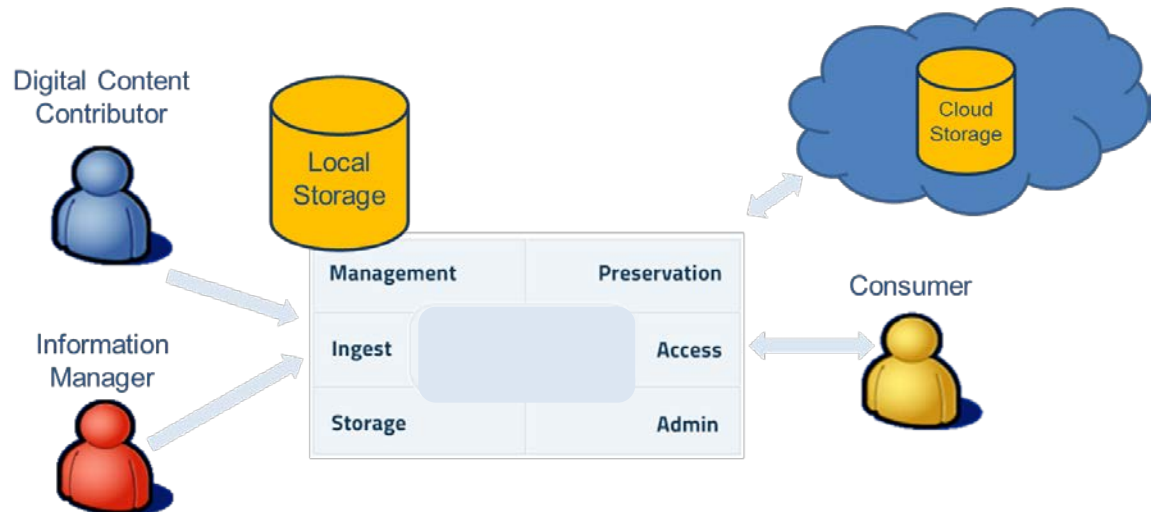


- **Transparency** via Access
 - Internal & external audiences
- **Connectivity** with other systems
 - Catalog synchronization
 - Automated bulk ingest
 - ECM, Email, Web, Digitization
- Need **flexibility** to adapt to changing needs and stay relevant

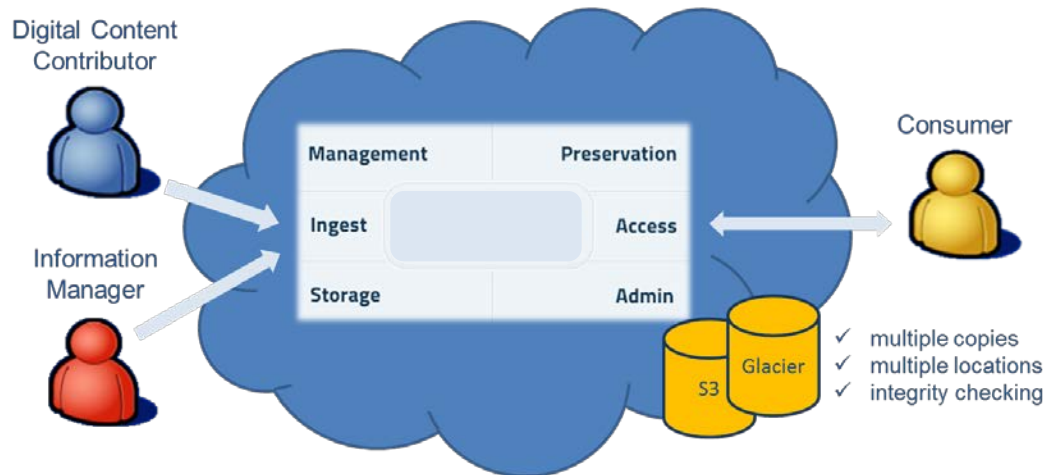


Storage Deployment: Examples

hybrid: on premise
& cloud







fully cloud hosted



Cloud Storage is Ideal for Digital Preservation

	Cloud Storage Considerations
✓	Affordable and Durable
✓	Intelligent Storage Choice for different content types
✓	Secure (ISO 27001) but also User Access Control
✓	Compliance
✓	Exit – no penalty to exit
✓	Ingesting large files and collections

AWS Compliance Example <https://aws.amazon.com/compliance/>

 AWS re:Invent Products Solutions Pricing Software Support Customers Partners Enterprises Startups More English			
<div>    </div>			
Certifications / Attestations		Laws, Regulations, and Privacy	
DoD SRG		DNB [Netherlands]	
FedRAMP		EAR	
FIPS		EU Model Clauses	
IRAP		FERPA	
ISO 9001		GLBA	
ISO 27001		HIPAA	
ISO 27017		HITECH	
ISO 27018		IRS 1075	
MLPS Level 3		ITAR	
MTCS		My Number Act [Japan]	
PCI DSS Level 1		U.K. DPA - 1988	
SEC Rule 17-a-4(f)		VPAT / Section 508	
SOC 1		EU Data Protection Directive	
SOC 2		Privacy Act [Australia]	
SOC 3		Privacy Act [New Zealand]	
UK Cyber Essentials Plus		PDPA - 2010 [Malaysia]	
		PDPA - 2012 [Singapore]	
		Spanish DPA Authorization	
		CIS	
		CJIS	
		CLIA	
		CMS EDGE	
		CMSR	
		CSA	
		EU-US Privacy Shield	
		FISC	
		FISMA	
		G-Cloud	
		GxP (FDA CFR 21 Part 11)	
		ICREA	
		IT Grundschutz	
		MITA 3.0	
		MPAA	
		NIST	
		PHR	
		Uptime Institute Tiers	
		UK Cloud Security Principles	

Speaking the Language of IT

An introduction to the terminology of data storage

Veronica Martzahl
Digital Records Archivist
Massachusetts Archives

Tape Libraries

DAS

NAS

RAID

Server

Back-ups

Cloud

SAN

UNIFIED STORAGE

Virtual Machine

Servers – not just for storage

- Application
- Web
- Database
- Print
- File
- Media
- Mail (Exchange)

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
- Information can not be accessed without a direct connection

NAS – Network Attached Storage

- Server storage that is dedicated to filing sharing
- Storage-centric design that is drawn on by other servers

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

Tape Drives/Libraries

- Write and Read on magnetic tape
- Typically off-line storage

Online vs Nearline vs Offline

- **Online:** Online usually means “connected.” In the storage arena, online means files that are immediately available. If you try to access them, they are there (depending on the speed of your connections).
- **Nearline:** As the word itself implies, “nearline” refers to files that are stored nearby. They are not immediately accessible, but they almost are. It takes a little effort to obtain the files, but you don’t have to wait for long. When a system uses nearline storage, infrequently used data is moved out of the main network into a secondary system. The secondary system is referred to as nearline.
- **Offline:** Offline means that the data is totally disconnected from the network. This is the data that would require the most time to recover, since physical measures may be necessary to get the data to the restore location. However, it is the safest data. Since it is totally disconnected from the network, it cannot be corrupted or accessed by hackers.
Offline storage allows you to store the data at locations separate from those that house the main servers. Tapes may even be stored by an offline data storage company, which safeguards tapes in its own secure warehouses. If physical disaster strikes the main office, the data is still safe if the tapes have been stored at other sites.
- <http://www.bench3.com/2010/03/different-types-of-data-storage.html> Accessed 2016-12-06

But we back everything up . . .

- Backup are part of a preservation program NOT all of it
- They are meant for disaster recovery
- They do not support the level of detail and attention that we need to pay to digital objects for their preservation

Don't forget about geographic distribution of your records!

Virtualized Storage

- Pulling together storage space from multiple storage devices so that they act and are managed like a single device

Learning about storage has helped me

- ask the right questions of software vendors
- anticipate issues of access speed and cost over time
- develop your preservation policy and plan
- talk with your IT staff to ask for what you need

A decorative graphic on the left side of the slide consists of a network of light blue lines and small circles, resembling a circuit board or a stylized tree structure, extending from the top to the bottom of the frame.

SELECTING AND NEGOTIATING STORAGE SOLUTIONS

JIM CORRIDAN

INDIANA ARCHIVES AND RECORDS ADMINISTRATION

INDIANA'S PROCESS - CURRENT & FUTURE NEEDS

- How much storage are we using?
- How much storage growth did we anticipate in various timeframes?
- Are the records surrogates or originals?
- Are the records confidential or public?

CONSIDERATIONS

Content Classification (Open, Confidential)

Content Uniqueness (Surrogates, Born Digital, Access copies)

User Need (High, Low, Negligible Demand)

Connectivity (Fast, Slow, Disconnected)

Security

CLOUD STORAGE SOLUTION

Amazon Glacier	\$.004 per gigabyte per month
Google Nearline	\$.01 per gigabyte per month
Archive Oracle Cloud	\$.001 per gigabyte per month

- All in, Indiana expects to spend about \$12,000 per year for 1 petabyte of dark storage with at least two cloud copies.
- Downloading one PB from Amazon would cost \$5,140.

QUESTIONS?

Thanks

Jim Corridan

Indiana Archives & Records Admin

jcorridan@iara.in.gov

317-232-3380



Storage and IT Analysis

DECEMBER 13, 2016

ELIZABETH PERKES

UTAH STATE ARCHIVES

Storage History

Data Storage was the Unsolvable Problem

- No funding
- High state costs
- Centralized IT, no place for unique needs
- Overhead for IT department (DTS) bundled with raw storage costs

Multiple Attempts to Solve Problem

- Begged DTS to reduce their rate
- DTS didn't understand our needs
- Tried to convince our department to support a building block
- Funding from Legislature to study the issue (\$100,000)
- Report to Legislature delivered right when 2008 recession hit

Moving Forward

Implemented Temporary Solutions

- M-disc to store permanent records, especially SIPs
- Portable hard drives used to store scanned images, and process electronic holdings

Problems with Temporary Solutions

- Hard drives failed regularly
- People would carry a hard drive from one office to another to share data
- We kept getting more and more data, with no place to put it other than hard drives
- M-discs are slow to write to, and media obsolescence is still a risk, although stable

IT Project Manager

Open Records Portal

- 2013 Open Records Portal Law openrecords.utah.gov
- Bill included data storage

IT Project Manager - 2014

- DTS to access a project manager (50 hours, \$3,700)
- Project manager provided a formal agreement of what was to be accomplished, and an evaluation at the end.
- End result was that DTS understood what we needed

Storage Project

Met with Project Manager Frequently

Identified Six Primary Storage Needs

- Quarantine store
- Local processing store
- Preservation data store
- Access copy store
- Migration space store
- Transitory digital image store

Storage Results

Local NAS Device Installed Onsite

- 10 TB for Local Processing Store
- 10 TB for Transitory Digital Image Store
- Tape Library for additional copies of Preservation Store AIPs

Forensic Workstation Set Up

Future Options

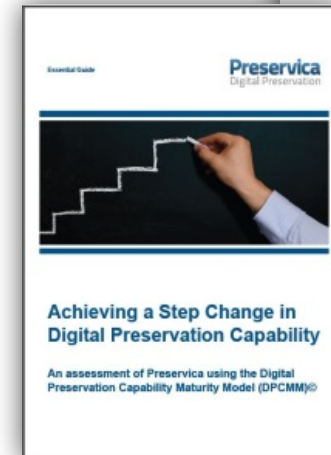
Summary : Digital Preservation in Cloud

- Make intelligent Storage choices – to optimise costs
- Flexibility and choice are key
- Understand the terms and cost to exit
- Preservation *with* Access is vital for long term sustainability
- Proven – the Cloud and Cloud storage are already being used for Digital

Preservation at a number of State Archives

Next Steps...

- **Next PDP How-To Session:** January 10, 2017 **[CANCELLED]**
 - Connecting Digital Preservation with Catalog Systems
- **What's on Tap for CoSA in 2017:** January 26, 2017
 - <https://www.statearchivists.org/programs/cosa-webinar-series/>
- **Next LIVE Demo** of Preservica v5.7 – December 15, 2017
 - Register at www.preservica.com
- **Discover** the Preservica Resources Center
<http://preservica.com/resources/>



Questions ?



Thank you!

www.preservica.com

info@preservica.com

@preservica

@dPreservation

www.statearchivists.org/