

# Library of Congress Designing Storage Architectures

March 24, 2025

Nina Smith  
Director, Digital Archive Management Operations



# Designing Storage Architectures Agenda

1	Who We Are
2	It's Not Just Storage
3	Challenges and Opportunities
4	What's Next

# Who We Are

The current collection of brands represents over 100 years of content.

That's a lot of assets to preserve.



# It's Not Just Storage

Costs to run a Digital Preservation Vault:

## People

Digital Archive Staff  
Quality Assurance and Control  
Engineer(s)  
Coding Specialist(s)  
Project Management  
Technical Metadata Quality Control  
Technical Support  
R&D

## Storage

Cloud (S3 tiering including Glacier Deep Archive)

## File Transfer

Aspera on Cloud  
Data Validation

# It's Not Just Storage continued

## Disaster Recovery (2 different systems)

- Second copy (minimum)
- Geographic Diversity

## Vendor Management

## Compute/AWS Management

- Lambda
- Kubernetes
- EC2
- S3
- GDA

## Documentation

- Runbook
- Search Query Language
- Logs and Reports
- File Naming Convention
- Taxonomies and Ontologies
- Unique Individual Asset Identification

## Security

- Information Technology
- Malware Protection

# It's Not Just Storage continued

## Software

- Enterprise and Individual e.g.:  
Microsoft Office Suite; Zoom;  
MediaPulse; Smartsheet; Title  
Management System; Confluence;  
Slack; Text Editor with XML  
capabilities; DPX Frame(s) Viewer;  
Hash Algorithm Generator; Proxy  
Viewer

## Software

- Media Asset Management System  
(includes development area, staging  
area and production area)
- FFMPEG Proxy Transcoder
- Data Aggregation Dashboards

# Challenges and Opportunities

## Challenges

- File size – 8K and beyond
- Media Asset Management system limitations
- Lower storage cost
- Format obsolescence → Digital and Physical
- New Media (Silica Glass, DNA, Archival Millennial Disc)

## Opportunities

- AI
- Time code matching from closed caption file to audio/video
- Automated processes
- Lossless Compression

# What's Next

## Third Copy Options

- Multi cloud
- Air-gapped system
- Secure 'bunker' data storage



# Thank you