



Designing a Cloud-Managed Tape Architecture for Long-Term Digital Preservation

Technical Talk | March 9-10, 2026 | Washington, DC

Nelson Nahum

CEO

nelson@geyserdata.com

Beyond Basic Storage: The Pillars of Preservation

What True Preservation Storage Entails

Preservation storage requires more than just data capacity.



Ingestion & Content Indexing

High Performance ingestion

AI Content Indexing

Proving content integrity over decades.



Migration & Refresh

Evolving media formats without breaking data references.



Security

Protecting against ransomware, insider threats, and operator errors.



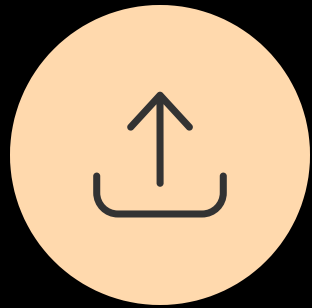
Sustainability

Ensuring operational longevity and energy efficiency.

THESE ARE THE KEY CONSIDERATIONS FOR EFFECTIVE PRESERVATION STORAGE.

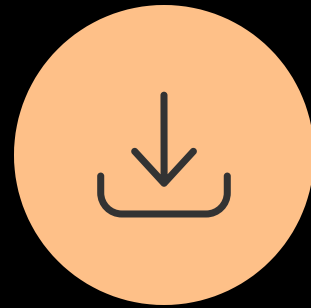
Workload Reality for Long-Term Digital Collections

Understanding the specific demands is crucial



Ingest Operations

Primarily write-heavy operations.



Read Patterns

Characterized by spiky and unpredictable read requests.



Object Characteristics

Large objects with long retention periods and occasional bulk recalls.



Key Priorities

Data certainty, throughput, and integrity evidence are more critical than time-to-first-byte.

The primary goal is to make data correctness a consistently boring operational aspect over 10–50 years.

Core Design Principles (Vendor-Neutral)

Ensuring flexibility and robustness through strategic design.



Decouple Interface from Media

Applications should interact with an abstract interface, not directly with the physical tape media.



Separate Control and Data Planes

Distinguish between policy and evidence management (control plane) and raw byte storage (data plane).



Immutable Inventories & Append-Only Logs

Utilize systems that maintain tamper-proof inventories and event logs for integrity.



Engineer for Predictability

Establish Service Level Objectives (SLOs), implement rate controls, and ensure repeatable operations for consistent performance.

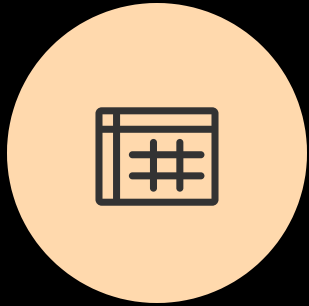
Reference Architecture: S3-Accessible 'Cold Bucket' Backed by Managed Tape

Combining modern accessibility with tape's durability

- **S3 API Front Door**
Provides compatibility for existing workflows and tools.
- **AI Content Indexing**
Acts as a quick locator for relevant data over the years
- **Policy/Orchestration**
Manages data placement, retention, validation, and restore queues.
- **Staging Cache (Disk)**
Buffers burst reads and facilitates efficient restores.
- **Tape Tier**
Provides durable, low-energy, long-horizon storage.

Ingest & Fixity Pipeline: Making Integrity Verifiable

Ensuring data integrity throughout the data lifecycle.



Ingest & Indexing

Establishes foundations for integrity verification, including checksums, manifests, and object identity. Use AI to Index the content of the data as it is ingested



Two-Phase Commit

A critical transaction protocol ensuring reliable data transitions from 'registered' to 'protected on tape' with guaranteed atomicity.



Periodic Audits

Ongoing verification processes including scheduled sampling and defined exception workflows for detected discrepancies.



Media Refresh

Proactive, policy-driven management of storage media for automated activity based on media health and age.

Retrieval & Operations Model

Predictable, Not Heroic

1

Async Restores

Utilizes job-based semantics for efficient bulk data recall.

2

Staging Cache

Absorbs sudden spikes in hot data demand, reducing repeated tape mounts.

3

Rate Controls

Actively manages system load to protect essential tape throughput.

4

Operational SLOs

Establishes clear, measurable targets for restore start time, throughput, and completion certainty.

Security, Isolation, and Resilience

Essential Measures for Long-Term Preservation



- **End-to-End Encryption**
Secures data throughout its entire lifecycle.
- **Immutable Retention Modes**
Implements WORM (Write Once, Read Many) policies where necessary.
- **Private Isolation Boundaries**
Ensures logical and operational separation of data and systems.
- **'Air-Gap Properties'**
Leverages offline media for enhanced ransomware resilience.
- **Audit Trails**
Records every lifecycle event, detailing who performed what action, and when.

Takeaways for DSA Practitioners

Implementing Effective Digital Preservation Strategies



Coexistence

Achieve cloud usability, tape economics, and preservation-grade evidence simultaneously.



Correctness

Ensure correctness through inventory management, policy enforcement, and routine audits, not just one-time validation.



Design for Decades

Plan for migration paths, comprehensive observability, and operational simplicity for long-term preservation.



Discussion Points

Key areas for discussion include fixity at scale, inventory schemas, restore SLOs, and automation opportunities.

Key takeaways emphasize achieving multiple goals simultaneously, ensuring ongoing correctness, designing for the long term, and focusing discussions on critical areas for effective digital preservation.