



Technology Module: Storage 101

Dorothea Salo

This POWRR Institute is generously funded by the



Expected Outcomes

- ✓ **Understand the necessity for periodic storage refreshes**
- ✓ **Understand the need for reliable, auditable storage**
- ✓ **Know how to estimate storage needs and costs**
- ✓ **Understand and work around risks of common storage types**
- ✓ **Know about added requirements for preservation-quality storage**
- ✓ **Know about consortial preservation-storage options**

Storage 101 Overview

- **Without storage, there is no digital preservation!**
- **Storage should be planned and budgeted for in cycles.**
- **Storage/backup goal: 3 copies, 2 types of media, 1 offsite**
- **Preservation storage should be reliable and auditable.**
- **Preservation storage entails requirements beyond backups.**
- **Explaining preservation storage to IT can be difficult.**
- **Preservation storage can be accomplished consortially.**

Like File Formats...

- ... the storage ecosystem changes constantly.
- Ergo we can't give you the One True Storage Solution for All Your Needs.
It will have changed by next Wednesday!
- So we'll give you an evaluative framework and suggestions.
Thinking About Storage 101.

Baseline Truth

- Digital preservation is a **relay race**, not a marathon.
- You're not looking for the Holy Grail storage medium that fits every situation, never fails, and lasts forever.
*Not least because **it does not exist and never will.***
- You're looking to “refresh” your storage every so often.
And you need to budget equipment money and staff time accordingly.
Silver lining: Storage tends to lessen in cost over time.
(For now, anyway.)

Regarding the Holy Grail...

- The everlasting digital storage medium?
- It's been tried. Every time, it didn't capture enough market share to survive **economically**.
- Eventually its equipment stopped existing. So much for the Holy Grail.
- Anybody promising this is selling snake oil. Do not buy!

CDs/DVDs vs Change

- Fewer computers have CD/DVD drives built-in.
- External drives are still easy to purchase, but how long will that last?
- What we think of as “USB” is on the way out!
- If you’re relying on CDs/DVDs, it’s time to work toward a storage refresh.

3-2-1: The Backup Mnemonic

THREE copies

on at least **TWO** different types of storage
medium

Examples of different media: a desktop/laptop, an external hard drive, a server, a tape backup system, cloud storage

with **ONE** offsite

Why offsite? So something that wipes out your building (fire, tornado, flood) doesn't wipe out every copy of your data.

Take a Moment:

How could you implement

3-2-1

where you are?

Backups

Reliability and **auditability** requirements are central.

- A backup system that doesn't work might as well not exist. **AUDIT IT:** yearly spot-checks are the absolute minimum acceptable.
- A backup system so complicated or cumbersome that nobody uses it is also useless.
- Speed of access? Unimportant.

Try to use a **different** type of storage medium for backups than for working data.

- Medium diversity reduces risk of data loss.
- Magnetic tape? Fine.
- IT's server? Sure.
- An external hard drive? Sure, if you'll actually use it.
- Cloud? Fine, with caveats (we'll talk about them).

Auditability vs. CDs/DVDs

- You will see LIS literature from the 2000s touting gold CDs. Ignore it.
- Everything they say about gold-CD reliability could be true, and they would **STILL** be a bad choice.
- The reason: auditing them takes much too much time and effort.
So when (!) one fails, nobody will know.
- Choose media that are computer-auditable with minimal human effort.

How Much Storage is Enough?

Part of the answer is time horizon.

Since digital curation is a relay race, you'll be rethinking the storage issue in a few years anyway. So don't think beyond that rethinking! You'll overspend!

Back-of-the-envelope calculation

- What you have + (how many items you believe you will have * amount of space per item)*3
- The *3 accounts for 3-2-1 storage.
- Then add some extra.
- If you have IT, ask if there are ways to reduce this burden. There might be! But wear your skeptic hat. ("Get rid of all those giant TIFFs; all you need is JPEG, right?")

“The Cloud” Has Risks!

Business risks

What if they go out of business, or get sued and have their servers confiscated? (*MegaUpload!*)

Security risks

Theirs AND Yours

(Yours: Not being careful enough with access permissions, such that sensitive data leaks)

Cost risks

This doesn't mean not to use cloud storage. Use it!

It means don't use it as your only backup. 3-2-1!

Cloud Storage Costs

Once you get above free tiers...

...you pay per GB/TB stored...

... but you **ALSO** pay for transfer bandwidth on the network.

This cost can bite you! Hard!

Use paid cloud storage only for data you will only access to audit data or retrieve lost data.

Preservation Storage

Baseline Truth

- **It's not the storage medium that counts.**
- **It's the extra work layered on top of the storage.**
- **That work includes:**
 - Geographic replication (like offsite storage, but across the country or the world, not just across town)
 - File-format assessment and (as needed) migration
 - Content auditing, also known as “fixity checking” (not just “does the drive work?” but “is every file still intact?”)
 - Logging things that happen to the data (“preservation events”) such as uploads, audits, changes, file-format migrations, deletions
 - Appropriate security and access controls (up to and including “dark archives” for material that cannot be made public)
- **It's not just backups!**

Explaining This To IT

- You may lose this fight. Repeatedly. Many have.
- Try “what if” scenarios, aimed at whatever IT is not considering. Back these up with real-life horror stories.
- Use the NDSA Levels of Preservation or “Preservation Storage Criteria” developed by the Library of Congress et al. as starting requirements lists.
PSC version 2 currently on Google Docs: <https://goo.gl/1Q9vDe>
“Requirements” often a language IT will understand.
- Consider not going it alone! Agitate for larger-than-local solutions.
Keep an eye on IMLS’s “National Digital Platform.”
- In the meantime, keep the relay race going. Your goal is for your data to exist still when a preservation solution arrives!

IT Offerings You Might Consider

Server storage and backups

As one of your 3-2-1 copies, IT's offerings may be just fine.

Ask them about fixity checking, content auditing, event logging!

Tape storage, tape library

Terrible for live data. Great for a 3-2-1 backup!

Ask about auditing and what happens to old tapes when IT refreshes its tape machines.

Data “parking” or “archiving”

This usually means storage for material that IT's clients won't be accessing very often. Can mean a tape library.

The cost per GB/TB will be less than for “live” storage.

The moment you stop paying, IT deletes your data. (So no, it's not a real archive.)

Consortial Preservation Storage

You don't have to go this alone. You probably shouldn't!

➤ **One option: “Private LOCKSS Network” (PLN)**

LOCKSS: Lots of Copies Keeps Stuff Safe

A Linux-based software platform for replicating and auditing digital content across (often geographically-dispersed) servers

All parties need to be able to run LOCKSS in-house!

➤ **Example: MetaArchive Cooperative, for any digital content needing preservation: *metaarchive.org***

Either run LOCKSS in-house OR pay yearly fee

➤ **Other options: Digital Preservation Network, Duraspace**

Membership-based systems with yearly fees

Act as broker between institutions and cloud-storage providers to ensure preservation-quality service-level agreements, vetted legalese

An Option For Many: The Internet Archive

If what you have is free of copyright-related restrictions...

... and can be shared openly with the world...

(that is, no donor restrictions or publicity concerns)

... is a book, image, audio file, video file, or website...

... the Internet Archive will take it. **Free.**

But Remember 3-2-1!

- Don't let the Internet Archive be your only copy.
They didn't implement geographic replication until late 2016!
- Keep two more copies, just in case.



Technology Module: Storage 101

QUESTIONS?