DIY Digital Preservation Tips for Independent Producers

Linda Tadic, Founder/CEO
Digital Bedrock
Email: Ltadic@digitalbedrock.com

IDA Workshop: Are Your Films Future-Proofed?
September 12, 2025
IDA office and on Zoom

What are *your* practices?

Q1. What element or file types do you save for future use?

- Video master (ProRes, MKV, other)
- Picture sequences (DPX, TIF, EXR)
- Audio masters (stems, MnE, printmasters)
- Audio mix
- DCP
- Camera & sound originals

Q2. How much total storage are you managing?

- less than 10 TB
- between 10-50 TB
- between 50-100 TB
- between 100-500 TB
- over 500 TB

Q3. How do you store your digital files?

- Hard drives
- SSDs
- LTO
- CD, DVD, BluRay
- Cloud hyperscalers (AWS, Google cloud, Azure, Wasabi, Backblaze)
- Cloud document storage (Google Drive, Dropbox, Box)

Q4. How many copies do you have on different storage media/platforms?

- 1
- 2
- 3
- 4+

Q5. Do you make checksums of your files?

- yes
- no
- sometimes
- what's a checksum?

Q6. How often to you check that your files are OK?

- Once a year
- More than once a year
- Once every few years
- Never

Digital Preservation Actions

1. Appraisal (identify what you NEED for future use)
2. Organize:
    inventory
    verify/make checksums
    create path and file naming conventions
    organize files with new structures
3. Catalog (metadata). Verify formats and technical characteristics of files
4. Store for long-term preservation
5. Manage over time (fixity checks, media migration)

Common problems in preserving independent films

1. Files are dispersed. Where are they?

- filmmaker/editor/producer home, studio, garage
- storage facility
- post house

2. Version control. Filenames, folder structure, labels on media not clear as to the content or version

Common problems in preserving independent films

3. **Files incomplete**, corrupted, mis-named

4. **Media failure** and no redundancy

    - Or, the redundant copies had errors in copying that weren't caught at the time of writing

Real-life problems case studies

Naming problems:

- DPX containers (folders) with only the sequence number as the filename (no prefix). Eg, 153810.dpx, 153811.dpx and the folder named “backup2”
- Duplicate filenames, but the checksums differ (what’s the difference between the files?)

Missing audio stems

You’ll want to save these for future re-mixes and restorations

More problems with DPX files

Errors in writing to storage media:

- Missing DPX files in a sequence (copy failed at some point)
- Scattered corrupt DPX files in a container

md5 checksum files were created with these errors, nobody verified post-writing, and the media was sent to storage with errors

Media failure and media obsolescence

- Unique content (and the sole copy) on LTO3, written in an unknown backup software
- Hard drives crashing
- Future problem: files are on Firewire drives, but FW to Thunderbolt (or other connectors) adapters are no longer made/available

DIY digital preservation: File organization is key!

First **list** what you should keep, then go looking for it.

Make a template that can be applied to all similar projects.

1. Appraisal: What do you need to keep?

- Consider what you'd need to restore/regrade/re-mix the work in the future (master elements, audio stems, effects, graphics, etc.)

- Can any content be re-purposed?

  - documentary productions often have interviews, b-roll that could be separately licensed

You don't have to keep everything, only what you **need**.

Missing Movies checklist

https://missingmovies.org/preservation-faq-for-filmmakers/

PICTURE

(Note: Sequential Picture Frames can be DPX, TIF, or EXR)

1. 2K or 4K ungraded Sequential Picture Frames. If scanned from a film original source, the frames should not be restored/cleaned up.
2. 2K or 4K Sequential Picture Frames and/or MOV of restored and/or graded, with final framing
3. 2K or 4K Sequential Picture Frames and/or MOV of restored and/or graded, full aperture
4. 2K, 4K or HD mezzanine MOV/MP4 files with specs
5. 2K or 4K DCP – unencrypted only
6. 2K or 4K DCDM (optional)
7. textless elements

Missing Movies checklist

AUDIO

8. Final sound mix
9. All audio/ mix stems
10. M&E mix

Missing Movies checklist

11. DOCUMENTS/PAPERWORK

- A READ ME document detailing any relevant Tech Specs for picture and sound
- Dialogue and Action Continuity script
- Full Chain of Title (COT) documents
- Music Cue Sheet
- Cue Sheets for all third-party licensed elements (photos, video clips)
- Abstract of all Distribution Deals listing the term, territory and any copyright provisions or assignments
- Production Notes and/or Press Kits

Missing Movies checklist

OTHER

13. Scans of Approved Key Stills, Poster
14. Project files for reference (AVID, Resolve or Final Cut files)
15. Documentary Films: full video interviews of all subjects

Discuss preparing these files with your Producer, Post-Production Supervisor and your Post-Production Vendors.

Pause for 2 minutes of questions

2. Organize: Find files, create an inventory of files on original media, and run checksums

- Collect your stuff: Search at your physical storage locations and post houses.
- If on hard drives, network attached storage, or LTO, create a file inventory of what you have on the different media, with the paths and filenames as-is. **Do this while the files are still on their original media.** [see list of inventory tools on the RESOURCES handout]

- Copy the needed files off to new disk media for organizing
- Verify the files you need to keep are OK. If the files already have checksums, verify the checksums after copying to new media.

Sidebar: What's a checksum?

A checksum or "hash" is an algorithm that generates a unique fingerprint for a digital file.

The checksum should be generated when the file is in its final state, and will not be changed. This becomes the master checksum (the fingerprint from birth).

It should be created "in situ" (eg, on the original media where the file is originally stored), or as soon as possible after the file has been moved off original media.

The file's path or filename can be changed later, and the checksum will still match *unless* the user accidentally open the file while changing the filename, and added errant bytes.

Sidebar: What's a checksum?

MD5 is fine for your purposes.

Hash algorithms in order of complexity:

MD5 Checksum: B52220359588B8C9D76C083B5ED58769

SHA-1 Checksum: 45A79CFE2D9930A4537E719EE893A43E0741C001

SHA-256 Checksum: CB29741D42846EB6599BA9B8B2C708255E16D190B54999065FC7D87D5A997875

SHA-512 Checksum:
3AC8F2C4BDA50A698A6AB74599F0B93AF59E21F7ECB840D408801B9A690184B07D48CEC94E6D1D2DBEA59E
86B709919A38A404BA23C78FE353B6F8BEF5624847

Sample free checksum generating tools are on the Resources handout.

You should always verify the checksum whenever a file is delivered, and after it is written to permanent storage. If the checksum doesn't match, something has changed in that file.

2. Organize: Folders and file naming conventions

Don’t think only about what makes sense to you.

You are doing this for the people who will sort through your data long after you are gone.

Make the organization and naming conventions clear and obvious.

2. Organize: Folders and file naming conventions

Create filename and folder **naming conventions** for a new file organization.

Be consistent, but be careful! Don’t change filenames/folders if a system requires it or would be looking for it (eg, audio files that ProTools would be looking for)

Organize the files in clear folders that describe the content (title, b-roll) and file function (versions, log graded master, ProRes masters, etc)

Sidebar: Naming conventions

Establish your own path and filenaming conventions, and follow it for every project. You likely won't have a database for your work, so you'll be finding files based on how they are organized and named. **Always think of how future users will parse through your data.**

Think hierarchically. The most important information to immediately identify an asset's content goes first.

Streaming platforms (Amazon, Apple, Hulu, Netflix) have their own naming conventions you might need to follow.

Sidebar: Naming conventions

Two streamer guides that you can adapt:

Netflix Picture Archival Assets: Folder Structure and File Naming Convention (picture only)

https://partnerhelp.netflixstudios.com/hc/en-us/articles/360000384727-Picture-Archival-Assets-Folder-Structure-and-File-Naming-Convention

Amazon Studios (includes audio as well as picture)

https://portal.amazonstudios.com/hc/en-us/articles/24167202468123-File-Naming-Convention

Sidebar: Naming conventions

Supervising sound editor Larry Blake* created this guide for audio files on the Academy's Digital Preservation Forum website:

https://academydigitalpreservationforum.org/audio-workflow-considerations/

Smithsonian Data Management Best Practices: Naming and Organizing Files (general good naming convention hygiene):

https://www.archives.gov/files/preservation/formats/pdf/naming-and-organizing-files.pdf

**Larry is writing a book on digital preservation for motion picture studios. Email him for more info: swelltone@aol.com*

Sidebar: Naming conventions

The Law: Only use letters, numbers, hyphens, and underscores. No spaces.

You should be thinking about the future, and make **interoperable decisions**. You don't know on which OS and file system your files could ultimately land. Be thoughtful.

Never use illegal characters such as these: ! , ☺ $: ; * [() ℗• ©

"/" can only be used in paths, never in filenames.

Never put a carriage return at the end of a path of filename. The file system in the OS will break up the path and think it's the start of a new file.

2. Organize: Final inventory

- Create a FINAL file inventory of what is being saved for preservation, with the new naming conventions (if names or organization changed).

- Create new MD5 file with new folders and filenames.

- Keep a record (audit trail) of the original paths/names, and what the files were changed to.

Pause for 2 minutes of questions

3. Catalog (metadata)

Catalog the files while organizing them, or after they've been moved to new media in their new organizational structure. This takes the longest time in the preservation process.

What would you need to know about the files in order to use them in the future?

1. Content (title, subject, creator names, shooting locations)
2. Information about the file itself

Metadata about the file's creation

The digital file's creation history to verify authenticity:

- WHO created it (who authorized? Who created -- vendor, artist, in-house?)
- WHAT created it (system, software, hardware, camera)
- WHEN (date created, modified)
- WHERE (geographic location)
- WHY (what's the purpose or function of the file?)

Metadata about the file's technical characteristics

The file's technical characteristics and embedded creation metadata:

- Basics: filesize, format, codec, creation date, compression, bit rate, bit depth, color, frame height & width
- Extended: camera model, UMID, lens, focal length, ……

[MediaInfo is a common tool used to capture technical characteristics. See the RESOURCES handout. If you have the funds, you could license a Media Asset Management system (MAM)]

Metadata to help future preservation

The object's preservation information:

- Original environment and context (filename, directory structure)
- Hash (algorithm, value)
- Validation (does it conform to a specified format)
- Storage media identifier (barcode)
- Storage media location

Pause for 2 minutes of questions

4. Store for long-term preservation

- Make three copies -- ideally on different storage media types/platforms (eg, hard drive, LTO, cloud)
- Run **fixity checks** (verify the checksums) according to a schedule on at least one copy using the new MD5s
- Geographically disperse the copies for disaster recovery.

4. Store for long-term preservation

Every storage media has pros and cons:

- Hard drives (whether desktop single drive, RAID, or NAS) are easy to use. They can fail within 5 years
- LTOs are stable, but should be migrated every 2-3 generations

  You can do your own LTO backup with desktop LTO drives, using software like YoYotta and Hedge. The software can create checksums and inventory your files too.

  Only use LTFS (Linear Tape File System) in writing data to LTOs. It's the open, interchangeable writing file system. NEVER put your files in any proprietary system.
- SSD has maximum writes and can crash without warning. DO NOT USE SSDs for digital preservation storage, period!!!!

Media migration (media refreshing) is unavoidable, no matter what media type.

4. Store for long-term preservation

Cloud storage

- **You don’t have to manage your own storage infrastructure.**
- **You can access the files anywhere in the world.**
- Don’t make it your only copy. Consider hacking, ransomware, AI, natural disasters, terrorist attack.
- Difficult to run fixity checks on files stored with a cloud provider. You have to trust them your files are OK.
- Can be hard and expensive to “exit” a cloud provider. The coldest tier storage is cheap, but to download (egress) is expensive.
- You have to use the cloud provider to get your data. You own the content, but they own the proprietary means to get your data out of their system.
- Consider the environmental impact of cloud storage and computing.

Pause for 2 minutes of questions

5. Manage over time

- Monitor bit health through fixity checks according to a schedule
- Migrate/refresh the media according to a schedule (depends on storage media type, eg HDD, LTO)
- Monitor format obsolescence

Digital preservation budget template

[in Resources packet]

Include preservation in your production budget.

Separated into 3 sections:

A. Metrics (information used to base the cost estimates)
B. Budget
C. Digital Preservation Work You Do

What if you don’t want to DIY?

- **Donate to an appropriate archive.**

  However, archives have limited resources. Have to find the niche where your work would fit their mission.

https://www.loc.gov/programs/national-film-preservation-board/resources/public-research-centers-and-archives/

- **Library of Congress Copyright Registration.**

  Submit a copy of the final work. They will have to keep it. It wouldn’t include your master elements, just a copy, but that is better than nothing.

https://www.copyright.gov/registration/motion-pictures/

Recommended formats: https://www.loc.gov/preservation/resources/rfs/

- **Use a vendor.**

  There are very few vendors who perform some form of digital preservation service (not just storage) and understand how to work with digital film files.

Digital Preservation Actions (re-cap)

1. Appraisal (identify what you NEED for future use)
2. Organize:
    inventory
    verify/make checksums
    create path and file naming conventions
    organize files with new structures
3. Catalog (metadata). Verify formats and technical characteristics of files
4. Store for long-term preservation
5. Manage over time (fixity checks, media migration)

Thank you.

Linda Tadic

Founder/CEO

ltadic@digitalbedrock.com

www.digitalbedrock.com