The Art of Self-Preservation: DIY Digital Preservation Tips for Independent Producers

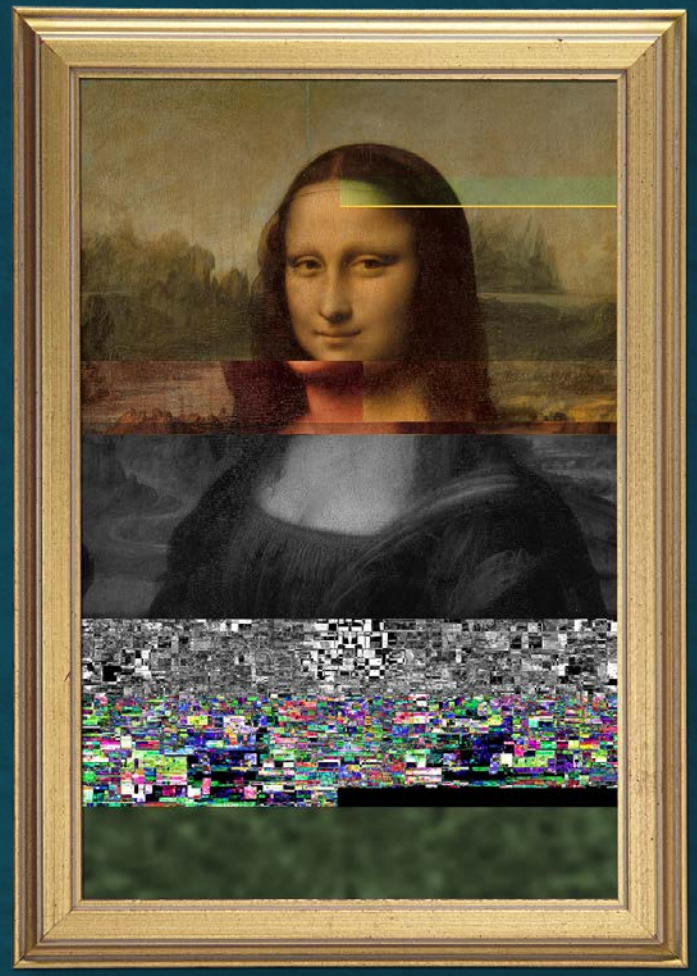
Linda Tadic, Founder/CEO
Digital Bedrock
ltadic@digitalbedrock.com

"Documenting the Present, Preserving the Future" panel
Getting Real 2018
September 25, 2018

Digital Bedrock

- Managed digital preservation services, built specifically for preservation actions.
- Born from decades in the archival moving image community, observing a perceived need – organizations and creators didn't have budgets or staff to build or buy infrastructure to perform digital preservation in-house.
- No license subscription, no hardware – we do the work.
- Any format and content type – not just media.
- No required metadata. Flexible database to support all types of data and metadata
- Affordable – pricing starting at $360 per year (plus one-time processing fee) for 3 copies geographically dispersed, on-going managed preservation, client portal access

What can go wrong?



Keeping digital content usable is a complex and on-going process:

STORAGE IS NOT SECURE.
Preservation is more than just storage – hard drives crash, files are destroyed, lost, or hacked.

DIGITAL BITS DEGRADE AND BECOME UNUSABLE OVER TIME.

FORMATS AND SOFTWARE ARE VULNERABLE TO OBSOLESCENCE.
Software and file relationship dependencies are proprietary, putting future access to digital content at risk.

“Store and ignore” is not an option.

Digital preservation requires active management of:

- The digital entity/entities (binaries/files/essence):
  - fixity (health) checks, redundancy for disaster recovery
- Extensive metadata about the data and its dependencies (format/codec, software, operating environment)
- Storage and migration (migrate storage media and sometimes objects)

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

Problems we've encountered with files

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

We've seen 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.)

- Any content that can 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**.

2. Find and organize

- Search at your physical storage locations and post houses
- Copy the needed files off to new disk media for organizing
- Verify the files you need to keep are OK (run fixity checks)
- Organize them in clear folders that describe the content (title, b-roll) and file function (versions, log graded master, ProRes masters, etc)
- Apply filenaming and folder naming conventions (be consistent, but be careful! Don't change filenames/folders if a system requires it or would be looking for it (eg, raw camera files))
- Create new md5 file with new folders and filenames.

3. Consolidate and store for long-term preservation

- Make three copies -- ideally on different storage media types (eg, hard drive, LTO, SSD, cloud)
- Run fixity checks on the copies using the new md5s
- Geographically disperse the copies
- Every storage media has pros and cons: external hard drives can fail within 3-5 years, LTO must be migrated every 2 generations (but with LTO8, that changed to one generation backwards readability), SSD has maximum writes and can crash without warning). **Media migration (refreshing) is unavoidable, no matter what media type.**
- Cloud storage? Your data on someone else's servers. Don't make it your only copy. Consider hacking, ransomware, lightning strike, how to run fixity checks on files stored with a cloud provider.

4. Catalog (metadata)

Catalog the files while organizing them, or after they've been moved to new media in their new organizational structure.

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, ……

Metadata to help future preservation

The object's original 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

Cataloging/metadata creation tool

PBCore Cataloging Tool

PBCore is a metadata standard developed by the public broadcasting community. A free open source cataloging tool using PBCore will be released by end of October. Watch this space: pbcore.org

Benefit: If you decide to donate your collection to a public archive, your collection will already be cataloged in a standard the archive can use.

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, SSD)
- Monitor format obsolescence

DIY Summary

1. Appraisal (identify what you NEED for future use)
2. Find and organize
3. Consolidate and store for long-term preservation
4. Catalog (metadata)
5. Manage over time

Thank you.

Linda Tadic

Founder/CEO

ltadic@digitalbedrock.com

www.digitalbedrock.com