IMLS Digital Product Form

"The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public."

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes.

1

Keeping with the sustainability theme, the topic for this meeting will be **the long-term storage and preservation of the digital content you're collecting or creating**. Part of your original grant proposal package to IMLS was a Digital Product Form -- question B.2 from the form is here on the slide: *Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes.*

This is a huge question with a complex answer. How many of you are doing exactly what you said you would do on this form? How many have changed your approach? How many of you maybe haven't looked at this form since you wrote your grant?

We want to introduce some language, tools, and strategies to help you address this question. What I'm going to cover today will likely be new information for some of you, totally old news for some of you, or it might be that there's somebody else at your organization that does this work.

What is digital preservation?

*"**Digital preservation** combines policies, strategies and actions to ensure access to digital content, regardless of the challenges of media failure and technological change. **The goal of digital preservation is the accurate rendering of authenticated content over time."***

Working group on Defining Digital Preservation, ALA Annual Conference, 2007

2

Digital preservation is active management of digital content over the long term with access as its ultimate goal.
Digital preservation is the set of activities that minimize loss or damage and protect information. The key here is "over time." However, due to the nature of digital content, we can't expect to set a digital file aside and then open it in 10 years much less 50 without active management. That's why digital preservation is often talked about as a relay race rather than a marathon. Active management in cycles of 3 or 5 years is how we can can continue to pass the baton from one generation of technology to the next.

And again, the unstated outcome here is ACCESS. We're not doing this for kicks, or because the standards say we have to, we're doing this work so current and future users can access and understand this digital information.

Digitization is not digital preservation…

"The goal should be that one feeds into the other. Scanning should be a first step in a longer process, and, as with many journeys, the first step can make all the difference."

The Signal, Library of Congress
https://blogs.loc.gov/thesignal/2011/07/digitization-is-different-than-digital-preservation-help-prevent-digital-orphans/

3

So continuing with some language, it can help to think about what digital preservation is NOT as well as what it is. All of you here with us today have probably heard this line many times – that digitization is NOT preservation. But for those who are not embedded in this kind of work, this is still a widespread assumption. You might need to explain to your board members or your partners, for instance, that just because we scanned it, or because it's available online, does not mean it's going to be around forever.

Backups are not digital preservation…

"Backups are solely designed to mitigate the risk of data loss in the event of an accident or attack, providing for business continuity. They are not designed to store data permanently offline for later retrieval."

Scott Prater, "How to Talk to IT about Digital Preservation," Journal of Archival Organization
https://minds.wisconsin.edu/handle/1793/78844

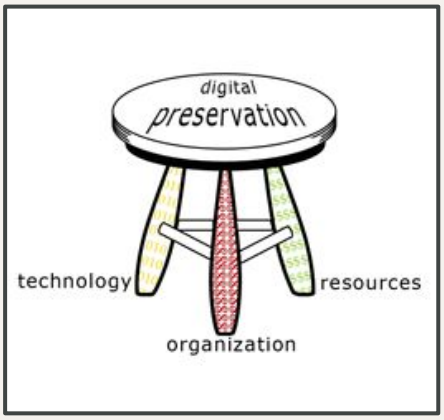
4

And here's another common assumption you may need to be prepared to address with your stakeholders. Backups are not preservation. Backing up your files is of course extremely important, and backups will be a part of your active management of digital content over time. But…

What you store can be corrupted or the storage media can fail causing data loss

- As technology and software change, data can become technologically inaccessible -- No amount of copies of a file is going to help if the software to open it doesn't exist anymore
- Files 'saved' on their own may not be findable - they need metadata to allow us to find them

Storage is not digital preservation…



- It's not the storage medium that counts. It's the work layered on top of the storage.
- There's no single tool or piece of software that "does" digital preservation.
- Preservation is enabled through skills development and training, dedicated resources, and development of and adherence to short- and long-term goals.

Nancy McGovern, Anne Kenney, "Digital Preservation Management: Implementing Short-Term Strategies for Long-term Problems"
http://www.dpworkshop.org/

5

Another misconception you might encounter is the idea that digital preservation is all about the technology – we just need to get the right software and then we're set. But software on its own cannot preserve anything. A repository is the sum of software, hardware, financial resources, staff time, and ongoing implementation of policies and planning to ensure long-term access to content. Preservation is the result of ongoing work of people and commitments of resources. The work is never finished.

One way to visualize this balanced management approach is as a three-legged stool, comprised of technology, the organization, and resources. If one leg of the stool is weaker than the others – or if it's not there at all – the whole thing will fall over. This graphic comes out of work by Nancy McGovern and Anne Kenney at Cornell University around 2003.

That work includes a lot of things we won't have time to get into today…

Geographic replication
File-format assessment and (as needed) migration
Content auditing, also known as "fixity checking"
Logging things that happen to the data ("preservation events") such as uploads, audits, changes, file-format migrations, deletions
Appropriate security and access controls

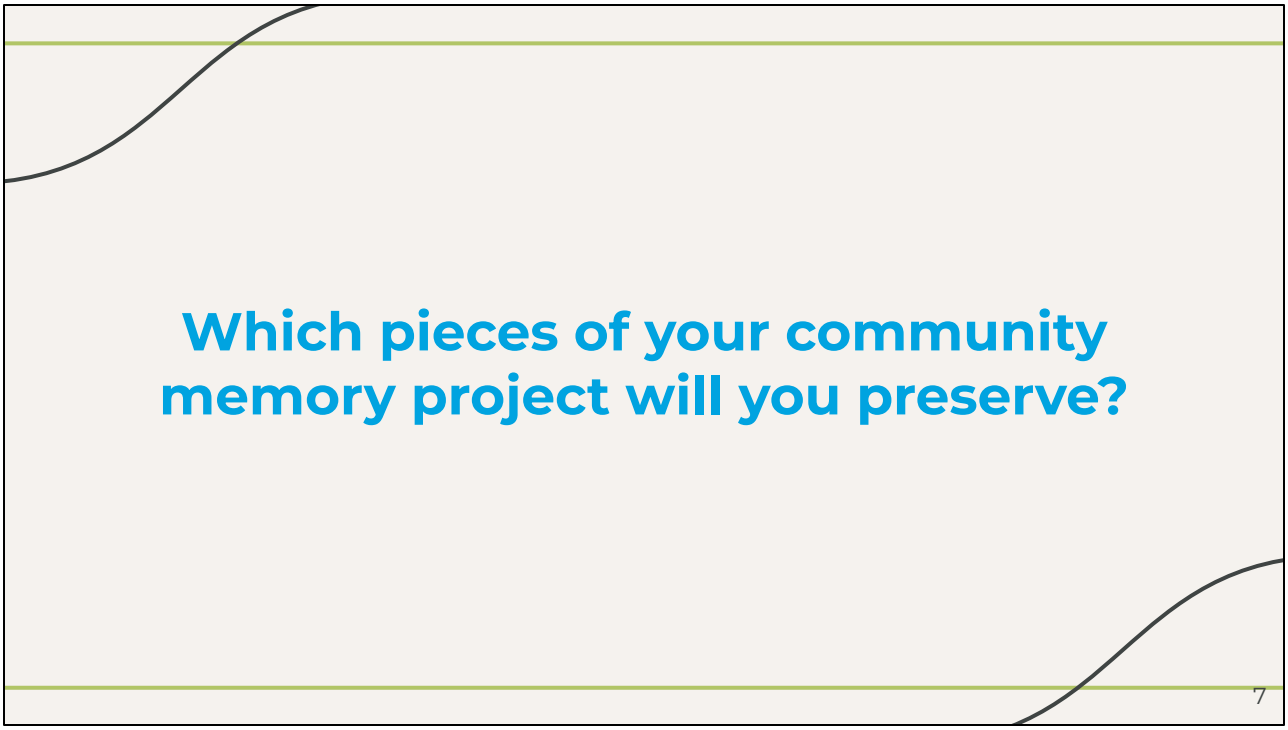
“Highly technical definitions of digital preservation are complicit in silencing the past.”

Trevor Owens, “Fifteen Guiding Digital Preservation Axioms” (2017). http://www.trevorowens.org/2017/06/getting-beyond-digital-hyperbole-tools-for-looking-forward/

“A highly technical framing of digital preservation has resulted in many smaller and less resource rich institutions feeling like they just can’t do digital preservation, or that they need to hire consultants to tell them about complex preservation metadata standards when what they need to do first is make a copy of their files.”

6

It’s long past time start taking actions. There are practical and pragmatic things everyone can and should do now to mitigate many of the most pressing risks of loss.
“get the boxes off the floor.”
the perfect is the enemy of the good

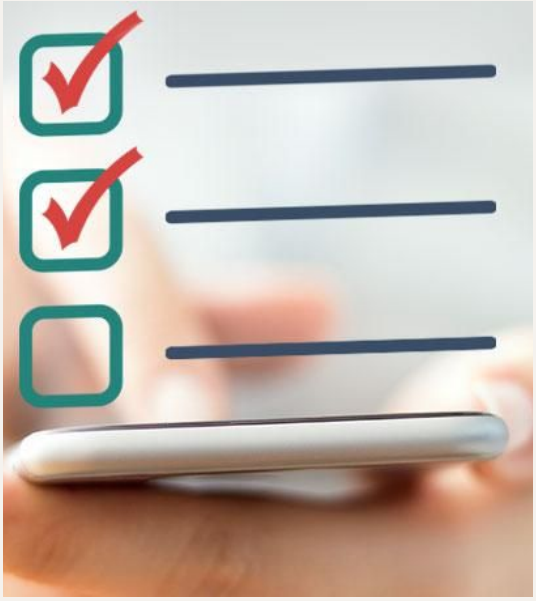


What are you preserving?
{these are likely options}

- Original (unedited) audio or video
- Edited audio or video
- Interview transcripts
- Permission forms
- Scanned photos or documents from community members or partner organizations
- Scanned photos or documents from your library's collection
- Web archives
- Metadata describing audio, video, photos, or documents
- Project documentation such as procedures, policies, or MoUs

Key actions

- **The "3-2-1" rule**
- **Digital collections inventory**
  - Identify what you have
  - Prioritize what you need to keep



8

The 3-2-1 rule provides guidance for storage, which is not digital preservation but is a foundation of being able to do preservation

A digital collections inventory is a good place to start documenting what you have within your collections

The 3-2-1 Rule

THREE copies

This is the LOCKSS principle: Lots of Copies Keeps Stuff Safe

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

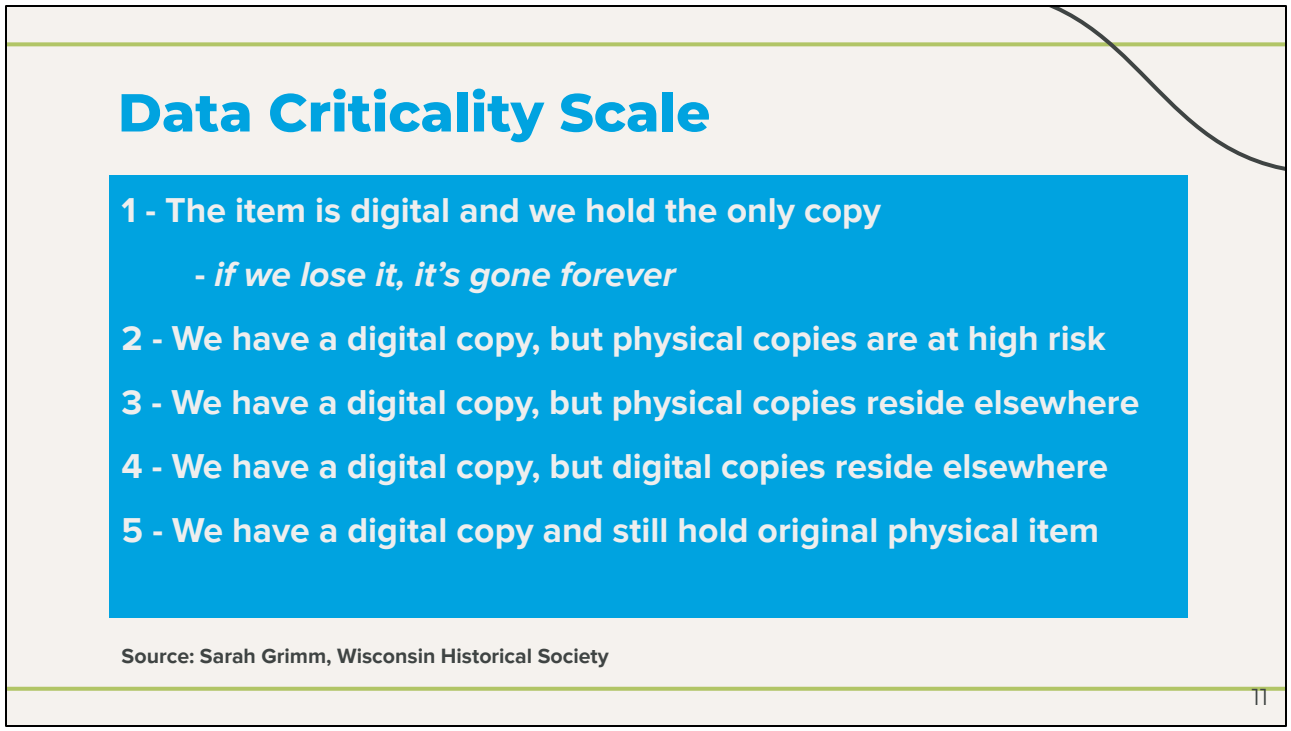
Examples of different media: a desktop/laptop, an external hard drive, a server, a tape backup system, cloud storage

with **ONE** offsite

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

Digital collections inventory information

- **Think in terms of groups or collections of content - not items**
  e.g. a digitization project, an oral history collection, a born-digital donation
- **Include:**
  - Collection title, creator, dates
  - File formats, # of files, collection size (in GB, MB)
  - Locations
  - Inventoried by, managed by and other roles
  - **Data criticality scale**



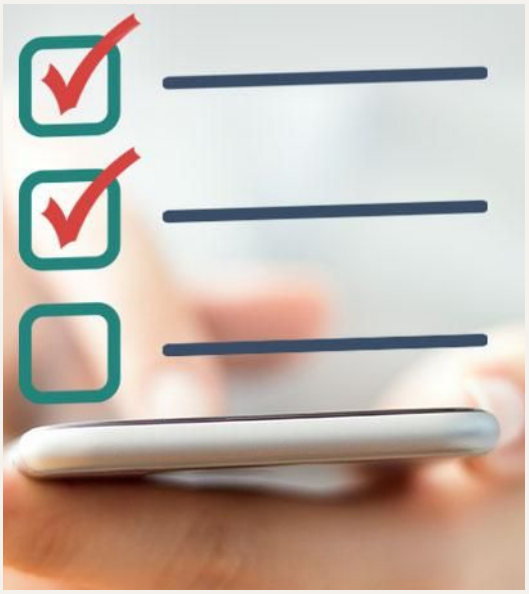
What is a data criticality scale? It's a field in your inventory where you document and rank a collection's preservation RISK.

A data criticality scale will you decide which collections to preserve first and at what level of preservation activity. This will also help you prioritize and allocate resources more effectively. You don't need to preserve everything!

So, for example, a born-digital oral history might be ranked "1" whereas a digitized college yearbook might be ranked "5".

Key resources

- National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation
- Northeast Document Conservation Center (NEDCC) Digital Preservation Assessment handbooks



12

The NDSA levels of digital preservation is an excellent resource that facilitates digital preservation assessment and planning, the process of which can promote digital preservation within your organization.

NDSA Levels of Digital Preservation

https://ndsa.org//activities/levels-of-digital-preservation/

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

This framework was created by National Digital Stewardship Alliance
There are five functional areas with four levels within each area that address storage, file integrity, information security, file formats and metadata.

This framework is useful for planning digital preservation work and setting goals

- Where are we now?
- Where do we want to be?
- What do we need to do to get there?
- Where are other orgs like ours?

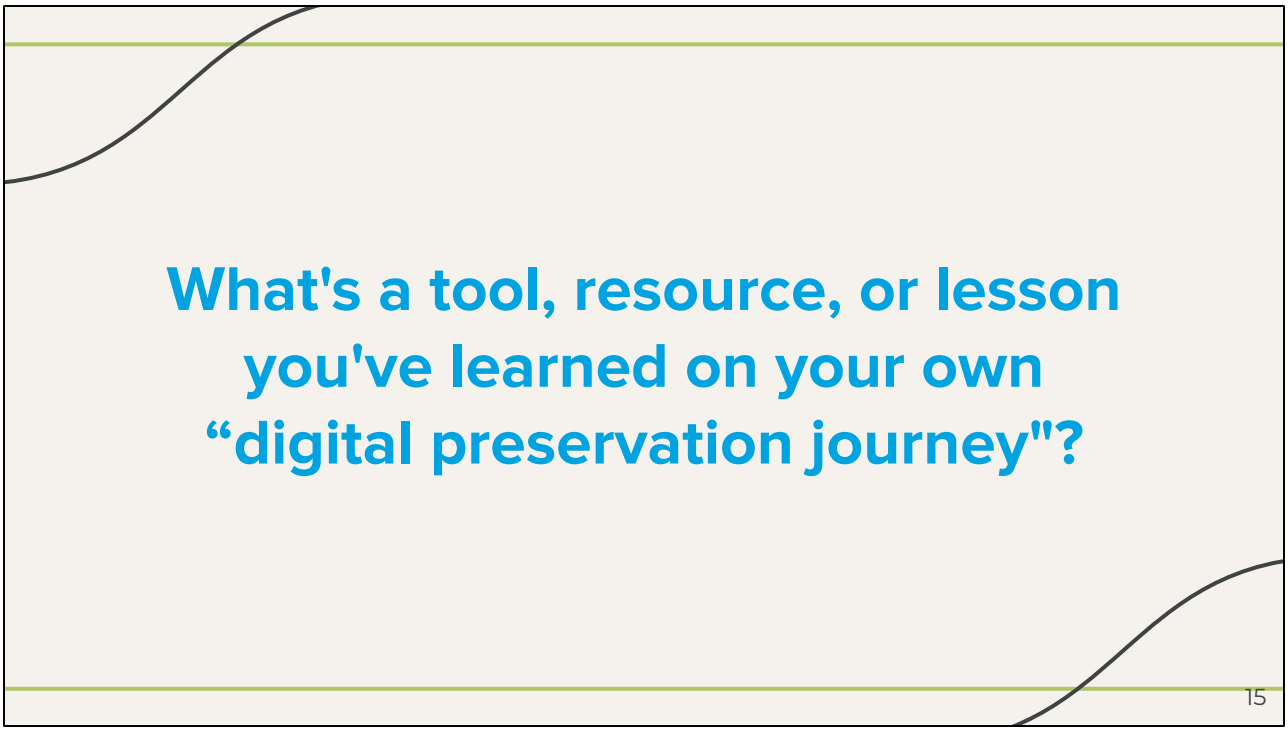
Incremental, not all or nothing.

NEDCC Assessment Handbooks

https://www.nedcc.org/preservation-training/digital-preservation-assessment-training

14

The following assessment template provides questions to prompt staff at cultural heritage institutions to think critically about their digital preservation activities. The goal of the assessment is to help an institution document digital preservation successes, recognize areas that need further growth, and identify challenges that stand in the way of that growth. It can also help an institution prioritize next steps for improved long-term access to digital collections with a digital preservation plan.



Now it’s your turn. What a tool, resource or lesson you’ve learned on your own “digital preservation journey”?