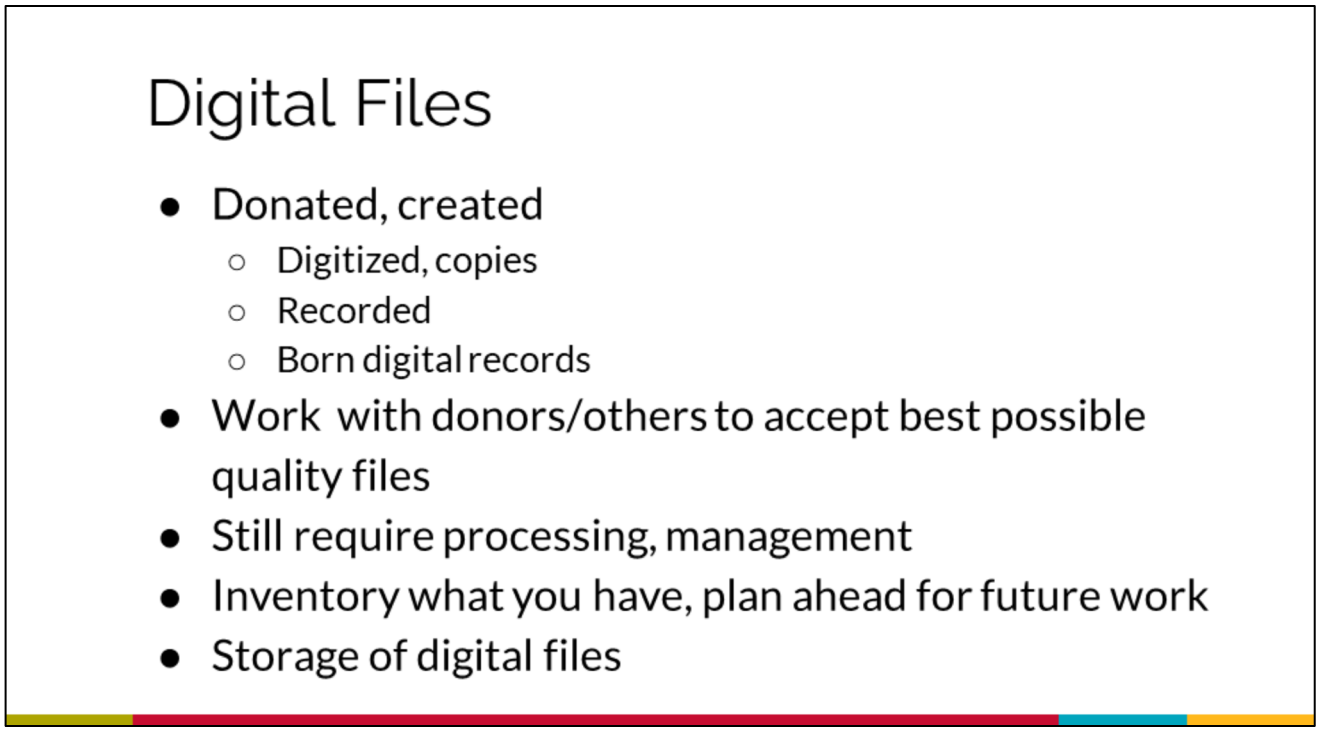
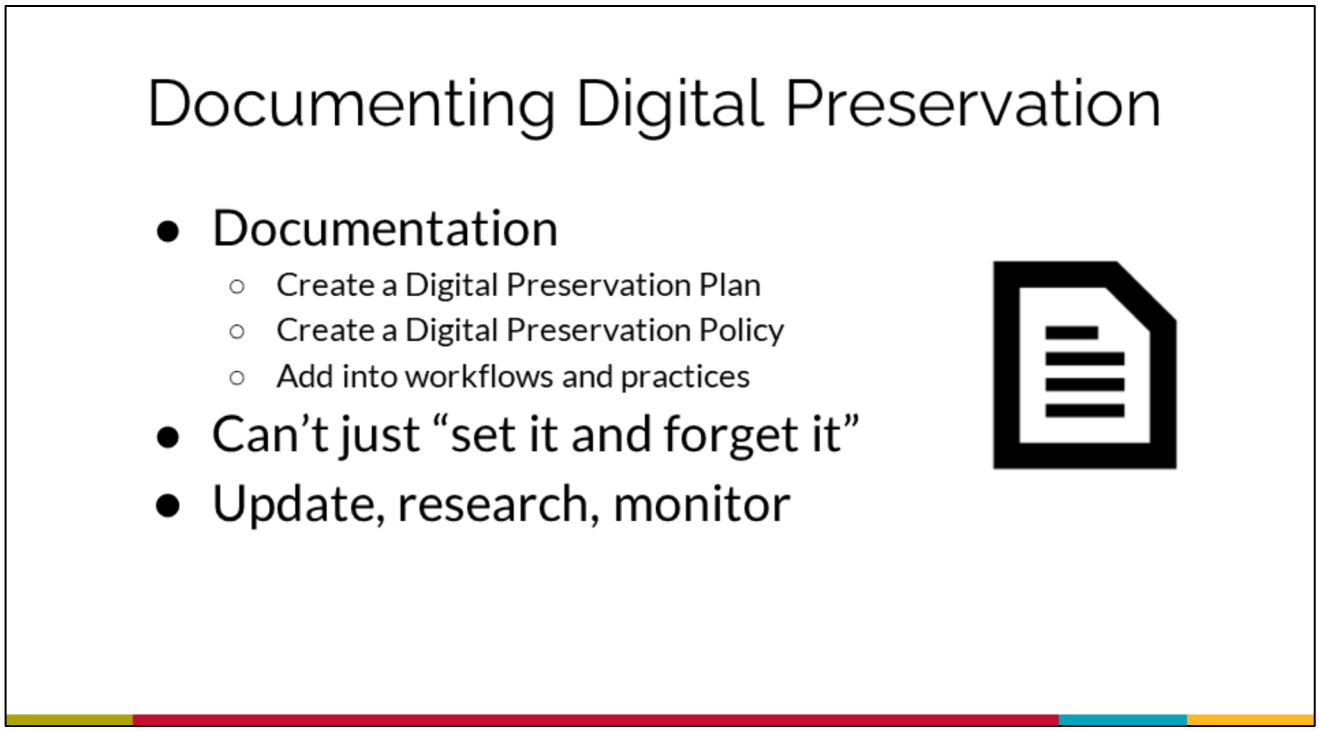
Introduction to Digital Preservation and Storage

Digital Stewardship Curriculum

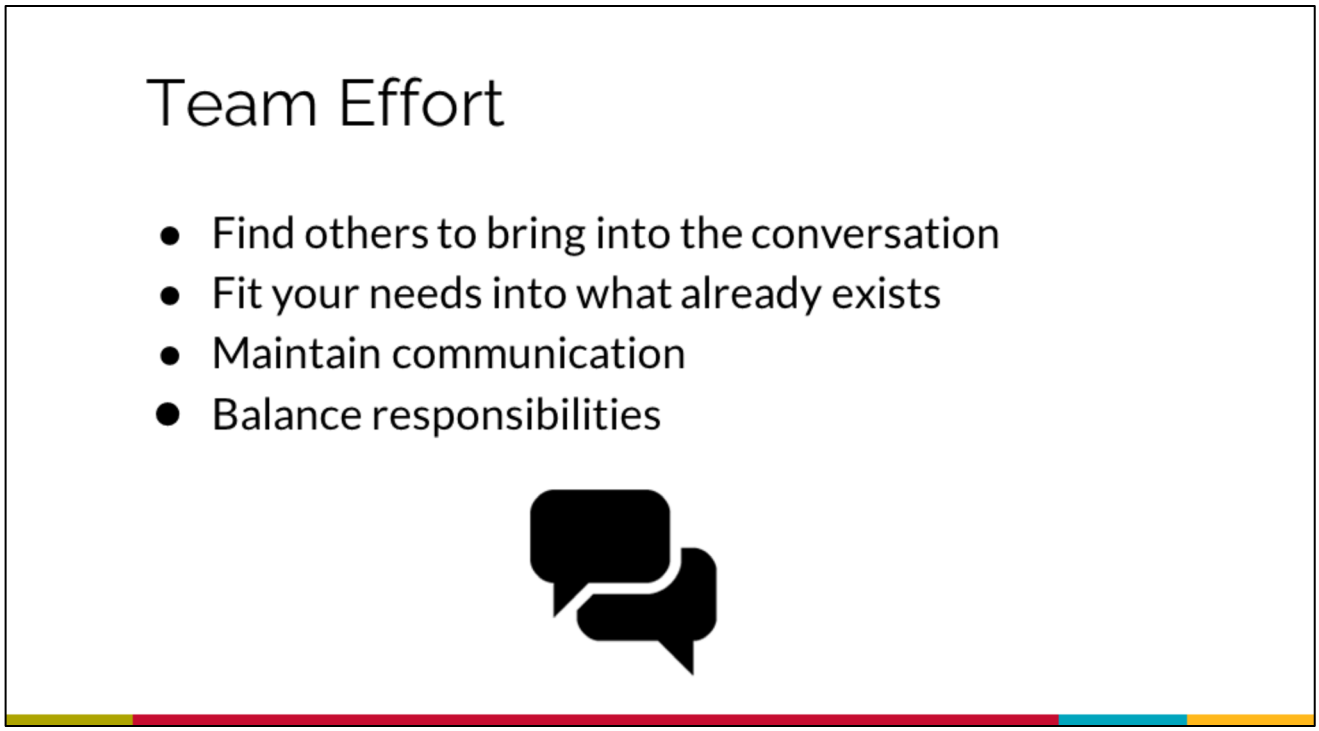


- In your department, or outside and donated
- May be digitized/converted
- Or may have started off as digital files
- Thinking about digital donations, working with donors becomes incredibly important. Easy to donate lots of unneeded or unorganized files. Sitting down with them and going through the files. BEST QUALITY - and your preferred formats. -working with donors to make sure you are getting ONLY what is needed
- Work to organize and make sense of digital files
- Think about physical and digital files in all policies, even if they are not being donated yet
- You need a place to store files, for the long term

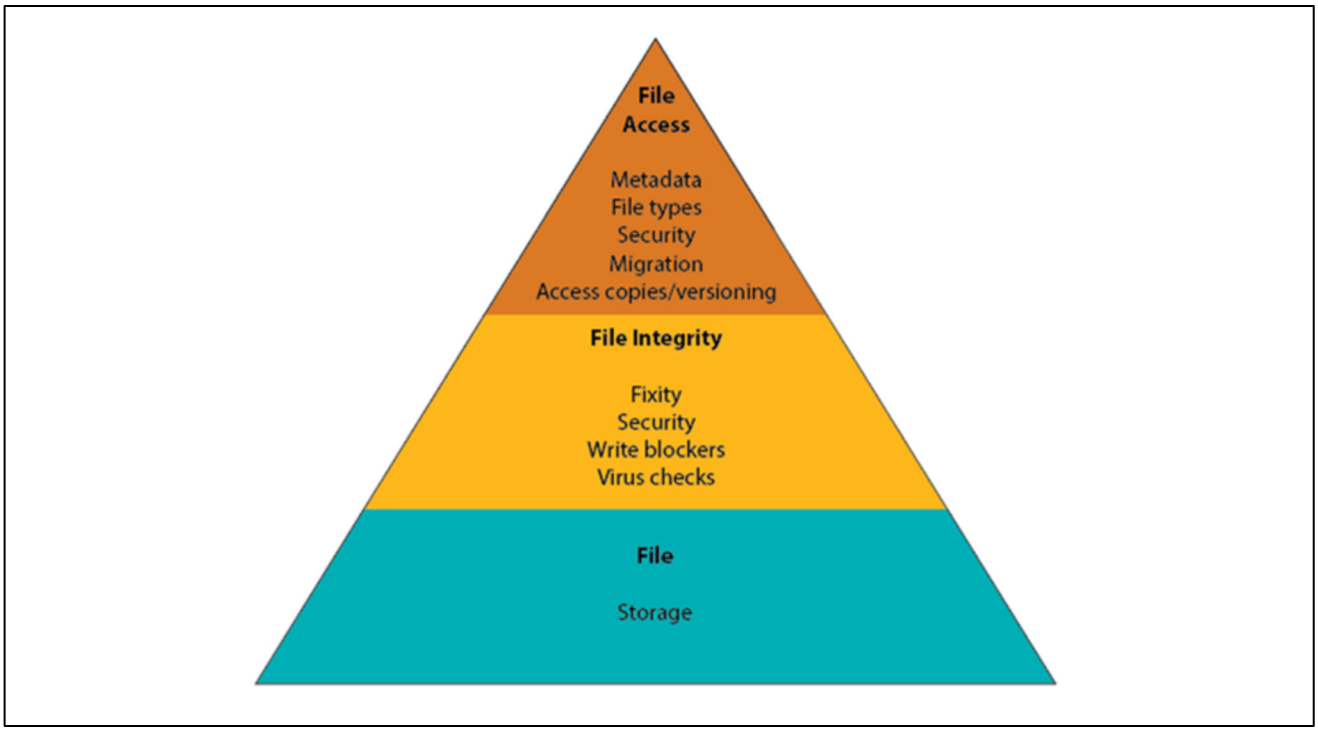
- What is digital preservation?
- Long term storage, management, and preservation
- Ensure that all the work you put into digitizing will be saved for the long term!
    - Digital preservation should be a conversation throughout your department/institution - if not, you will have to start small and keep at it
- Should be considering digital preservation in general, and with every digital project that you start



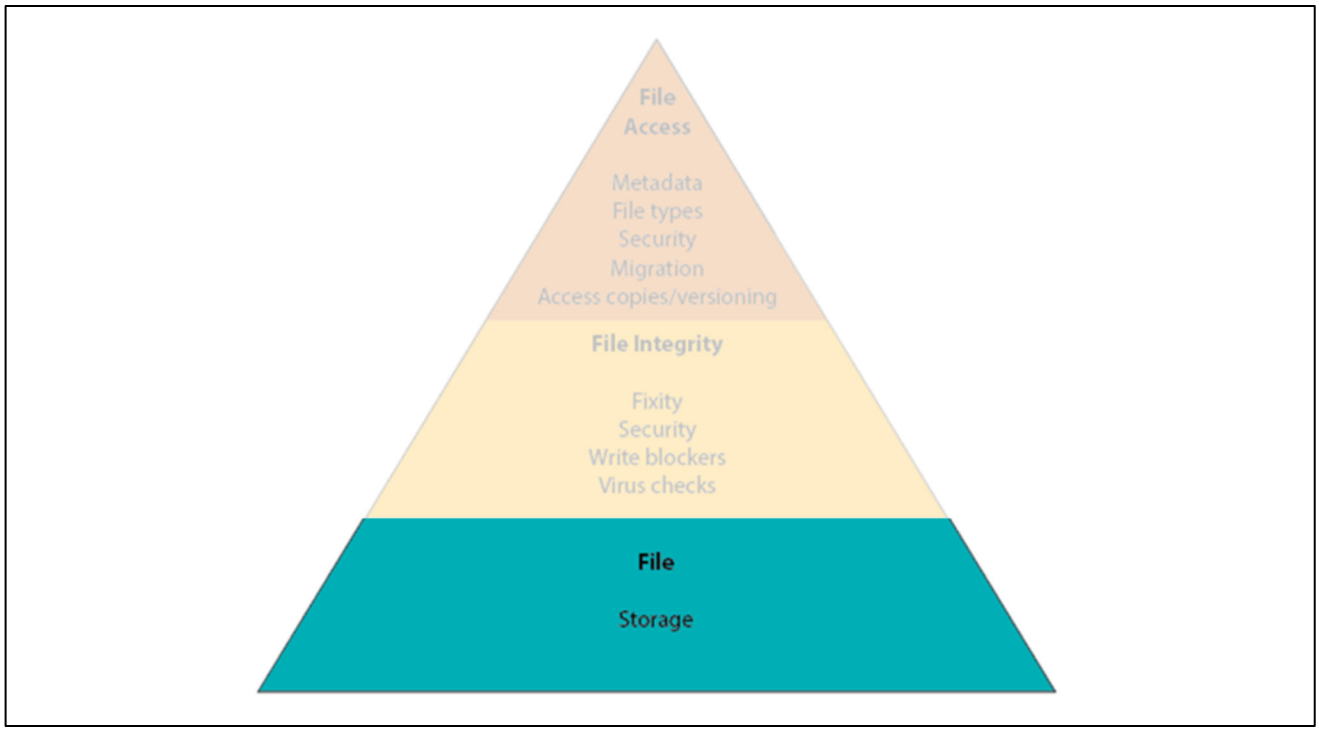
- Documentation
  - A Digital preservation plan that includes all parts of saving and preserving files, managing files, checking files - making sure it all works together and is carried out
    - See the SHN Resources *Activities to Include in a Digital Preservation Plan* and *Digital Preservation Plan Worksheet*
  - A policy is a written version of this information, that ties into institutional and departmental goals
    - See the SHN resource *Developing a Digital Preservation Policy*
  - Your workflows and practices are what gets carried out day-to-day, the information from your plan and policy must be applicable to daily/weekly/monthly/yearly actions to implement effective digital preservation
  - All of this documentation and implementation must be updated as technology and approaches change and evolve
- Similar to digitization projects - the most time goes into the planning (this planning requires time up front, but will help sustain the project)
- Your plan - not just created once and complete...like with your other policies, it must be revisited (especially with changing technology- updates)
- As hardware, software, security changes, your plan must also stay up to date



- Luckily, you aren't alone! You either have other people in your institution that should be involved OR you can reach out to others in similar situations (institutions of all sizes struggle with these issues) - make a team!
- You might also be joining a conversation that has already been started
- To do this,...need to FIND OUT what already exists
    - When going to get funding from outside agencies or within your own organization it is imperative have a conversation about what your organization already has in place, so any new funding can be put to the best use.
- Don’t want to re-do anything that already is working well!
    - Or duplicate work done in two separate departments
- Have initial conversations, and then continue -- - how do people check in (meetings, conference calls?)
- Want to make sure all responsibilities are covered - who is in charge of what? Make sure the long term preservation needs are understood
- Since DP can get very complicated, and is unique to each situation, we created a Pyramid of Digital Preservation...



- Broke down digital preservation into 3 essential parts that any Digital Preservation Plan will have to involve
- FIRST, the most important foundational piece

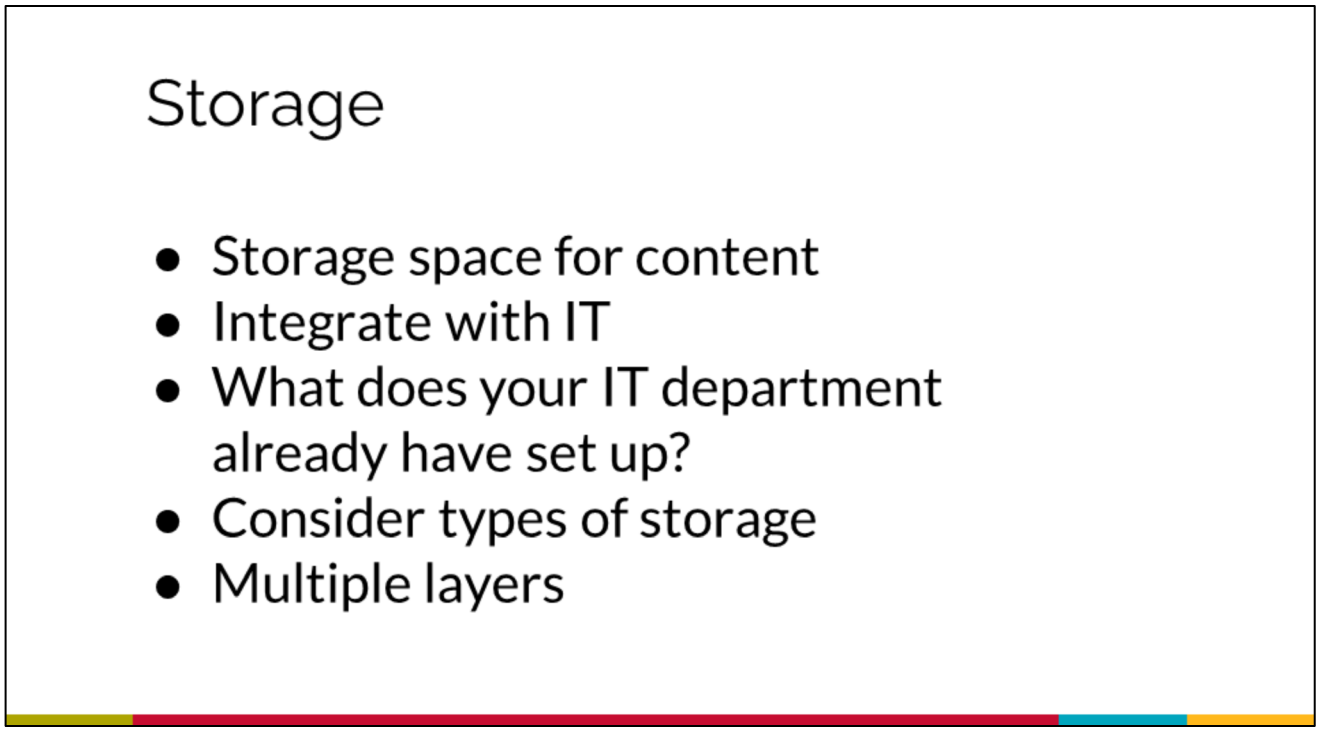


- Storage is the base of our pyramid -
- Making sure that content is stored safely and securely for long term preservation
- You might not be doing this work! You might have an IT department, or a person on your staff who is in charge of storage, technology, computers….But that doesn't mean that you don't have a responsibility to make sure that it is being carried out, and your files are securely preserved.

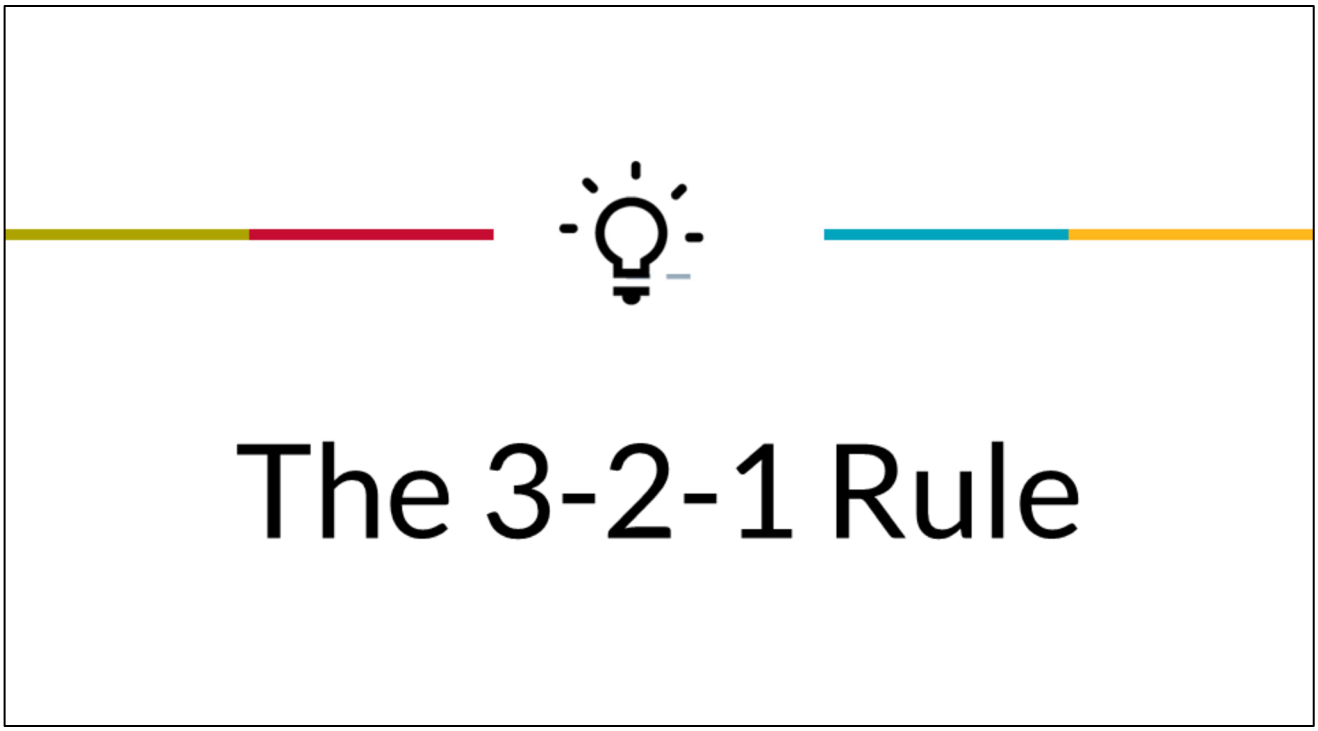
Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have 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

- Another wonderful resource for understanding digital preservation, starting out with some basics and improving over time is the National Digital Stewardship Alliances Levels of Digital Preservation

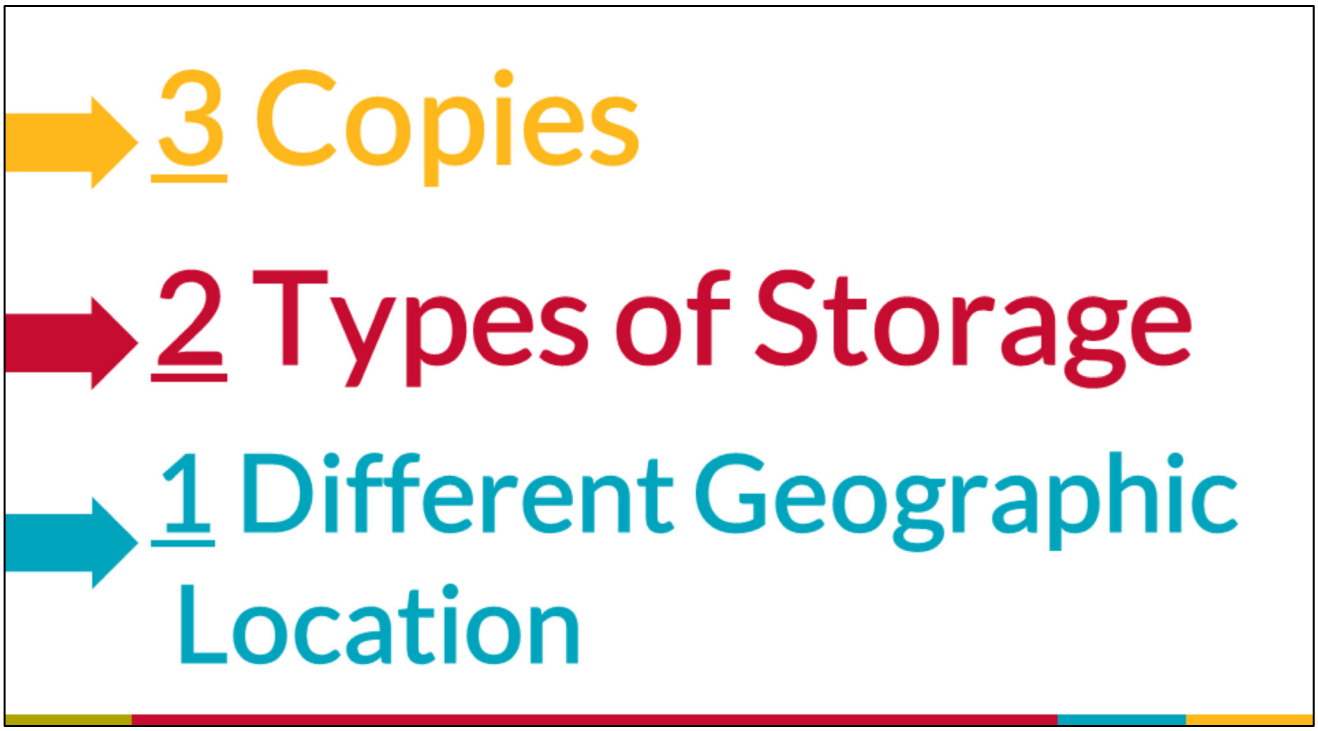
- For more about the different levels of preservation, check out: https://ndsa.org/publications/levels-of-digital-preservation/



- Get what you need for your digital content, AND what will be created in the future
    - Ideally all of your long term preservation files can be stored together and organized in a meaningful way
- Talk to IT, they may be the ones setting you up, or may prefer a certain type of storage, figuring out funding
- Many options for storage - all storage options have pros and cons - make a decision that will fit best for you (and try to avoid flash drives/single hard drives)
    - Think about how your backups fit in with each other
- Workflows for backing up content, multiple places for backups - fail safes if one is destroyed

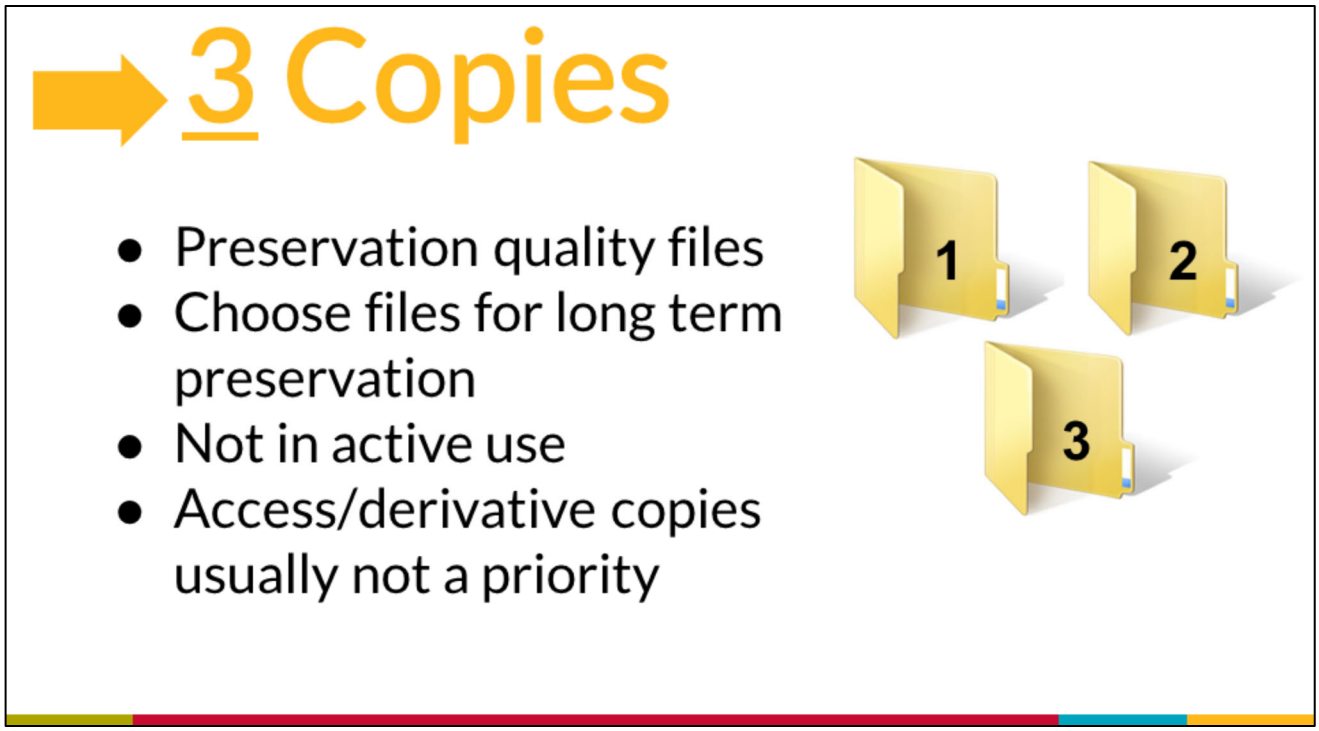


- This concept is something that we like to repeat as a baseline for setting up digital storage. “The 3-2-1 Rule”

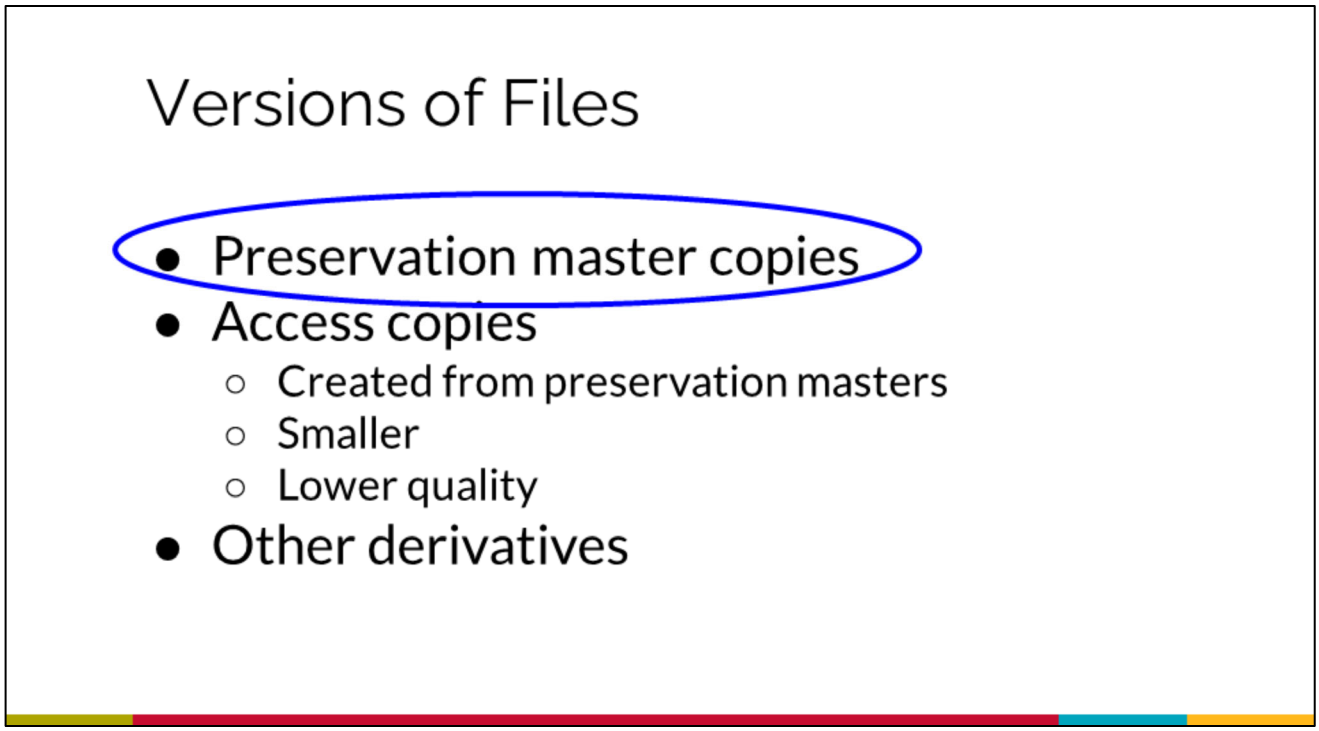


- Have 3 copies of any valuable information
- Store 2 on at least different types of media
- Store 1 in a different geographic location

- We will explain each of these more

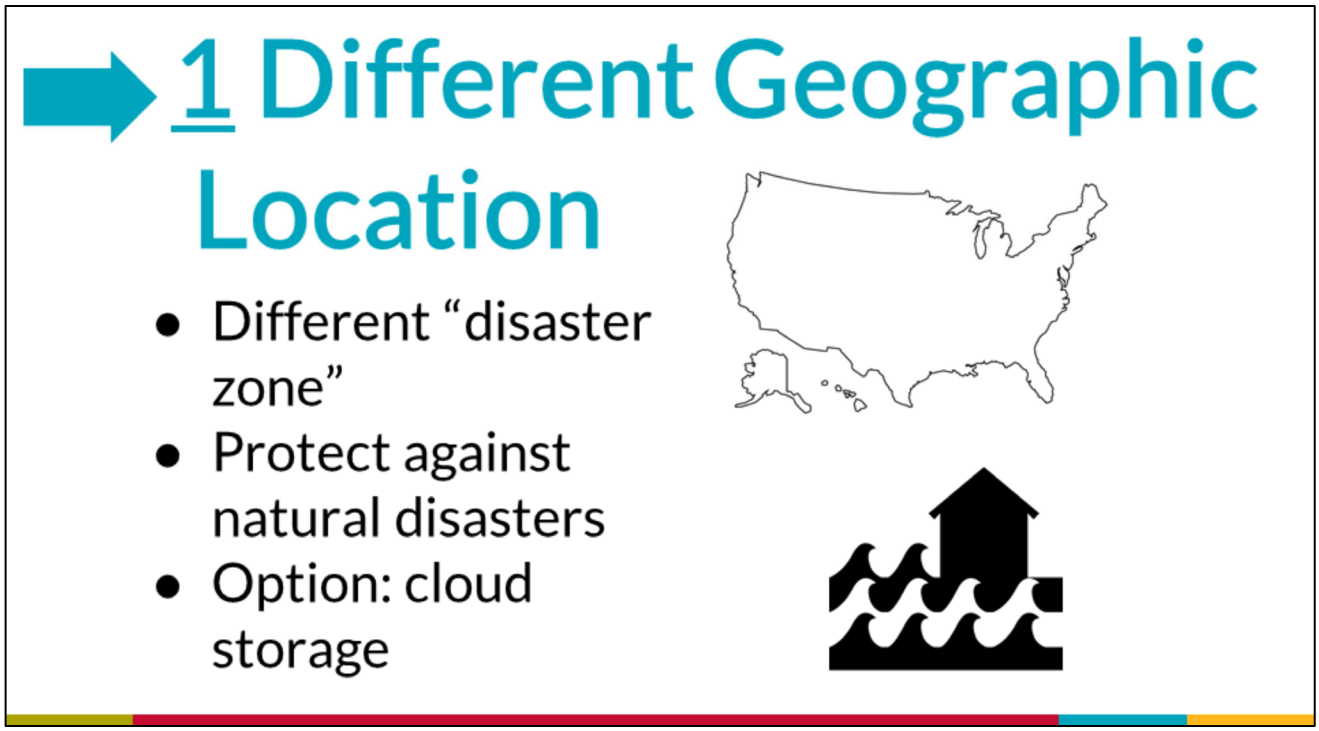


- Have 3 copies of any valuable information
  - These are your highest quality, preservation files that need to be saved in the long term
  - Not just office files
  - Select what gets saved
  - These files should not be in active use
  - If you have multiple versions of files - like access copies for web upload, these are not usually preserved - because you are prioritizing that space for the PRESERVATION files



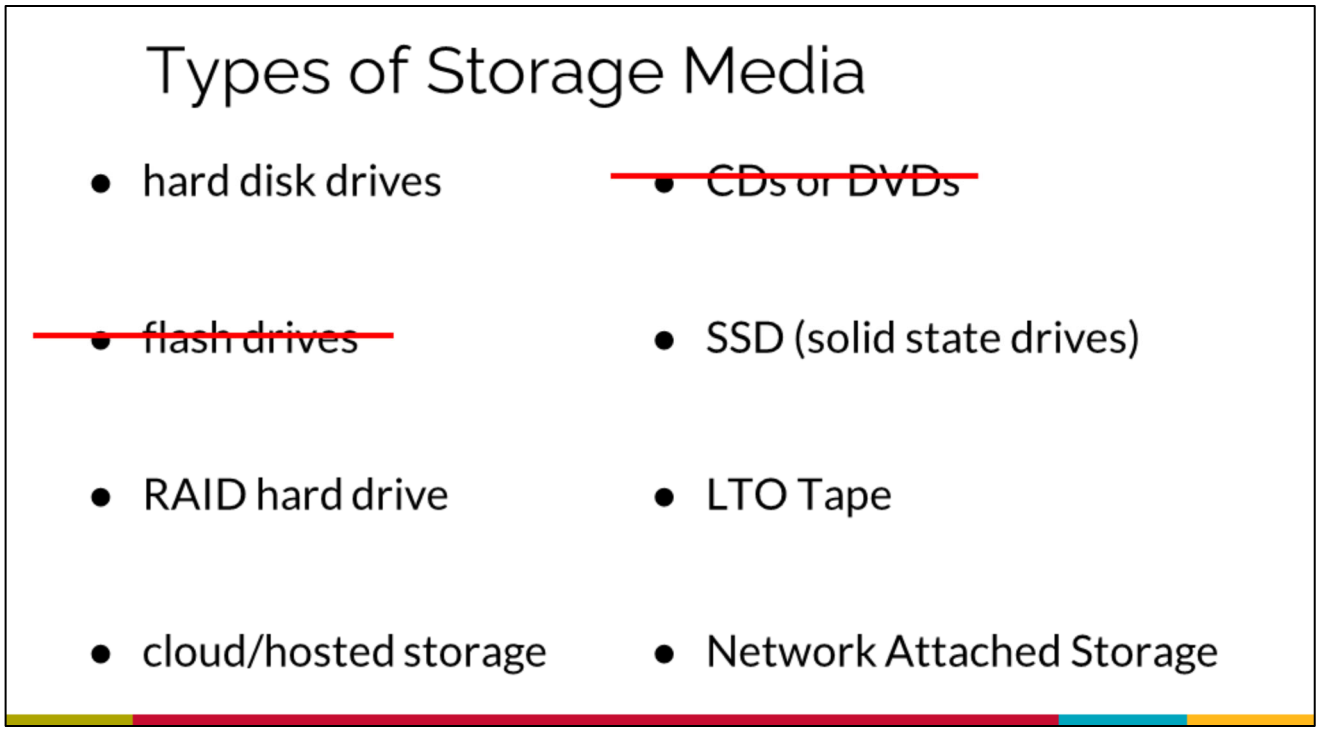
- When thinking about LONG TERM storage - having those different versions of your files (master file, access, copy, edited copies) organized is really important. So you know which files you need to add to long term storage.
- Long term storage, best quality that you can afford -- those are your preservation masters - these files will dictate how much space you need to plan for
- Access copes
- Derivatives - exhibits, flyers, thumbnails online
- AGAIN - your preservation copy is the one you are worried about in the long term

- 2 types of storage
- At least two different types of storage media
    - For your three storage locations - you can set it up however you like, but having at least two types of storage makes it a safer setup, in case a certain format turns out to be not viable
    - So you might have two raid hard drive systems (1st and 2nd copy), and then also back up to cloud storage (3rd copy),
    - Or a network server (1st copy), and two RAID hard drive systems (2nd and 3rd copy)
    - Or as pictured in the image - RAID storage (1st copy) LTO tape (2nd copy) and cloud storage (3rd copy)
- Manageable - you should be able to easily maintain and oversee the storage, or easily be able to work with IT to manage
- Fault tolerant - you don't want to worry about drive failure, so pick storage that has been tested, reviewed highly, and you can trust
    - And swap out according to how long the storage media lasts

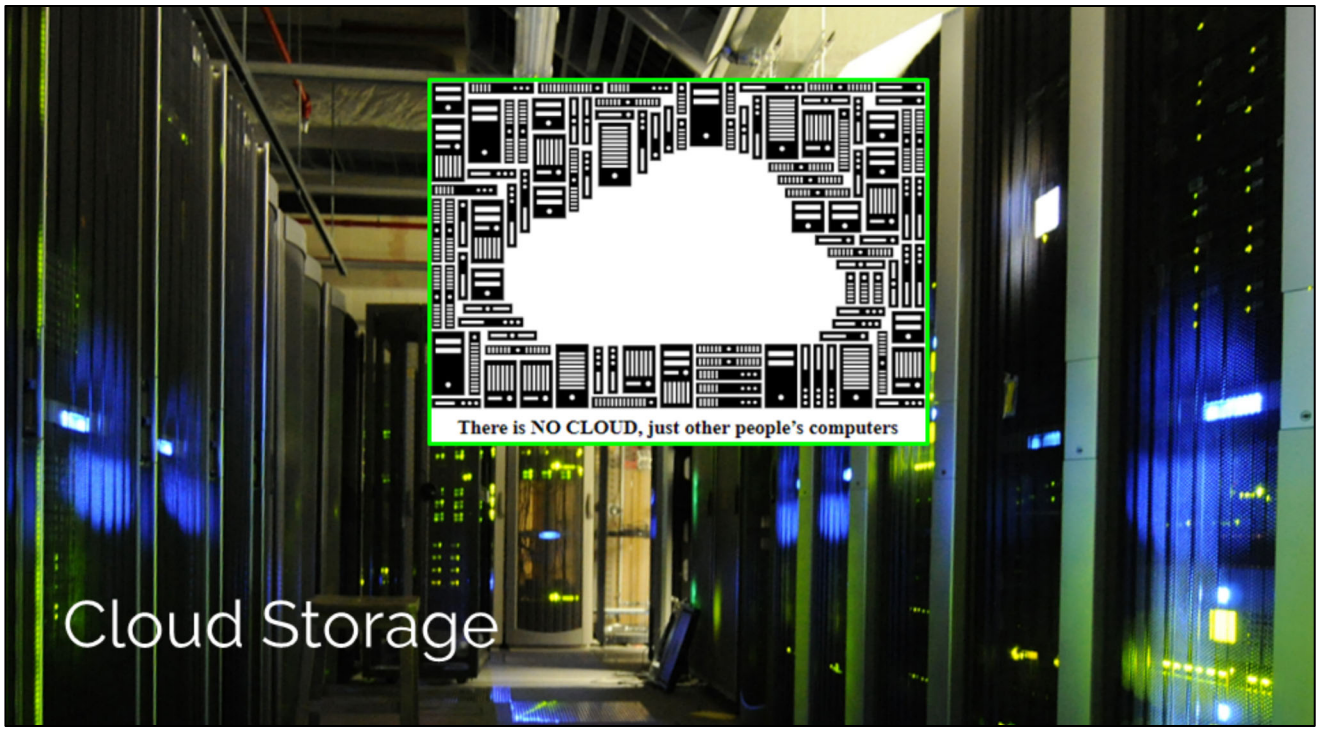


- Then, if possible, have one storage location in a different disaster zone (avoiding a natural disaster that might destroy other storage)
    - Think about what can affect your area

- Cloud storage can be a good option for this, but you are **not** required to use cloud storage
    - Some consortiums, Universities and other institutions look into partnering with each other to exchange storage space
    - Other more simple ideas might be bringing a hard drive home with you, if you live in a different location from the other storage



- Here are some different types of storage,
- External hard drives - not a matter of if they will fail, but when(and need to be replaced every 2-3 years)
- Flash drives - temporary storage - like the ones we have for this program [avoid using these!]
- RAID hard drive - Redundant Array of Independent Disks - multiple hard drives together, more fault tolerant
- Cloud storage AKA hosted storage,
- CDs, DVDs - these are a legacy format, were used for preservation at one point, but not anymore [avoid using these!]
- SSD - similar to hard drive, with no spinning disc -longer life 3-5 years
- LTO tape - systems that IT would need to manage
- NAS - server
- Knock a few off the list right away:
    - Flash drives - fail easily, carry viruses easily (and are easily lost or misplaced)
    - CDs or DVDs- obsolete, not used anymore - many new laptops no longer have an optical drive built in, media can degrade without warning



- Cloud storage and hosted storage
- Paying someone else to host your storage - contract
- These files still take up space and energy to maintain, have environmental impact and cost
- May have **restrictions** on this in your community - of storing files and information
- PROS/CONS
    - Pros: can be cheap to store, can pick a storage location in another area (geographic region/disaster area different from yours.
    - Cons: depending on company can be expensive to get data out again on a quick turnaround.
- You may have heard of Google Drive or Dropbox
- Amazon glacier
- Microsoft azure
- Carbonite

- Cloud image: Markus Meier, FSFE / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) https://commons.wikimedia.org/wiki/File:FSFE_There_is_no_cloud_postcard_en.svg
- Background image: The National Archives (UK) / CC BY (https://creativecommons.org/licenses/by/3.0) https://commons.wikimedia.org/wiki/File:A_view_of_the_server_room_at_The_National_Archives.jpg

- External HDD - hard disc drive
  - Relatively inexpensive
  - But if you just have one, and it **fails**, you are out of luck
  - Riskier than others, need to REPLACE 3-5 years
  - Plug in, drag and drop - easy enough to use after setup

- Armchair / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) https://commons.wikimedia.org/wiki/File:Typical_External_Hard_Drive.JPG

External RAID enclosure

- External RAID enclosure.
  - Multiple hard drives
- More expensive
- Redundant backups - less risk of failure -- if one fails, another will have backed up (still need to replace)
  - Ex: Synology diskstation (2 8tb drives = 4 tb)
  - Western Digital
  - Lacie
- Kevin Lim Unboxing OWC Mercury Elite Pro RAID enclosure
  https://creativecommons.org/licenses/by-nc-sa/2.0/
  https://www.flickr.com/photos/inju/193066937

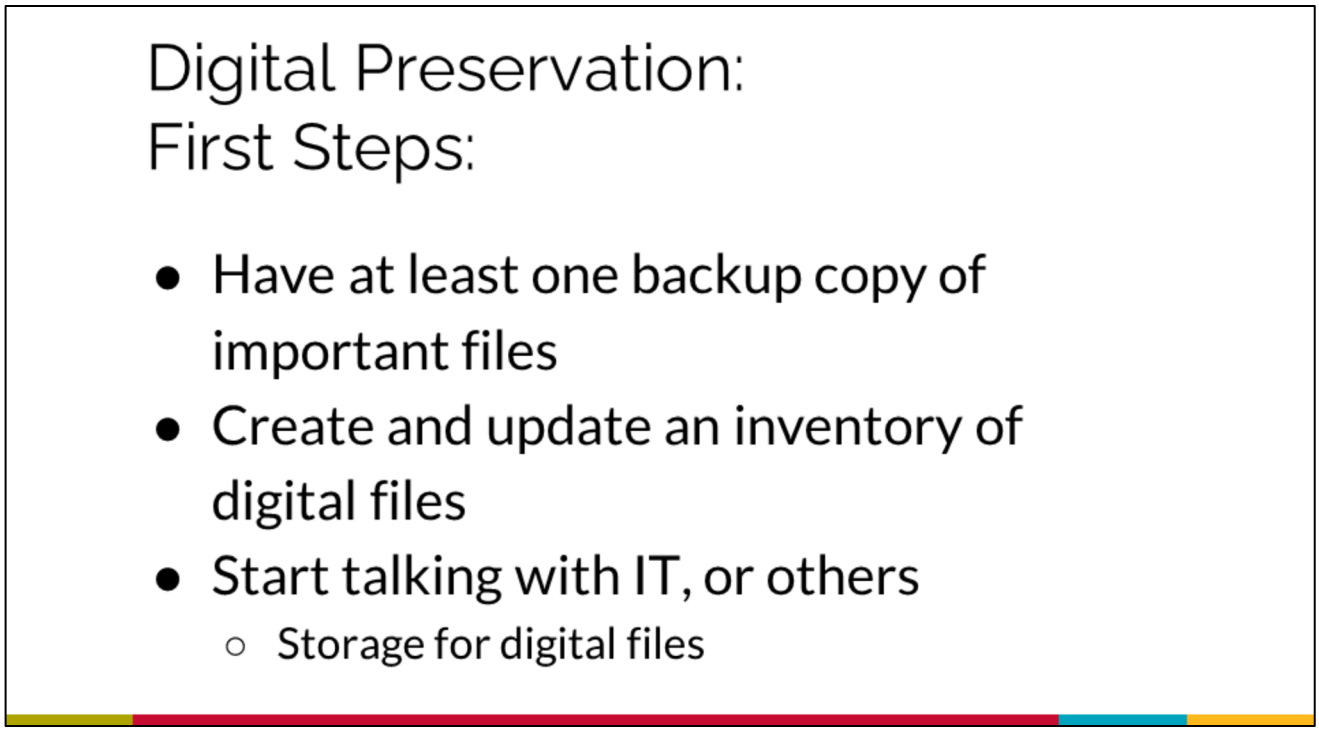
LTO Tape Storage

- LTO Tape storage
- Cheap media, need equipment to read/write,
- Need someone who knows how to run the equipment (tape reader) - that staff is expensive infrastructure
- Would probably want IT to handle
- Austinmurphy at English Wikipedia / CC BY-SA (https://creativecommons.org/licenses/by-sa/3.0) https://commons.wikimedia.org/wiki/File:LTO2-cart-wo-top-shell.jpg
- User Trimbo on en.wikipedia / CC SA (http://creativecommons.org/licenses/sa/1.0/) https://commons.wikimedia.org/wiki/File:LTODriveWithTape.jpg

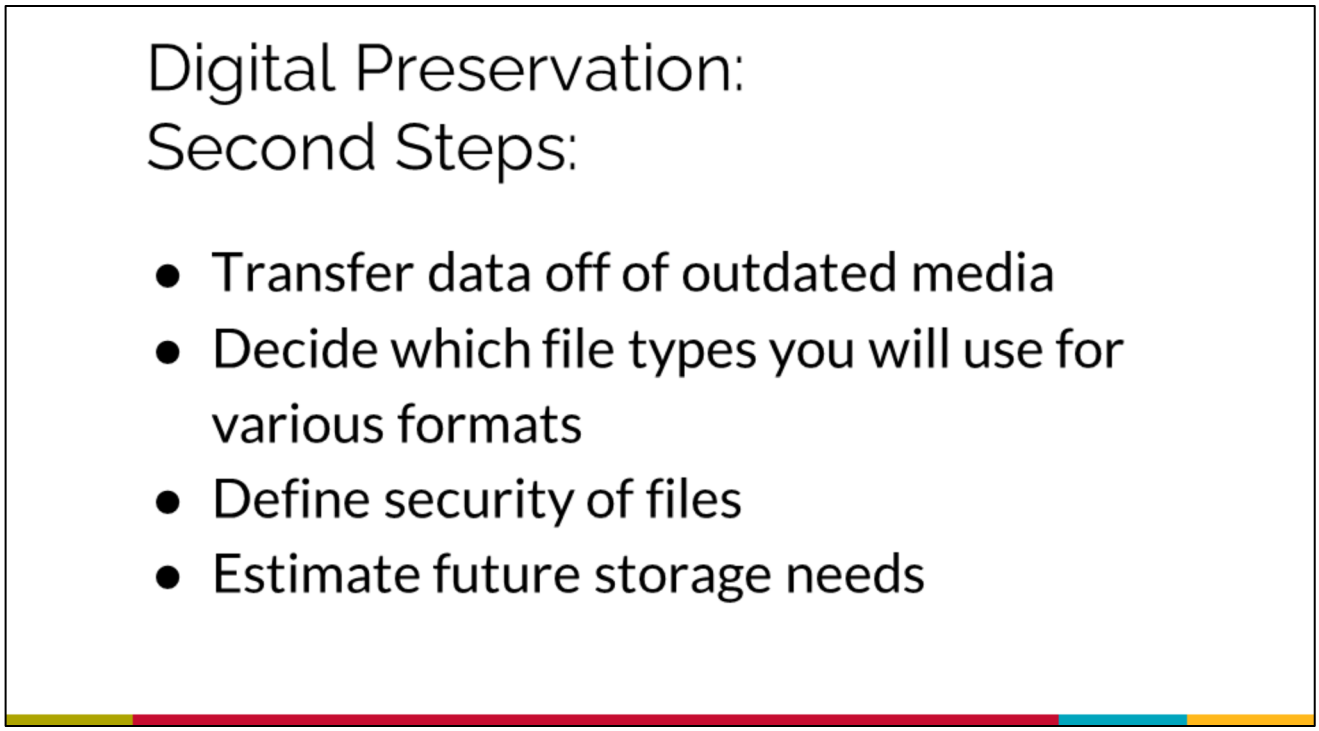
Network Attached Storage, or Storage Server

- Network Attached Storage or a Storage Server
- IT would probably handle

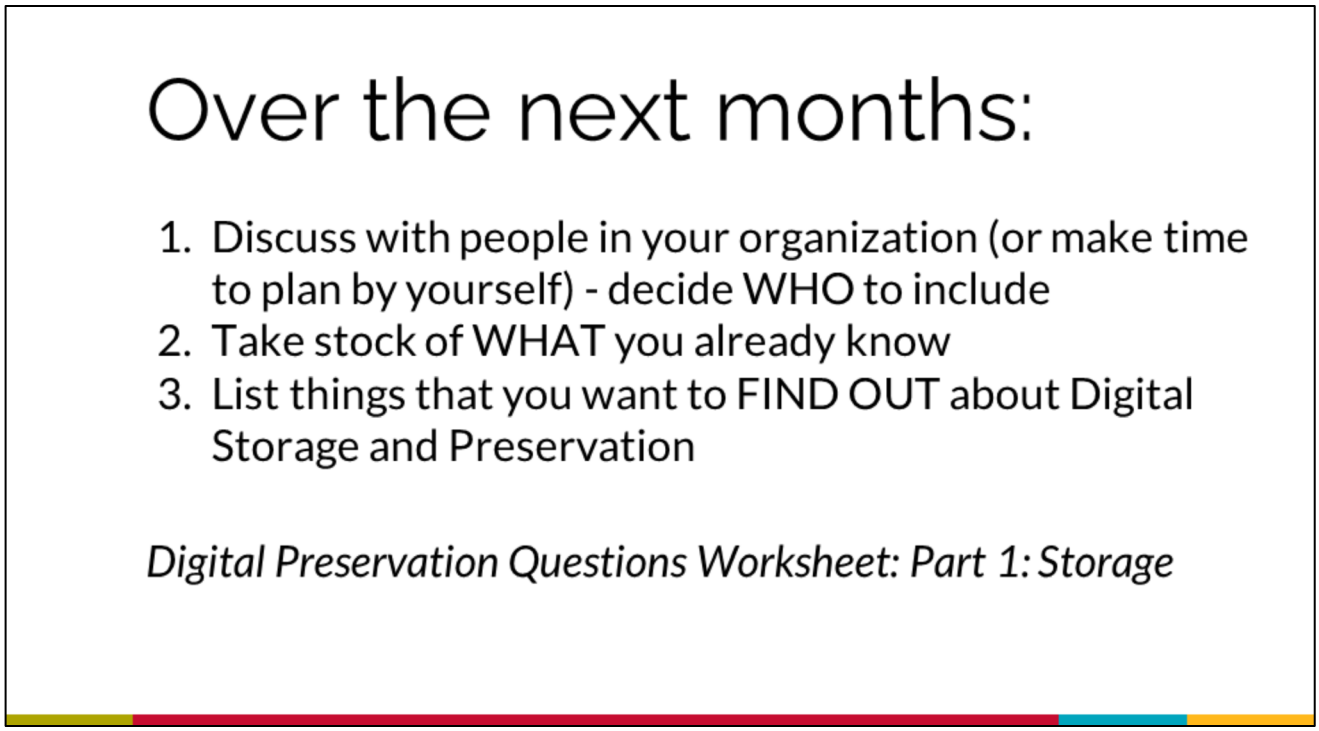
- VIA Gallery from Hsintien, Taiwan / CC BY (https://creativecommons.org/licenses/by/2.0) https://commons.wikimedia.org/wiki/File:Network_Storage_Server-_VIA_NSD7800_(angle_view_with_case_open)_(3179536746).jpg



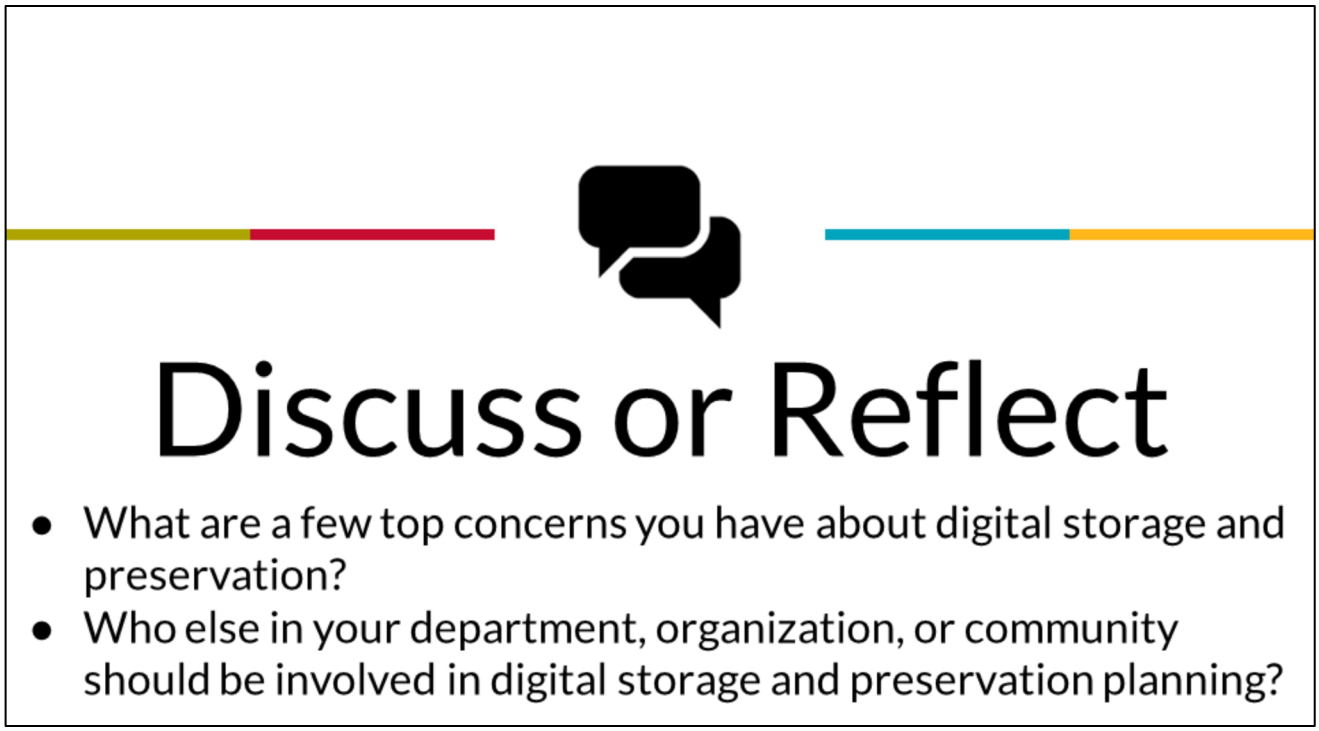
- Think about what you currently have - is there anything that you can't afford to lose? Backup
- Document what digital collections you have - create an inventory
- Start conversations - your department, your supervisors, administration, IT department
    - In an ideal world everyone could have all those resources to support your work
    - We understand that things aren't always ideal
    - Think of WHO is involved with storage of digital files, think about how to talk and with them
    - If that is just you, that's ok too, because you have us to talk to
- Levels of Digital Preservation Preparedness document



- Get your files off of thumb drives, CDs, floppy disks - these are at risk!
- Figure out file types - what are the recommended file types for the formats you have
- Think about internal security and access to files - who currently has access? Who should have access?
    - Prevent human error
    - Prevent malicious actions
- Forecast what you will create - understand how much space you need currently and in future
    - Can be hard to know how much storage you need, but there are ways of estimating…
    - Digital Rebellion
    - Audio - The Audio Archive
    - Photograph - NAU



- Complete the SHN resource Digital Preservation Questions Worksheet: Part 1 to get some helpful questions and discussion started around this topic. Bring others into the conversation



- Take 20-30 minutes and discuss with others, or reflect by yourself and take notes
    - What are a few top concerns you have about digital storage and preservation?
        - Does your current digital storage setup pass the “3-2-1” rule?
        - Is there anything you are particularly worried about?
        - Are there any past events like data loss or corruption?
    - Who else in your department, organization, or community should be involved in digital storage and preservation planning?
        - Who are your allies?
        - Who has skills that would be helpful?
        - Who has access to funding or grant writing skills?
        - Who in leadership or administration do you need to speak with?

Credits: Images

- Slide 3: Save icon Font Awesome by Dave Gandy https://fortawesome.github.com/Font-Awesome Data Center CC Amber Case; Flickr https://creativecommons.org/licenses/by-nc/2.0/
- Slide 6-7: Center for Digital Scholarship and Curation, Lotus Norton-Wisla, Michael Wynne, Alex Merrill
- Slide 8: NDSA image https://ndsa.org/publications/levels-of-digital-preservation/
- Slide 14: External hard drive CC Armchair; Wikimedia Commons https://creativecommons.org/licenses/by-sa/4.0/deed.en
- Slide 17: Cloud image: Markus Meier, FSFE / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) https://commons.wikimedia.org/wiki/File:FSFE_There_is_no_cloud_postcard_en.svg Background image: The National Archives (UK) / CC BY (https://creativecommons.org/licenses/by/3.0) https://commons.wikimedia.org/wiki/File:A_view_of_the_server_room_at_The_National_Archives.jpg
- Slide 18: Armchair / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) https://commons.wikimedia.org/wiki/File:Typical_External_Hard_Drive.JPG
- Slide 19: Kevin Lim Unboxing OWC Mercury Elite Pro RAID enclosure https://creativecommons.org/licenses/by-nc-sa/2.0/ https://www.flickr.com/photos/inju/193066937
- Slide 20: Austinmurphy at English Wikipedia / CC BY-SA (https://creativecommons.org/licenses/by-sa/3.0) https://commons.wikimedia.org/wiki/File:LTO2-cart-wo-top-shell.jpg User Trimbo on en.wikipedia / CC SA (http://creativecommons.org/licenses/sa/1.0/) https://commons.wikimedia.org/wiki/File:LTODriveWithTape.jpg
- Slide 21: VIA Gallery from Hsintien, Taiwan / CC BY (https://creativecommons.org/licenses/by/2.0) https://commons.wikimedia.org/wiki/File:Network_Storage_Server-_VIA_NSD7800_(angle_view_with_case_open)_(3179536746).jpg

Credits: Presentation

- Presentation template by SlidesCarnival.
- Minicons by Webalys
- *This template is free to use under Creative Commons Attribution license.*
- These slides contain changes to color scheme and content.

Using this Resource

The Digital Stewardship Curriculum is an Open Educational Resource created by the Center for Digital Scholarship and Curation.