Storage and Maintenance

Storage and Maintenance

Where and how digital files will be stored is an important subject that needs to be discussed at a project's outset. Long term-storage capacity, file security and maintenance, number of copies, back-up procedures, and file and technology migration plans all become more important the longer the files are to be retained. It is essential that you discuss your post-digitization storage and preservation needs with whomever will be responsible for maintaining the digitized records **prior** to beginning any digitization project.

Agencies are responsible for ensuring that records remain "accessible, accurate, authentic, reliable, legible, and readable" per Administrative Rule 12. Addressing these key concepts ensures that digital records will be properly stored and maintained. Specific protective measures may include, but are not limited to:

- Establishment of security protocols, and approved administrators and users.
- Employment of system checks and error-checking utilities.
- Implementation of back-ups and disaster preparedness measures.

Planning for Storage of Digitized Records

While storage of the records comes at the end of a digitization project, this is a critical topic that needs to be established from the beginning to ensure your agency can commit the staff and resources to the long-term preservation of the digitized records. This is especially important for records that will be held permanently in digital format where the paper has been discarded. Your discussion should include:

- **Where records will be stored both during and after the digitization project.** It may make sense to have a limited, temporary storage area for the project while the records are being digitized and QA'd and then transfer to longer-term storage.
- **How long the records need to be kept.** This is important both to plan for storage space as well as to plan for the destruction or transfer of records from your organization. For records that will be stored in a content management system, the system should be help you manage the disposition of those records through technological reminders you place on a folder. For records that are stored on a network drive, you will need to establish some other way of tracking the disposition date of the records which may implement some combination of inventories and folder naming conventions to help you remember when the records need to be disposed of.
- **Plans to address management of files over time, including:**
    - File formats become obsolete over time. Today's formats will need to me monitored and migrated as needed
    - Storage systems become obsolete over time and digital materials should be migrated regularly between storage systems.
    - Migrations to new software platforms when current technology is no longer supported.

Storage Considerations

- What is the current storage capacity of your organization?
    - Is there currently enough space for the project?
    - Will there be space for future projects?
- What is the current file management system?
- Are all your files stored in one place or do you have geographically disperse storage locations?
    - Are you able to produce, manage, and store back-up copies of the files or will you need outside help?
    - How often are backups done?
    - Who is in charge of them?
    - How are they documented?
- Do you have a disaster recovery plan?

Discussing these topics and setting consistent expectations between IT, agency management, project staff, and content owners will go a long way toward making your project successful. The budget, staffing and resources of your organization will help you determine how you can best accomplish these goals.

Preservation vs. Backups

Today's IT systems can provide an excellent option for mass digitization projects. They are highly scalable, and can provide quick access to digitized content, however, traditional IT systems and backups are primarily geared toward active content. The transition to the long-term / permanent preservation of digital records requires a higher level of IT management that should be understood by those managing the digital files to help protect your organization in case of natural disasters, cyberattacks, computer hijacking, accidental deletion, or file corruption.

The National Digital Stewardship Alliance has developed a technology-neutral Levels of Preservation framework to provide guidance on preserving digital content at four progressive levels. It ranges from Level 1 which provides the bare minimum of requirements to minimally protect your data to Level 4 which provides guidelines that will provide the highest likelihood of successful long-term preservation. In reviewing this document, you will likely find your organization is at different points in each of the 6 categories, but you can use this framework to identify where you are currently operating and work toward Level 4 to the greatest extent possible. Of particular importance to the IT / Storage aspect are the sections related to "Storage and Geographic Location", "File Fixity and Data Integrity" and "Information Security".

Storage and Geographic Location

In a traditional IT paradigm, backups periodically perform either incremental or full scheduled backups of files. If at some point a file becomes corrupt, that corrupt file will become part of the backup and by the time it is discovered, there may not be an uncorrupted version available. This is particularly risky for files that are being held permanently by an organization and for this reason IT backups do not take the place of keeping multiple copies of digitized content.

The 3-2-1 rule is often cited as best practice for storing digital files. Ideally, your organization should have 3 copies of each master digital item scanned at the resolution described in the "*Digitization Guidelines*". Realistically, your organization needs to evaluate the costs, staffing and technological infrastructure to evaluate what your capabilities are. It is strongly recommended that the 3-2-1 rule is used for permanent digital records to minimize the risk of losing them over time.

- Copy of a Master Image
  - An exact duplicate of the file at the image's creation.
  - Can be kept in "dark storage" or otherwise not accessed unless there is an issue with the Master Image.
  - A copy of a master image should be migrated and refreshed the same as the Master Image.

- Backup of a Master Image (IT generated)
  - Does not take the place of records management; records should not be retrieved from back up tapes for use or to fulfil an open records request.
  - Backups periodically overwrite files, if a Master Image becomes corrupt, at some point, the backup will rewrite the "good" master with the corrupt file.

Traditional IT backups are appropriate for derivatives of master files that are created at a lower resolution for access copies or thumbnails.

File Fixity and Data Integrity

Every file is made up of bits and bytes that are arrange in a certain way that produces what you see on your computer screen. One of the greatest threats to digital files is the loss of those bits and bytes over time that lead to the corruption of the files so you can no longer access it. This is very different from traditional paper files. If a page in a paper document is damaged in some way, you can likely still read the other pages….there is some loss, but it is minimal. With a digital document, if you lose enough bits and bytes, the entire document is gone. Again, a high risk when you are holding the only record in digital format. One way to monitor file corruption is through performing integrity checks which monitor files over time. These programs calculate a checksum or digital signature of the original file. If over time, the file was to change in some way (corruption, manual change by a person, etc), the program will notify you that something has changed and you can replace it with one of your copies. If you have a robust IT network, you may already have system tools available to perform integrity checks on a periodic basis. Similarly, if one of you geographically distributed copies are with cloud providers, that may be a service they can provide you. If you need to perform integrity checks on your own, there are several free tools that can be implemented at your organization like AV Preserve's Fixity (https://www.avpreserve.com/tools/fixity/).

Types of Storage

Determining the best storage solution for your agency's digital assets involves evaluating likelihood of access, overall cost in maintaining them, and how access will be provided to the digitized records. Every storage type has advantages and disadvantages.

- **Online Storage:** Allows immediate access to records to anyone on your organization's network. Online storage maintains the greatest functionality but is more expensive than other storage options.

- **Near-line Storage:** Uses a system that is not a direct part of your network, but that can be accessed through your network. Files are accessed using an automated process that selects the correct disk/tape from a disk/tape library and makes it accessible.

  Near-line storage is less expensive than online storage, but requires extra time to manipulate both the files and media to access the records. Near-line storage is often used for backups as large quantities of data can be managed quickly.

- **Offline Storage:** Files are not accessible through your network. They may be saved on removable media like external hard drives or magnetic tape. Offline storage is a good option for records that do not need to be accessed frequently.

- **Storage with a Third-Party**: You may also consider using a third-party storage that can store, access, and deliver records to you. As part of their offerings, they may also have the infrastructure to perform integrity checks and store multiple copies of your files in geographically diverse locations. The Wisconsin Public Records Board has issued guidance on the Use of Contractors for Records Management Services and Guidance on the Use of Contractors For Records Management Services in Cloud Computing Environments. The State of Wisconsin Department of Technology (DET) has already vetted cloud service providers and may be a good resource for exploring this option.

A note about removable storage: Removable storage devices include CDs, DVDs, thumb drives, and other types of technology. These devices should NOT be used to store Master Image Files as they have the lowest life expectancy and highest fail rates.

Single external USB storage devices are not ideal being the sole copy of Master File images. Multiple external devices can be used as part of your overall storage plan for smaller institutions in conjunction with other options, but these must be rotated out and replaced at least every 5 years so that cost must be accounted for in long term plans.

Which Option Should I Choose?

Traditional IT procedures tend to backup all items, regardless of content, under a set time schedule. This is done by necessity given the amount digital content they are required to manage. Long-term preservation of digital records may require that different records may be treated differently over time. As stated earlier, it is strongly recommended that the 3-2-1 rule is used for long-term and permanent digital records to minimize the risk of losing them over time. Your organization may also want to implement a more robust plan for records that may not be permanent, but would result in negative public backlash should they be lost. This could be achieved through a combination of copies via on-line, near-line and dark storage with a cloud provider. Records that are digitized to provide access and the paper still remains could have a less robust storage plan since the records could be re-digitized. The storage of the records and any derivatives should be discussed with all relevant parties and documented as part of your digitization plan.

Maintenance

Long-term storage of digital objects demands greater planning and attention than the storage of paper records. Additionally, the expense of storing records electronically exceed the costs associated with storing paper records once the costs of multiple copies and derivatives are calculated.

- All digital media and hardware have a limited life expectancy based on factors such as manufacturing quality, age and condition, handling and maintenance, frequency of access, and storage conditions.
- Hardware and software may be subject to rapid advances in technology or changes in standards.

Due to the life expectancy of both software and hardware, no single digital storage medium or format can be considered “permanent” for the long term storage or preservation of records. The most generous estimate of physical obsolescence is thirty years, while technological obsolescence can be expected within five to ten years. Therefore, assume files will need to be migrated to a new storage medium at regular intervals and periodically to a new format.

Retention

Images must be stored, maintained, and remain accessible for the entire length of the required retention period. When designing your storage and maintenance plan for digitized records, you need to take into account the retention “trigger.” Retention triggers can be easy to determine and implement like the end of a calendar year or the close of a fiscal year. Others are more difficult to determine and implement such as the close of a case or issuance of a final report.

Event based retentions like these can pose difficulties for systems monitoring retention and disposition. They add another layer of necessary metadata, and they require someone to designate the event date in the system so that records can be disposed of or transferred according to record schedules.

Disposition

The majority of records held by local governments will eventually reach a disposition date that involves either destroying the electronic records or transferring them to the Wisconsin Historical Society for permanent preservation. Disposition is a vital step in the records lifecycle and cannot be overlooked when planning and implementing a digitization project. Planning for the disposition of non-records that an agency decides to scan is equally important, and perhaps more challenging as this content does not have a schedule-driven disposition date. It is critical that the project plan includes a review date at a minimum to ensure the content is periodically reviewed and deleted when no longer of use to the agency.

Failure to plan for the disposition of digitized content (both records and non-records) will result in an ever increasing number of items, which will exponentially increase your agency's storage costs over time, make it more difficult to find records, and burden IT staff with the migration of digital content through storage and format migrations over time. Local governments need to remember that digital records are still public records and will need to go through the written notification process before they can be destroyed. Please consult with the Wisconsin Historical Society regarding the proper procedures on notification and destruction of public records.

Resources referenced for this document

Digital Preservation Coalition. Digital Preservation Handbook - Storage. 2017. http://www.dpconline.org/handbook/organisational-activities/storage (accessed June 2017).

LeBlanc, Suzannee. Digital Preservation and Technological Obsolescence: A Risk Assessment Strategy. 2011. http://fiq.ischool.utoronto.ca/index.php/fiq/article/view/15411/12442 (accessed June 2017).

Minnesota Historical Society. Electronic Management Guidelines. March 2012. http://www.mnhs.org/preserve/records/electronicrecords/erpreserve.php (accessed June 2017).

National Digital Stewardship Alliance (NDSA). Levels of Digital Preservation. 2013. http://ndsa.org/activities/levels-of-digital-preservation/ (accessed June 2017).

National Ditital Stewardship Alliance (NDSA). The NDSA Levels of Digital Preservation: An Explanation and Uses. 2013. http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf (accessed June 2017).