

Digital Storage Guide



About this guide

This guide should help you to learn some basic information about digital preservation and what practices are recommended by the Indigitization team. You will learn how to determine the amount of digital storage required for your preservation files and the most common options of digital storage. Lastly, you will be presented with some advice about adopting a digital preservation strategy and selecting storage hardware that matches the needs and technical capacity of your organization.

Digital Preservation and Digital Storage

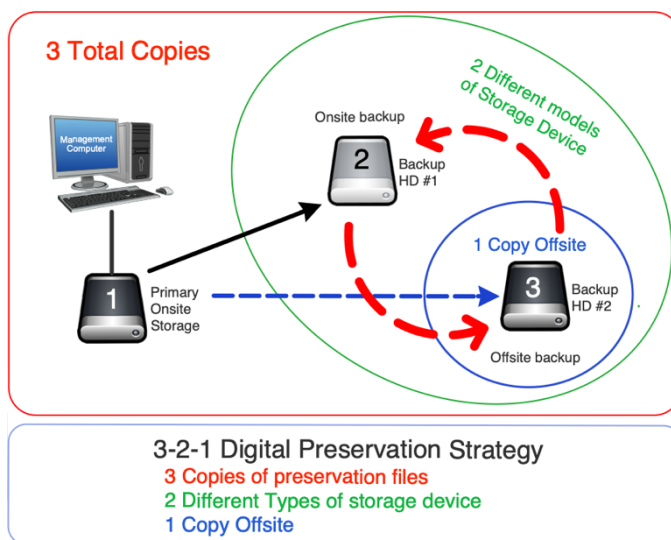
Digital preservation is an important concern for all organizations and should be considered beyond a single digitization project. Most organizations have some way of backing up important office documents but may not be considering how to preserve digital information for the long-term. Digitization projects can be a wonderful entry point to building a long-term digital preservation strategy.

3-2-1 Preservation Strategy

There are many ways that digital files can become lost or corrupt. Files can accidentally be deleted or over-written, storage devices can malfunction, be stolen, or even destroyed. In order to make sure your important files are safe you should have multiple copies of them. Indigitization recommends the “**3-2-1 backup model**” (US-CERT 2012) and the lots of copies keep stuff safe (LOCKSS) approach. You should have 3 copies of your important files, on 2 different types of storage device, with one copy being stored offsite.

Why 3 copies?

Digital preservation files are generally only accessed to make sure that they are readable and unchanged. For this reason, such files may not be checked on for extended periods of time. This makes it possible for multiple storage devices to become lost or malfunction. Having a primary collection of preservation media, with 2 backups makes the collection much safer than if only one backup is created.



Why 2 different types of digital storage?

All electronic equipment has a useable lifespan and will eventually fail as they age. Some storage devices unfortunately will malfunction much earlier than others. While such failures are not common, they do tend to happen to certain models due to design flaws or manufacturing issues. By using more than one type of storage device, you can avoid the risk of having all of your preservation copies on faulty storage models.

Why have 1 copy offsite?

Digital storage equipment can be lost to a disaster, such as fire flood, or electrical surge. It can also be lost to theft or vandalism. At least one copy of your data should be kept offsite, optimally distant enough so that a single natural disaster would not affect both locations.

How Much Preservation Storage Do I Need?



In order to calculate how big your digital preservation collection will be, you will need to first estimate the likely runtime of the audio-visual tapes that will be digitized. This is done by taking an inventory of the tapes and recording the advertised runtime, or estimating the runtime of tapes that are not labeled.

You will also need to decide the quality to which you will digitize your analog media. Depending upon the quality you can easily calculate or find the amount of storage that is required per hour of content. The Indigitization team recommends 24bit/96kHz quality digitization for analog audio, 480X720 10bit YUV uncompressed for analog video, and DV25 files for MiniDV and Digital8 video formats. The amount of storage requirements for these types of files are listed below:

24bit/96kHz audio (stereo)	2.07 GB/hour
480X720 29.97fps 10bit YUV uncompressed	90 GB/hour
DV25 video	11 GB/hour

It is common practice to digitize audio tapes completely and retain the whole recording, including blank sections where there isn't any content. This is not a big storage expense as audio doesn't take up very much digital storage space. Video on the other hand is comparatively very large. In general, blank sections are deleted from digital video preservation files. For this reason, you can consider the advertised runtime for audio to be the storage requirement for your project. For digitized video, the required storage will be somewhat less than the advertised runtime. You can either try to add up the actual runtimes from label information (if available), or you can just consider video content to be a fraction of the advertised runtime. Just make your best estimate based on your knowledge of those recordings.

As you can imagine from the different storage requirements, audio preservation files take up much less storage than video files. You can easily manage the digitized audio preservation files for a project on average sized external hard disks. Video preservation files take up much more

storage space and will generally require some thought be put into very large capacity storage devices, or multiple external hard disks.

What Type of Digital Storage Do I Need?

How does an organization decide on a digital media storage hardware?

Your storage solution will depend on:

- The amount of storage that will be needed for digitized files from this project
- The amount of storage that will be needed for all of your digital preservation files
- Your current budget
- Your organization's technological infrastructure and staffing

As mentioned above, you need to know how much storage is required for the digitization project that you are currently planning. You should also think about how much storage you will need longer-term. This includes both digitized files and recordings that are born-digital. All of your important recordings will need preserving. Don't be afraid to use a fairly simple storage solution if that's what right for your organization. Start with a simple solution as a way to gain digital preservation skills and improve your systems over time.

Main types of digital storage

Internal hard drives:

An internal hard **drive** is a hard **disk** that is mounted directly in a computer. There are two kinds of hard drives: **hard disk drives (HDD)** and **solid-state drives (SSD & NVMe)**:

1. HDDs are an older technology, the most common form of storage, and one of the least expensive types of storage. They use spinning mechanical platters and a moving head to read and write data. These drives are differentiated by the speed that the platters spin. If transfer speed is a high priority, then 7200rpm hard disks are recommended when buying HDDs.



At the time this guide was written HDDs could store up to 20TBs of data. This has risen dramatically over the past few years and makes these drives much better candidates for backing-up large video files.

2. An SSD is generally considerably faster than spinning hard disks. SSDs store data on memory chips, uses less power and can withstand accidental drops. SSDs are more expensive, but prices are dropping. An NVMe drive is just a solid-state drive configured as a computer expansion card.



At the time this guide was written SSDs could store up to 4TBs of data. These devices

are currently recommended for computer operating system storage, and project working storage, but not for long-term storage.

Both types of drives will fail eventually and need to be replaced regularly. However, Indigitization currently recommends HDDs rather than SSDs for **long-term preservation**. Solid-state drives have been shown to suffer from file corruption when not in use. SSDs however are recommended for temporary storage of files during digitization, and other digital media projects.

Internal hard disks are recommended as one solution to holding primary copies of preservation files. These drives should be inside of a secure workstation that is assigned to a person who is authorized to work with the collection. As computers are often stolen during break-ins you should consider the risk of the loss of the digital collection, and the availability of backups. You should also consider the danger of having sensitive files stolen and employ good disk encryption to help mitigate that risk.

External hard drives:

An external hard **drive** is a hard **disk**, as mentioned above, in a case that is external to the computer. It connects to a computer through a **USB** or **Thunderbolt** cable and can be easily moved from computer to computer. External drives can use either HDDs or SSDs for storage. See the recommendations above for which type to use for which purpose (HDD for long-term storage and SDD for fast, short-term storage).



External hard disks are recommended as a fairly simple solution to holding primary, or backup copies of preservation files. Low transfer speed was once an issue for external drives but is mostly alleviated with newer technologies. As these drives can be easily moved and secured in cabinets, or secure storage areas, they are well suited as backup storage. This does change as collection grow in size. It can be difficult to juggle several backup hard disks, and keep track of what files are where.

Some external hard disk enclosures can accommodate 2 hard disks and operate similar to RAID units below. These units can be configured to write to two disks at once, helping to always keep one onsite backup of preservation files.

Redundant array of independent disks (RAID):

In a RAID setup, the data is copied on multiple disks (drives), so that if a single disk failure occurs no data is lost. There are various ways to arrange these disks for your needs.

Some RAID units will connect directly to a computer and act as a large, fast external hard drive. Other RAID units, called Network Addressed Storage (NAS) connect to your network so storage can be accessed by multiple staff members at once. One downside to NAS storage is that it adds a level of risk and network security becomes very important. File transfers across older ethernet networks can also cause bottlenecks in your digitization workflow.



RAID units used to be extremely expensive and really were not a viable option for small organizations. As consumer versions of RAID enclosures have become available, and hard disk capacities have risen, RAID units are becoming a very good option for reasonably priced, large storage devices. These devices are not as efficient in cost per Terabyte as simple hard disks, but they do offer some valuable protection when a drive malfunctions. If the project budget allows, we recommend RAID storage as a good way to store primary preservation files. Having an additional RAID can be a good way to storage offsite backups. This can especially effective where fast internet service, with unlimited bandwidth is present. With some technical knowledge a RAID can be configured to automatically backup to another RAID during off-peak hours.

Note: RAID storage units are sometimes confused as backup solutions, as they provide resiliency against hard drive failures. However, RAID drives are still vulnerable to file corruption, malware, user error, as well as natural disasters or fires. RAID units still should be backed up to 2 other storage devices.

Linear tape open (LTO):



LTO is a **magnetic tape** data storage technology that was first released in 2000, starting with LTO-1. The newest generation is the LTO-9, which was released in 2020. An LTO-9 tape holds 18 TB of **uncompressed** data. The deck is expensive, but the tapes themselves are relatively

inexpensive compared to external hard drives. They are projected to last for over 20 years. Although the tapes have a long life span, the playback decks are only capable of playing tapes two generations back. So, if you have an LTO-9 deck, it can only play LTO-7, LTO-8, and LTO-9 tapes. Therefore, you will have to invest in a new deck frequently, and the tapes must be migrated to the new version. This is a good long-term storage option for organizations that have a mandate for digital heritage preservation and know that their digital collection will be larger budgets. If several organizations are located close to each other, they could pool resources and share an LTO deck.

It may seem odd to digitize audio and video recordings from magnetic tape formats, only to store the digital version back on a magnetic tape format. LTO tape will eventually suffer from the same problems of degradation, and equipment obsolescence that analog audio-visual collections face. LTO storage however does have a very good lifespan compared to other digital technologies. LTO is also a very good value for storage as digital collection sizes increase.

Cloud storage:

Cloud storage is a type of data storage that relies on a hosting company (a cloud storage provider) to keep the data available and accessible. The data is stored remotely on multiple **servers** and is managed by the hosting company.



That company generally owns and maintains both the servers and the physical space where the servers are located. When using cloud storage, there is a risk of data breaches due to a variety of web security issues. There is also a risk that the storage provider could go out of business and your data could be inaccessible or lost. You need to choose stable storage providers, with good security infrastructure and security practices.

There are various types of collection management software that rely on cloud storage. They could be useful for managing and making digitized video available. Mukurtu, Arca, CollectiveAccess, Omeka, and AtoM are all web-based software that support collections management and access. These software developers maintain relationships with cloud storage providers, and these services are used in conjunction with the software.

The following table compares the various storage options:

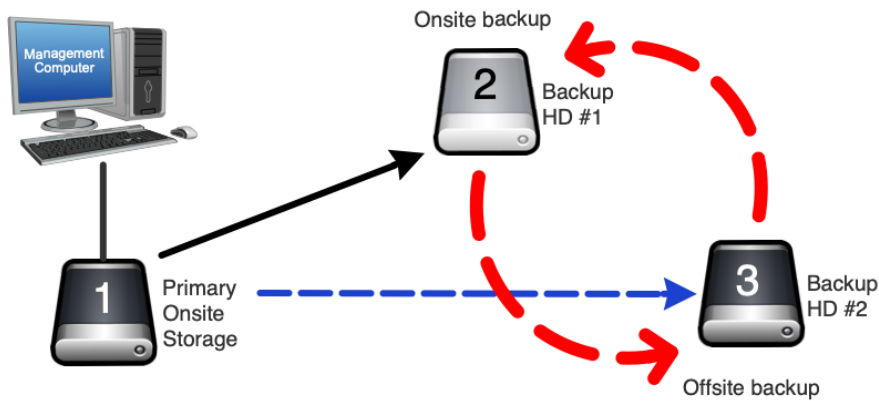
Type of storage	Cost	Technology requirements	Capacity
Internal hard drives	\$ Least expensive, but need to be replaced, duplicated, and backed up regularly. Will suffer from faster equipment wear due to greater power-on time.	Requires computers that have multiple hard disc bays (not small form factor computers or laptops. Works well for organizations with little technical infrastructure, but requires management and scheduling to	This solution works best for small collections of digital materials.

		back up and ensure the drives do not fail. The drives will need to be replaced on a regular basis (at least every three to five years).	
External hard drives	\$ Relatively inexpensive, but need to be replaced, duplicated, and backed up regularly.	Works well for organizations with little technical infrastructure, but requires management and scheduling to back up and ensure the drives do not fail. The drives will need to be replaced on a regular basis (at least every three to five years).	This solution works best for smaller collections of digital materials.
RAID drives	\$\$ More expensive than external hard drives to implement, but a safer storage option. A RAID system automatically backs up materials and alerts the user to drive failures.	Works well for organizations that have some technical infrastructure or an IT person who can help (if needed) with the initial set up. There are many guides and tutorials online showing how to set up a RAID.	This solution works well for most organizations. The RAID setup can be configured to meet the data storage needs of the organization.
LTO (magnetic data tapes)	\$\$\$ Initial cost for the deck is quite high, but the tapes themselves are affordable and stable. Sharing an LTO deck between several organizations is a possibility.	Works well for organizations that have some technical infrastructure or an IT person who can help (if needed) with the initial set up.	Works well for organizations that are planning on digitizing many TB of video. The storage capacity of an LTO-9 tape is 18 TB, and the tape lasts longer than an external hard drive.
Cloud storage solution	\$\$\$ This is an ongoing cost and can be quite expensive for storing larger preservation files. It works well for access files.	Works well for organizations that are planning on using a web-based software for managing or publishing digital media files. Setting up this option would require building a relationship with the developer. If you choose this option, you will want to also store the access files remotely with one of the other storage options.	The capacity of cloud storage depends on how much storage is paid for. Capacity can grow as the number of digital objects increases.

Recommendations

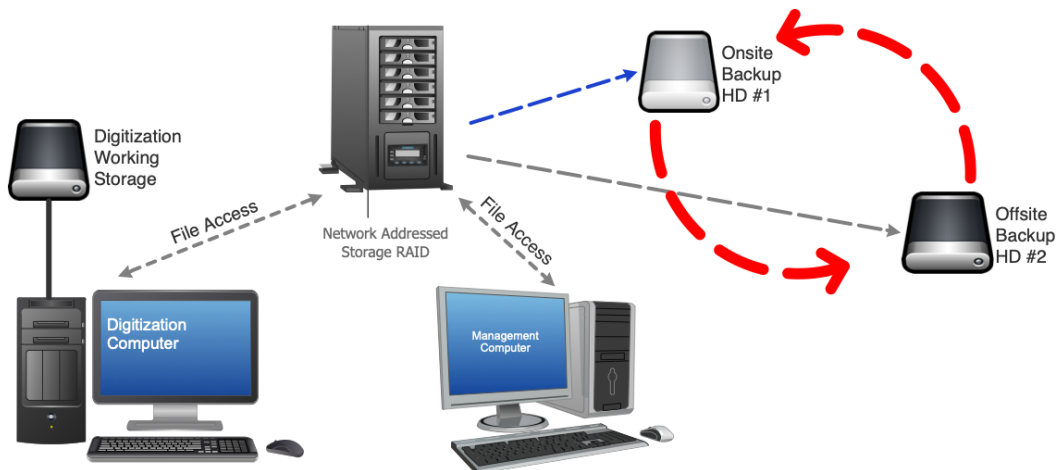
It's hard to create recommendations that work for all organizations. You really need to envision what system you are working towards, but also be realistic about what you can manage at this time.

Simple: External hard drive primary storage with rotating external hard drive backups



If you do not have technical support available to you, or if this support is fairly new to your organization, then we recommend keeping your storage solution as simple as possible. This might be a large internal hard, or external disk drive, or attached RAID enclosure, to store your primary preservation files, and two external hard disks that you rotate between onsite, and offsite as you add more files. This works particularly well for audio preservation projects, where files are relatively small. The availability of larger hard drives is making this solution more practical for video preservation projects where files are much larger.

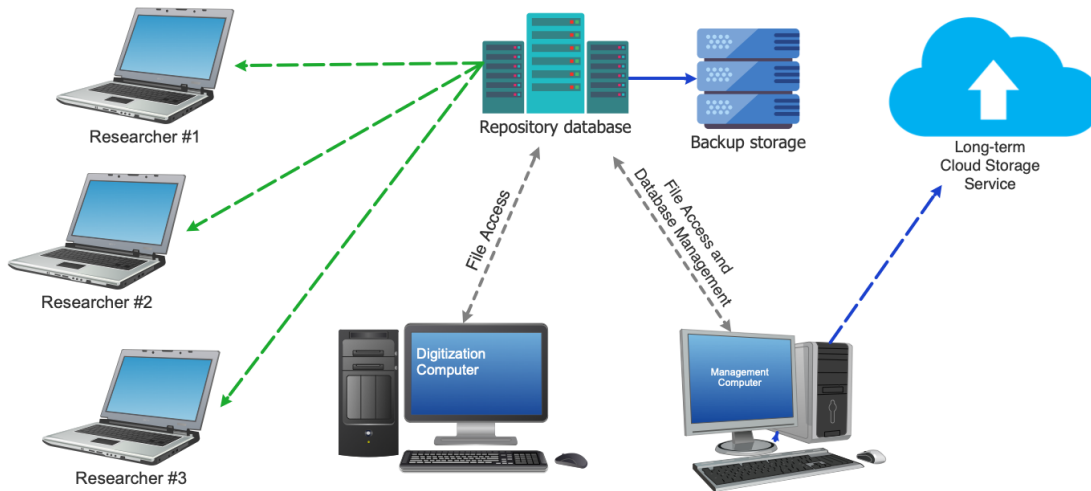
Better: NAS RAID primary storage with rotating external hard drive backups



If you have good technical support in your organization and want to implement a solution that has some advantages over simple hard disks, then a NAS RAID solution, with external hard drives as backup, is a good fit. NAS RAID systems have come down in price dramatically over the past several years. You can now purchase a reasonable sized NAS system, with several hard drive bays, for a few thousand dollars. This solution does take some additional management of network security, but offers advantage in the safety of your onsite files, and the ability to share files with any staff who needs access. Most new NAS RAID enclosures can be

upgraded to enable faster ethernet connections. With a similar upgrade to cabling and relevant workstations, much faster file transfers can be achieved.

Complex: Content Management System or Repository with network and cloud backups:



If you have a very professional IT department, or support, then you can plan for storage and backup systems that might include a Content Management System, a RAID or other network storage system, and perhaps cloud storage for long-term backups. Such systems do take a long time to plan and implement so even if this is your goal, you may wish to purchase an interim storage solution as you work towards this.