

# Digitization Folder Structure Guide

## Overview

There are many steps that every digitized media file must go through. You will create and process many files, and it can be hard to keep track of all the processing stages. This guide will help you to create folders in various areas to assist in keeping your precious digital media recordings organized.

## Digital files created during a digitization project

The number of files contained in digital media collections grows rapidly during digitization but are very consistent in what kinds of files are created. Throughout your digitization work you will typically create the following digital media files:

1. **preservation file (pf)**—a high quality digitized version of an analog audio, or video tape.
2. **primary access file (paf)**—a high quality, possibly edited, copy of the *preservation file*.
3. **access copy (ac)**—you may wish to create versions of the *primary access files* at a variety of qualities to allow for differing methods of distribution (i.e. email, flash drive or Internet download).
4. **subclips**—subclips media files that contain portions of the content of files listed above.

You will create one, or more, preservation files for every tape that is digitized, and then multiple versions of that file to enable efficient access to the content. Through the digitization, processing and moving of files, it is easy to lose track of some of your data. By keeping the files in specific folders during different parts of the digitization process, it is easier to manage each file through the various stages of digitization.

## Establishing a clear folder structure

As mentioned above, establishing a clear folder structure will help you to maintain consistency in your digitization processes. Clear folder structuring also allows you and other users to find and retrieve digital media files efficiently after digitization.

For communities that do not have an established file plan, consider adopting elements of the folder structures from this guide. If you already are using structured folders for your other files then integrate these new files in a manner that is compatible and consistent to the existing arrangement.

For the purposes of a digitization project, you can organize files in four separate areas:

1. **digitization work area**—this is where files live during the digitization process. It is where digitized files are captured to, processed for preservation and access and where quality

control procedures are performed. Folders in this area will be created and named to follow steps in your digitization workflow. This helps to keep the files organized as they are digitized and processed.

2. **preservation area**—once your digitized files have been processed and checked for quality, they can be considered the “preservation file. The “preservation area” is for the long-term storage of these files. This folder structure will be created and named to allow for efficient organization and retrieval of your collection of preservation files.
3. **access file collection**—files that have been processed for access are moved to this location for easy ongoing retrieval for project or research use. These folders should be created and named to allow for efficient retrieval when needed.
4. **Project records**—these are the administrative files that support your digitization efforts and preservation.

The first three areas are for your digitized audio and video files. The project records area is for organizing spreadsheets and textual records relating to the administration and execution of your project. Please see the [Administrative File Management Guide](#) for information on the Project Records area.

## General folder naming practices

There are a few elements that you will want to include to create effective file names for your digital assets, and there are a few things to avoid. The following are helpful practices in naming your digital files:

1. keep folder names relatively short
2. separate words in folder names with an underscore
3. avoid the use of most non-alphanumeric characters
4. add a number to the front of a folder name to force it to appear in the order that you wish

For more information on these file naming practices, please refer to the [Administrative File Management Guide](#).

# 1. Working Area: Folder structure for digitization workstation

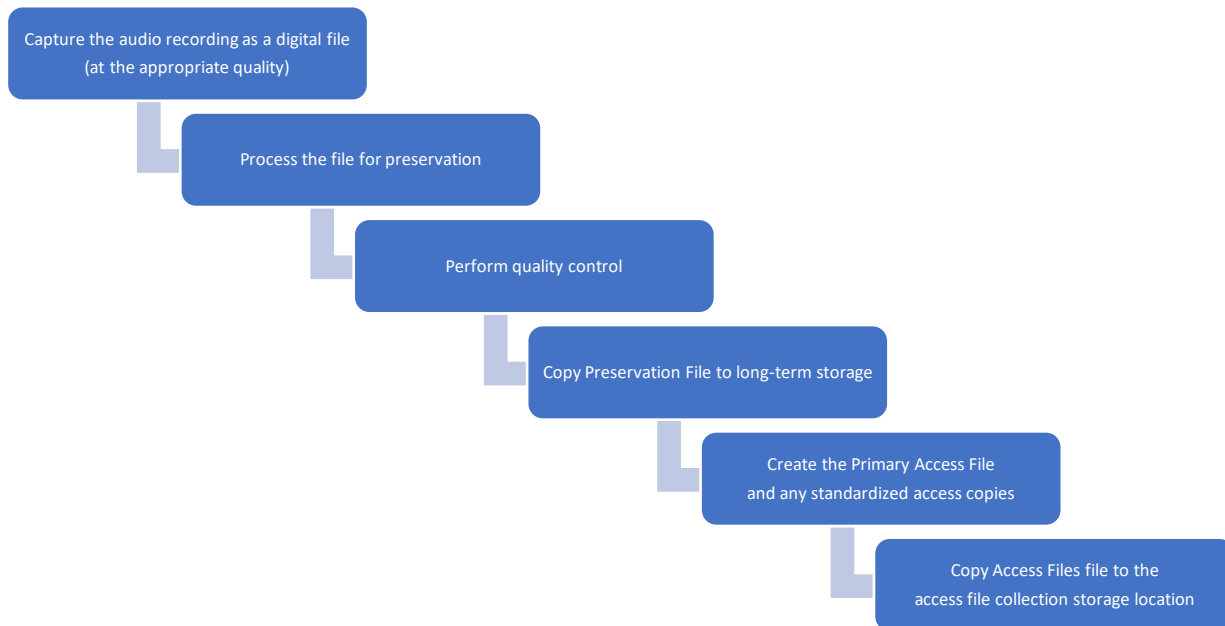
The digitization working area is your workspace for capturing and processing digitized media. Your digitization computer should be purchased with consideration to having enough fast, digital storage available, to process digitized files. In many cases this storage will be large internal hard drives, and in other cases it will require fast external storage devices.

Note: If you are using an external storage drive to hold these folders then you may run into problems digitizing directly to this drive. If this is the case then you may need to digitize files directly to an internal storage drive, and then move those files to the external drive. See the [Digital Storage Guide](#) for more detail about storage devices.

An example structure is provided below. We recommend creating folders based on the major steps in the digitization process. As different formats may have different workflows, it is generally a good idea to create a set of folders for each format that you are digitizing. Detail on each step/subfolder is then provided to explain the workflow associated with this folder structure. Steps may be added or removed from this folder structure if your workflow differs from the one set out below. Use the Indigitization guides for each format to establish your own digitization workflows.

## Folders for audio cassette digitization

The following diagram shows a typical audio cassette digitization workflow. These steps also reflect some of the phases that are used in our Digitization Tracking Spreadsheets.



By creating a folder structure that matches our major workflow steps, and moving digitized files to the next folder when a step is complete, then we have a system where it's easy to keep track of files throughout the digitization process.

📁 01\_Recently\_Digitized\_Audio

- 📁 02\_Processing\_For\_Preservation
- 📁 03\_Ready\_For\_QC
  - 📁 Failed\_QC
- 📁 04\_Preservation\_Files\_Passed\_QC
- 📁 05\_Processing\_For\_Access
- 📁 06\_Primary\_Access\_Files

## Open reel digitization

Digitizing open reel audio usually requires that you perform a playback assessment before digitizing the audio tape. Even though digitizing open reel tapes can be much more complicated than audio cassette digitization, you will find that the folder structure is quite similar.

- 📁 01\_Playback\_Assessment
- 📁 02\_Recently\_Digitized\_Audio
- 📁 03\_Processing\_For\_Preservation
- 📁 04\_Ready\_For\_QC
- 📁 05\_Preservation\_Files\_Passed\_QC
- 📁 06\_Processing\_For\_Access
- 📁 07\_Primary\_Access\_Files

## Video Digitization

Video digitization generally requires that a compressed copy of the preservation files be created in order to easily perform quality control (preservation files can be too large and unwieldy to easily review). In many cases this compressed copy can also be used as the primary access file. For this reason, the Processing\_for\_Access might be earlier in the video digitization workflow, than in audio workflows.

- 📁 01\_Recently\_Digitized\_Video
- 📁 02\_Processing\_For\_Preservation
- 📁 03\_Processing\_For\_Access
- 📁 04\_Ready\_For\_QC
- 📁 05\_Preservation\_Files\_Passed\_QC
- 📁 06\_Primary\_Access\_Files

## DV25 (MiniDV and Digital8 Video)

DV25 video is a compressed format and is more reasonably sized than digitized analog video. DV25 workflows are generally similar to audio cassette and share a similarly named folder structure.


- 📁 01\_Recently\_Transferred\_Video
- 📁 02\_Processing\_For\_Preservation
- 📁 03\_Ready\_For\_QC
- 📁 04\_Preservation\_Files\_Passed\_QC
- 📁 05\_Processing\_For\_Access


📁 06\_Primary\_Access\_Files


## The "Working Area"

The files that occupy these folders are actively being processed. They should not stay in these folders for extended periods of time. These files are generally considered ephemeral until they are ready to be used as primary preservation files, and primary access files. Due to their size, these files are generally not backup up, in favor of finalizing the processing and moving them to more secure locations quickly. **You should never keep more files in this area than you can afford to lose if this hard disk fails.**

Digitization computer – Working Hard Drive
📁 Digitization_Working_Area
📁 Audio
📁 01_Recently_Digitized_Audio
📁 02_Processing_For_Preservation
📁 03_Ready_For_QC
📁 Failed_QC
📁 04_Preservation_Files_Passed_QC
📁 05_Processing_For_Access
📁 06_Primary_Access_Files
📁 Open_Reel_Audio
📁 01_Playback_Assessment
📁 02_Recently_Digitized_Audio
📁 03_Processing_For_Preservation
📁 04_Ready_For_QC
📁 05_Preservation_Files_Passed_QC
📁 06_Processing_For_Access
📁 07_Primary_Access_Files
📁 Video
📁 01_Recently_Digitized_Video
📁 02_Processing_For_Preservation
📁 03_Processing_For_Access
📁 04_Ready_For_QC
📁 05_Preservation_Files_Passed_QC
📁 06_Primary_Access_Files
📁 DV_Video
📁 01_Recently_Transferred_Video
📁 02_Processing_For_Preservation
📁 03_Ready_For_QC

 04\_Preservation\_Files\_Passed\_QC

 05\_Processing\_For\_Access

 06\_Primary\_Access\_Files

## 2. Folder structures for the Preservation Area

The Preservation Area of storage is the place that your permanent preservation collection is held. This should be placed on a relatively stable storage medium, backed up and regularly checked for file integrity. We recommend creating folders that allow for efficient location and retrieval of the individual files. It is a good idea to use a folder structure that can separate your collection into manageable numbers, and sizes for each folder. This will make it easier to find files within the folder and run integrity-checks by folder. You should consider how many files your collection will likely contain and if your collection will continue to grow.

Note that you must maintain a collection of your *preservation files*, but can also keep versions of access files here as well (as seen in the example below).

External Hard Disk or Network Addressed Storage (NAS) RAID  
(primary preservation drive): Simple file naming

- 📁 **Preservation\_Area**
  - (optional) 📁 **Audio\_Cassette**
    - 📁 **cc0001**
      - 📄 cc0001-1a\_ac.wav
      - 📄 cc0001-1a\_paf.wav
      - 📄 cc0001-1a\_pf.wav
      - 📄 cc0001-1b\_ac.wav
      - 📄 cc0001-1b\_paf.wav
      - 📄 cc0001-1b\_pf.wav
    - 📁 **cc0002**
      - 📄 cc0002-1a\_ac.wav
      - 📄 cc0002-1a\_paf.wav
      - 📄 cc0002-1a\_pf.wav
      - 📄 cc0002-1b\_ac.wav
      - 📄 cc0002-1b\_paf.wav
      - 📄 cc0002-1b\_pf.wav
    - 📁 **cc0003**
      - 📄 cc0003-1a\_ac.wav
      - 📄 cc0003-1a\_paf.wav
      - 📄 cc0003-1a\_pf.wav
      - 📄 cc0003-1b\_ac.wav
      - 📄 cc0003-1b\_paf.wav
      - 📄 cc0003-1b\_pf.wav
  - (optional) 📁 **Images**
    - 📁 Sub-folder1
    - 📁 Sub-folder2
    - 📁 Sub-folder3
  - (optional) 📁 **Video**
    - 📁 Sub-folder1
    - 📁 Sub-folder2
    - 📁 Sub-folder3





External Hard Disk or Network Addressed Storage (NAS) RAID  
(primary preservation drive): File naming with sub-collection identifier.

📁 **Preservation\_Area**

(optional) 📁 **A2021-01**

📁 **A2021-01cc01**

- 📄 A2021cc01-1a \_ac.wav
- 📄 A2021cc01-1a \_paf.wav
- 📄 A2021cc01-1a\_pf.wav
- 📄 A2021cc01-1b\_ac.wav
- 📄 A2021cc01-1b\_ paf.wav
- 📄 A2021cc01-1b\_pf.wav

📁 **A2021-01cc02**

- 📄 A2021cc02-1a \_ac.wav
- 📄 A2021cc02-1a\_paf.wav
- 📄 A2021cc02-1a\_pf.wav
- 📄 A2021cc02-1b\_ac.wav
- 📄 A2021cc02-1b\_paf.wav
- 📄 A2021cc02-1b\_pf.wav

📁 **A2021-01cc03**

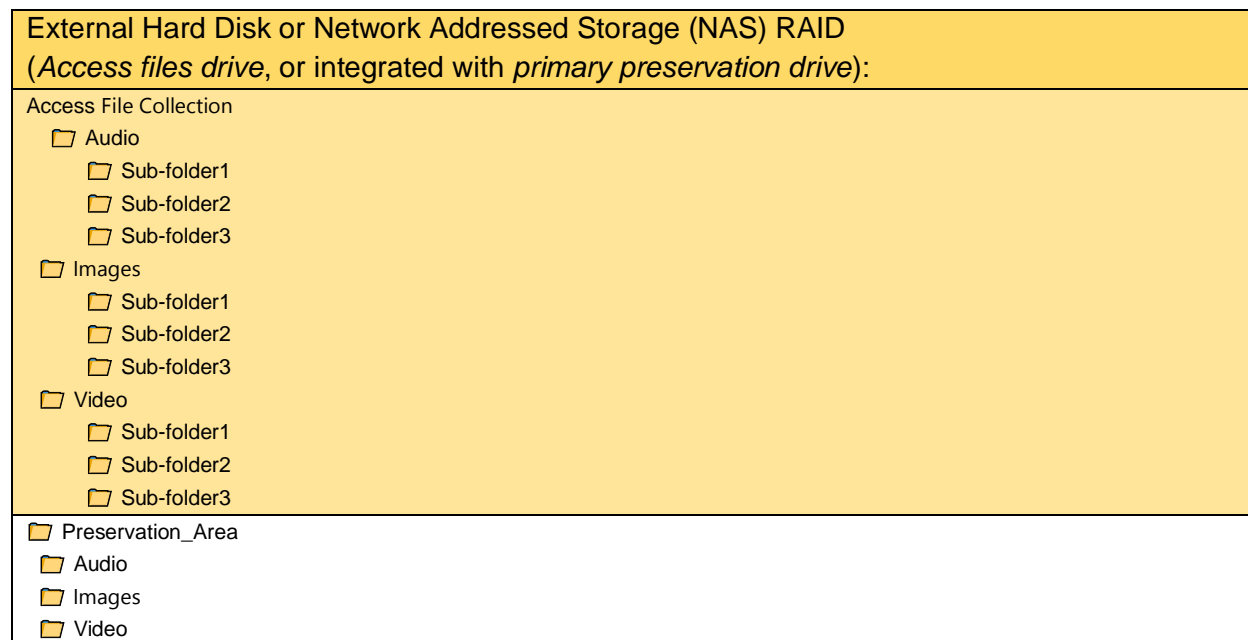
- 📄 A2021cc03-1a\_ac.wav
- 📄 A2021cc03-1a\_paf.wav
- 📄 A2021cc03-1a\_pf.wav
- 📄 A2021cc03-1b\_ac.wav
- 📄 A2021cc03-1b\_paf.wav
- 📄 A2021cc03-1b\_pf.wav

Within the folders for Audio, Images and Video you can create sub-folders to further separate your files based on sub-collection, year digitized or other characteristic of the media.

These files should follow the 3-2-1 backup method. There should be at least three copies of your preservation files. Use at least two different types of storage (i.e. hard disk brand/model) so that they would not be subject to similar manufacturing defects. Lastly, store one copy off-site, far enough away that it would not be subject to similar disaster effects (i.e. fire, flood etc...)

### 3. Folder structures for the Access File Area

Your access file collection is the group of media files that can be provided to authorized users for research or media projects. These files should be on a storage medium that is compatible with other digital files in use at your organization. For efficiency you can keep these files on the same external hard disk as your preservation files. Alternately you can keep them on a different hard disk to increase the longevity of your *primary preservation drive*. You should keep at least one backup of these files as recreating them would require a great deal of expense in time and labour. You do not need to place the same level of effort into long-term file-integrity as you do with your *Preservation File* collection. We recommend creating folders similar to your preservation file collection to allow for efficient location and retrieval of these files. It is a good idea to use a folder structure that can separate your collection into manageable numbers, and sizes for each folder.



Within the folders for Audio, Images and Video you can create sub-folders to further separate your files based on sub-collection, year digitized or other characteristic of the media.

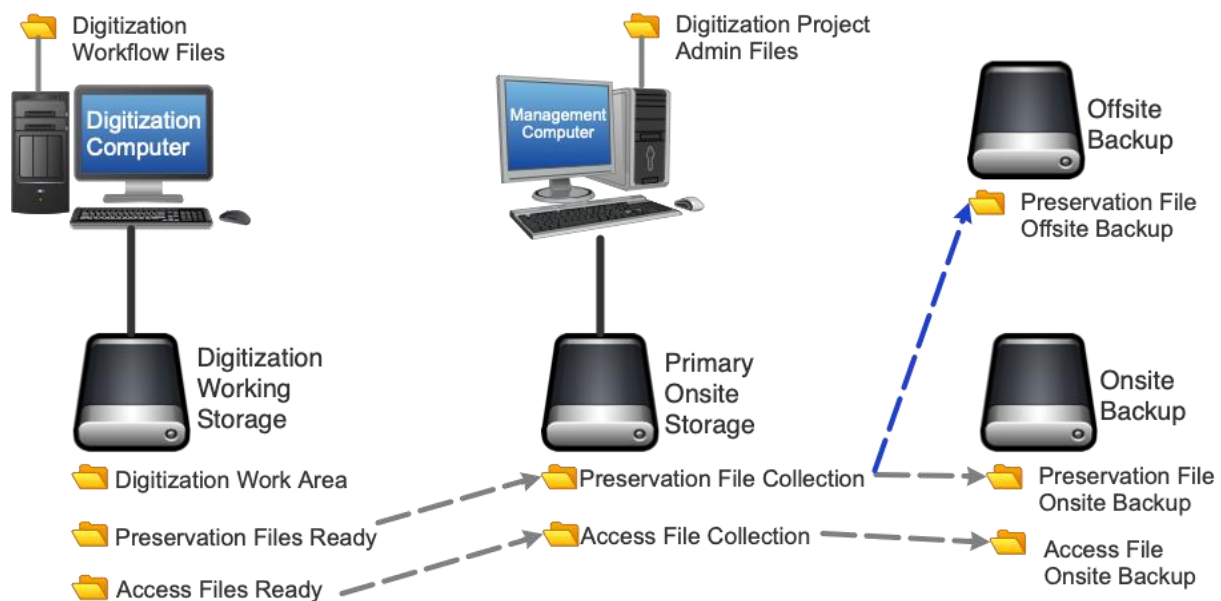
These files should follow the 3-2-1 backup method. There should be at least three copies of your preservation files. Use at least two different types of storage (i.e. hard disk brand/model) so that they would not be subject to similar manufacturing defects. Lastly, store one copy off-site, far enough away that it would not be subject to similar disaster effects (i.e. fire, flood etc...)

## Folder structure areas and storage hardware

Where these four folder structure areas will reside in your digital media storage hardware will depend on what storage hardware you purchase, which and how many staff will need to access which files, and security/privacy considerations. Your selection of storage hardware should be informed by which formats of analogue media, and how many hours of each, you will be digitizing. Refer to the [Digital Storage Guide](#) for more details on storage options.

If your organization does not have a file server already in use, then setting up the folder structures on external hard drives is likely the best approach (see diagram below). Storage for your digitization area files can take the form of an internal hard drive built into the digitization computer, or as an external storage drive. Holding this storage on an external drive allows it to be passed between your team members easily as needed.

If you use internal storage, then these files will need to be copied to an intermediary external hard disk to share with team members as needed. It is generally best to have the preservation area on its own drive, whether internal, or external, to limit the possibility of inadvertent changes being made to your **preservation files**.



You may notice that some of the project records from the project records area have been placed on the digitization area internal hard drive. This is because the person doing the digitization will need access to these folders in their digitization workflow. Consider who will need primary access to which folders when dividing folders between storage hardware.

<b>Digitization Computer</b> <b>Internal HD:</b>	<b>External HD:</b> <b>Digitization Working Drive</b> <b>Digitization Area</b>
<ul style="list-style-type: none"> <li>📁 Digitization_Project <ul style="list-style-type: none"> <li>📁 -10_Inventories</li> <li>📁 -11_Tracking_Spreadsheet</li> <li>📁 -12_Condition_Assessments</li> <li>📁 -13_Digitization_Logs</li> <li>📁 -14_Collections_Management</li> <li>📁 -15_Transcriptions</li> <li>📁 -16_Backups</li> <li>📁 -XX_Reference</li> <li>📄 RM_Readme</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>📁 Digitization_Area <ul style="list-style-type: none"> <li>📁 01_Recently_Digitized_Audio</li> <li>📁 02_Processing_For_Preservation</li> <li>📁 03_Ready_For_QC</li> <li>📁 04_Preservation_Files_Passed_QC</li> <li>📁 05_Processing_For_Access</li> <li>📁 06_Primary_Access_Files</li> </ul> </li> </ul>
<b>Manager Computer</b> <b>Internal HD:</b>	<b>External HD:</b> <b>Primary Preservation Collection</b> <b>Preservation Area</b>
<ul style="list-style-type: none"> <li>📁 Digitization_Project <ul style="list-style-type: none"> <li>📁 -00_Policies_Procedures</li> <li>📁 -01_General</li> <li>📁 -02_Project_Planning</li> <li>📁 -03_Reports</li> <li>📁 -04_Correspondence</li> <li>📁 -05_Project_Meetings</li> <li>📁 -06_Finances</li> <li>📄 RM_Readme</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>📁 Access File Collection <ul style="list-style-type: none"> <li>📁 Audio</li> <li>📁 Images</li> <li>📁 Video</li> </ul> </li> <li>📁 Preservation_Area <ul style="list-style-type: none"> <li>📁 Audio</li> <li>📁 Images</li> <li>📁 Video</li> </ul> </li> </ul>
<b>External HD:</b> <b>Onsite Preservation Backup</b> <b>Preservation Area</b>	<b>External HD:</b> <b>Offsite Preservation Backup</b> <b>Preservation Area</b>
<ul style="list-style-type: none"> <li>📁 Access File Collection <ul style="list-style-type: none"> <li>📁 Audio</li> <li>📁 Images</li> <li>📁 Video</li> </ul> </li> <li>📁 Preservation_Area <ul style="list-style-type: none"> <li>📁 Audio</li> <li>📁 Images</li> <li>📁 Video</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>📁 Preservation_Area <ul style="list-style-type: none"> <li>📁 Audio</li> <li>📁 Images</li> <li>📁 Video</li> </ul> </li> </ul>

If you have Network Addressed RAID storage on your office network, you might set up your filing areas in the following manner:

