

# Digital Media Formats Guide

## About this guide

There are a variety of different digital media formats that have been created for specific purposes. These differences cause some types of files to be better, or worse, for preserving digitized media or making media accessible to researchers. This guide describes some of the differences between media files and provides recommendations for media formats you might use in your digitization projects.

## Types of media files

There have been dozens of different media file formats created over the years. Formats that allow for “high-quality” media tend to be large and may not work well for consumer applications, such as streaming media, or portable media playback. File formats that do support consumer demands for media may not be good for audio or video editing, or for long-term preservation. It is important to choose file formats that support the intended purpose of those files.

## Files in a digitization project

Typically, a few different types of digital media files are created during digitization. Each type of file should be maintained in a format that has appropriate qualities.

- *preservation file* — a high quality digital copy of an analog audio, or video tape. This file will have little or no alterations, acting as a faithful representation of the original tape.
  - The format should retain all of the detail captured during digitization.
  - The format should be “open” and not require software purchase or licencing.
  - The format should be stable. It should be widely used and supported so that it doesn’t become quickly obsolete.
- *primary access file* — a high quality copy of the *preservation file* intended for access by a person who is allowed to listen to or view the content. This file may be edited to improve the visual or audible clarity of the media. You may also edit, or redact, sensitive information to allow for different levels of access.
  - The format should retain enough detail from the preservation file to create access copy files as needed.
  - The format should ideally be “open” and not require software purchase but could be proprietary if the benefits are compelling.
  - The format should be reasonably stable. All formats should be evaluated for stability and important files migrated to new formats if necessary.

- *access copy (ac)*— you may wish to create copies 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).
  - The file format should be able to maintain a good representation of the primary access file at the desired file-size and quality settings.
  - The format usually should be “open” and not require software purchase to play. An exception to this would be if your organization is using a specific distribution platform to accommodate monetization or rights-management (not common for community organizations).
- *subclips* — subclips are media files that contain portions of the content of files listed above. These files are generally created when that content is requested and can be ephemeral or managed with your digital media collection.
  - These files can be created in any format that serves the intended use.

## Properties of media files

Media files are made up of metadata, or information about the media, and a number of audio-visual streams. The types of information and media that can be in a file are determined by the media “container” type, which is often the same as the file extension (i.e. WAV, AVI, MOV).

### Container/wrapper

The container, or wrapper, refers to the structure of a digital media file. Containers are designed to hold specific information about the digital media, and specific kinds of digital media data. All containers store information that software needs to play the audio-visual data properly. Other containers allow users, or software, to embed information about the content of the file.

Containers also are designed to accommodate only certain types of audio and/or video data.

- **Open vs. proprietary:** One of the main differences in containers is whether or not the container is open (free, and well-described), or owned by private interests.
  - If a container is open then you will always be able to play, and convert, such media files.
  - If a container is proprietary, then there is a risk that the owner of that media type will enforce expensive licencing in software to access that media type.
- **Codec support:** Codec stands for **code-decode**. This is the way that audio, or video information is digitally recorded. Each media container supports a set list of codecs. Different codecs will store audio-visual information as either uncompressed or compressed data. Good digital preservation practices require media be digitized and stored as uncompressed data, or with lossless compression.
  - **Uncompressed:** Media files that are encoded in an uncompressed format retain all of the available media information. These files represent the highest integrity for the media, but also the highest digital storage requirement.

- **Lossless Compression:** Some codecs can employ lossless compression technologies to reduce the storage requirements without sacrificing the integrity of the original media. At present these file types tend to require a lot of processing power to play back and are not widely supported by common media editing programs.
- **Lossy Compression:** Most compression types that are widely used, employ forms of lossy compression that are designed to sacrifice quality for file size, according to the needs of the end-user. Examples of this are video files that are intended to stream from the internet favour small file sizes, while video files that are used by professional video editors favour visual quality.

## Preservation file and primary access file formats

Before starting an audio or video digitization project, you must choose **preservation file formats** that work for your organization. A **preservation file** (referred to in the quote below as a “preservation master file”, Indigitization is choosing to not use this term) is the file that will be saved for preservation purposes. They are large files meant for long-term storage and aren't meant to be played for access. According to the U.S. National Archives:

Preservation master files are created at high to maximum capture specifications and can therefore serve a variety of purposes, including satisfying long-term preservation needs as well as fulfilling most researcher requests for high-quality files.

Preservation master files may capture additional information about the original beyond the content itself. Because they are created to high capture standards, preservation master files could take the place of the original record if the original was destroyed, damaged, or not retained. Preservation masters generally do not undergo significant processing or editing.

Analog preservation masters are usually created from the most original record copy available. Analog preservation masters are used to create reproduction master copies that satisfy most customer requests, as well as protect the original from loss of information. In most cases, the preservation master is not accessed directly unless the reproduction master is damaged, destroyed, or otherwise compromised.

Preservation masters are often used to make other copies including reproduction and distribution copies. Both physical format and digital preservation masters are managed and maintained in a secure storage environment and are retained over the long term.<sup>1</sup>

---

<sup>1</sup> “Preservation Master,” National Archives, The National Archives and Records Administration, <https://www.archives.gov/preservation/products/definitions/preservation-copy.html>.

An **access file** is a version of your preservation file that is intended to be used by researchers to view/listen to the content. It's a good idea to create a **primary access file**, which is sufficient quality for the vast majority of your access needs. This file is used to make other access copies. For example, a web platform (such as YouTube or Vimeo) might ask for a particular file format, you would use the primary access file to transcode to the new format, instead of the preservation file. The primary access file can be stored, copied and accessed regularly. They are meant to be easily played, shared, and accessed (according to your organization's access protocols).

Unlike preservation files, which are meant to be placed in storage and accessed infrequently, access files are meant to be played and used. They can be uploaded to a collections management software, shared using hardware such as USB drives, and/or sent via file sharing software such as Dropbox or WeTransfer. Because these files need to be used and accessed, it is OK for them to be smaller and have a lower resolution.

## Sustainability and risk

When deciding on formats to use for preservation purposes you should examine the qualities of different formats to determine if they are suitable for use in your organization.

- **Cost:** Some formats must be licenced in order to use or are only available if expensive software packages are purchased. What does it cost to use a format? Can you justify purchasing the format licence or software in the future when computers are updated or upgraded.
- **Complexity:** Some formats are very suitable for preservation, or access purposes, but may not be supported by common media editing programs. If this is the case, then will the lack of tools affect your ability to efficiently use your media files.
- **Risk:** If a format is proprietary, and the company suddenly went out of business, then you could lose the ability to activate licences and access your files. A company could also suddenly raise the price of their licences, which could adversely affect organizations with meager budgets.

Audio files have clear, defined, and widely adopted recommendations for preservation file formats. However, there has been no consensus in the archival community to select one file format for digitized video.

Each organization must select a file format for the preservation of digital video that suits their resources, capacities, and structure.

## Recommended Audio File Formats

The following formats are recommended for digitizing audio collections utilizing the Indigitization guides.

Audio Preservation File: Broadcast WAV (BWF) file – uncompressed, mono, or stereo

The BWF audio standard was developed by the European Broadcast Union and is an expansion of the popular WAV format. BWF was created to allow content producers to exchange high-quality files, while maintaining file-integrity and intellectual property information. To this end, metadata capacities were increased and checksums could be embedded. Any standard WAV file can be converted to BWF via the free, open source application *BWF MetadEdit*.

- Broadcast WAV container
- Microsoft uncompressed audio codec (1 or 2 streams)
- Recommend 24bit, 96kHz sampling (24bit, 48kHz minimum)
- File size limit of 4 Gigabytes(GBs). For files that are very long and require more than 4GBs use RF64 version of broadcast WAV)

Audio Primary Access File: Broadcast WAV (BWF) file – uncompressed, mono, or stereo

As audio files do not generally require large file sizes it is recommended to maintain Primary Access Files in the broadcast WAV format Primary to preserve quality. While you could choose a compressed format to save on digital storage space the dollar value saved would not be very much for small or medium sized collections. You might however choose to store these files at 24bit, 48kHz rather than the 24bit, 96 kHz recommended for preservation. This would cut the storage space required in half.

Audio Access Copy: WAV or MP3

Access copies can be made in any format that suits the research needs.

- A copy of the Primary Access File (WAV) could be provided if file size is not prohibitive for the delivery method. These would generally fit onto portable storage such as hard drives, or thumb drives.
- MP3 files can be much smaller than uncompressed WAV files and yet maintain very good quality. These files are generally much faster to make copies of, or transfer via the internet.

## Recommended Video File Formats

The following formats are recommended for digitizing video collections utilizing the Indigitization guides.

## Video Preservation file: .avi or .mov wrapper with an uncompressed V210 codec

*Note: The default file format for a Windows computer is .avi, and the default file format for a Mac computer is .mov.*

This preservation file format is recommended for the following reasons:

- Both wrapper and codec are stable, well **disclosed**, and widely adopted.
- This is a medium-cost solution. Although the files are large, they maintain all the necessary data and are simple to implement with minimal training. This is a sustainable choice for most organizations.
- The .mov and .avi containers are proprietary file formats (developed by Apple/Microsoft). However, they are well documented and have the ability to support timecodes and closed captioning.
- They offer good support for high-quality video.
- These file formats can be captured using the Blackmagic Media Express software with the settings provided.
- The formats are easily understood and identified by free file analysis and playback software like MediaInfo, ffmpeg, and VLC.
- .avi files can be embedded metadata using AVI MetaEdit (<https://mediaarea.net/AVIMetaEdit>).

## Video Primary Access file: .mp4 wrapper with an H.264 codec

This primary access file format is recommended for the following reasons:

- Both the wrapper and codec are widely adopted and can be played back on most computer or mobile platforms.
- This file format can be easily uploaded and streamed online to provide access through various platforms.
- The H.264 codec contains more video information than the basic mpeg4 compression. The file is high quality, but small enough to be streamed, copied, and stored efficiently.
- Shutter Encoder (an open source and free software) can create an H.264/.mp4 file from the preservation file using its built-in and user-friendly interface.

## Video Access file: .mp4 wrapper with an H.264 codec

This access file format is recommended for the following reasons:

- Both the wrapper and codec are widely adopted and can be played back on both Mac and Windows machines.
- This file format can be easily uploaded and streamed online to provide access through various platforms.

- The H.264 codec contains more video information than the basic mpeg4 compression. The file is high quality, but small enough to be streamed, copied, and stored efficiently.
- Shutter Encoder (an open source and free software) can create an H.264/.mp4 file from the preservation file using its built-in and user-friendly interface.

## Digital Video tape (MiniDV, Digital8) file format recommendations

The following formats are recommended for capturing video collections from NTSC digital videotapes utilizing the Indigitization guides.

### Digital Video Preservation file: .mov or .avi wrapper with a DV25 codec

*Note: .avi is the default file format for a Windows computer, .mov is the default file format for a Mac computer*

This preservation file format is recommended for the following reasons:

- The original video was recorded as DV25, and can be captured from tape formats without loss of detail. This should be maintained in the preservation file.
- The wrapper is well disclosed and widely adopted.
- This is a medium-cost solution. Although the files are relatively large, they maintain all the original data and are simple to implement with minimal training. This is a sustainable choice for most organizations.
- The .mov and .avi containers are proprietary file formats (developed by Apple/Microsoft). However, they are well documented and have the ability to support timecodes and closed captioning.
- They offer good support for high-quality video.
- The formats are easily understood and identified by free file analysis and playback software like MediaInfo, ffmpeg, and VLC.
- .avi files can be embedded with metadata using AVI MetaEdit (<https://mediaarea.net/AVIMetaEdit>)

### Digital Video Access file: .mp4 wrapper with an H.264 codec

This access file format is recommended for the following reasons:

- Both the wrapper and codec are widely adopted and can be played back on both Mac and Windows machines.
- This file format can be easily uploaded and streamed online to provide access through various platforms.
- The H.264 codec contains more video information than the basic mpeg4 compression. The file is high quality, but small enough to be streamed, copied, and stored efficiently.
- Shutter Encoder (an open source and free software) can create an H.264/.mp4 file from the preservation file using its built-in and user-friendly interface.