Proposal for Processing Atlas Digitized Video Collection

Prepared by Eric Silberberg
June 17, 2021

Recommendations In Brief:
For purposes of providing reference images within the Inmagic Genie catalog, an MOV copy of all video files should be made using FFmpeg and placed in the *reference image\video* directory. ARM's installation of Inmagic Genie appears to only support downloads of MOV files from the catalog. The MP4 files created by Atlas should be treated as archival master files and accessed on a limited basis. They should be stored elsewhere and preferably in at least two different locations.

Table of Contents

1. File Formats

The following tables were compiled based on file manifests of directories within the Atlas digitization project. See directory_manifest_summary.xlsx for more information. Though there is a combination of audio and video formats, this proposal will focus on video items.

Table 1.1: AV items in VIDEO-AUDIO directory

Format	No. of Items
AIFF	89
M4V	42
MOV	6
MP3	70
MP4	463
WAV	19
TOTAL	689

Table 1.2: AV item in VIDEO-AUDIO directory and subdirectories

Format	No. of Items
AIFF	103
M4V	42
M4A	28
MOV	18
MP3	70
MP4	469
WAV	28
TOTAL	758

Reference images (RI) of videos attached to records in Inmagic Genie catalog are accessible (downloadable from the catalog entry) when these RIs are in MOV format. Atlas provided digitized video items in MP4 format. This raises the question for ACLU Archive and Records Management Department (ARM) if MOV is a wise choice for digital video archiving.

1.1 Codecs

Uncompressed video files provide full video data and are extremely large, as would be expected. To create manageable file sizes and make video capture and reproduction time-efficient, a video codec is applied to compress the file. In other words, a video codec is a computer program that applies a specific algorithm (compression) to encode and decode video during capture and reproduction.[1] This compression comes in two principal types: lossless and lossy.

---

[1] Weise, M., & Waynand, D. (2007). *How video works: From analog to high definition* (2nd ed). Focal Press, 173.

1. Lossless compression, offered by the V210 and FFV1 codecs among others, results in a more manageable file without sacrificing quality. It accomplishes this by removing redundant information such as static regions of an image. The file is still quite large, but this is generally the type of compression preferred by AV archives as it strikes a balance between file size and data preservation.

2. Lossy compression, such as the H264 codec chosen by Atlas during the digitization project, creates considerably smaller files than lossless codecs. It provides an approximation of the uncompressed file, which means that some information is lost during compression. However, this difference is difficult to detect when watching a lossy and lossless compressed file side by side on a computer monitor. Transcoding from lossy and lossless will not recuperate the lost data. H264 is widely supported by media players and video editing software. Furthermore, the relatively small file size means that it is cost-effective way of preserving the video collection.[2]

Recommendations In Brief: The files have been digitized using the H264 codec (lossy compression). This is perfectly acceptable for ARM's needs when addressing reference or media requests. If another opportunity arises to digitize the collection, a lossless codec such as FFV1 should be considered if there is funding and sufficient disc space.

1.2 Containers

While codecs are applied to compress video information, a video container is the envelope used to package together all the different streams associated with a video. These include the video, audio, metadata, subtitles, chapter markings, etc. Atlas encapsulated the digitized videos in MP4. This container is widely supported and is an international standard. The QuickTime File Format (MOV) is Apple's proprietary format, and it is also widely supported. The Library of Congress recommends this format.[3] MP4 and MOV share the same functionality.

Recommendations In Brief: MP4 files should be converted to MOV as this is the preferred format of Inmagic Genie. The MP4 versions should be saved as the archival master file outside of Inmagic Genie to safeguard against human error or network issues.

1.3 Other Considerations

1. Directory **10511_P_ACLU_PSA_Outlawed-torture-disappearances_2006_16044_media** is one element within CatID 16044. It is an informational CD archived in HTML and a DOC.

---

[2] Weise & Waynand, 157

[3] https://www.loc.gov/preservation/resources/rfs/moving.html

2. Directory **10510_P_ACLU_AUDIO_Civic-Duty-Movie-Screening_2007_16044_code**
contains two .cda files, which do not contain audio information; they are track metadata
from a CD.

2. Transcoding

This section details how the files in MP4 containers can be converted to MOV containers while
preserving the H264 codec.

2.1 FFmpeg

This software offers the simplest method to convert the 469 MP4 files to MOV. With small
adjustments, this can be used to convert audio files as well. FFmpeg is a suite of free and open-
source libraries and programs for handling audio and video files. It also comes with FFplay, a
simple media player useful when playing uncommon file formats, and FFprobe, a tool to query
technical metadata. These three programs run off the command line and follow a straightforward
syntax:

ffmpeg -i "*file path*" output.mov

This code snippet would result in the creation of a MOV copy of the input (-i) file. Furthermore,
FFmpeg can be incorporated into a Powershell script that can batch convert MP4 files to MOV.

FFmpeg documentation: https://www.ffmpeg.org/ffmpeg.html

Download: https://amiaopensource.github.io/ffmpeg-artschool/#windows

2.2 VLC

This method is appropriate if a user prefers a graphical user interface. However, this process
requires that the user pay attention to many of the fine details. Conversion of multiple files is
possible, but not to MOV. Thus, the user would have to convert Atlas' digital files one by one.

1. Open VLC
2. Media > Covert/Save
3. Add the file you would like to convert
4. Click Convert/Save
5. Create profile of desired video and audio codecs
6. Create output file destination
7. Save as type: All (*.*)
8. Manually type .mov at the end of the filename.
9. Click Save, then click Start.

3. Digital Preservation

A guiding principle for the preservation of audiovisual collections is the NDSA Levels of Digital Preservation. This document identifies storage and geographic location, file fixity, information security, metadata, and file formats as distinct areas that deserve attention when formulating a digital preservation plan.[4]

The following chart takes the NDSA recommendations as a point of reference for ARM's digital preservation options.

Focus	NDSA Recommendation	ARM Option
Storage	At least three copies in different geographic locations.	MOV access file in the Inmagic Genie directory. Copy of MP4 on ACLU network, external hard drive in office, off-site network storage with TRIM.
Integrity	Check fixity of content at fixed intervals.	Powershell can generate SHA256 hash value by default. Also can generate SHA1 and MD5 among others. Simple script can be written to audit fixity and compare to previous year's.
Security	Maintain logs of who performed what actions on files, including deletions and preservation actions	Place user restriction on directory containing files.
Metadata	Store standard technical and descriptive metadata	In Inmagic Genie catalog.
File Formats	Monitor file format obsolescence issues. Perform format migrations, emulation and similar activities as needed	H264 codec and MP4 container are widely supported. This will make eventual migration easy.

---

[4] Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). *The NDSA levels of digital preservation: An explanation and uses.* National Digital Stewardship Alliance.

4. Proposed Workflow

1. Install FFmpeg via scoop.
2. Create and test FFmpeg script **Batch2MOV** to batch convert MP4 to MOV files.
3. Get network permission from IT to run Powershell scripts on Virtual Machine.
   a. Without permission: Run scripts from Powershell ISE.
4. Apply **Batch2MOV** working through sections of Atlas folder copy in Eric's folder in case of error. Script saves these MOV copies to directory MOV_converts.
5. Create CSV with Atlas files names in one column and ARM-approved names (CatId_RI) in the other. **DirectoryManifest** script can create list of Atlas files. Archivist needs to enter ARM-approved names. This CSV file must be of MOV files in the MOV_converts directory.
6. For each MOV file in MOV_converts:
   a. Rename MOV according to ARM-approved name. Use **RenameFromCSV** to look up ARM-approved name and replace the file name.
   b. Pass metadata from Atlas master log spreadsheet to Inmagic Genie batch upload CSV file along with administrative note:
      i. Catalog record updated to link file digitized by Atlas in connection with ACLU100 documentary project that was never realized. K. Palm, June 2021
   c. Move renamed MOV file to **reference image\video** and place file path within Inmagic Genie batch upload CSV.
7. Run Inmagic Genie batch upload CSV file.
8. Place administrative protections on directory containing archival master MP4s.
9. Place second copy of MP4s on external hard drive in office and investigate options with TRIM for off-site storage of third copy.

4.1 Script References

DirectoryManifest Lists file name, type, and size (MB) for all files in a given directory and saves this information to a CSV file.

Batch2MOV Creates a MOV copy of all MP4s in a directory. These MOV copies are saved to a directory called MOV_converts.

RenameFromCSV Looks up the file name of a MOV file in a directory within a CSV table. If matched, the script replaces that original name with ARM-approved name presented in CSV table.

5. Appendix: Comparative Metadata

The following is the technical metadata for asset 5183 as MP4 and MOV for comparative reference:

5183_RI.MP4
Duration: 00:09:49.29, start: 0.000000, bitrate: 1579 kb/s
  Stream #0:0(eng): Video: h264 (Main) (avc1 / 0x31637661), yuv420p(tv, top first), 720x480 [SAR 8:9 DAR 4:3], 1447 kb/s, 29.96 fps, 29.97 tbr, 29970 tbn, 59.94 tbc (default)
    Metadata:
     creation_time   : 2018-11-29T16:37:05.000000Z
     vendor_id   : ElGt
     encoder   : H.264/AVC
  Stream #0:1(eng): Audio: aac (LC) (mp4a / 0x6134706D), 48000 Hz, stereo, fltp, 127 kb/s (default)
    Metadata:
     creation_time   : 2018-11-29T16:37:05.000000Z
     vendor_id   : [0][0][0][0]
  Stream #0:2: Video: mjpeg (Baseline), yuvj420p(pc, bt470bg/unknown/unknown), 512x748 [SAR 72:72 DAR 128:187], 90k tbr, 90k tbn, 90k tbc (attached pic)

5183_RI.MOV
Input #0, mov,mp4,m4a,3gp,3g2,mj2, from 'C:\Users\intern_admf_esilb\Desktop\5183_RI.mov':
  Metadata:
    major_brand   : qt
    minor_version   : 512
    compatible_brands: qt
    title   : 10001_P_ACLU_Taking-Liberties-2_5183
    date   : 2018-11-29
    encoder   : Lavf58.76.100
  Duration: 00:09:49.29, start: 0.000000, bitrate: 1252 kb/s
  Stream #0:0(eng): Video: h264 (High) (avc1 / 0x31637661), yuv420p, 720x480 [SAR 8:9 DAR 4:3], 1116 kb/s, 29.97 fps, 29.97 tbr, 11988 tbn, 59.94 tbc (default)
    Metadata:
     handler_name   : VideoHandler
     vendor_id   : FFMP
     encoder   : Lavc58.134.100 libx264
  Stream #0:1(eng): Audio: aac (LC) (mp4a / 0x6134706D), 48000 Hz, stereo, fltp, 128 kb/s (default)
    Metadata:

handler_name   : SoundHandler
vendor_id   : [0][0][0][0]