

Moving Images Procedures

Version 1.18

Created date:	26 January 2012
Last updated:	03 July 2017
Review Due:	24 August 2018
Authors:	Steward Waller, Jen Mitcham, Kieron Niven, Ray Moore and Jenny O'Brien
Maintained by:	Digital Archivists
Previous version:	Live

1 Purpose of this document

This page is designed to document current ADS procedures for production of dissemination and preservation copies of digital video files. It contains a list of current dissemination/preservation formats and how to migrate files to required formats. While holdings of digital video files are currently limited it is worth highlighting at the outset the “importance of regular digital-to-digital transcoding to moving image collections cannot be overstated” (Anon., 2009) due to the fast changing developments within this sector. **As/when we begin to receive more A/V files a thorough re-evaluation of preservation policy should be considered.** More information can be found in the *Guides to Good Practice for Digital Video*.¹

- Anon. (2009). Transcoding Digital Video. JISC Digital Media.²
- Heydegger, V (2008) 'Analysing the impact of file formats on data integrity' *Proceedings of Archiving 2008, Bern, Switzerland, June 24-27*. pp. 50-55. Society for Imaging Science and Technology.³
- Pearson, G and Gill, M (2005) 'An Evaluation of Motion JPEG 2000 for Video Archiving', *Proc. Archiving 2005 (April 26-29)*, pp. 237-243.⁴
- van der Knijff, J (2011)
- Wright, R (2011) *Audiovisual Digital Preservation Status Report 2*. PrestoPrime Deliverable D7.1.4.⁵
- Wright, R (2012) *Preserving Moving Pictures and Sound. DPC Technology Watch Report 12 – 01 March 2012*. Digital Preservation Coalition: Great Britain.⁶

¹ http://guides.archaeologydataservice.ac.uk/g2gp/Video_Toc

² <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/transcoding-digital-video/>

³ <http://www.ingentaconnect.com/contentone/ist/ac/2008/00002008/00000001/>

⁴ https://www.prestocentre.org/system/files/.../MJ2_video_archiving%5B1%5D.pdf

⁵ <https://www.prestocentre.org/resources/audiovisual-digital-preservation-status-report-2-2010>

⁶ <http://dx.doi.org/10.7207/twr12-01>

2 Formats

Offered format	Accepted	Preservation	Presentation	Notes
MPEG 1 .mpg, .mpeg	YES	MPEG 1 .mpg, .mpeg	MPEG1, 2 or 4 .mpg, .mpeg, .mp4	
MPEG 2 .mpg, .mpeg	YES	MPEG 2 .mpg, .mpeg	MPEG1, 2 or 4 .mpg, .mpeg, .mp4	
MPEG 4 .mp4	YES	MPEG 4 .mp4	MPEG1, 2 or 4 .mpg, .mpeg, .mp4	MPEG 4 is suitable for data preservation and dissemination although a higher quality MPEG format should be used where appropriate.
DivX .divx, .avi	YES	MPEG 1, 2 or 4 .mpg, .mpeg, .mp4	MPEG1, 2 or 4 .mpg, .mpeg, .mp4	

Ideally depositors wishing to include moving image/video files with their archive should be notified as early as possible that the accepted formats for deposition are largely confined to the MPEG format. At the same time depositors should be encouraged to archive uncompressed versions of files. We can accept other video file formats; however the onus should be placed on the depositor to supply files in a suitable MPEG archive format alongside these other formats. As long as this format is suitable it can be used for dissemination.

If the depositor is unable to supply suitable preservation versions they should be advised that any conversions are carried out on a 'best efforts' basis (particularly as we lack the commercial software's required to carry out many of these conversions). If the creation of preservation versions is not possible the depositor should be informed that we can archive the file in its current form, **but that we are unable to migrate them to newer versions of that format.**

Our current policy should really be considered a temporary solution to the problem of A/V files, so there should be a monitor on the obsolescence issue within our current preservation formats, so that future migrations can be enacted. Ideally files should be preserved in an uncompressed format, but our current policy relies on the use of compressed, lossy formats (MPEG).

N.B. Motion JPEG 2000 (MJ2), an open/ISO standard, has been adopted by the Library of Congress (and NASA) as their format of choice for preserving moving images, using a MXF wrapper (MXF Operational Pattern 1a). This may well be a better, lossless preservation format which we may wish to consider in the future.

Compression

Compression can be used on any kind of file, but comprehension has particular significance for audiovisual content: 1) compression is widely used; 2) compression greatly affects resistance to small corruptions in a file (bit rot); 3) audiovisual content produces such large files that bit rot becomes a real concern... With compression, an altered bit changes a number that is a parameter in a calculation, not just one pixel or one sound sample. The calculation could apply to a whole image, and so the effect of the loss is magnified. Heydegger (2008) found that a one-bit error in a compressed file could affect 105 bits or more or render the whole file unreadable, depending upon the type of compression. (Wright 2012: 16).

Embedded metadata

Like other file types metadata is often automatically generated on creation/editing, but the quality and nature of this metadata varies between file creation software packages. The above table shows the current minimum recommendations for A/V files. During conversion care should be taken to preserve any existing metadata found within the audio file, in an ideal world we should ask for metadata to be supplied separately in an TXT / XML form but this may not always be possible. If this metadata is embedded within the file then it should be extracted and saved in a suitable preservation format. Whether supplied by the depositor, or extracted by the digital archivist, this metadata should be preserved and, if appropriate, disseminated alongside the relevant files (see below for notes on storage). Information on extracting this metadata can be seen below.

Associated metadata

It is important that any copyright information/permissions are also stored alongside the requisite file in a suitable preservation format.

3 Documentation / Metadata

The current *Guides to Good Practice* recommends that the following documentation is stored for each file⁷.

Element	Description
Software, version and platform	The software (or hardware if taken directly from a device) used to create the video.
Video Codec	The name and version of video codec (where appropriate).
Video Dimension	The video dimension (in pixels).
Frame Rate	Frame rate per second (fps).
Bit rate	The video bit rate.
Audio Codec	Name and version of audio codec.
Audio Sample Frequency	

⁷ http://guides.archaeologydataservice.ac.uk/g2gp/Video_3

Audio Bit-rate	
Audio Channels	Channels used e.g. Stereo.
Length	Length (hours, minutes, seconds) of file.
File Size	Size of the file in MB.

4 Accessioning checks

- Do we have the necessary documentation (see below)
- Depositors should be encouraged to supply preservation versions of their files, with a more compressed version for web delivery.
- The necessary rights and clearances have been attained from contributors with regard to copyright
- Embedded metadata
- Associated project/file level metadata as per Guidelines for Depositors.⁸

Significant properties

The most significant property for video and film is image quality unfortunately this element remains 'ultimately subjective' (NB digital tools are being developed to automatically detect any disruptions to quality these are currently), but as Wright observes "the only process that could reduce quality is compression, or successive applications of compression" (2012: 17, also discussed below). The properties of digital video files that should remain unchanged when preserving or storing data as recorded in the *Guides to Good Practice* are:

- the length and size of the file (e.g. 5min 31secs / 150MB)
- the frame rate in frames per second (e.g. 25 for PAL or 30 for NTSC)
- the frame size / video resolution (e.g. 720 x 576 pixels)
- the bit-rate (in kbps)
- audio bit-rate (kbps)
- audio frequency (kHz)
- audio channels used (e.g. stereo)
- associated metadata, documentation and file size.

Deposited files should be compared with any extant metadata/documentation prior to conversion.

5 How to convert files

Software

The ADS has little/no software which can effectively deal with A/V files as a consequence of the rarity of their deposition, consequently we should endeavour to place the onus on the depositor to create preservation and dissemination versions, however there are a number of freeware programmes available which can be utilised should the need arise. NCH software distributes *Prism Video Converter*⁹ which can be used freely (for non-commercial purposes),

⁸ <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>

⁹ <http://www.nchsoftware.com/prism/index.html>

is relatively easy to use, but has a degree of flexibility for controlling encoder and video options. Download and run the executable file:

1. Add the file you wish to convert using the + *Add file(s)* button, or by selecting *File > Add File(s) to List*
2. Set the destination folder in the *Save to Folder* box.
3. Set the output/dissemination format you wish to create using the *Output Format* box.
4. You can alter the video settings using the *Encoder* button. From where you can set the compressor/bitrate of the video, along with sample/bitrate of the audio. When you happy click 'OK'.
5. Press *Convert* and the file will magically appear.
6. Check the video runs and the audio, if there is any, is working correctly.
7. It is also worth using the metadata tool (see below) to check that whatever metadata should be the same (e.g. length) remains the same.

MPEG4IP, a command line tool, has conversion capabilities but its support of the Window platform is limited. At the same time we have not extensively used it, so it is difficult to endorse.¹⁰

Extracting Embedded Metadata

Any embedded metadata should be regarded as a significant property of A/V files consequently the Digital Archivist should take care to extract and save this in a suitable preservation format. Often dedicated video editing software packages contain metadata editors/extractors which allow for the export of metadata. Unfortunately we currently do not have any dedicated A/V software, however there are a number of freeware alternatives. For example, GSpot.¹¹

1. Download and unzip GSpot to an appropriate place. There are some useful 'idiots' instructions available on the website.¹²
2. Open GSpot, go to *Options > Export*, check *Enable Export* and click *Save as default*.
3. Check the format as *Generic default as readable NFO*.
4. The metadata should appear as a TXT document in the same folder as the file.

GSpot is useful as it also allows the batch process video files. Another useful alternative is AVIcodec which allows for the extraction of technical metadata which can be exported in .csv form.¹³ Any embedded metadata that is extracted should be stored using the requisite file structure noted below.

File naming

Where possible files should retain the same name as the original. On occasion (and normally for dissemination), it may be necessary to create different versions of the same file. In these cases a logical naming strategy should be used, and should be accompanied by explanation in the Processes section of the CMS. Extracted metadata should also be named consistently, for example

myvideo_meta.xml

These files should then be placed in prescribed location noted below.

¹⁰ <http://mpeg4ip.sourceforge.net/>

¹¹ <http://www.headbands.com/gspot/>

¹² <http://www.headbands.com/gspot/index.htm>

¹³ <http://www.videohelp.com/tools/AviCodec>

6 Post-migration checking

We should ensure data consistency by undertaking a check of files post migration. The number of files to be checked is at the discretion of the Digital Archivist. Specific attention should be paid to:

- length of the movie is consistent with the original.
- frame rate has been maintained.
- video dimension has been maintained.
- bit rate has been maintained.
- any embedded metadata has been preserved

Compare the metadata for both preservation and dissemination formats to make sure the conversion has preserved the significant properties of the original file.

Storage

Data should be stored in appropriately named folders, as described in the ADS Repository Operations manual.¹⁴ Any directory structure from the SIP should be retained in the AIP. In some cases editing/restructuring may be necessary, but such restructuring should be recorded in the Processes section of the CMS.

Otherwise, store data in one of the following directory structure:

```
/preservation
  /{original_structure}
    myvideo.mpg
    myvideo2.mp4
```

```
/dissemination
  /{original_structure}
    myvideo.mpg
    myvideo2.mp4
```

Storing file metadata

File metadata (copyrights, documentation, etc) should be stored in an appropriate archival format with the preservation/dissemination files in a *documentation* folder within the requisite file type, for example:

```
/preservation
  /{original_structure}
    mymovie.mpg
  /documentation
    mymovie_metadata.docx
```

```
/dissemination
  /{original_structure}
    mymovie.mp4
  /documentation
    mymovie_metadata.pdf
```

¹⁴ <https://archaeologydataservice.ac.uk/advice/RepositoryOperations.xhtml>

Embedded metadata

For preservation purposes, any 'embedded' metadata should be stored separately in a "documentation" subfolder as CSV, XML or TXT files, for example:

```
/preservation
  /{original_structure}
    mymovie.mpg
  /documentation
    mymovie_metadata.docx
    mymovie_embedded_metadata.xml
```

For dissemination, any embedded metadata can be left within the file, although care should be taken to make sure this metadata remains unchanged, particularly following conversion. If this metadata is supplied by depositor separately, then this should be presented with the dissemination data. For example:

```
/dissemination
  /{original_structure}
    mymovie.mp4
  /documentation
    mymovie_metadata.pdf
    mymovie_metadata.xml
```

If metadata has been extracted from an existing file the Digital Archivist should make sure that these files are named logically (i.e. relate them to the file(s) and what they contain). Any metadata extraction should be recorded in the Processes section of the CMS.