# Born-Digital Video Preservation: A Final Report

By Andrea Shahmohammadi, Archivist

---

Born-digital videos bring about a number of multifaceted and still unanswered questions amongst archivists when discussing best-preservation practices. While digital audio files have gradually evolved some accepted standards from which to build a framework for preservation, this is unfortunately not the case with digital video. Moving images are now primarily made in digital formats, and, as such, the archival community needs to better address how to preserve and make accessible incoming collections for future generations.

**Issues**

The fragile nature of born-digital video means that archives cannot postpone preservation practices until standards are universally accepted. Preserving digital information, especially multimedia material, is complicated. Computer hardware, software and documentation are needed to interpret/read the digital bits and format specifications. Subjection to intellectual property rights is another complication, but it is one that will not be discussed in detail in this report. Active commitment and cooperation amongst archives, IT professionals and professional organizations is necessary to develop appropriate policies, plans and implementation of sound best practices to preserving digital video for the future.

Like all born-digital records, digital obsolescence is a major challenge for long-term access to video files. While completing my research I came across dozens of file formats and wrappers and more than 300 codecs specific to moving image and audio. A video file format is often in the form of a wrapper (also known as a container), which is a compressed multi-file format that contains the specifications of different data elements and metadata that co-exist within the file and is used to identify and interleave the various data types.[1] Advanced formats can contain multiple audio and video streams, subtitles, chapter-information, and synchronization information needed to play back the various streams together. When the file is bundled together the compressed data streams need to be encoded, so each file also has at least one codec. A codec[2] is a program that identifies the method used to compress data into fewer bytes, and does the opposite when a video file is played back, decompressing it.[3] With faster, more capable and less expensive storage and processing devices being developed, older versions quickly become replaced. As software and decoding technologies become abandoned, or hardware devices are no longer produced to make room for the newest version, born-digital records created with such technologies can become obsolete in a matter of years. This becomes more complicated as moving images are composite artifacts, meaning that they have several inter-related elements all of which need to be readable and match up to be viewable. All it takes is one component of the video file to be obsolete to make the whole file obsolete.

Compression and encryption are other challenges prevalent amongst digital video. Storage constraints and narrow bandwidth result in a large percentage of video file formats to use compression schemes. Due to the size and application of multimedia data, lossy compression is most often used. Lossy

---

[1] "Container format (digital)." *Wikipedia*. Web. <http://en.wikipedia.org/wiki/Video_file_format>.
[2] The word *codec* is a portmanteau of 'compressor-decompressor' or, more commonly, 'coder-decoder'.
[3] Roberson, Mark. "Video File Container Formats, Compression And Codecs – Oh My! ." *ReelSEO*. N.p., 2010. Web. 24 Feb 2011. <http://www.reelseo.com/file-formats-containers-compression/#>.

compression is defined as a technique that does not decompress data back to 100% of the original. Lossy methods provide high degrees of compression and result in relatively smaller files than the original, and unfortunately, there is a certain amount of data loss when the files are uncompressed.[4] Archival standards for digital materials suggest an uncompressed or reliable lossless compressed object for an archival master to ensure longevity and preservation of the highest quality. Lossless compression means that the output from the decompressor is bit-for-bit identical with the original input to the compressor; the decompressed video should be indistinguishable to the original.[5] For born-digital objects, the best preservation copy would be to the exact specifications of the original. Since archival institutions rarely have a say in the format in which they receive artifacts and no improvements in quality can be gained by upsampling, compression schemes prove to be just one more obstacle in the preservation course. Encryption and password-protection also add to the complexity of the preservation process and should be avoided at all costs. Without the specific encryption key associated with the schema, access to the video is lost.

**Codecs/File Formats**

There are hundreds of types of file formats, wrappers and codecs used in multimedia digital files adding to the complexity of both digitized and born-digital records for preservation. Media players need to be able to read all of the formats included in the file to successfully play and transcode files. It is like a jigsaw puzzle; all the pieces must be identified and matched up to be able to see the big picture. As a result, archivists need to address the sustainability of all relevant formats in their digital collections to ensure access, authenticity and readability of digital videos in their holdings. The Library of Congress identifies seven sustainability factors of digital formats which should be considered for preservation actions: disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents, and technical protection mechanisms.[6] Their Sustainability of Digital Formats website does a good job assessing popular files against these seven factors. Unfortunately this website does not appraise codecs or uncommon formats. Risk assessment surveys against all known digital formats, wrappers and codecs should be completed by the repository when collections come to them and every few years to ensure best preservation practices. A format's quality, stability, potential longevity and industry acceptance need to be established and addressed.

Digital video files that are not sustainable should be transcoded into files with formats and codecs which are. Transcoding is the process of converting one type of encoding or data stream into another. This is usually done in archives to incomplete or obsolete data formats to make them more suitable by converting them into more accessible and modern formats. When transcoding files, it is important to understand that transcoding lossy compression files to other lossy compressions results in generational data loss. In addition, the transcoding of lossy to lossless or uncompressed files results in

[4] "lossy compression ." *PC Magazine Encyclopedia*. Web. <http://www.pcmag.com/encyclopedia_term/0,2542,t=lossy+compression&i=46335,00.asp
[5] Dávila, R., and Ian Gilmour. "Lossless Video Compression for Archives: Motion JPEG2k and Other Options." *Media Matters, llc* (2006): 8. Web. 24 Feb 2011. <http://www.media-matters.net/docs/WhitePapers/WPMJ2k.pdf>.
[6] Library of Congress . "Sustainability Factors." *Sustainability of Digital Formats*. N.p., 2007. Web. 24 Feb 2011. <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.

no information loss in the conversion, but the process is irreversible. There are a number of transcoding software options available in both proprietary and free and open source with a variety of different source codes and libraries. The more obscure the wrapper, file format or codec the harder it will be to find acceptable software. This also holds true for custom created codecs or formats.

**Proposed Plan/Guideline Recommendations**

The lack of professionally established standards, protocols and proven methods has resulted in a failure to achieve consensus in the archival community about how to preserve digital video. As a result, archives are left to create their own guidelines and standards. Below are my recommendations developed after researching this topic for my National Archives and Records Administration Archivist Development Program rotation. As always, the guidelines presented here are recommendations, and there may be cases where judgment calls will need to be made about objects that would be better preserved by modifying the recommended guidelines for this purpose.

While some file formats become obsolete in a short period of time, others remain viewable/accessible over the long-term. The identification and sustainability of a collection's file formats and codecs are critical for preserving digital information. Due to a mix of technical and practical issues some digital files cannot just be put on a virtual shelf for safekeeping. An internal risk management document identifies and describes a number of digital video formats, codecs (both audio and visual) and player to determine which files are at risk and which ones are sustainable for the time being. This spreadsheet should be maintained and updated by the repository when new formats, codecs or additional information is identified. File interoperability for successful transcoding and usability across platforms is crucial for long-term access.

Until universally accepted professional standards are in place, the first question we as archivists should be asking is "are these files at risk?" If the answer is no, then there is no reason for institutions to waste resources and the possible authenticity of digital video records through unnecessary transcoding and migration. An analysis of the file formats, wrappers, codecs and available players is necessary to make this assessment. A master preservation copy to the specific frame rate, resolution, and bit rate specifications of the original and a transcoded reference copy will be needed. Every three to five years a file interoperability and risk assessment of such collections should be re-evaluated. If the files are at risk for obsolescence, they should be transcoded into a lossless compressed file with a format, codec and wrapper sustained by the archival institution. A general rule is to use platform-independent, vendor-independent, non-proprietary, stable, open and well-supported formats.[7]

Archival master preservation files should be saved with uncompressed or lossless compression and in a widely used file format, with maximum likelihood of continued support. Uncompressed and lossless compressed files do demand a large amount of storage space per file, but they are necessary to best ensure stability. Additionally, access copies of digital video may be saved in downsized and compressed formats that will allow for end users to access such videos through reference copies or streaming online.

[7] Florida Digital Archive. "Recommended Data Formats for Preservation Purposes in the Florida Digital Archive." FLCA, 2008. Web. 24 Feb 2011. <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.

These reference copies are to be maintained in accessible formats that are easy to view, ones that use file formats and codecs that are compatible with multiple computer platforms.

This begs to ask, what formats are recommended or acceptable for preservation at the current time. A number of archival institutions recommend JPEG 2000 as the ultimate preservation format for digital video. JPEG 2000 is an International Standards Organization (ISO) open-source image compressive standard and coding system. It is usually paired with an MXF wrapper because together they are able to place the still image codec into a high quality video while retaining complex metadata.[8] For preservation this is great, but as archivists we also need to think about access. There are few viewers to play it and limited tools for extraction and accessibility. Until the format is better supported with open-sourced, stable and high performance software libraries easily available, it is unnecessary at this time to spend resources transcoding digital video into the JPEG2000 format. Instead, the repository should keep preservation copies of files in their current format if they are .AVI, .JP2, .MOV, .mp2, .mp4, .MXF and .WMV with sustainable codecs. If the original has a lossy compression scheme then the preservation copy should be copied with a lossless compression.

Metadata, central for long-term access, is an essential component of most digital preservation strategies. Preservation metadata includes, but is not limited to, information on creation, access rights, restrictions, preservation history, and rights management, which supports and documents the digital preservation process. The fragile nature of digital materials and their dependency on software, hardware, platforms, formats and codecs makes preservation metadata vital to the longevity of the file. Brian Lavoie and Lorcan Dempsey explain that "while a print book with a broken spine can be easily re-bound, a digital object that has become corrupted or obsolete is often impossible (or prohibitively expensive) to repair".[9] There are a number of models and tools available to collect and maintain preservation metadata institution wide, but the key is consistency. As video archives grow, metadata become increasingly important.

**Conclusion:**

Today moving images are primarily being created digitally. Examples can be seen everywhere: broadcasting has shed its analog system and now requires digital only transmission, the Internet has an ever growing number of podcasts and videos on sites like YouTube, and handheld devices to view and record video are owned by individual consumers. As hopeless as the current system appears, standards will eventually be adopted by both private and the archival community. In the meantime, it is important that archival institutions work to preserve what digital videos are currently in their possession in a sustainable format.

[8] "Choosing a Digital Video File Type." JISC Digital Media, 03MAR2009. Web. 24 Feb 2011. <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/choosing-a-digital-video-file-type/>.
[9] Lavoie, B. & Dempsey, L., Thirteen ways of looking at...digital preservation, *D-Lib Magazine*, 10 (7/8), 2004.