

R. Sumi Matsumoto
INFO 655
Project Summary
6 May 2020

My research question was very open-ended, but I hoped that through the act of researching, I would have a better sense of which direction to narrow my focus. I wanted to look into video container formats and codecs to better understand the options for video archiving since there is currently not a widely accepted standard. The ambitious, and expectedly unattainable, goal was to choose a standard. However the more realistic impetus was to be able to make more educated decisions about the tradeoffs between formats when working with digital video in the future. Of course audio streams and codecs are an important part of a video file, but I chose to omit this factor in my research as a way to slightly narrow my research topic.

I started by reading the current literature on digital video preservation. Most of the current literature is the public listings of what standards a particular institution uses (National Archives, n.d.; National Archives of Australia, n.d.; *Recommended Formats Statement 2019-2020*, 2019; Smithsonian Institution Archives, 2017). There were some articles that discussed various formats more in depth, but they were mostly comparisons rather than an argument for any particular format (Escobedo, 2011; Hodge, 2000; Library of Congress, 2019; Marsh, 2015). The most in depth analysis I found was the case study analysis done by the Federal Agencies Digital Guidelines Initiative (FADGI). FADGI published information on both born digital video formats as well as reformatting digital video (Federal Agencies Digital Guidelines Initiative, 2014).

Although many articles seemed optimistic about a forthcoming video standard, there has yet to be any clear options being put forth (Shahmohammadi, 2011). Notably, Library of

Congress touts Motion JPEG2000 codecs with MXF wrappers due to their high performance (Lazorchak, 2011). FADGI also endorses these high-performance formats, but is more reticent in encouraging them as a widespread standard. Other institutions have fewer resources to support these formats, and each institution has unique needs that require various exceptions from the ideal (Federal Agencies Digital Guidelines Initiative, 2014). FFV1 is also becoming more popular with cultural institutions as it allows for lossless compression with smaller file sizes, but it has a very narrow scope of compatibility (Marsh, 2015).

Most assessments of video formats compare the features of each format. Important aspects of wrappers include the availability of documentation or proprietary status, level of adoption and playback compatibilities, compatible codecs for audio and video, support for multiple video and audio streams, support of chapters and captioning, and imbedded metadata capabilities. Regardless of format, there is also always a tradeoff between video quality and storage cost. Additionally institutions must assess their technical capabilities, and formats that require more complex tools to generate or manipulate may not be within the abilities of the institution's archivists or would be too much of a burden to add to IT staff purview (Federal Agencies Digital Guidelines Initiative, 2014; Library of Congress, 2019). The current lack of standard appears to come from the twofold problem of institutions' large variance in digital video collection extent and storage/playback capabilities as well as the general lack of video format standardization within video content creators and video playback systems (Escobedo, 2011).

I personally am more inclined towards non-proprietary formats. Proprietary formats encourage the use of proprietary media players, whereas open formats rely on open source or general-purpose players. This is important because open source media players tend to have

better support for legacy media. When I was working with videos, the open source player VLC was able to play a 2011 MOV file without issue. However Quicktime player, the intended playback tool, had to convert the codecs to update them before playing the video. Further, Apple states that macOS 10.14 (Mojave) will be the last version to convert/support any legacy media in Quicktime player (*If QuickTime Player Converts Legacy Media Files before Playing Them*, 2019).

Throughout my cursory research, I realized that there was little to no discussion of video quality assessment tools. Although the capabilities and limitations of formats is an important consideration, end result quality is also very important. Some case study reports discussed the use of file normalization briefly, but no clear measurement to assess the change in video quality was used (Federal Agencies Digital Guidelines Initiative, 2014). I looked further into this aspect, and there seems to be little interest in discussing the option of file normalization. Most studies on file conversion are from the angle of analog to digital, rather than on updating legacy digital formats (Hodge, 2000).

This avenue of research was very fascinating to me, but seemed to be a much larger project than the scope of this assignment. I would be interested in testing a large sample of videos, or perhaps a more controlled sample of videos if such a thing is possible, with a video quality assessment tool in order to receive more concrete scores and calculations on a video's quality. I could then use those quantitative scores to compare quality between different transcoded, transwrapped, and fully converted versions of the video. Unfortunately, most tools for creating concrete video quality metrics are built for commercial use (Li et al., 2017; *NTIA/Vqm*, 2019/2020). I am curious which transcoding and transwrapping mappings create the least loss in quality and how codec compatibilities impact this. Or perhaps I would find that

there were more important factors in quality retention than codecs. These other factors might be resolution, compression type, or initial bit rate. I also don't know if I would find one format that would clearly stand out as the "best" or if different formats would best transcode to various formats. I would also have to take into consideration what the human threshold for perception in quality loss is.

I do not have the tools to do a large-scale quantitative assessment of video quality in normalized videos. Instead, I took a more exploratory approach to the same topic. I looked at small-scale comparisons of various fully converted and transwrapped videos using recommended file formats.

I worked with a small collection of videos recorded between 2011 and 2018 that I already own. The videos range in container format, recording device, and video codec. Although they do not represent every available format, there is a good diversity. My videos notably lacked a Motion JPEG 2000 (MJP2/MJ2). I also lacked OGG, and had a high number of MOV and MP4 files. However, I believe they are a good representation of what formats are commonly used and generated by individuals not thinking about the archival value of recordings. My sample collection also reflects my experience from surveying the NYPL Dance Division Original Documentation collection, which heavily favors MOV and MP4 with a secondary emphasis on MXF.

The two main tools I used to work with these video files were FFmpeg and MediaInfo. Both are open-source, command-line tools. I used MediaInfo, which we used in class when discussing audio and video encoding, to evaluate the video files and FFmpeg to transcode and transwrap files. FFmpeg allowed me to specify codecs when transcoding, and although I did play around with this a little bit it added a new dimension of complexity that I did not have time

to fully explore. The codec specification flag, `-c`, did allow me to transwrap files by copying the streams rather than decompressing and re-compressing them (*FFmpeg*, n.d.).

I tried transcoding and transwrapping files to see which formats were compatible and which weren't ("Comparison of Video Container Formats," 2020). For example, AVI files and MOV files could not be easily converted to MXF, but MP4 and MPG did convert. I used very simple techniques for transcoding and transwrapping, which are probably not representative of archival quality reformatting techniques. I also didn't dig too deeply into failed conversions to try and find a work around, but I did receive detailed feedback from *FFmpeg* about the conversion attempt. This is certainly an avenue of further research.

It was important to view "successfully" converted files, since some ended up with weird playback issues. There were more issues with just transwrapped files than fully converted files. It was also interesting to visually see the shift in quality. Although many fully converted videos had dramatically reduced size and bit rate, I was surprised at how little visual impact it had on the quality. Although the quality reduction was noticeable, it was not nearly as egregious as I would have guessed. Playback also brought up an additional aspect, which is the player used. I worked with Quicktime Player and VLC: one proprietary and one open source. The same video in two different players sometimes created a distinct shift in coloring. This means the player being used has an impact on the video and thus must also be taken into consideration, beyond a technical or pragmatic level, when thinking about choosing formats. Again, this as another area of further exploration, and one I have yet to see discussed at all.

Although transcoding allowed more wrapper options for a file, the process often created a lower resolution file. Just transwrapping, when possible, made a much smaller impact on the file's quality. Seeing how much of an impact on file quality full conversions usually made, I

understand why file normalization is not a widely used practice. However, I do think that transwrapping may be a useful tool in creating a normalized viewing format without too much quality loss or file manipulation. Of course, this prompts the question of which wrapper would best be fitted to this purpose. Two main considerations outside of quality assessment would be broadest codec compatibility and media player compatibility (both generally as well as with whatever software is in use institutionally).

I did not have a pointed research question on the outset, but I do think the research process answered some of my questions. It certainly gave me a better understanding of video wrappers, codecs, and file normalization. Despite my discoveries, I only scratched the surface of this vast topic. My research was very casual, but I think a more qualitative assessment of video codec and wrapper formats, video quality, media player impact, and effects of transcoding and transwrapping might help clarify viable options for cultural institutions more than the current anecdotal data. Unfortunately the topic is complicated by its many variables. A qualitative study would be a vast undertaking as it would be difficult to isolate each variable.

Sources

Comparison of video container formats. (2020). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Comparison_of_video_container_formats&oldid=953278112

Escobedo, K. (2011, August 16). Digital Video Preservation: Further Challenges for Preserving Digital Video and Beyond. *Smithsonian Institution Archives*.

<https://siarchives.si.edu/blog/digital-video-preservation-further-challenges-preserving-digital-video-and-beyond>

Federal Agencies Digital Guidelines Initiative. (2014, December 2). *File Format Comparison Projects: Still Image and Audio-Visual Working Groups*.

http://www.digitizationguidelines.gov/guidelines/File_format_compare.html

FFmpeg. (n.d.). Retrieved April 28, 2020, from <https://ffmpeg.org/>

Hodge, G. M. (2000). Best Practices for Digital Archiving: An Information Life Cycle Approach. *D-Lib Magazine*, 6(1). <https://doi.org/10.1045/january2000-hodge>

If QuickTime Player converts legacy media files before playing them. (2019). Apple Support.

<https://support.apple.com/en-us/HT202884>

Lazorchak, B. (2011, July 5). Whither Digital Video Preservation? | The Signal [Webpage]. *The Signal*.

[//blogs.loc.gov/thesignal/2011/07/whither-digital-video-preservation/](https://blogs.loc.gov/thesignal/2011/07/whither-digital-video-preservation/)

Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., & Manohara, M. (2017, April 19). Toward A Practical Perceptual Video Quality Metric. *The Netflix Tech Blog*.

<https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652>

Library of Congress. (2019, March 29). *Moving Images* [Webpage]. Sustainability of Digital Formats: Planning for Library of Congress Collections.

<https://www.loc.gov/preservation/digital/formats/content/video.shtml>

Marsh, A. (2015, May 8). The Pros and Cons of FFV1. *Bitstreams: The Digital Collections Blog*.

<https://blogs.library.duke.edu/bitstreams/2015/05/08/the-pros-and-cons-of-ffv1/>

National Archives. (n.d.). *Digital File Types*. Retrieved April 26, 2020, from

<https://www.archives.gov/preservation/products/definitions/filetypes.html>

National Archives of Australia. (n.d.). *Long-term file formats* | *naa.gov.au*. Retrieved April 15,

2020, from [https://www.naa.gov.au/information-management/storing-and-preserving-](https://www.naa.gov.au/information-management/storing-and-preserving-information/preserving-information/preserving-digital-information/long-term-file-formats)

[information/preserving-information/preserving-digital-information/long-term-file-formats](https://www.naa.gov.au/information-management/storing-and-preserving-information/preserving-information/preserving-digital-information/long-term-file-formats)

NTIA/vqm. (2020). [MATLAB]. Code from NTIA. <https://github.com/NTIA/vqm> (Original work published 2019)

Recommended Formats Statement 2019-2020 (p. 34). (2019). Library of Congress.

Shahmohammadi, A. (2011). *Born Digital Video Preservation: A Final Report* (p. 4).

Smithsonian Institution Archives.

Smithsonian Institution Archives. (2017, April 26). *Recommended Preservation Formats for*

Electronic Records. Smithsonian Institution Archives. [https://siarchives.si.edu/what-we-](https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records)

[do/digital-curation/recommended-preservation-formats-electronic-records](https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records)