Caring for Born Digital Video Camera Original Formats

*CONSIDERING INTENTIONAL CHANGE*

|  |
| --- |
| **Crystal Sanchez** |
| *Smithsonian Institution*  *United States*  *sanchezca@si.edu* |
|  |  |

**Over time, we have seen the exponential growth of born-digital files, specifically those created by consumer, prosumer, and professional cameras. This paper is about a new issue that is specifically rooted in today's digital workflows: the maintenance of born-digital camera original video formats in an archival setting and the intentional irreversible change that may be required during processing to stabilize them for future access.**

**Keywords – born digital; digital preservation; file-based; video production**

**Conference Topics – community**

# Introduction

The Smithsonian Institution (SI) has been working steadily towards understanding and building approaches to responsibly care for born digital video files created by production teams. Born digital file care has been an active area of the field for many years but approaching files that originate from video cameras brings its own unique challenges, and the archive is beginning to see their growth due to a variety of reasons rooted in changing production tools and workflows. This paper is about a new issue that is specifically rooted in today's digital workflows: the maintenance of born-digital camera original video formats in an archival setting and the *intentional irreversible change* that may be required to stabilize them for future access.

The Smithsonian Institution centralizes digital collection asset storage with an enterprise Digital Asset Management System (SI DAMS) [1]. As part of the Smithsonian DAMS team, we work closely with archivists, collection managers, registrars, and conservators across the Smithsonian’s many units, as well as advancement, communication, and video production professionals. Over the past ten years, the SI DAMS team has worked hard to support video in the system with implementation of tools, metadata extraction, and technical documentation for all files in our care. Over time, we have seen the exponential growth of born-digital files, specifically those created by consumer, prosumer, and professional cameras.

## A Shift to Camera Original Files

Work with all of these stakeholders has shifted over the years. At first, video files were predominantly received as ProRes wrapped in Quicktime (MOV). It was most common for creators using Final Cut Pro to transcode files from camera originals into ProRes codecs during transfer to computers for editing. In this workflow, ProRes [2] is the Master file; camera originals were deleted when cards were re-formatted for reuse. Jonah Volk, writing in 2009, but anticipating the future, expresses this, “While current workflows generally involve transcoding media to QuickTime for use in editing software, it is not inconceivable to imagine that Apple might in the near future allow for native MXF editing in QuickTime, as Adobe Premiere already does.” [3]

As the Smithsonian Institution rapidly shifted to Adobe products around 2015, video producers also shifted to Adobe Premiere for editing. Adobe supports a wide variety of codecs, allowing users to edit with most camera original formats without transcoding. This saves producers time in the workflow but creates new challenges for archivists. As Adobe themselves say in their guide, “If you could only edit native formats, you probably would” [5].

As a result of this shift in production practices, we surveyed the video production practices at the Smithsonian in the Spring of 2018. The survey asked staff to describe their current production tools, practices, and pressing archival needs.

Sixteen participants from ten Smithsonian museums discussed their current practices. Overall, most producers reported maintaining a variety of camera original file formats as raw masters, only exporting lower resolution derivative edit masters for delivery. We also saw trends in the cameras and tools they used; producers referenced the same handful of cameras, and most reported using Adobe products to create produced pieces. We anticipated receiving a variety of camera original born-digital files in the near future. The survey results highlighted the need to develop a risk-assessment approach to analyze specific, common, born-digital, camera original video formats to develop some recommended practices for our community [6].

Technical Data Research & Analysis was conducted in 2019 on SI DAMS repository’s 42,823 video files, findings of which suggest this growth trend. In analyzing video codec data, ProRes codecs made up 12.5% of the total in 2019 (39.5% AVC/h.264; 22% unknown; 9.5% DV; 7% MPEG-2; 6.5% Motion JPEG; 2% Uncompressed 10 bit;.5% misc) [4]. In looking at file format data, totals showed 73% QT(MOV); 17% MPEG; and just 7.5% MXF.

The preceding years indicate an increase in total video files to 76,520 files at the beginning of 2022, up 44% from 2019). MXF wrapped files increased from 7.5% in 2019 to 10% in January 2022. This indicates to us that the shift in production practices will likely increase the deposit of raw footage to the archive from edit masters formats (ProRes) to camera original formats.

## Researching Camera Original Formats

After this analysis, my colleague Taylor McBride and I began a crowd sourced project to inventory & research a dozen formats encountered at SI and connect them to cameras, encouraging shared documentation of format specifics through open tools (ie, Google Sheets) [7]. We presented this work to the field at AMIA 2018 to gain community feedback and encourage participation [8].

The SI DAMS team turned this work into a Supported File Formats Guide for our internal audience (SI), documenting a handful of specific wrappers and video codecs our community encountered and what actions we recommend for their long-term care [9].

For example, AVCHD, a camera original format developed by Sony and Panasonic in 2006, stores the video content using a commonly used AVC video codec, but wrapped in proprietary MTS wrappers split into 2GB files (spanned clips). Video clips and other camera files are stored in a proprietary packaged directory structure that is only natively accessible on a Mac OS. It is a risky structure, with a good video codec in a largely inaccessible wrapper. We recommend rewrapping and combining any spanned clips but maintaining the original video codec data when possible.

P2 is a Panasonic camera original format with a proprietary and complex directory structure, separating audio and video streams into separate directories and spanning larger clips into multiple files. Since the format is highly dependent on retaining directory structure, we recommend processing the files to concatenate any spanned clips and output a master asset at as close to the technical specs of the original as possible.

Other formats come off the camera as single files with common codecs and wrappers, and they are extracted from their sidecar camera files and saved as is (XAVC, XDCAM).

Considerations on how to build a model for approaching the long-term care of camera original born-digital file-based video formats has led to more questions. Normalizing files to one defined codec or format often creates larger files than the originals. Saving everything as native formats as deposited doesn’t address the inherent risks some formats embody. Stabilization and processing of specific formats requires one to address certain key technical aspects.

## Key Aspects to Consider

The 2014 Federal Agencies Digitization Guidelines Initiative Guide to Creating and Archiving Born Digital Video is a good framing of key aspects archivists may encounter [10]. The Advice for File Archivists section lists important concepts like “Document the Original Order” (RP 2.1) [11], especially camera created file structures, and “Identify the file characteristics at the most granular level” (RP 2.3). The recommendation to “Determine and document criteria for when (if ever) it is appropriate to change the video’s technical properties (RP 2.5) asks you “Is the file "at risk" in its current form?””

Other recommendations in the guide, although noteworthy and important, are not feasible for the Smithsonian’s scale and tools, for example: “Retain the original file when transcoding” (RP 2.6). Taking into account our resources, if the data is not deemed critical, we do not keep original files after processing or even “Retain all the data from the original files if the video file structure has changed” (RP 2.10). Some technical and origination data is changed and lost as a result of processing. The recommendation “Select appropriate technical characteristics for the video encoding if transcoding, normalizing or otherwise changing the video stream to meet business needs” (RP 2.7) highlights the reality that as archivists and digital preservationists, our business needs include changing files when they are deemed “at risk”.

The FADGI recommendation “Determine and document criteria for when (if ever) it is appropriate to change the video file’s technical properties” (RP 2.5) might be our most difficult challenge, and over the years we have been grappling with what this criteria might consist of. If we determine the files are at risk, is it irresponsible not to act, even when decisions we make create irreversible change to our collections. The process has led to the following insights in our growing efforts to build this criteria.

## Some Insights

# Building a Shared Language

In talking with video producers at the Smithsonian, we discovered we were approaching the same questions with different experiences and assumptions. Producers referenced cameras and wrappers when discussing formats, ie the GoPro makes mov, while we were talking in the language of codecs and structure, ie this file is AVC split into multiple clips in this subdirectory. Working to build a bridge where cameras and codecs are mapped is an important step in understanding the technical aspects of the products shared between these two stakeholders. As archivists, it is our responsibility to build those relationships into a more permanent and sustainable solution.

# Understanding Production Workflows

Starting these conversations with producers led to expeditions to production shoots. How do they choose settings on the camera? What aspects of the workflow are important for them in their work? How do they transfer from camera to computer to editing suite? Are there any data management practices in place, even if not formalized, that we can use to understand file history as they move through the production workflow and into the archive? This process is an important component of building those relationships and allows for not only a better understanding of the files but also the goals and priorities creators have for them during and after archival deposit.

# Defining Risk Tolerance by Analyzing the Stability of Format Structure

File format research for camera original video formats is not a well-developed field, and we must depend on our own examples from the cameras we encounter or from commercial sites listing out technical information. In addition to using the Library of Congress File Format Sustainability Factors [12] as a guide, we also include the structure of the camera original format to weigh stability, as seen in P2’s separate audio and video sub directories [13] and AVCHD’s [14], spanned clips. These camera original format structures were deemed too unstable to retain, taking into account a risk analysis matrix drawing from the Sustainability Factors, internal tool requirements, structural format dependencies, and future management and access needs.

# Remembering Archival Practice

We are always bound by the ethical code of our archival training, especially as cultural heritage professionals, which include concepts like original order, provenance [15], minimizing loss, and do no harm [16]. The SAA Core Values Statement and Code of Ethics lists out many bullet points, one of which is “Develop and follow professional standards that promote transparency and mitigate harm” [17]. Andrea Shahmohammadi’s Smithsonian Institution Archives 2011 paper details these archival approaches with great clarity [18].

# Embracing a New Archival Perspective – Intentional Change

Staying true to archival core concepts might mean new models within the frame of digital formats that require processing where intentional change occurs, ie not retaining original camera structures and not maintaining a copy of the original after processing. Can we define intentional irreversible change essential to stabilize these collections? We must take action that creates change that will lead to irreversible loss but essential to current stability and future access.

# Requiring Minimal Processing

And this leads us to the action we must take. Our format research on specific camera original formats has led us to recommend action for formats that we deem too unstable. We must minimally process these files to stabilize them according to the risks we determine they have, and we have normalized this at the Smithsonian for all of our stakeholders. Some formats are simply extracted, others are re-wrapped, some are concatenated, some are flattened and re-transcoded, some are completely changed to new formats and streams [19]. FADGI AV Group’s Significant Properties for Digital Video, now in draft, serves as a guide to define the most significant technical properties to retain when migrating files with intentional loss [20].

# Sharing File Format Research

At-risk and stable and processing are all variable parameters defined according to organizational needs and tools and resources, but file formats created by video cameras and copied off card, with all their various sidecars and technical data and packaging standards, can be inventoried and documented and shared for all of us to access as a field. In looking forward, allow for a call to action to find a space to gather all of the happenings already occurring in research and practice [21].

In conclusion, the body of work that has been done in the last 10 years in building frames of thought and documentation for approaching born digital video is key to moving forward with the consideration of born digital camera original formats. It may take a new approach to video that allows archivists to face the riskiness of these formats, and it will take irreversible action towards intentional change. As Claire Fox wrote in 2020, archivist-driven research on born digital camera original formats "aims to shed light on what the ingredients of these formats are, the conditions of their creation and use, a look into historical context, and – most importantly – what responsibility archivists have to preserve them, whether to the highest standard, or *maybe something different*” [22].

## Acknowledgements

The discussion here is impossible without all of the work that has come before and all of the people working on video and digital preservation, but notably: Isabel Meyer for critical edits, support, and confidence. Dave Walker who helped fill in some research holes. Claire Fox for her careful review, edits, and excellent questions for me on expanding concepts. Taylor McBride’s research and collaboration. The SI DAMS Team and all of my SI OCIO colleagues. The countless SI video creators who let me follow them around and chatted with us about cameras. The Smithsonian AV Archivists, formalized in the AVAIL group.. All our FADGI AV colleagues, led fearlessly by Kate Murray and Carl Fleischhauer. The Smithsonian Center for Folklife and Cultural Heritage Folklife Festival Documentation Team, for so much on hands experience, notably Stephanie Smith, Cecilia Peterson, Dave Walker, and Charlie Weber. Digital archivist colleagues who I call up randomly and ask codec questions. And all of the great work cited in this paper.

# REFERENCES

1. Smithsonian DAMS. <https://si.edu/dams>. Accessed March 2022.
2. ProRes, as referred to here, is a family of codecs, all proprietary formats owned and developed by Apple. Apple ProRes white paper. <https://www.apple.com/final-cut-pro/docs/Apple_ProRes_White_Paper.pdf>
3. J. Volk. “Digital Preservation Workflow: Wrapper Formats”,. 2009. Accessed March 2022. <https://zdoc.pub/digital-preservation-workflow-wrapper-formats-jonah-volk-dec.html>
4. *C. Sanchez. What’s in your Repository? Facing Legacy Data in Smithsonian DAMS*. Crystal Sanchez. AMIA 2019. <http://www.amiaconference.net/poster-presentations-2/>
5. Adobe. *Best practices in working with native formats*. Last updated on Dec 22, 2021. Accessed March 2022. <https://helpx.adobe.com/premiere-pro/using/best-practices-formats.html>
6. Smithsonian Video Production Discussion Notes. Led by Crystal Sanchez. April 4 & April 11, 2018. <https://docs.google.com/document/d/1R0qk9Bi8ymQiI5hiEwCYvG6w-bC1asy-K9YAJe-qnnE/edit?usp=sharing>
7. Born-digital camera-original video format inventory. Presented to AMIA 2018. Taylor McBride and Crystal Sanchez. <https://docs.google.com/spreadsheets/d/1OvZkGkizNnx_nZ9OVDOkJJIVFuIMK_7FYKC77YhoUac/>

<http://www.amiaconference.net/wp-content/uploads/2019/01/2018-Program-Web.pdf>

1. Harvard did this beginning in 2016 with common digital video formats to be preserved in the library’s repository; not primarily file-based camera originals but serves as a nice model for this work. Accessed March 2022. <https://wiki.harvard.edu/confluence/display/digitalpreservation/Video+Formats>

<https://docs.google.com/spreadsheets/d/1rR7HNoQswcOrl66yeRRI2qMGDKzYQxitrOmD7nfVFGQ/>

1. *SI DAMS Supported File Formats Guide for Digital Still Images, Audio, and Video Files.* Authored by the SI DAMS Team. Published 2019. Accessed March 2022. <https://www.si.edu/sites/default/files/unit/OCIO/si_dams_supported_file_formats_2019.pdf>

*FADGI. Creating and Archiving Born Digital Video Part III. High Level Recommended Practices.* 2014. Accessed March 2022. Author note: she was a contributor to this report. <http://www.digitizationguidelines.gov/guidelines/video_bornDigital.html>

Author disclosure here that the SI DAMS system does not support complex nested file structures well so initially this discussion took into consideration tool requirements but led to a more in-depth tool agnostic diagnosis.

*Sustainability of Digital Formats: Planning for Library of Congress Collections* includes the factors Disclosure, Adoption, Transparency, Self-Documentation, External Dependencies, Impact of Patents, & Technical Protection Mechanisms. Accessed March 2022. <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>

Panasonic. How to handle data recorded on P2 cards. Accessed March 2022. <https://pro-av.panasonic.net/manual/html/VARICAM_35(VQT5K88A-10(E))/chapter04_04_07.htm>

AVCHD Format Specific Overview. Accessed March 2022. <http://www.avchd-info.org/format/>

Society of American Archivists. “Core Archival Functions.” Accessed March 2022.

Society of American Archivists. “Preservation”. Dictionary of Archives Terminology. Accessed March 2022. <https://dictionary.archivists.org/entry/preservation.html>

Society of American Archivists Core Values Statement. Accessed March 2022. <https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>

Andrea Shahmohammadi. *Born Digital Video Preservation: A Final Report*. 2011. Accessed March 2022. <https://siarchives.si.edu/sites/default/files/pdfs/bornDigitalVideoPreservation2011.pdf>

For more details on specific format choices, see the Video section *SI DAMS Supported File Formats Guide for Digital Still Images, Audio, and Video Files.* Authored by the SI DAMS Team. Published 2019. Accessed March 2022. <https://www.si.edu/sites/default/files/unit/OCIO/si_dams_supported_file_formats_2019.pdf> More work is in development to define significant properties for these specific formats for our internal audiences.

FADGI. *Significant Properties of Digital Video*. 2019. In Draft. Author note: she was a contributor to this report. <https://www.digitizationguidelines.gov/guidelines/sigpropvideo.html>

Promising work is happening with IASA’s TC-07.  *Guidelines for the Preservation of Born Digital Video.* <https://www.iasa-web.org/tc-07-guidelines-preservation-born-digital-video>

Claire Fox. *Not Normalized: Born-Digital Camera Original Video Formats in the Archives.* May 2020. Accessed March 2022. Emphasis mine. <https://miap.hosting.nyu.edu/program/student_work/2020spring/20s_thesis_Fox_deposit_copy_y.pdf>