



**Federal Agencies
Digitization Guidelines Initiative**

Creating and Archiving Born Digital Video

Part I. Introduction

December 2, 2014

The FADGI Audio-Visual Working Group
<http://www.digitizationguidelines.gov/audio-visual/>

Creating and Archiving Born Digital Video I: Introduction

By the Federal Agencies Digitization Guidelines Initiative Audio-Visual Working Group
<http://www.digitizationguidelines.gov/audio-visual/>

Version 1.1, December 2, 2014

TABLE OF CONTENTS

<i>What Is This Document?</i>	3
<i>Understanding the Case Histories</i>	3
<i>Understanding the Recommended Practices</i>	5
<i>A Snapshot of Current Practice</i>	7

WHAT IS THIS DOCUMENT?

This is one of four documents examining aspects of the current practice for creating and archiving born digital video at selected institutional members of the Federal Agencies Digitization Guidelines Initiative Audio-Visual Working Group. The three companion documents are:

- *Creating and Archiving Born Digital Video II: Eight Federal Case Histories (Version 1.1, 12/2/14)*
- *Creating and Archiving Born Digital Video III: High Level Recommended Practices (Version 1.1, 12/2/14)*
- *Creating and Archiving Born Digital Video IV: Resource Guide (Version 1.1, 12/2/14)*¹

Creating and Archiving Born Digital Video is intended to serve memory institutions, especially in the U.S. federal sector. We also hope that this report will serve the broader cultural heritage community, who may receive and ingest materials that range from high-end professional productions to more modest (but culturally important) grass-roots footage. *Creating and Archiving Born Digital Video* also offers insights and guidance to organizations that produce non-broadcast classes of content recording such as oral histories. When planning such production projects, organization have an opportunity to develop born digital video specifications from the start. Although practical and artistic aspects must be considered, such specifications can also take account of the long term needs of preservation and access.

As the preceding paragraph suggests, this report speaks in two voices. First is the voice of the archive that is receiving born digital video and seeks to preserve it for the long term. What is the range of arriving materials? How should they be assessed? What treatment, e.g., re-wrapping or transcoding, is appropriate when materials are ingested? The second voice is of the organization that oversees production and, in many cases, that organization is itself an archive. Thus one of the goals of this document set is to encourage dialog between stakeholders involved in creating born digital video files and those responsible for protecting the legacy of those files. In the end, these documents will provide practical technical information for both file creators and file archivists to make informed decisions about born digital video files and understand the consequences of those decisions for future actions. Dialog between producers and archivists is essential to sustainability and interoperability of born digital video; this document set aims to broach that topic in earnest in looking at thoughtful approaches and helpful practices.

UNDERSTANDING THE CASE HISTORIES

Context for the Case Histories

Born digital video is a growing area of responsibility for much of the cultural heritage community, including federal agencies responsible for both creating and archiving it. Born digital video formats are volatile--new ones emerge with some regularity--and archival and preservation practices are, at best, emergent. For these reasons, the FADGI Born Digital Video subgroup felt that it was premature (or perhaps even impossible) to draft a clear, definitive guideline, a cookbook for the production and archiving of these forms of content. Instead, the group decided to assemble a representative cross-section of case histories to represent the current range of practices and to use these case histories to begin the compilation of general principles and guidelines.

This report presents a diverse set of eight case histories to document the aspects of current practice in six federal agencies working with born digital video. The case history framework not only documents deliverables and specifications but also tells the story of the project, including the background of the institution and the collection but also the goals and lessons learned. The case histories create the opportunity to tell the story of a specific project so that others may learn from the experience.

¹ The URLs for the three documents are:

(II) http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p2_20141202.pdf

(III) http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p3_20141202.pdf

(IV) http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p4_20141202.pdf

As the case histories developed, a set of recommended practices emerged from the collective project experiences. The case histories and the Recommended Practices are intertwined in some sense because they impact one another. The project decisions evolved into the Recommended Practices but not all Recommended Practices are followed in the case history projects.

One important fact to note is the eight case histories are the stories of specific projects within larger institutions. Other projects within the same institutions may have different requirements, outputs and workflows. The case histories included in this document set should not be understood to be monolithic institutional policy but rather as a reflection of a specific project at a certain point in time.

In addition, these eight case histories might be considered an initial set. They are not reflective of all the FADGI members nor do they cover the myriad of projects within the contributing federal agencies. More case histories might be added in the future to address different problem sets or solutions.

Organization of the Case Histories

The eight case histories are organized into two groups:

- Creating Born Digital Video
- Archiving Born Digital Video

Breaking up the case histories along these lines is somewhat arbitrary because each of the three *Creating* case histories actually extend into the *Archiving* space as well. The *Creating* case histories might be best read as bridging across the lifecycle. The creation of new born digital video files is just part of the project because the files are later ingested into managed repositories. The *Creating* case history projects often reflect Recommended Practices aimed at archiving workflows in addition to file creation workflows.

The goal of the three *Creating* case histories is to encourage a thoughtful approach from the very beginning of the video production project, before even shooting the video, which takes sustainability and interoperability into account because choices made during the file creation process will have impacts on the long term archiving and distribution processes. They reflect Recommended Practices which illustrate the advantages of starting with high quality data capture from the very start.

The five *Archiving* case histories tell the story of bringing the born digital video files into managed data repositories for long term retention and access. These case histories explore the issues which emerge when the born digital video objects arrive at the archive. They cover topics including the challenges of dealing with diverse formats, understanding and documenting relationships among the video files and related objects, and metadata. A major topic for this case history sets is the technical characteristics of file formats: how to identify and document what formats comes in to the archive, when are changes to the file attributes needed, and what are the impact of changes to the format and encoding.

Key to Case History Abbreviations

- **LC-AFC-CRHP:** Library of Congress American Folklife Center Civil Rights History Project
- **LC-NAVCC-VEF:** Library of Congress Packard Campus of the National Audio-Visual Conservation Center Video Evergreen Format
- **LC-WebArch-YouTube:** Library of Congress Web Archiving YouTube Harvesting
- **NARA-BRCC:** National Archives and Records Administration Base Realignment and Closure Commissions project
- **NOAA-OkEx:** National Oceanic and Atmospheric Administration *Okeanus Explorer*
- **SIA-DVD:** Smithsonian Institution Archives Authored DVD project
- **SI-DAMS:** Smithsonian Institution Digital Asset Management System
- **VOA-MMAM:** Voice of America Metadata for Media Asset Management

UNDERSTANDING THE RECOMMENDED PRACTICES

Limits to the Recommended Practices

The Recommended Practices emerged from the projects detailed in the eight case histories from FADGI member agencies. They are high level by design. They could just have easily been called *Emerging* Recommended Practices because they are not intended to be comprehensive, mandatory or dictatorial. Instead, the Recommended Practices reflect the choices encountered in the projects listed but do not yet cover every issue that other projects might encounter when creating or archiving born digital video. The Recommended Practices aim to encourage informed decision-making and guide file creators and archivists as they seek out processes, file characteristics, and other practices during the production process that can lead to greater possibility that those files that follow these practices can be maintained over time.

In addition, the Recommended Practices all carry the implied preface of “if you have the option... .” Sometimes, options simply aren’t possible for a myriad of reasons, such as a defined “house style” or equipment limitations or digital storage constraints. Hardware and software limitations constrain options for file capture. Purchasing or inheriting equipment with restricted compression schemes options or only supports 8-bit capture – with no option for 10-bit and above – will impact the quality of the born digital video files you can create. When a choice is possible, however, the Recommended Practices strive to highlight the advantages of selecting one option over another.

Finally, the Recommended Practices are tightly scoped to issues essential to creating and archiving born digital video. Concerns common to digital preservation as a whole, such as consistent file naming protocols or repository actions including establishing and verifying file-level fixity or implementing write-blockers to limit access to altering file data, are all important - information on these and other topics are included in the Resource Guide – but they are not unique to born digital video files and are not addressed in these documents.

As we studied the eight case histories in order to assemble the list of Recommended Practices, we observed an interesting and instructive dissonance. Many of the case histories do not adhere to all of the Recommended Practices. We saw that the Recommended Practices are informed suggestions on file outputs, metadata choices, and workflows based on the experiences of the case history projects; they are not hard-and-fast unyielding rules. The Recommended Practices are drafted as if a working context is ideal, but not all environments are ideal. Every organization must assess their own capabilities and circumstances and make decisions and compromises based on what is desired and what is actually possible. Waiting for circumstances to align with the ideal inhibits forward movement and the cost of inaction may be far higher than any consequences to a less-than-ideal preservation work plan.

In some cases, the different goals of the case history projects necessitated different choices in order to meet business needs. We saw that an inability to follow a specific Recommended Practice did not equate to a bad decision or an inferior result. It simply means that there was a point of compromise. The central take-away is to understand the impact of the choices presented in the Recommended Practices and have a mitigation plan if needed.

Nevertheless, some of the Recommended Practices seemed especially important and came closest to “must do” status. All of the Creating case histories, for example, have options to create and capture metadata at the time of capture. All of the Archiving case histories transfer video files from unstable physical delivery carriers to more stable storage media as soon as possible. These Recommended Practices often are extensions of sound archival practice applied to the specifics of born digital video.

Finally, certain Recommended Practices may not apply to all the case histories projects. The case histories cover different problem sets so there may not be strong overlap in all cases. A case history not being listed under a Recommended Practice should only imply that the project is not directly related to the intent of the Recommended Practice.

Compression Decisions

The use, or more accurately the *lack of use*, of uncompressed video encodings is one marked example of how the case history projects deviated from the Recommended Practices. In each case history project, the need for smaller files and/or systems-specific compressed formats outweighed the need for uncompressed video. Two common rationales for compression are practical and financial restrictions. Practical considerations include system requirements (for streaming, etc.); financial considerations include costs associated with data storage and transmission.

On paper at least, the Recommended Practice for selecting uncompressed video when there's the option to do so is still a valid one because uncompressed video has advantages over compressed video including the retention of all the visual information captured at the selected resolution as well as less processing complexity. Archival repositories will want to receive and retain data-rich files which will allow the most flexibility in the future. However, uncompressed video encodings also have disadvantages such as often very large files which can be more expensive to store and difficult to transfer. For these particular projects, the compromises were made in order to meet the project goals so compressed encodings are appropriate.

Another example is the use of lossy compression in many of the case history projects when the Recommended Practice advises using lossless, specifically *mathematically* lossless, compression when compression is needed. Lossy compression, according to the FADGI glossary, "results in the [permanent] loss of information [so that] the decompressed data will not be identical to the original uncompressed data."² Lossless compression, on the other hand, uses "a lossless compression technique will allow the decompressed data to be exactly the same as the original data before compression, bit for bit. The compression of data is achieved by coding redundant data in a more efficient manner than in the uncompressed format."³ The compression ratios that can be achieved with lossless compression are generally much lower than those that can be achieved using lossy compression techniques. This means that lossy compressed files generally will be smaller than losslessly compressed files of the same content. These smaller file sizes resulting from lossy compression can translate to increased savings in storage costs but at the price of permanent data loss in the compression process.

There's a further distinction between visually lossless and mathematically lossless compression. Visually lossless compression is essentially lossy compression because "the data that is lost after the file is compressed and decompressed is not detectable to the eye; the compressed data appearing identical to the uncompressed data."⁴ The decompressed data appears to be identical but in fact, there is data loss. In mathematically lossless compression, however, no data is lost during the compression and decompression. The resulting decompressed data is identical to the original uncompressed video stream.

Of the five case history projects that implement compression, only one implements mathematically lossless compression. The remaining four use various forms of lossy compression, including visually lossless. In the LC-NAVCC-VEF case history, the Packard Campus implements a reversible and mathematically losslessly compressed profile of JPEG2000 in order to save digital storage space and by extension, cost. The JPEG 2000 core coding, lossless compression scheme documented in ISO/IEC 15444-1:2004 averages 2.5 to 1 compression ratio. With collections at the scale of those at the Packard Campus, this translates into considerable savings in digital storage space and costs compared to storing uncompressed files. Lossy compression is not a viable option for NAVCC at this time because its inherent data loss does not support the institution's mission to preserve essentially permanent data set of digital content for the full length of the Copyright term. Because the Copyright term is 150 years for most content, born digital videos created in 2013 need to be preserved until at least 2163. In addition, lossless compression does not significantly add to processing complexity over uncompressed files because the more data present, the more complex the processing. For NAVCC, the savings in the volume of data justifies the complexity of the file structure.

Lossy or visually lossless compression is the informed choice for the four other case histories. For LC-AFC-CRHP and NOAA-OkEx, implementing lossy compressed ProRes files make sense. The relatively uncomplicated content – oral history interviews – in the LC-AFC-CRHP case history project don't warrant the large file sizes from uncompressed or losslessly compressed files. Space is a premium concern in the NOAA-OkEx project in which the digital storage capacity is limited to the space on board the ocean-going ship. The smaller ProRes files allow the scientists to collect and store larger amounts of data. For the NARA-BRCC project, visually lossless MPEG-2 files are the solution because lossless compression would have resulted in larger file sizes that weren't justified by the source material. Moreover, MPEG-2 is an established product at NARA for other projects so it made sense to choose

² <http://www.digitizationguidelines.gov/term.php?term=compressionlossy>

³ <http://www.digitizationguidelines.gov/term.php?term=compressionlossless>

⁴ <http://www.digitizationguidelines.gov/term.php?term=compressionvisuallylossless>

an encoding that had strong internal support. Finally, the house standard at VOA is DV25. While DV25 is a lossy codec, it meets VOA-MMAM's business needs and is supported by essential internal VOA-MMAM systems.

Organization of the Recommended Practices

The Recommended Practices are organized into three groups:

- Advice for File Creators,
- Advice for File Archivists, and
- Advice for File Creators and File Archivists

Advice for File Creators, also known as “advice for shooters,” focuses on providing video content producers, including videographers and, by extension, the project managers within cultural heritage institutions who are responsible for the creation new born digital video files, with a set of practices that emphasize the benefits of aiming for high quality and planning for archival repository ingest from the point of file creation. File creators are encouraged to make born digital video files at the highest quality that the institution can afford to make and maintain over the long term. While current limitations are important to understand and acknowledge, file creators should look beyond present constraints towards future, and most likely more advanced, capabilities. The files should suit the immediate business needs but have a long life expectancy beyond the primary need. Capturing high quality, data rich files offers information on file/camera characteristics with the goal of identifying sustainable production practices that will allow a seamless transfer of materials to an archival repository.

Advice for File Archivists seeks to provide guidance about video-specific issues which come into play when ingesting the files into a managed storage repository. These Recommended Practices are aimed at digital preservationists including archivists, librarians, digital asset managers, and other staff within cultural heritage institutions that receive born digital video from creators and inherit the responsibility of describing, preserving, and providing access to those files. The guidance focuses on possible technical changes to acquired files, such as normalizing file formats, as well as metadata requirements needed to track those changes.

Advice for File Creators and File Archivists are grouped together because they transcend specific lifecycle points. This guidance focuses on selecting sustainable encodings and wrappers whether at initial file creation or during normalization upon ingest. Both file creators and file archivists should seek to use born digital video formats that are well-structured, transparent and contain the potential for rich data to facilitate long term retention and access.

A SNAPSHOT OF CURRENT PRACTICE

As evidenced in our case history projects, there is no one answer to the question “what format should I use to ensure sustainable long term access for my born digital video files?” The variety of target file formats in our case histories demonstrate a range of solutions based on the *fitness for purpose* concept where the workflows and deliverables achieve the specific goals set out for the project within the existing constraints and circumstances. There is variety in the current practices. Several wrapper and encoding combinations currently have support across the broader digital video preservation community including FFV1 in Matroska,⁵ JPEG2000 lossless in MXF OP1a,⁶ Uncompressed YUY2 Video Encoding (or other uncompressed encodings) in either AVI⁷ or QuickTime.⁸ Each of these, and other combinations, has strengths and weaknesses which make them viable in some environments and not viable in others.

In addition, FADGI is near completion on the MXF Application Specification for Archiving and Preservation which will offer strong support for the preservation needs of the cultural heritage communities and beyond.⁹ Once published, this application specification is expected to help increase the adoption of MXF especially within large

⁵ <http://www.digitalpreservation.gov/formats/fdd/fdd000343.shtml>

⁶ <http://www.digitalpreservation.gov/formats/fdd/fdd000206.shtml>

⁷ <http://www.digitalpreservation.gov/formats/fdd/fdd000366.shtml>

⁸ <http://www.digitalpreservation.gov/formats/fdd/fdd000369.shtml>

⁹ http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html

federal agencies and increase the tool set options available from vendors. But even with the FADGI-sponsored AS, MXF may not be for everyone in all situations, especially for smaller institutions.

The documents comprising *Creating and Archiving Born Digital Video* are intended to illustrate the current practices in selected memory institutions in the U.S. federal sector when creating and archiving born digital video. The project goals, constraints, workflows, file format selections, and other parameters all represent a “slice in time” view of what institutions are doing with individual collections right now. These are very likely to change over time as practices advance and mature. FADGI is interested in exploring these topics again in the future to evaluate the progress.