Smithsonian Institution

OCIO Digital Asset Management System (DAMS)

Smithsonian DAMS Supported File Formats

This document is intended to provide guidance in choosing file formats and digital archival workflow steps for storage in the Smithsonian Digital Asset Management System. You can find Smithsonian DAMS supported formats for Images, Audio, and Video formats outlined below. Often, file formats acquired or captured are not stable enough for long term storage, and yet other times normalizing file formats has its own consequences. This guide outlines the most common formats and troubleshooting issues encountered at the Smithsonian Institution by the DAMS team. For additions or questions, please contact your DAMS team.

1. Digital Archival Still Images
   - Introduction
   - Still Image File Formats for Preservation and Access
   - Common Issues in Image Processing
2. Digital Audio Files
   - Introduction
   - Digital Audio File Formats
3. Digital Video Assets
   - Introduction
   - Video Wrappers
   - Video Streams/Capture Formats
4. Related Resources
   - Digital Images
   - Digital Audio
   - Digital Video

Smithsonian DAMS Supported File Formats
   I.   **Introduction: Archival Still Images (Two-Dimensional)**

When capturing and processing still images, there are a variety of file formats available depending on the camera/capture devices and software versions utilized. Below are some of the most common still image file formats currently in use at Smithsonian. These format types are the files currently supported by the SI DAMS[1].
If you have a proprietary Camera Raw format, you may be required to convert the file to a supported format upon ingest to ensure eventual sustainability and migration of assets.

   II.   **Still Image File Format Recommendations for Preservation and Access:**
            Color key:

	Most commonly seen formats
	Consider generating additional image alongside capture original
	Conversion is necessary from the capture/original format / Decision on data
	DAMS does not support/Consult 3D repository

Capture/Original Format	Preservation Master	Access Derivative	File Considerations/Recommendations	File Attributes/Vulnerabilities
.DNG	.DNG	.TIF	May consider generating .TIF or .JPG for working or access copy. Generates DAMS Proxy: YES	• Preferred RAW format • Developed by Adobe systems • Based on TIFF/EP standard • Three types: In- camera RAW, Converted DNG RAW, Converted DNG Linear • Widely accepted • Embedded Metadata

---

Smithsonian DAMS Supported File Formats

				• Contains Raw image data and metadata (self-contained) may contain JPEG preview • Lossless and Lossy Compression **Pros:** • Can adjust range of parameters without data loss • Openly published • Submitted to ISO for standard[2] • Most common RAW format of choice for long term preservation **Con:** • Must process file before usable
.TIF[3]	.TIF	.TIF or .JPG	Ingest to DAMS Can ingest high bit depth TIFF masters (in lieu of RAW for reprocessing purposes depending on analog materials and cost analysis of project) Generates DAMS	• Lossless compression • TIFF/EP (ISO 12234-2:2001) ISO Standard • Needs plug-in or external application for web display • Tag with Basic Metadata (TIFF Header) • Embedded Metadata • Adobe –proprietary • Most common Image file format chosen for sustainability/ long term preservation • Used for files up to 64-bit depth • Wrapper: with encodings that include uncompressed, LZW compressed, LZW lossless compressed (20-30% storage space savings with no drawbacks) or bitonal-Group 4 compression. Often preferred: uncompressed raster data

---

[2] DNG has been submitted to ISO. Current Documentation: Digital Negative(DNG) Specification Ver. 1.4.0.0 June 2012
[3] Officially .TIF is the preferred file extension

Smithsonian DAMS Supported File Formats

			Proxy: YES	in a TIFF wrapper. LZW lossless compressed also growing in use. **Pro:** • Widely accepted for long term archiving, low implementation cost, can save high bit TIF **Con:** • May want more control and need to revisit original shot (i.e. Need for .DNG)
.JPG	Ingest to DAMS as master if no TIFF or RAW format available.	Ideal format for accessibility and quick, easy downloads.	Use highest settings available. May degrade quality; ensure image parameters and compression are minimal. Generates DAMS Proxy: YES	• Lossy compression (ISO/IEC 10918) • Standard for web display • Native to web browsers • Embedded metadata • Can be used for files up to 24-bit depth • Rarely used for masters in digital preservation • Used in most Digital Cameras **Pros:** • Reduced size versus TIFF and RAW • Can adjust degree of compression • JPEG/EXIF most common digital camera format **Con:** • Lossy compression, may degrade quality
.JP2	Ingest to DAMS as master if preferred or if TIF or		Use highest settings available. May degrade quality; ensure image parameters	• JPEG 2000 standards include three encodings; *core* encoding is free of patents • JP2; JPF; JPX JPF is an extended JPG2 format to include image

Smithsonian DAMS Supported File Formats

	RAW format are not available.[4]		and compression are minimal. .JP2 may need to convert to .JPF to pass JHOVE if .JP2 fails. Contact DAMS for setting conversion Generates DAMS Proxy: YES	transparency. Can save with ability to include standard support for .JP2 and display extended .JPX • Lossy and lossless compression • Offers options such as tiling, scalability • May be good for large oversized materials **Pro** • Reduced size of files • Reduced storage cost **Con** • Universal adoption has been mixed in cultural heritage[5] • Limited support in imaging software; tool cost may be high and limited availability
.EIP or .IIQ (Phase One)	Ingest to DAMS when available or special use case that proprietary RAW	Consider generating TIFF for access alongside ingest of. EIP RAW	May want to retain for reprocessing purposes which allows for relative colorimetric interpretation. Generates DAMS	• Proprietary format • Packages RAW file to contains: RAW file (i.e. IIQ, CR2, or NEF) +Settings (Capture One Settings) + ICC Profile (International Color Consortium) +LCC (Lens Correction Calibration) file • .ZIP technology • Cannot be converted to .DNG without important data loss • Embedded Metadata

---

[4] .JP2 is not a preferred format for long term preservation, but in some practical use cases and some vendor projects has been used. TIF and/or RAW files may not be available.

[5] Please refer to FADGI's *Raster Still Images for Digitization: A Comparison of File Formats* section on JPEG 2000 and its adoption for more information.

	should be maintained		Proxy: NO	**Pro:** • Contains RAW information **Con:** • Proprietary format
.3FR and/or. FFF (Hasselblad)	Ingest to DAMS when available or special use case that proprietary RAW should be maintained	Consider generating TIFF for access alongside ingest of .3FR and. FFF RAW	May want to retain for reprocessing purposes which allows for relative colorimetric interpretation. Generates DAMS Proxy: NO	• Proprietary format • Cannot be converted to .DNG without important data loss • Embedded Metadata **Pro:** • Contains RAW information **Con:** • Proprietary format
.NEF	Convert to DNG	Consider generating TIFF for access alongside ingest of converted. DNG master	Recommend conversion to DNG prior to DAMS ingest for long term sustainability[6] Generates a DAMS Proxy: No/YesDependent on Processing	• Proprietary format • Consider conversion to. DNG prior to ingest • Embedded Metadata **Pro:** • Contains RAW information **Cons**: • Proprietary format • High-risk long term • Long term compatibility risk

---

[6] Refer to RAW to DNG Conversion section for conversion parameters.

Smithsonian DAMS Supported File Formats

.CR2	Convert to DNG	Consider generating TIFF for access alongside ingest of converted .DNG master	Recommend conversion to DNG prior to DAMS ingest for long term sustainability[7] Generates a DAMS Proxy: No/Yes-Dependent on Processing	• Proprietary format • Consider conversion to .DNG prior to ingest • Embedded Metadata **Pro:** • Contains RAW information **Cons**: • Proprietary format • High Risk long term • Long term compatibility risk
.XMP sidecar files	N/A		Legacy files may have .XMP associated files ingested. Consider whether data is contained record and if .XMP files are necessary to maintain.	Contain metadata

---

[7] Refer to RAW to DNG Conversion section for conversion parameters.

Smithsonian DAMS Supported File Formats

.DICOM	May be legacy versions across SI, consider conversion to DICOM 3.0 before DAMS ingest	Convert to DICOM-3 for SI wide use.	Consult with DPO 3D repository development may lend itself to stacked data. OpenText does not have module support. Reported Units with DICOMS: NASM, NMNH, SIL Generates a DAMS Proxy: NO	• Specialized format (i.e. CT Scans) DICOM1, DICOM 2, DICOM 3[8][9] • DICOM headers include information on hardware and software associated with the individual file. All versions contain these types of embedded technical metadata, though DICOM 3 is the most robust. • Embedded Metadata **Pros**: • DICOM-3 easier to maintain than prior DICOM formats • Keeping files in DICOM format preserves key data points that are not preserved when converted to image formats. **Cons:** • More complex format to maintain/ Legacy formats include variations • Older DICOM formats High-risk long term
.DPX	.DPX		Recommend to embed profile upon ingest for generation of accurate thumbnail and screen view in DAMS. Generates a DAMS Proxy: NO	• Still picture digitization/ based on Kodak Cineon (.CIN) format • Maintains more information than TIFF format for still images from motion picture film • Embedded Metadata **Pro:** • Can hold 48-bit RGB or RAW data **Con:** • Support for .DPX can be problematic

---

[8] There exists a variety of legacy DICOM formats at SI, please consult the DAMS teams or DPO should you hold DICOMS.
[9] DICOM format created by NEMA, and is ISO 12052:2007

Smithsonian DAMS Supported File Formats

Overall, you may consider a workflow where you convert proprietary RAW formats to .DNG for long-term preservation (i.e. .NEF and .CR2). Exceptions with .EIP and .3FR and .FFF are noted above. If you have TIFF files in some use cases, the high TIFF file may want to be maintained. Overall the use case of the analog material and the requirements of the project will inform your decisions of which formats will work best for the project. Always consider long-term sustainability and migration of the file as well as the need to reprocess the master.

III. Common Issues Seen at SI DAMS: RAW to DNG Conversion

When saving a RAW file directly from a capture device, the processing of that file and combination of software settings utilized in converting to DNG, may generate a DAMS proxy that has a green or magenta cast, upon ingest to DAMS. Likewise, the file icon, prior to ingest may or may not accurately represent the file. **However, once the file has been ingested into DAMS it will download (export) from DAMS correctly.** Based on testing, the issue does not appear to be generated or caused by the DAMS. Rather, it appears to occur in the initial processing of the RAW image off the camera prior to ingest into DAMS. Please review file settings /conversion settings and icons prior to ingest.

Some Conclusions:
1. CR2 RAW file converted to .DNG with Embed Fast load setting = magenta cast .DNG proxy in DAMS. May also occur with .NEF files
   Recommendation: Do Not use Embed Fast Load setting; No JPG Preview
2. .IIQ or .FFF convert to .DNG = green cast may occur as loss of data occurs in conversion
   Recommendation: Do not convert .IIQ or .FFF to .DNG. Consider producing a .TIFF file alongside ingest of the .IIQ /.FFF file.
3. Setting: Embed Original RAW in DNG. There may be a use case/ practical implementation where you would want to embed the original RAW in DNG. Doing so will increase file size. More often than not, the conversion to DNG appears sufficient.
   Recommendation: This is currently under review. Use cases. Note: NMNH never has had in instance to do this.

Smithsonian DAMS Supported File Formats

4. **Report issue.** If you experience an issue. Please document the following. Name of file(s). Software (including version) and workflow utilized in creation of image.

Software and Recommended Settings

The following applications help assist in conversion to DNG. They provide ability to convert camera RAW files from proprietary formats to the more accepted/ usable DNG.

Adobe Camera RAW (ACR) – This is a free plug-in but requires a subscription to an Adobe imaging product (i.e. Adobe Photoshop). https://helpx.adobe.com/camera-raw/using/supported-cameras.html

Adobe Digital Negative Converter (ADC) – Free download from Adobe to convert to DNG. https://helpx.adobe.com/photoshop/using/adobe-dng-converter.html

II.   **Some Common Issues in Image Processing:**

Common Issues	Description of Issue	Correction of issue
Saving Parameters	Save parameters are set for Byte order for **Macintosh**	Save parameters should be set for Byte order for **IBM PC** (This is most cross platform)
Conversion from one format to another, i.e. TIF to JPG, not JPG to TIF	Some files have been converted from a lossy format to a lossless format. Converted from lower byte depth to higher. Creation from lower quality file to higher quality file. I.e. JPG to DNG; JPG to TIF	Always convert from a Lossless format to a lossy format; not vice versa. Convert from higher byte depth to lower byte depth. Creation from higher quality file to lower quality file. i.e. DNG to TIF; TIF to JPG

Smithsonian DAMS Supported File Formats

Embedded Profile	Image file has Untagged Color Profile. Seen often with legacy images	Embed a Profile (Adobe RGB 1998; Gray Gamma 2.2, SRGB)
Commonly Supported Color Profiles	Color Profiles that are more widely used. Project specifications/nature of original capture may dictate which profile is utilized	-Adobe RGB 1998; Gray Gamma 2.2, SRGB -ProPhoto and eciRGBv2 (offer wider gamut
Stitched images-maximum issue (saving parameters)	Ingesting very large image files to DAMS for delivery via IDS	max IDS size as 65000x65000 Size limit set by JPEG: JPEG/JFIF supports a maximum image size of 65,535×65,535 pixels,[5] hence up to 4 gigapixels for an <u>aspect ratio</u> of 1:1
IPTC-X Digital Source Type field	Field is populated and fails on ingest; -Values are populated with free text and do not adhere to standard drop down values -Invalid domain value (Selected Value) for field. The (Selected Value) of the field has not been populated correctly. - Notes: Drop down Values for this IPTC-X standard are correct in pointing to IPTC-X standard URL link. i.e. <Iptc4xmpExt:DigitalSourceType> http://cv.iptc.org/newscodes/digitalsourcetype/digitalCapture </Iptc4xmpExt:DigitalSourceType>	-Suggest review of files and IPTC-X field in imaging software. Either remove incorrect value, or if value appears to be correct and **is already populated**, **manually make the value change** then resave the file. This clears error. -If need to remove from Large batch of files, EXIF TOOL option to remove in larger batches Contact DAMS team.

Smithsonian DAMS Supported File Formats

III.   Embedded Metadata:

Depending on what file format you choose you should always consider the quality of your embedded metadata. You can review the EXIF, IPTC, and XMP metadata. You may choose to add additional embedded metadata but be aware that some metadata is already generated when you take the picture. Consider how this will work in conjunction with the DAMS level metadata and your existing workflow to create a more robust and sustainable digital asset.

IV.   JHOVE Image/Audio Validation:

JHOVE is a file format identification, validation and characterization tool. It is currently being implemented in all SI imaging projects that utilize VFCU (Volume File Copy Utility). The software analyzes image file formats for characteristics such as "well-formed and valid." It also provides format characterization for format specific files. VFCU determines the expected file format for each file passing though VFCU based on the filename extension (final part of the filename following the last period '.') If the filename extension is listed in the following table, VFCU will then use JHOVE to verify if the file is indeed the expected file format. Filenames are not case sensitive (for example, a file will go through the same JHOVE validation if the filename extension is TIF or tif). Below is a list of formats at SI that are run through JHOVE as part of the VFCU process:

Filename Extension	Expected File Format
tif, tiff, dng[10]	TIFF (http://jhove.openpreservation.org/modules/tiff)
.jpg, .jpeg	JPEG (http://jhove.openpreservation.org/modules/jpeg)
.jp2, .jpf, .jpx	JPEG2000 (http://jhove.openpreservation.org/modules/jpeg2000) ; current BUG in JHOVE that causes .JP2 failure in some files; conversion to .JPF has passed in JHOVE.
.pdf	PDF (http://jhove.openpreservation.org/modules/pdf)
.aif, aiff	AIFF[11] (http://jhove.openpreservation.org/modules/aiff)
.wav	WAVE[12] (http://jhove.openpreservation.org/modules/wave

---

[10] Although it is not the same as a tiff, DNG files DO meet tiff specifications as far as file format layout.

[11] Audio format described in next section

Smithsonian DAMS Supported File Formats

Introduction: Audiovisual Materials

When capturing, analyzing, or processing audio and video files, there are a variety of file formats available depending on the camera/capture devices, digitizing decisions, and/or donor provided files. Although constantly changing, below are some of the most common digital audio and video file formats currently in use at Smithsonian, with pros and cons outlined for each choice. The breakdown below of format types outlines what formats are currently supported by the SI DAMS, and what is recommended for formats that are not. Remember if you have a proprietary Capture/Camera Original Format, you may be required to convert the file to a supported format before ingest to ensure eventual sustainability and migration of assets.

I. Digital Audio File Formats for Preservation and Access

Original Format	File Considerations/Recommendations	File Attributes/Vulnerabilities
WAVE (.wav)	Ingest to DAMS to keep as your preservation master file. Recommended Audio Preservation file format. Generates DAMS Proxy: YES	WAV is a widely adopted Microsoft audio format with lossless compression. The Broadcast Wave Format (BWF) is based on the Microsoft WAVE audio file format, with additional required metadata. **Pros:** • WAVE supports embedded metadata using BEXT and INFO chunks, which DAMS extracts and indexes.

Smithsonian DAMS Supported File Formats

		• For preservation, most common audio stream is PCM (Pulse Code Modulation); Linear PCM is an uncompressed format. • LPCM wrapped in BWF is considered the standard preservation format audio files. For recommendations on detailed technical audio specification, see the document on AV Digitization Format Recommendations. • DAMS will make an MP3 proxy upon ingest so ingesting derivate MP3 files in addition to a master WAVE files is not recommended.
MP3 (.mp3)	If there is no WAVE file, ingest MP3 to DAMS to keep as your master file. However, this is not considered a preservation format. Generates DAMS Proxy: YES	MP3 is a widely adopted file format and generally regarded as an access format. **Pros:** • MP3 supports embedded metadata using ID3 tags, which are extracted & indexed. • The DAMS creates MP3 proxies for all audio files ingested to the DAMS that can then be exported out for access/researcher needs, so if you have a higher quality WAVE file, we don't recommend ingesting MP3 into DAMS. **Cons:** • Proprietary, lossless compression

Smithsonian DAMS Supported File Formats

FLAC	If conforming to preservation standards, ingest to DAMS to keep as your master file; if considered an access file, only ingest if no access file is available. Generates DAMS Proxy: YES	Free Lossless Audio Codec (FLAC) is a lossless compression open source format, typically used for access, but now growing in support as a preservation format in developing standardization through the CELLAR group. **Pros:** • Lossless Compression; FLAC supports linear PCM • Free and open source • Undergoing standardization **Cons:** • No widespread use; growing in popularity as a preservation format, but still not a lot of tools, vendors using it. • Until standardization, risk of interoperability issues with early uses.
AIFF	If there is no WAVE file, ingest AIFF to DAMS to keep as your master file. Consider transforming into a more open format like WAVE if storing uncompressed data. Not considered a preservation format. Generates DAMS Proxy: YES	Audio Interchange File Format was developed by Apple in 1988. Can store uncompressed data or can be compressed. **Pros:** • Widespread use especially in audio production • Supports uncompressed PCM **Cons:** • Proprietary Apple format. • Compressed data is lossy.

Smithsonian DAMS Supported File Formats

WMA, WMA-9, WMA-8, WMA-7	Consider transcoding into a more widely used codec, retaining original specs as possible. No longer a DAMS supported format.	Windows Media Audio. **Cons:** • Proprietary Microsoft format. • A compressed format used for distribution. • Not good for long-term preservation. • Sparsely used and becoming more obsolete.

Smithsonian DAMS Supported File Formats

II. Digital Video File Formats for Preservation and Access:

Digital Video assets are constructed of a video *wrapper*, also known as a container, which contains multiple *streams*, or tracks, inside it, including video, audio, text, and other supported types. The following breakdown lists common wrapper formats first and then common video streams & capture formats. For full support, the video wrapper and all contained streams should be considered for full DAMS support and sustainability analysis.

When acquiring or producing born digital video materials for preservation and ingest to the DAMS, there are many file format options. SI DAMS team recommendations acquiring and storing the highest quality available of finished production materials. Pre-production versions and raw footage may be stored as appropriate. Born digital video file technical specifications change rapidly over time as technology changes, so these guidelines are general recommendations with some current examples below. SI DAMS cannot ingest complex sets of files, so archiving camera original formats that are received in nested folder structures is not supported.

For produced content:
- Receive final production versions over or in addition to pre-production versions.
- Request files that reflect original production resolution and frame rate, or a high-resolution output of the produced material.
- Receive stable, individual files as opposed to many clips of one work, when possible. Choose from a variety of codecs supported by DAMS transcoders (see list in Video User Guide)

For raw footage or camera originals:
- Request original raw footage if it can be played and supported. If not, request format technical specs that retain as much of the specs as the original files, especially in color space, bit rate, and resolution.
- Only normalize files to more standard formats when the original formats are at risk and cannot be ingested as individual files in DAMS. Prioritize retaining of original technical specifications, especially in color space, bit rate, and resolution.

Smithsonian DAMS Supported File Formats

Original Format	File Considerations/Recommendations	File Attributes/Vulnerabilities
VIDEO WRAPPER		
QuickTime File Format/MOV (.mov)	Check video and audio streams to be sure they are supported formats (listed below). Ingest to DAMS if it is the wrapper of the master file. Consider adoption of MXF or MKV for archival masters in the future.	MOV is a video wrapper format initially developed by Apple but then standardized under ISO. **Pros:** • Very widespread use and supported by many editing suites and cameras. • Can contain a wide number of compressed and uncompressed video formats. • Familiar with users and supported by many players. **Cons:** • Proprietary • Data streams are stored in an mdat atom and the information for reading the data is stored in a separate moov atom either at the front or end of the file, making this format more susceptible to loss if this moov atom is missing, truncated, or corrupted.

Smithsonian DAMS Supported File Formats

MPEG-4 (.mp4)	Check video and audio streams to be sure they are supported formats (listed below). Ingest to DAMS if it is the wrapper of the master file.	An ISO standard, MPEG-4 Part 14 is based on the QuickTime File Format, published in 2003. **Pros:** • MP4 is a widely adopted video container that can store video, audio, still images, subtitles, and other data. • Capable of containing a wide range of video streams, most commonly AVC/h.264 at this time. **Cons:** • Considered an access format, as it is the most common video streaming container with largely compressed data inside. • Cannot store timecode.
MXF (.mxf)	Ingest to DAMS if the master file is OP1A. If not, rewrap as appropriate. Check video and audio streams to be sure they are supported formats (listed below).	The Material Exchange Format (mxf) container was developed for video producers/broadcasters with 10 varieties standardized by SMPTE to increase interoperability. **Pros:** • MXF version OP1A was standardized by the archival community as the standard for archiving (Library of Congress uses this as their standard). Operational Pattern 1a (OP1a) defines the file as a single file self-

Smithsonian DAMS Supported File Formats

		contained comprising of a single essence or interleaved essence elements (one file). **Cons:** • MXF containers can contain a variety of audio and video formats, many camera originals, and sometimes in separate spanned or split streams, creating unstable packages for long term storage. Be sure to store and create OP1a and follow archival guidelines (formerly AS-07 at the Library of Congress).
MPG2/MPEG-2 (.mpg/.mpeg)	Ingest to DAMS if it is the master file. Maintain as much of the original specifications as possible if transcoding (ie DVDs).	Commonly known as MPEG-2, the ISO specification defines the video format developed by the Motion Picture Experts Group (MPEG). Originally developed in 2006 for broadcast/transmission purposes. **Pros:** • A history of wide adoption by many distribution outlets (broadcast, etc), leading to uses in many common practices, notably for mezzanine/edit master files. • Can be seen in very high bit rates- ie 50mbps for optimal editing.

Smithsonian DAMS Supported File Formats

		• You may see MPEG-2 PS (program stream) files or TS (transport stream) files. • The DVD authored video format primary uses the MPEG-2 format. **Cons:** • The standard defines the bitstream and decoder, but not the encoding, leading to a variety of slightly different encoding options, increasing the risk that files will not be interoperable. • Lossy, compressed format often used as an option for high quality derivatives/editing. Not considered a preservation format. • MPEG-2 files cannot accept additional embedded metadata. • Closed wrapper can only contain a very limited amount of and strictly defined video and audio streams.
AVI (.avi)	Ingest to DAMS if it is the master file. Consider re-wrapping if loss can be minimized.	AVI (Audio Video Interleaved) is a proprietary format developed by Microsoft and IBM as part of RIFF (the Resource Interchange File Format). **Pros:**

		• Supports a wide range of video codecs, with AVI_DivX and AVI_Indeo being two of the most common. • AVI was first specified in 1992 and is still widely used, most often as an access format. **Cons:** • AVI files previously used as preservation masters have been retired and replaced at most large collecting institutions. • Proprietary and large data packages • May contain WAVE and MP3 Constant Bit Rate audio but cannot contain some types of Variable Bitrate (VBR) data reliably, such as MP3 audio at sample rates below 32 kHz.
Matroska (.mkv)	Ingest to DAMS if it is the master file.	An open source, widely adopted wrapper that has been in use since 2002. **Pros:** • Widespread adoption as an internet video format. Current standardization for archiving being undertaken by the CELLAR group. • Data is stored in clusters with hash values stored at the beginning of each clusters,

		allowing for less risk in loss of the entire file in the event of corruption or data loss. • Allows for structured metadata and file attachment, as well as subtitle and chaptering capabilities. Metadata manipulation and other preservation and file validation activities through the MKVtoolnix suite of open source tools. • Broad support of a variety of video encodings and is currently used in preservation workflows with the FFV1 codec. **Cons:** • Still new as an option for archiving. Many vendors and tools do not yet support it. • Before standardization, risk of interoperability issues.

Smithsonian DAMS Supported File Formats

Original Format	File Considerations/Recommendations	File Attributes/Vulnerabilities
VIDEO STREAM/CAPTURE FORMAT		
AVC/h.264	Ingest to DAMS if master file. If a supported master file exists and a low quality AVC copy exists, there is no need to retain. Generates DAMS Proxy: YES	Advanced Video Coding, or AVC, also known as h.264. **Pros:** A very common and widespread video stream that can be found in an MOV or MP4 wrapper, among others. - It is a common capture format, coming directly off camera. - This flexibility and widespread use make it a good stable video stream supported by DAMS. If retaining as a master file, consider retaining a high-quality copy and/or specifications original/close to the capture settings. - DAMS proxy files are created with a low res AVC stream inside an MP4 wrapper with AAC audio. **Cons:** - Lossy compression. - It can be stored in various bit rates, causing it to be either very high quality (and large data size) or very low quality. It

Smithsonian DAMS Supported File Formats

		can be uncompressed (very high quality) or very very compressed (the most common choice for video streaming often paired with AAC audio streams or Advanced Audio Coding).
ProRes	Ingest to DAMS if master file. Generates DAMS Proxy: YES	Proprietary Apple codec optimized for Editing; not a camera original file. **Pros:** - A very common format due to the need to transcode files for editing, leading to most camera originals being turned into ProRes and then discarded. **Cons:** - Proprietary, not an open format. - Very large data sizes due to intraframe compression. - A variety of very high resolution but still lossy compression ProRes formats with standard technical specs include: ProRes 422, ProRes 422, ProRes 422 HQ, Prores 444.
DV, HDV, DVC-PRO	Ingest to DAMS if master file. If digitizing or transferring/rewrapping, preferred to stay within the DV family.	All are part of the DV (Digital Video) family of video codecs, developed in 1996. **Pros:** Very widespread use, open simple format.

	Generates DAMS Proxy: YES	• DV uses intra frame compression or compression in each frame, not between frames, making it ideal for editing. • Varieties developed off DV include similar features but with variety in technical specs: DV(25mpbs); DV-50 (50mbps); DVCPro (100mbps); HDV (HD&MPEG2) **Cons:** • Lossy compression
MPEG-2	Ingest to DAMS if master file. Generates DAMS Proxy: YES	See MPEG in container section about. MPEG-2 is an entire package that has to have specific mpeg video and audio streams. **Pros:** • Widespread use **Cons:** • Because the standard is built on decoding and not encoding, interoperable issues exist • Do not support embedded data • Lossy compression
AVCHD	Cannot be ingested to DAMS in its native package format. Recommendation to span split clips and create single files for each clip or export one clip per sequence (multiple related clips) as	Widespread Camera Original format developed by Sony and Panasonic in 2006. **Pros:** • A very common capture format due to the high quality produced for small data packages.

	appropriate. Retain close to original video stream specs as possible (h.264/AVC) and rewrap into a common format, most likely .mov Generates DAMS Proxy (mts video stream files): NO	• Supported AVC video stream (see above) **Cons:** • lossy compression built to maximize storage of high definition video on limited camera storage/flash drives. • Files are packaged in a proprietary directory AVCHD directory structure. • Only natively viewable on Mac OS. • Format highly dependent on retaining directory structure. • Video files are stored in a proprietary MTS wrapper with a supported AVC/h264 video stream but they get split into 2GB chunks in a Stream folder. • Spanning clips can make audio drift if not done correctly; codecs predominant but MTS wrapper is not. • Initially built to be compatible with Blue-ray Disk format for access.
XDCAM	Ingest to DAMS if the master file is OP1A. If not, rewrap as appropriate. Generates DAMS Proxy: YES	A Sony developed production level camera original format developed in 2003 and with a variety of types that vary in technical specs: XDCAM SD, XDCAM HD, XDCAM EX and XDCAM HD422 **Pros:** • Widespread camera original.

		• Camera original package has a nested folder structure, but video files are standalone and sidecar files most often do not contain archival relevant data and can be discarded. • Video codec is mostly MPEG2 for HD or can be DV for SD, wrapped in MXF. Both supported formats.
P2	Cannot be ingested to DAMS in its native package format. Recommendation to create single files for each clip or export one clip per sequence (multiple related clips) as appropriate, but retain close to original video stream specs as possible and transcode file for stability. Generates DAMS Proxy: NO	Camera Original Panasonic Format. **Cons:** • Complex directory structure; P2 individual media files use the FAT32 file system and are therefore limited to 4 GB. Sequences can be spanned across multiple files. • The format is highly dependent on retaining directory structure; audio and video streams stored in separate folders making them very unstable.
XAVC-I; XAVC-L; XAVC-S	Ingest to DAMS if master file is wrapped in MXF OP1a or other supported container formats. Generates DAMS Proxy: YES	Camera Original Sony prosumer format **Pros:** • XAVC-L and XAVC-I use the MXF wrapper; XAVC-S use an MP4 wrapper. • XAVC is based off the AVC prediction and is widespread use and supported by DAMS.

Smithsonian DAMS Supported File Formats

		• Gaining in popularity so the amount of files we are seeing is steadily increasing. **Cons:** • Huge data packages. A common format for 4k productions creating very large files. • Lossy compression. • A very new camera original format carries its own risk.
REDCODE/ R3D	Create something else. Do not keep proprietary files that we cannot open and view. Generates DAMS Proxy: NO	Camera Original RED Digital Cinema Camera **Cons:** • Very Proprietary, need special tools to render.
m2v1	Either keep native files or transform into an open format when resources are available. Generates DAMS Proxy: YES	Codec used as part of the Tricaster, a tool used by SI for livestreaming and capture. **Cons:** • Newtek proprietary codec **Pros:** • Although not an open codec, the codec pack was given to us freely by the vendor and we can access and transform the files. Due to scale of this collection, we support storing these files.

Smithsonian DAMS Supported File Formats

FFV1	Ingest to DAMS if master file. Preferred format is ffv1.3. Not recommended to turn already lossy compressed files (ie born digital camera originals) to ffv1. Generates DAMS Proxy: YES	Widely used as a standard for capture in analog to digital video preservation workflows. **Pros:** • Open source codec with lossless encoding, used with a Matroska (.mkv) wrapper. Created through the ffmpeg project and standardizing through the CELLAR group. • Properties conducive to archiving include smaller data size due to lossless compression (about 1/3 of size of uncompressed); the ability to create a crc hash for each frame or each slice of a frame; and self description. • FFV1 files with MKV wrappers allow for embedded metadata and other preservation and file validation activities through the MKVtoolnix suite of open source tools. **Cons:** • Not widely used but growing in popularity. Most tools are command line.
Uncompressed Video	Ingest to DAMS if master file. Preferred format is v210. Not recommended to turn already lossy	Widely used as a standard for capture in analog to digital video preservation workflows.

	compressed files (ie born digital camera originals) to uncompressed video. Generates DAMS Proxy: YES	**Pros:** • No compression allows for less loss if there is any corruption. • Widespread use at the Smithsonian. Video is commonly created at 10bits with the v210 fourcc, but there are other varieties. **Cons:** • Very large data packages • Interoperability could be a problem due to decoder differences without a codec defined.
Motion JPEG2000	Ingest to DAMS if master file. Not recommended to turn already lossy compressed files (ie born digital camera originals) to Motion JPEG2000. Generates DAMS Proxy: NO	Widely used as a standard for capture in analog to digital video preservation workflows. **Pros:** • Lossless compression creates smaller data packages than uncompressed. • Standardized as one of the few open standards for video digitization standardization and used by the Library of Congress (wrapped in MXF). **Cons:** • Not widespread use and limited tools for creation and troubleshooting • Early tools exhibit interoperable issues.

Smithsonian DAMS Supported File Formats

Related Resources

1. **Digital Images**
- FADGI (Federal Digitization Guidelines Initiative) Still Image File Format Comparison
  http://www.digitizationguidelines.gov/guidelines/File_format_compare.html
- Sustainability of Digital Formats Planning for Library of Congress
  http://digitalpreservation.gov/formats/intro/intro.shtml
- NARA Reformatting Approaches
  http://www.archives.gov/preservation/products/definitions/reformatting.html
- Digital Negative (DNG) Specification Ver. 1.4.0.0 June 2012
  https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf
- JHOVE
  http://jhove.openpreservation.org/

2. **Digital Audio**
- FADGI (Federal Digitization Guidelines Initiative) Guidelines: Embedded Metadata in Broadcast WAVE Files, Audio-Visual Working Group
  http://www.digitizationguidelines.gov/guidelines/digitize-embedding.html
- BWF MetaEdit Tool:
  http://sourceforge.net/projects/bwfmetaedit/
- *Essential Resources for Audio Preservation*, ARSC
  http://www.arsc-audio.org/pdf/ARSCTC_resources.pdf
  ARSC Guide to Audio Preservation
  http://www.clir.org/wp-content/uploads/sites/6/pub164.pdf

Smithsonian DAMS Supported File Formats

- IASA-TC04: *Guidelines on the Production and Preservation of Digital Audio Objects*, International Association of Sound and Audiovisual Archives
  https://www.iasa-web.org/tc04/key-digital-principles

3. **Digital Video**

- *File Formats and Compression Guide: Digital Video Formats*, JISC Digital Media:
  http://www.jiscdigitalmedia.ac.uk/guide/file-formats-and-compression/#video
- Format Descriptions, *Sustainability of Digital Formats: Planning for Library of Congress Collections*:
  http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml
- IASA-TC 06 Guidelines for the Preservation of Video Recordings
- https://www.iasa-web.org/tc06/guidelines-preservation-video-recordings
- FADGI (Federal Digitization Guidelines Initiative) Creating and Archiving Digital Video:
  Part 1: Introduction http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p1_20141202.pdf
  Part II: Eight Federal Case Histories: http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p2_20141202.pdf
  Part III: High Level Recommended Practices:
  http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p3_20141202.pdf
- Codec Encoding for LossLess Archiving and Realtime transmission (cellar)-
  https://datatracker.ietf.org/wg/cellar/charter/
- Born-Digital Camera-Original Video: Practices and Risks, SI, OCIO, DAMS Survey
  https://docs.google.com/spreadsheets/d/1OvZkGkizNnx_nZ9OVDOkJJIVFuIMK_7FYKC77YhoUac/edit#gid=0
- ProRes
  https://www.apple.com/final-cut-pro/docs/Apple_ProRes_White_Paper.pdf