

DATA CURATION NETWORK

Audiovisual Data Curation Primer

Authors: Madina Grace, Meg Jerrild, and Lauren Phegley

Mentor: J.A. Pryse

Affiliate Contributors (Peer reviewers): Talya Cooper and Courtney Dalton

Suggested Citation: Grace, M., Jerrild, M., & Phegley, L. (2023). Audiovisual data curation primer. Data Curation Network. [GitHub Repository](#).

Overview

Topic	Description
File Extension	.mkv, .jp2, .mp4
MIME Type	video
Structure	Audiovisual files (file wrapper and encoded bitstream)
Versions	
Primary fields or areas of use	Multidisciplinary
Metadata standards	DublinCore, PREMIS, METS, AudioMD and VideoMD
Key questions for curation review	<ul style="list-style-type: none">• Evaluate ethical issues• Determine if documentation and metadata meet reuse needs• Assess staff and institutional ability to undertake a project• Consider preservation actions• Inventory and review deposit
Tools for curation review	MediaInfo, VLCPlayer or general media player

Date Created	2023
Created by	<ul style="list-style-type: none"> • Madina Grace - madina.grace@duke.edu • Meg Jerrild - mjerrild@umich.edu • Lauren Phegley - lphegley@upenn.edu
Date updated and summary of changes made	

Suggested Citation: Grace, M., Jerrild, M., & Phegley, L. (2023). Audiovisual data curation primer. Data Curation Network.

Table of Contents

Audiovisual Data Curation Primer	1
Overview	1
Table of Contents	1
Description of format	2
Key Terms	3
Key clarifications to get from researcher	4
Curators Log	4
Workflow	5
Ethical issues	5
Documentation	6
Metadata core elements and standards	6
Administrative metadata	6
Technical metadata	7
Descriptive metadata	7
Preservation metadata	7
Rights and access metadata	7
Technical issues	8
Check Files	8
Preservation Actions	9
Preservation Actions for Audiovisual Media	10
What to look for to make sure this file meets FAIR principles	12
Appendix A: filetype CURATE checklist	12
Appendix B: Software and Resources	12
Appendix C: Additional Information and Materials	13
Special Considerations	13
User Purpose	13
Editing	14
Staffing Requirements	14
	2

Unresolved issues and questions that are beyond the scope of this primer	14
References	14
Resources	16

Description of format

This primer reviews the practices of curating audiovisual data. Data is defined as “facts, ideas, or discrete pieces of information, especially when in the form originally collected and unanalyzed” (Society of American Archivists). Audiovisual data is then discrete pieces of information captured in signals and sound waves that when given context allow the user to create meaning. We believe that what makes something research data is the way it is utilized, as not all data was created initially for research purposes. While audiovisual materials are not a common form of research data in all fields, they have been used in the social sciences, such as behavioral psychology and anthropology, and we have seen a new movement towards audiovisual data in the sciences as well. The increasing affordability of storage space and the continual development of quality recording equipment seem to be driving this new enthusiasm for audiovisual data.

The technically correct way to describe audiovisual data is video payload, as a video signal includes components such as audio, captions, and time code (IASA, 2019). We will be focusing on file based digital media that is media independent, and either born digital or previously digitized. File based media exist only in the digital form as bitstreams, which do not need to rely on a carrier (material holding the signal) for it to be usable (IASA, 2019). Media independence frees digital materials from the issues that plague analog materials that are confined to a certain carrier, such as an ethnographic interview confined to a U-matic ¾-inch videocassette (Wright, 2012). You can migrate file based content between storage systems for preservation purposes in a way that you can’t with media dependent video where the carrier and signal depend on one another (IASA, 2019).

Before getting into how to curate audiovisual data, we need to define a few key terms. Within a digital audiovisual file, the video payload includes both the file wrapper and the encoded bitstream. A wrapper/file wrapper stores and structures the data represented in the encoded bitstream while usually also storing technical and descriptive metadata (IASA, 2019). Example wrappers include Matroska (.mkv), Motion JPEG 2000 (.jp2), and MPEG-4 (.mp4). A codec, the abbreviation for coder/decoder, is the specification details of how to encode raw signal into a compressed format and decode a compressed format back into a raw signal (Robitza, 2018)(Wright, 2012)(IASA, 2019). Common codecs include H.264, FFV1, and lossless JPEG 2000. An encoding is, according to IASA-TC 06, “any set of rules that governs the conversion of any kind of information into another form for communication or mediated storage, e.g., Morse code for the alphabet (some would say the alphabet is itself an encoding) in a telegraphic system” (B-16). This is the actual software that outputs a codec-compliant bitstream, such as AAC (Robitza, 2018). There are multiple glossaries and dictionaries available in the references and resources sections of this primer for a deeper explanation of these topics and other terms.

Key Terms

Term	Definition	Example
Wrapper/File Wrapper	<p>A container that stores and structures encoded bitstreams (data) while also storing technical and descriptive metadata.</p> <p>Occasionally referred to as a container format.</p>	Matroska (.mkv), Motion JPEG 2000 (.jp2), MPEG-4 (.mp4)
Codec	Short for coder/decoder. It is the specification details of how to encode raw signal into a compressed format and decode a compressed format back into a raw signal.	H.264, FFV1, lossless JPEG 2000
Encoding	Any set of rules that governs the conversion of any kind of information into another form for communication or mediated storage, e.g., Morse code for the alphabet (some would say the alphabet is itself an encoding) in a telegraphic system	AAC (native FFmpeg AAC encoder), libx264 (most popular free and open source H.264 encoder)

Key clarifications to get from researcher

Due to the highly complex nature of audiovisual data, before you agree to ingest and curate this type of data, it would be beneficial to begin a conversation with the depositor. You need to outline the services you do and do not offer since, in most cases, you will not be doing digitization or converting to a standard format, known as normalization. For example, do you offer mediated access to data, offer a certain amount of storage space per deposit, or transmux (the act of converting from one format to another, such as .mp4 to .webm)? This type of information would be helpful to outline in a policy.

In terms of technical metadata, there are parts of the audiovisual creation process that don't automatically create embedded metadata but can be incredibly important to future users and for

preservation purposes. Get as much information as you can about the editing software used, the equipment that it was created with (including all audio and visual components if coming from different sources), any major adjustments made with the software, and why certain choices were made - for both aesthetic and technical reasons. Some of this information might be best associated with the audiovisual data as a README file. Depending on the differences between files, you might need a README file at the file level, record level, and project level.

Curators Log

The curators log is information that the data curator should be tracking throughout the process of working with a deposit. These notes are most likely something you notate internally for yourself or the curation team, so you all are able to access this information. Some information, such as metadata accompanying the files, should be kept in a separate location and linked in the log. This log should be kept long term with the data.

- Issues that you might come across or questions you want to follow up on
- Technical changes made to the materials
- Original metadata that accompanied the files
- Any additional metadata added to the original metadata, such as that resulting from communication with a researcher
- Quality evaluation
 - Is this in a container format you know is dead upon arrival, such as Windows Media files? Did they deposit standard definition recordings or 4K recordings, which need much more space?
- Ethical issues that you may have come across
- Notes from correspondence with the depositor
 - It can be helpful to attach email chains and notes taken during consultations to the log for detailed documentation.

Workflow

- Ethical Issues: does the depositor have permission to share these materials? Does your institution provide mediated access to materials that cannot be fully open access?
- Documentation: does the README provide enough information to support reuse?
- Metadata Evaluation: is there sufficient accurate metadata accompanying the file, including detailed technical metadata?
- Technical Evaluation: is there sufficient storage in your network drives? How do you transfer files from a depositor?
- Check Files: do the files open and contain the intended contents? Do the file names accurately describe the contents? Are there any issues within the file(s) that affect their future use?

Ethical issues

Depending on the nature of the video data, some datasets may classify as human subjects research. In this case, there are a plethora of ethical issues to consider. Unlike quantitative data, audiovisual data often display visual images of faces and recordings of voices, and can be more difficult to monitor for disclosure risk due to their size and complexity. For highly sensitive data of this nature, the best option is often to only release data under an appropriate data use agreement (DUA) to select researchers (Databrary, 2020). For more information on human subjects research, see the [Human Participants Data Essentials Primer](#).

There are **three agreements** that must be considered when working with identifiable human subjects research:

1. An informed consent agreement between the subject and the data collector,
2. A data transfer agreement between the data collector and the data archive, and
3. A data use agreement between the data archive and the secondary user.

Ask yourself:

1. Is there documentation of IRB approval?
2. What does the IRB protocol allow with respect to making the data available to others? Ensure that the deposit, along with any redaction or de-identification already performed on the data, meets the standards described in the IRB approval.
3. How will the data be utilized? Will different data usage agreements be necessary for different types of user research?
4. What kinds of data has the researcher been given permission to share and at what level (e.g., transcript-only, audio-only files, audio-video files; at a public or restricted level)?
5. Have subjects provided informed consent that is in line with what the researcher wishes to release?

Video data that do not meet your institution's minimum requirements for release or re-use should not be made publicly available (e.g., videos missing evidence of appropriate informed consent should NOT be made available for re-use).

Some repositories allow for flexible levels of data use. Consider Databrary, a data repository which allows subjects to agree to public viewing of short clips in classes and professional meetings (U.S. Department of Health and Human Services, 2016).

Documentation

Review any accompanying documentation files. Do these documents contain information that is necessary to release in order to understand the video data? Is there sufficient documentation provided in order for the data to be meaningful and useful? If this is not the case, either recommend or create further documentation (e.g., README template).

README files should include any applicable software, format, file structure, size of files, dates of major edits to the files, and any other relevant software characteristics.

Metadata core elements and standards

A curator should consider any ethical or technical challenges that could be a part of metadata, e.g. user agreements, preservation of large files, etc. It's preferable to store some of the metadata components mentioned below in a secure location accessible to curators only. Core categories of metadata for discovery, preservation and curation purposes include (Mazé, 2012):

Administrative metadata

- Data Use Agreements (DUA) for any human subject audiovisual data provided by a depositor.
- Curation log documenting any edits and additions
- User agreements for any legal access requirements
- Manifests
- Any preservation documents, e.g. checksum logs

Technical metadata

Technical metadata describes technical elements and processing information of items. Most of this metadata is produced by video and audio recording and editing software and is part of the digital files' description. For audiovisual data, one should consider hardware and software documentation and technical digitization information (if applicable), including size, formats, color, compression ratios, etc.

Descriptive metadata

This includes the description of an item, its elements and their versions, including but not limited to title, keywords, tagging and relationship between elements.

Preservation metadata

It documents all preservation processes performed on audio visual files. Digital files should go through a fixity check by generating checksums, which document the integrity of the ingested files. The metadata should also record any actions and/or changes made during migration, digitization, and preservation, e.g. reformatting or resizing files.

Rights and access metadata

Information about ownership and copyright. Consider discussing all available options for accessing the files including licensing information, e.g. [Creative Commons](#), with a depositor prior to the submission process.

Suggested metadata standards:

[DublinCore](#)

DublinCore is one of the most used schemas for description.

[PREMIS](#)

Preservation Metadata Maintenance Activity schema provides digital preservation information including logs of fixity checks and ingest data.

[METS](#)

Metadata Encoding and Transmission Standard schema describe descriptive, administrative, and structural metadata for audiovisual digital objects.

[AudioMD and VideoMD](#)

XML Schemas were developed by the Library of Congress to provide detailed technical metadata for audiovisual resources. Most often they are used along with METS or PREMIS as a part of their administrative metadata.

Technical issues

Managing, curating and storing audiovisual data can be challenging without considering possible technical issues and infrastructure requirements (Dunn, 2021):

- *Is there sufficient storage in your network?*

Size and storage of files is one of the major challenges of audiovisual data. Video files are very large, and require a considerable storage capacity, including higher level computer processing specifications.

- *Is there a way to transfer files from a depositor?*

A curator should have a system or a way to receive files from the depositors. Files transfer could be challenging with extra large formats and sizes. Online file sharing applications like Box have a limit and require a paid subscription.

- *Are there any edits to the original files?*

In cases when a depositor sends large sized files a curator should note if the files were compressed or changed by a software. There could be a possible loss of data when files are received in a compressed/processed format, including but not limited to editing, resizing, color edits, etc (Richardson, 2018).

- *What level of curation is needed?*

A curator needs to evaluate the level of curation necessary including any labeling, cleaning, organizing and converting files to applicable formats.

- *Are files available to end-users through the repository's user interface?*

Ideally, the user interface in the repository would have an embedded video and/or audio player; however this requires constant development and software support.

Check Files

After you have completed the preceding sections of the workflow, it comes time to check the files. You should only get to this step after you make sure that any of the preceding ethical, documentation, metadata, or technical requirements are met and any issues have been dealt with. This reduces unnecessary burden on the curator.

- Before you evaluate the files, make sure you have a bit level duplicate copy that you do not touch. This way if something happens to the files during curation, you have a backup.
- Open and evaluate the file directories. Check to see if there are any files missing that are listed in the documentation. Do the file names accurately reflect the content of the files?
- Evaluate files. Check to see if there is any distortion in the audio, video, captions, or transcripts that the depositor did not describe in the documentation. This may involve watching or listening to files at a fast speed to ensure that there are no critical issues. Critical issues are problems that are not documented by the depositor and/or significantly reduce the usefulness of the files for future users. Document any inconsistencies or errors in your curators log with the exact time or location in the recording.
- There is an overlap between checking files and preservation actions. Steps such as virus scans, fixity checks, and evaluating the quality of the files are described in detail in the Preservation Actions for Audiovisual Media list.

Preservation Actions

While this section talks about the basic preservation actions that can be done with audiovisual content, we would be remiss if we didn't address one of the main questions - who is doing the preservation of audiovisual data? Much of the audiovisual data ingested into an institution are materials that have been digitized, and thus are automatically digitized into a quality file format in such a way that it won't need immediate preservation actions.

Audiovisual materials that are deposited within a repository complicate the curation process as the responsibility of the data curator has traditionally been to curate the materials and give the

feedback to the depositor so they can update the dataset themselves. The depositors' ability to create the audiovisual content does not equate to their ability to transmux (convert formats/containers), transrate (change bit rate), or create uncompressed copies. This is also beyond the scope of most data curators.

What we have found upon interviewing multiple audiovisual researchers about where to draw the line at curating audiovisual media is that it depends on a host of factors. It comes down to capacity, technical knowledge, infrastructure, and commitment. Audiovisual preservation takes time, lots of storage space, experienced technical knowledge ranging from subjects like electrical engineering to computer science, and people. If you have the ability to add an experienced audiovisual media specialist to your workflow or have the knowledge yourself, then you might be able to take on actions like normalization in your workflow. For others, the complicated audiovisual preservation actions are generally out of scope for data curation.

We suggest “good-enough” audiovisual data preservation practices. Normalization will be discussed below but this is not a required action for the curator to undertake unless stipulated by an institution's policies. If an institution is committing to offer long term access, storage, and preservation of audiovisual materials, there may come a time when materials need to be migrated into another format to ensure continued access. An institution's preservation policies may focus on bit rate preservation, which preserves the submitted file's bits, rather than promising to migrate the content to a new format. Front-loading preservation actions, such as normalizing all audiovisual containers into Matroska containers with FFV1 codecs, does save some of the hassle down the line by ensuring continued access.

Preservation Actions for Audiovisual Media

- Scan for viruses
 - a. Before you bring anything into your system, you should run a virus check. This is standard in most workflows for established repository systems.
- Evaluate files
 - a. Before you can understand what the preservation actions are that you need to take, you need to evaluate the state of the deposited data. For the majority of situations where you need to review technical metadata, you can use ExifTool or MedialInfo (described in the Software and resources section). For instances when you need more in-depth quality control information or to evaluate what might be causing an issue with a file, QC Tools can be used.
- Fixity check
 - a. Defined as “a mechanism to verify that a digital object has not been altered in an undocumented manner” (NDSA). A checksum is an example of a tool used to run fixity checks. Fixity checks are an established practice in the digital preservation and archival communities, and are crucial to the success of digital audiovisual materials. The vast amount of steps between recording equipment and deposit increase the risk of accidental corruption. A best practice is to create fixity information at the earliest time possible after the creation of the file (Kim et al., 2021).

- Normalize
 - a. One of the key decisions that curation and digital audiovisual experts need to make is whether they will be accepting deposits in their current format or normalizing (Wright, 2012). There are solid arguments for both activities: archival principles dictate that materials should be preserved in their native format while normalization allows for files to exist in established file formats that should be accessible for a longer term. Another variable to consider is if files should be normalized all into one common format or one of a set of trustworthy formats depending on the use case. Video preservationists have not reached consensus as to a single file format for preservation of born-digital video or digitized content, although it is agreed that it should be in an uncompressed open standard format (IASA, 2019).
 - b. The choice to normalize files is up to you and your institution, but there are some key questions to ask yourself.
 - i. Do you have the capacity to undertake normalization of audiovisual files, at this point or even in the future as service burden grows?
 - ii. Does your institution have a standard format that you would normalize files into?
 - iii. If you choose more than one format to normalize into, are you going to be able to map the metadata difference between the files?
 - iv. If you won't be normalizing during ingestion, will you be periodically evaluating files for loss and at what point will you commit to normalizing in the future?
- Levels of file preservation and access:
 - a. A common archival preservation technique is to save an uncompressed high quality preservation copy (sometimes called a master copy) of a file in a preservation data management system. The preservation copy can also occasionally be the original file copy submitted for deposit, depending on normalization policies, or saved separately as well. A lower quality, accessible version would then be made available for patrons to use (such as .MP4). However, many of the repository systems that are ingesting audiovisual data as research objects were not built to store tiered preservation copies. You need to not only support the mission of preserving these audiovisual materials but also allow easy access by patrons.
 - b. The ability for an institution or organization to undertake this workflow depends on the amount of audiovisual data they have coming down the pipeline and if they have the capacity, storage space, and budget to support this method. The best practice is to have higher quality preservation copies as well as access copies, but if that is not possible then do your best to balance preservation and access.
 - c. We recommend making a compressed video file available in a common, widely used format. If you intend to support large amounts of audiovisual data, you may want to explore installing an in-page player for your repository, such as Kaltura,

as patrons may expect to be able to preview the files in the repository before they download them. Installing or developing a video player can be costly.

- Compression
 - a. It is a good practice to keep audiovisual files in an uncompressed format as compression exacerbates the effects of bit rot (Wright, 2012). Uncompressed files have redundant information and the rot of one bit in the audio or image area of the file won't destroy the file. Bit rot in compression changes part of the mathematical compression calculation, which magnifies the loss in the file (Wright, 2012). Despite this, it is often an unavoidable reality that compression must be done in order to store the file. While network and storage costs are continuing to decline, they are far from free. A good practice is to keep your preservation level file uncompressed, and make access files available using lossless compression.
 - b. One of the biggest issues that can reduce quality of audiovisual materials is the successive process of compression, where lossy compressed materials are decoded and recoded in a cycle that leads to quality loss (Wright, 2012). For this reason, avoid multiple uses of compression and try to keep materials in lossless compression rather than lossy compression.

What to look for to make sure this file meets FAIR principles

All curation work should aim to meet the four FAIR principles, in that the data ought to be findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). For video data, this means creating rich, robust metadata and the inclusion of any supplementary materials that would aid in a user's complete understanding of a video dataset.

Appendix A: filetype CURATE checklist

The following CURATED checklist model is based on the Data Curation Network's standardized set of [C-U-R-A-T-E-D steps](#).

Check files/code and read documentation

- Review if your repository is able to handle the type/volume/required preservation level of the data
- Check to see if files open

Understand the data (or try to)

- Check to see if audiovisual materials have obvious technical issues
- Review the video payload information to check formats, codecs, etc.
- Determine if there are ethical issues that need to be addressed

Request missing information or changes

- Are the changes that need completion within the abilities of the depositor?

Augment metadata for findability

- Appropriately describe content of the video, such as with keywords
- Link to existing publications, project information, and other datasets

Transform file formats for reuse*

*Only do this if this is an established practice in your institution

- Transform into a well-established open access file format. If you compress the file, use lossless compression.
- Retain original format file

Evaluate for FAIRness

- Review your actions thus far to see how you can improve the dataset's FAIRness

Document all curation activities throughout the process

- Record actions from metadata changes to depositor communications in your Curators Log

Appendix B: Software and Resources

All listed software are open source and free to use. There are some tools that require more technical knowledge than others to use, and this will be noted in the description. The tools that you use may depend on your institution's workflow and what they support, especially when it comes to ingesting audiovisual materials.

- Tools that for extracting or reviewing metadata
 - a. [MedialInfo](#) - a tool that you can run on an individual file or a directory that evaluates technical and tag information about video and audio files. It shows you all of the most important information about a file, such as format, codec id, frame rate, etc., all in one place. The app has a GUI interface that makes it easy to use for those who don't want to use the command line. You can export the metadata into open file formats like .csv and .txt for tracking purposes.
 - b. [ExifTool](#) - a detailed tool that allows you to read, write, and edit metadata. While it supports a wide variety of file formats and metadata fields, this command line utility is overkill for many curators' needs. Stick with MedialInfo first, and if that doesn't suit your needs, give this a try.
- Tools for viewing videos
 - a. [VLC Media Player](#) - a free, open source, and easy to use media player software. It supports a variety of audio and video compression methods and file formats.
- Tools for quality assurance
 - a. [QC Tools](#) - a tool for evaluating the technical information in a video file. Provides in depth information such as YUV values, temporal outliers (TOUT), and vertical line repetitions (VREP). The preview window is an analytical preview environment that allows you to review a video through multiple filters simultaneously. Unless you have a need to find out why there is an issue with a digital audiovisual file, this tool is too technical for the general curator.
- Tools for encoding/converting/metadata evaluation
 - a. [FFmpeg](#) - a multimedia framework that is able to encode, decode, transcode, stream, filter, etc., just about any file format. This command-line tool is one of the golden standards when working with audiovisual materials. FFMpeg is most likely

out of the scope of the normal curation process, but it is important to know about if you need to make any changes to digital audiovisual materials.

- b. [Handbrake](#) - a GUI video transcoder that can convert between audiovisual containers and codecs.

Appendix C: Additional Information and Materials

Special Considerations

User Purpose

When curating any type of data, it is important to keep the end-user's purpose in mind. For video data, there are three significant types of use to consider: annotation, machine learning, and classification.

- *Annotation*: the process of tagging or labeling features of a video for later analysis. Annotations correspond with time points or segments, and may also indicate a region of the viewing area or an object or person. Annotation is done manually by coders (e.g., by research assistants). There are annotation software packages available to assist with this process.
- *Machine Learning*: Algorithms are trained on a set of videos (usually annotated) and develop a mathematical model of characteristics. ML can be very computer intensive and may require substantial computational resources.
- *Classification*: applying an established model to new videos so that they can be tagged and labeled.

Editing

Video data is most often curated without necessitating any video editing. However, in special cases, manual censorship of faces or trimming of videos may be necessary to comply with data use agreements and IRB requirements. A case such as this would be considered especially intensive and outside the scope of the curation process.

Staffing Requirements

During the planning period before curation even begins, it is important to consider whether there is an appropriate amount of staff and whether they have the skills required for video curation work. If there is an expectation for normalization tasks to be performed, then it is essential to make sure a staff member with normalization experience will be available. Additionally, the curation of video data can be extremely time consuming, and therefore there must be an appropriate allotment of staff time available to dedicate to the project.

Unresolved issues and questions that are beyond the scope of this primer

- Digitization/Analog Materials - we are focusing on materials that have already been digitized into an appropriate format or that were born digital.
- Technical Infrastructure - developing a technical infrastructure should be done by developers. This also includes figuring out if your system allows for different copies of the videos (access copy, etc.).
- Obtaining IRB approval and organizing Data Use Agreements (DUA) - just double check that these exist but understand that you are not the one doing this work
- How to de-identify human subject videos
- How to install a video player in your repository infrastructure
- Accessibility - please see the [Accessibility Data Primer](#) for more information
- Legal Issues - special considerations for human subjects
- Beyond the scope of the curator -> improving the quality, making it preservation quality, etc. It should either already be good enough, or it needs to be handled by someone else.

References

Databrary. "Access Agreement -- Annex III -- Databrary Access Guide." *Databrary.org*, 29

Apr. 2020, databrary.org/about/agreement/agreement-annex-III.html.

Darragh, J., Hofelich Mohr, A., Hunt, S., Woodbrook, R., Fearon, D., Moore, J., & Hadley, H. (2020, October 6). *Human Participants Data Essentials Data Curation Primer*.

GitHub. Retrieved March 24, 2023, from

<https://github.com/DataCurationNetwork/data-primers/blob/master/Human%20Participants%20Data%20Essentials%20Data%20Curation%20Primer/human-participants-data-essentials-data-curation-primer.md>

National Digital Stewardship Alliance (n.d.). *Glossary*. Retrieved February 17, 2023, from

<http://ndsa.org/glossary/>

Robitza, W. (2018). *FFmpeg Encoding and Editing Course*.

<http://slhck.info/ffmpeg-encoding-course/#/>

Society of American Archivists. (n.d.). *SAA Dictionary: Data*. Retrieved March 7, 2023, from

<https://dictionary.archivists.org/entry/data.html>

U.S. Department of Health and Human Services. (2016, March 8). *Databrary: Enabling more research through sharing Rich Video Data*. National Institutes of Health.

Retrieved March 24, 2023, from

<https://obssr.od.nih.gov/news-and-events/news/databrary-enabling-more-research-through-sharing-rich-video-data>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016, March 15). *The Fair Guiding Principles for Scientific Data Management and Stewardship*. Nature News. Retrieved March 24, 2023, from

<https://www.nature.com/articles/sdata201618#Abs1>

Wright, R. (2012). *Preserving Moving Pictures and Sound: DPC Technology Watch Report 12-01*. Digital Preservation Coalition. <https://doi.org/10.7207/twr12-01>

Kim, J., Colloton, E., Finn, D., Fraimow, R., Lin, S.W., Sanchez, C., & Schweikert, A. (2021). Audiovisual Quality Control and Preservation Case Studies from Libraries, Archives, and Museums. *International Association of Sound and Audiovisual Archives (IASA) Journal*, 51, 23–40. <https://doi.org/10.35320/ij.v0i51.111>

Richardson, I. (2018). Compressed Video Quality. *International Association of Sound and Audiovisual Archives (IASA) Journal*, 48. <https://doi.org/10.35320/ij.v0i48.41>

Dunn, J.W., Feng, Y., Hardesty, J.L., Wheeler, B., Whitaker, M., Whittaker, T., Averkamp, S., Lyons, B., Rudersdorf, A., Clemenet, T., & Fischer, L. (2021). Audiovisual Metadata Platform Pilot Development (AMPPD) Final Project Report.

<https://hdl.handle.net/2022/26989>

International Association of Sound and Audiovisual Archives Technical Committee. (2019).

Guidelines for the Preservation of Video Recordings, Part B. Video Signal,

Preservation Concepts, and Target Formats. (2nd ed.)

<https://www.iasa-web.org/tc06/guidelines-preservation-video-recordings>

Gilmore, R. O., Adolph, K. E., & Millman, D. S. (2016, August). Curating identifiable data for sharing: The databrary project. In *2016 New York Scientific Data Summit (NYSDS)* (pp. 1-6). IEEE.

Mazé, E. A. (2012). Metadata: best practices for oral history access and preservation. *Oral History in the Digital Age*.

Resources

The Activists' Guide to Archiving Video. (n.d.). *WITNESS Archiving*. Retrieved December 20, 2022, from <https://archiving.witness.org/archive-guide/>

Audio Glossary. (n.d.). Oral History in the Digital Age. Retrieved December 19, 2022, from <https://ohda.matrix.msu.edu/gettingstarted/glossary/audioglossary/>

AV Artifact Atlas (n.d.). Retrieved March 17, 2023, from

<http://www.avartifactatlas.com/index.html>

Awais, A. (2021). My Command Line Audio Video Workflows with FFmpeg. *Ahmad Awais*.

<https://ahmadawais.com/ffmpeg-command-line-audio-video-workflows/>

Blackden, C. (2021). *Everything You Ever Wanted to Know about ExifTool*.

<https://adamtheautomator.com/exiftool/>

Corrado, E. M., & Moulaison Sandy, H. (2016). Digital Preservation of Audiovisual-Based Materials: The State of the Art. *Archiving Conference*, 13(1), 161–166.

<https://doi.org/10.2352/issn.2168-3204.2016.1.0.161>

Elnabli, S. [alctsce]. (2019, March 13). *Digital Preservation: Audiovisual Edition, Part II*.

<https://www.youtube.com/watch?v=23h7vIcQf-k>

- Elhabli, S. [alctsc]. (2019, February 27). *Digital Preservation Audiovisual Edition, Part I*.
https://youtu.be/L6cM6H_pGnE
- Federal Agencies Digital Guidelines Initiative. (n.d.). Retrieved March 17, 2023, from
<https://www.digitizationguidelines.gov/>
- FFmpeg buddy. (n.d.). Retrieved March 17, 2023, from
<https://evanhahn.github.io/ffmpeg-buddy/>
- FFmprovisr. (n.d.). Retrieved December 19, 2022, from
https://amiaopensource.github.io/ffmprovisr/index.html#extract_audio.
- Hee, J. (2022). *A Quick Guide To FFmpeg*. CodeX.
<https://medium.com/codex/a-quick-guide-to-ffmpeg-bed33bc0bc03>
- Kromer, R. (n.d.). *FFmpeg Cookbook for Archivists*. Retrieved March 17, 2023, from
<https://avpres.net/FFmpeg/>
- Matroska File Format with FFV1 video encoding*. (2022). Sustainability of Digital Formats:
Planning for Library of Congress Collections.
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000343.shtml>
- Moreira, L. (2023, January 5). *FFmpeg libav tutorial—Learn how media works from basic to transmuxing, transcoding and more*. GitHub.
<https://github.com/leandromoreira/ffmpeg-libav-tutorial>
- Padilla, T., & Baker, J. (n.d.). *The sourcecaster*. Retrieved March 17, 2023, from
<https://datapraxis.github.io/sourcecaster/>
- Pryse, J. A., Harp, M., Mannheimer, S., Marsolek, W., & Cowles, W. (2022, December 2).
Oral History Interview Data Curation Primer. GitHub.
<https://github.com/DataCurationNetwork/data-primers/blob/c6ec438e76fea49eaaf2806bc79ec2c8c12de7f3/Oral%20History%20Interviews%20Data%20Curation%20Primer/oral-history-interviews-data-curation-primer.md> (Original work published 2018)
- Rao Vijayanagar, K. (2021). *Recipes in FFmpeg*. Ottverse.

<https://ottverse.com/recipes-in-ffmpeg/>

Video Glossary. (n.d.). Oral History in the Digital Age. Retrieved December 19, 2022, from

<https://ohda.matrix.msu.edu/gettingstarted/glossary/videoglossary/>

Walsh, T. (n.d.). *Brunnhilde*. Retrieved December 19, 2022, from

<https://www.bitarchivist.net/projects/brunnhilde>

Xuan Nguyen, N. (2022, July 19). *Back to the Basics: Use and Install FFmpeg*. Adam the

Automator. <https://adamtheautomator.com/install-ffmpeg/>