

Oral History in the Digital Age

What are organizations saying and doing about digital video?

Steve Cherry
University of Michigan, Student
MATRIX at Michigan State University, Intern
sccherry@umich.edu
<http://www.stevecherry.net>

Introduction

As digital video recording becomes less expensive it likely will be increasingly used in oral history projects. "Oral History in the Digital Age" is an IMLS funded project led by MATRIX at Michigan State University that seeks to develop a set of best practices that can unify the work done by groups involved with oral history such as oral historians, libraries, archives, and museums. However, developing standard practices for recording and using digital video is complicated because these different groups use oral history for different purposes (e.g. information gathering, research, exhibition, broadcasting).

As a first step in developing best practices it is important to get a sense of current practices. Thus, I have collected and reviewed best practice documents that have been published by organizations working with oral history and/or digital video. While many subjects will be explored during the course of the project, this poster focuses on preferences for digital video formats and codecs.

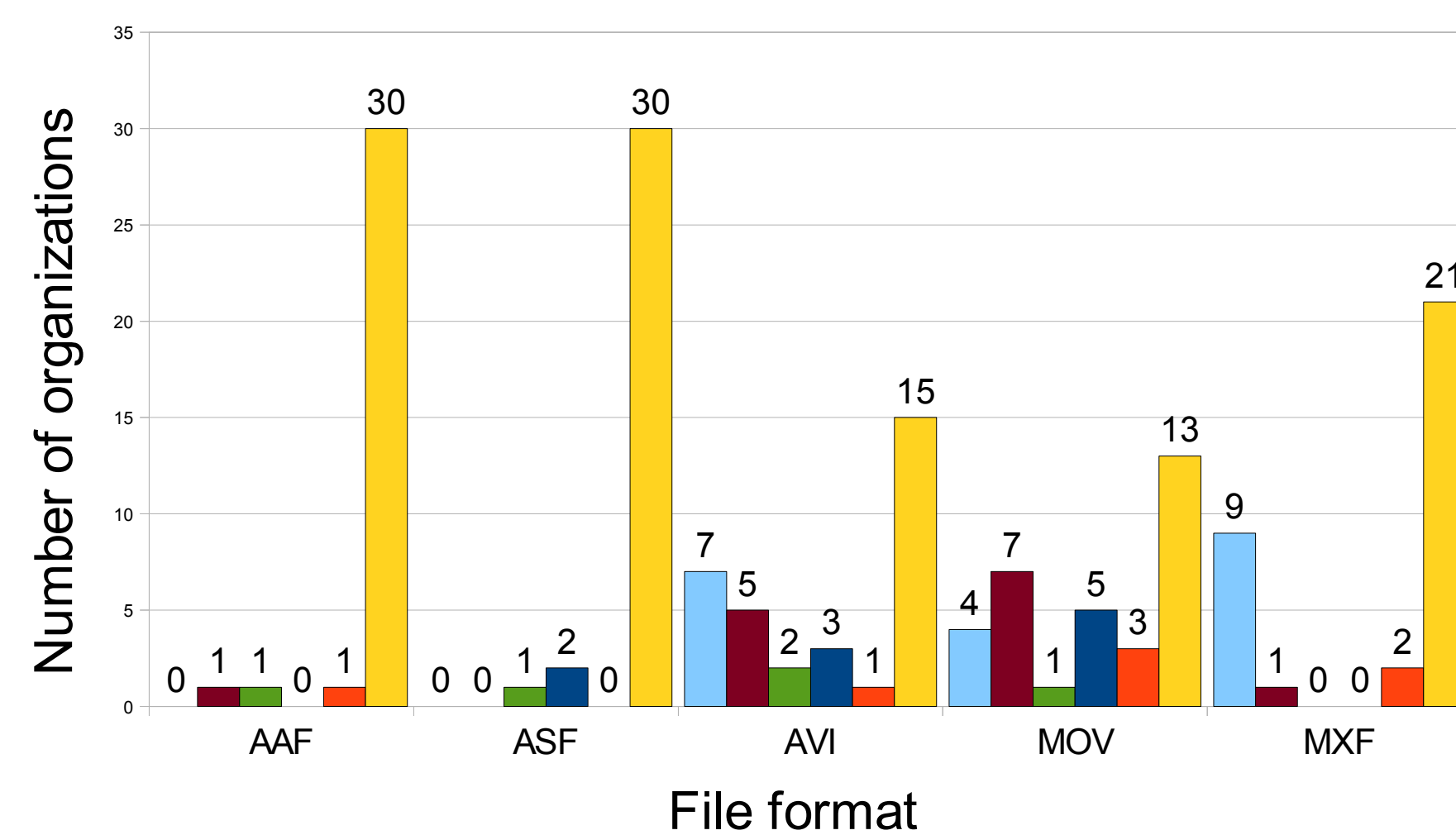
Methods

- **Find documents:** either institutional guidelines or articles describing current practices that were published online. Most of the available documents came from library and archive digitization projects or national standards bodies and were not specific to oral history.
- **Group documents:** according to the organization that was associated with the document. In a few cases where multiple documents were associated with a single organization, documents were treated together as if they were one document.
- **Create a matrix:** listing each organization and each file format and codec mentioned in each document.
- **Note preferences:** for archiving and access file formats and codecs that were mentioned in a given document.
- **Group preferences:** into categories that facilitate the comparison of preferences across documents. This step was complicated due to a lack of consistent terminology for stating preferences and sometimes a lack of clarity about which specific file format or codec was being discussed.
- **Note reasons for preferences:** for file formats and codecs in order to understand better why certain formats and codecs were preferred. These reasons were compiled separately.

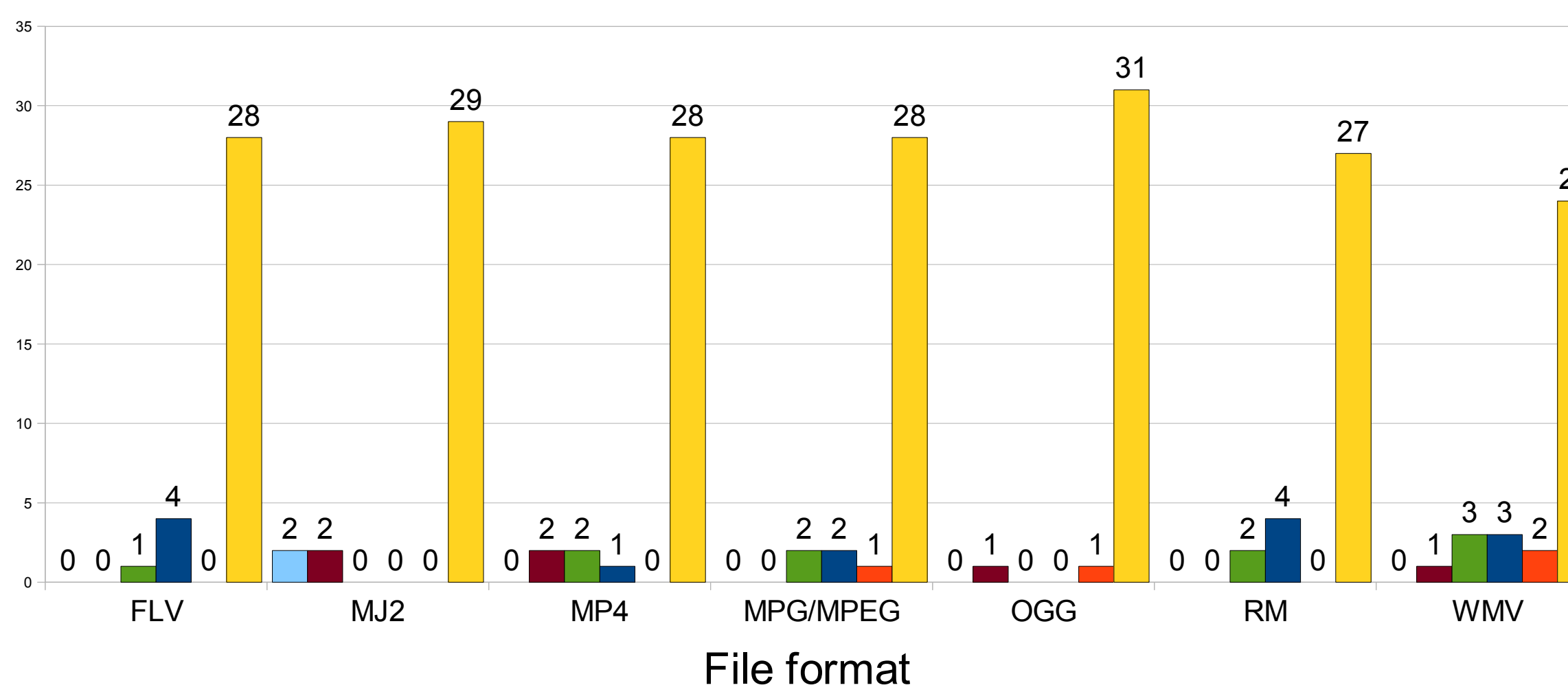
Explanation of terminology used

- **File format:** also called a container or wrapper, it is used to hold together video and audio codecs and related metadata. In the chart these are labeled according to their file name extensions.
- **Codec independent:** these file formats can hold a wide variety of codecs.
- **Codec dependent:** these file formats generally only hold one or a few types of codecs.
- **Codec:** the algorithm used to compress a digital video signal.
- **Lossless:** codecs that will return exactly the same picture each time the video is decompressed. Also called mathematically lossless.
- **Lossy:** codecs that change the picture in order to achieve higher levels of compression. Also called perceptually lossless.
- **"Uncompressed":** this codec technically is not a codec, but raw uncompressed video. Nevertheless, it made the most sense to group it with the lossless codecs for comparison.
- **MPEG (unspecified):** some documents recommended using MPEG, but did not specify MPEG-1, -2, or -4. I did not want to assume which specific codec these documents recommended.
- **MPEG-4:** like MPEG (unspecified), documents did not clearly distinguish between MPEG-4 (part 2) and MPEG-4 AVC (part 10), which are two different codecs. When this occurred I put the data under MPEG-4 (part 2).

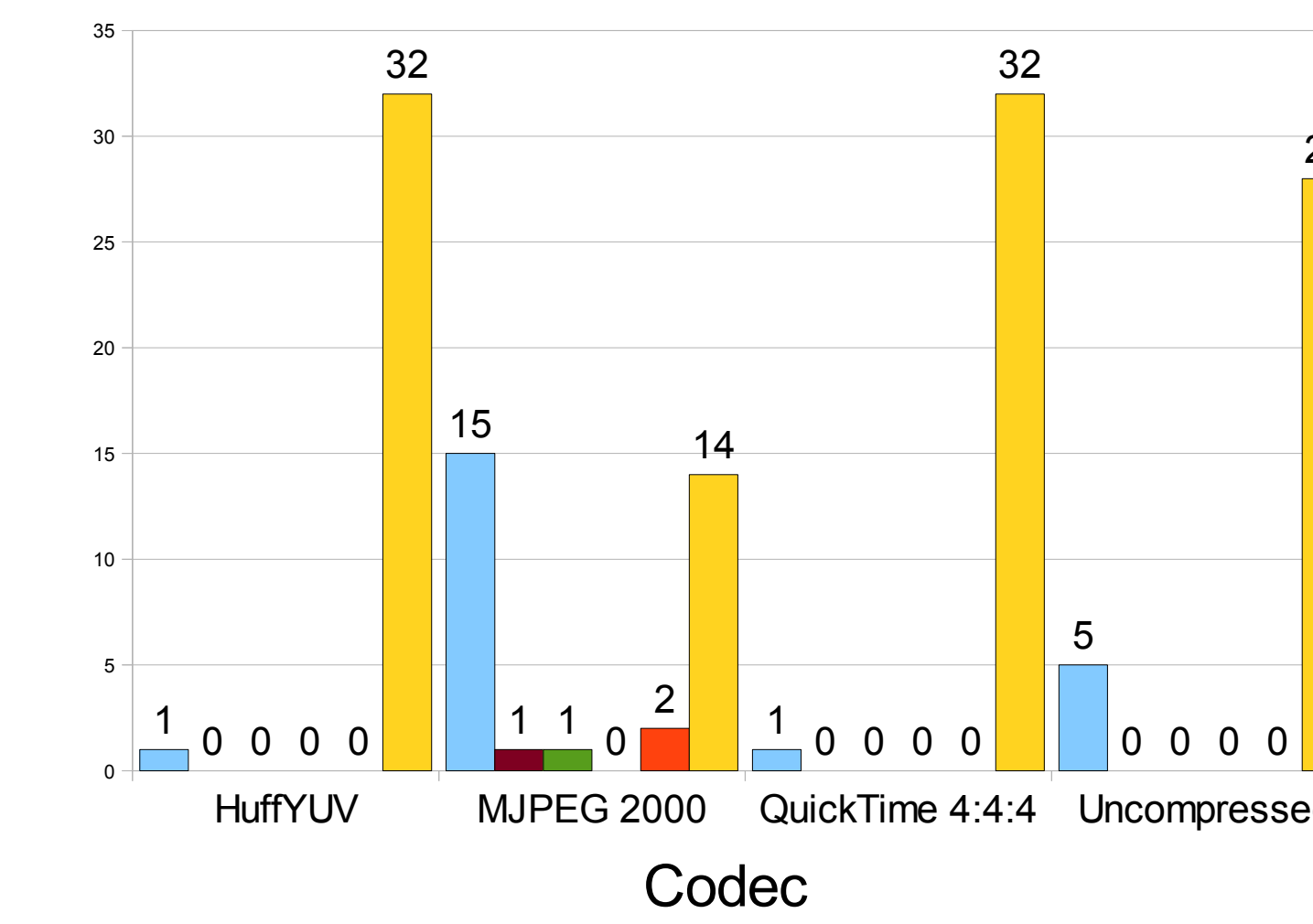
Codec independent formats



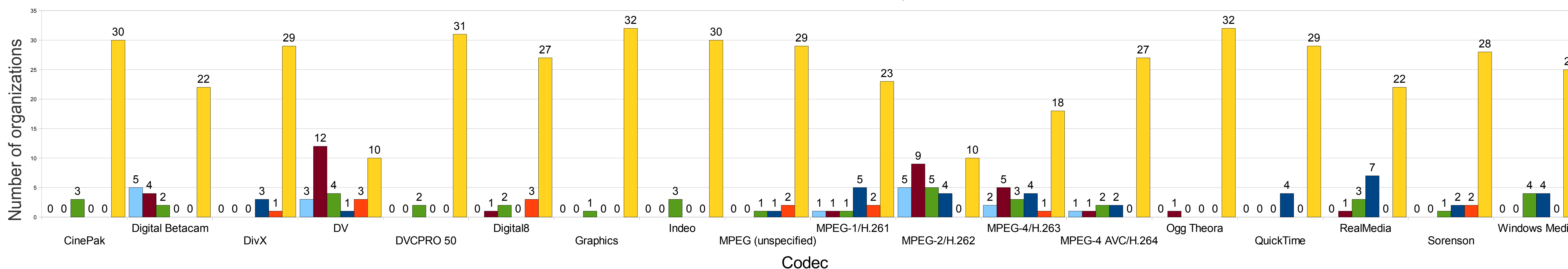
Codec dependent formats



Lossless codecs



Lossy codecs



- Preferred or used for archiving
- Acceptable alternative for archiving
- Not preferred for archiving
- Suitable for access
- No preference stated
- Not mentioned

Common criteria for choosing a digital video format

- **Open standards:** rather than proprietary ones so that a video does not depend on the whims or fortunes of a single company for continued support.
- **Wide adoption:** indicates a format or codec will have a long life before becoming obsolete.
- **File-based formats:** rather than tape-based.
- **Compatibility:** with a wide variety of hardware and software across different platforms.
- For archiving:
 - **Lossless codecs:** will not change with successive compressions.
 - **Spatial compression:** means that each frame is compressed separately without reference to other frames, which can help with editing in the future. This might also be referred to as Intra-frame only compression.
 - **Higher bitrate:** generally means that less compression is used, which can be helpful if a lossy codec needs to be used.
- For access:
 - **Lossy codecs:** provide reduced quality in exchange for higher levels of compression.
 - **Temporal compression:** is an important way of achieving high levels of compression. In addition to Intra-frames, this would use Predictive and Bidirectional frames.
 - **Lower bitrate:** means reduced bandwidth for transmitting videos.

Findings

- Documents recommend preserving born-digital video as is. There will be no gain in picture quality by migrating to a lossless codec, but there will be increased loss by migrating to a lossy codec. Thus it would be advantageous initially to record oral history video using file formats and codecs that are preferred or acceptable for archiving.
- AVI, MOV, and MXF are the most preferred file formats for archiving. Codec dependent formats are generally not recommended, probably because they are not as flexible and tend to be associated with proprietary, lossy codecs.
- MJPEG 2000 is the most preferred lossless codec for archiving, but it is seen as an emerging standard that is not widely adopted yet. Cost and storage space are additional factors working against adopting MJPEG 2000 or other lossless codecs.
- MPEG-2 and DV are the most acceptable lossy codecs for archiving and can be configured to only use spatial compression. They are also common, open formats so it might be a long time before they need to be migrated. During that time lossless video archiving may become more viable and standardized.
- While MPEG-4 is less accepted than MPEG-2 or DV some documents stated that it will likely supersede MPEG-2, especially for High Definition video.
- Digital Betacam is also well accepted for archiving, but has the disadvantage of being tape-based.
- There is less agreement regarding file formats and codecs for access, which is probably due to the less stringent criteria and for the multiple possible ways of providing access (e.g. making DVD copies, streaming, or online downloading).

Organizations

Documents from the following organizations were consulted for this poster:

Arts and Humanities Data Service, Bay Area Video Coalition, Baylor Institute of Oral History, British Library Endangered Archives Programme, Library and Archives Canada, Concordia Centre for Oral History and Digital Storytelling, Dance Heritage Coalition, E-MELD School of Best Practice, Electronic Arts Intermix, Evia Digital Archive, Florida Digital Archive, Florida Electronic Library, Folkstreams.net, National Library of France, Joint Information Systems Committee, Maine Community Heritage Project, MATRIX, University of Michigan, Nantucket Historical Association, National Archives and Records Administration, National Information Standards Organization, Library of Congress, New York University, Oral History Society, PrestoSpace, State Library of Queensland, Rutgers University Community Repository, Video Preservation Website, Training for Audiovisual Preservation in Europe, UKOLN, Veteran's History Project, Yale University Library

Acknowledgements

Thanks to Dean Rehberger for providing feedback on an earlier version of this poster.