DIGITAL PRESERVATION: AUDIOVISUAL EDITION PART 2

ALCTS Webinar Series
3/13/2019

Facilitator
Stefan Elnabli (UC San Diego)
ABOUT ME

Stefan Elnabli
UC San Diego

Media Curation Librarian \ Manager, Digital Reformatting Operations

Hosted by ALCTS, Association for Library Collections and Technical Services
<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mins</td>
<td>Introduction and Context of Part 2</td>
</tr>
<tr>
<td>10 mins</td>
<td>Digital Archiving and Preservation</td>
</tr>
<tr>
<td>10 mins</td>
<td>Considerations for AV</td>
</tr>
<tr>
<td>10 mins</td>
<td>Case Studies</td>
</tr>
<tr>
<td>5 mins</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

Hosted by ALCTS, Association for Library Collections and Technical Services
CONTEXT OF PART 2
Isn’t something that is digital already preserved?

Is digital archiving different from management of analog collections?

What is a digital archive?

What is the difference between Born-Digital and Digitized AV?

How can I start a digital archiving program for my AV?
What is Born-Digital?

“Relating to or noting documents, images, etc., that are created and managed in electronic form.”

Cameras recording to file formats on solid state drives or removable media like the Panasonic P2 card

Computers, cell phones, tablets, floppy disks

Optical media like CDs, DVDs, Blu-rays

"as research libraries are discovering, born-digital materials are more complicated and costly to preserve than anticipated"
Examples at UCSD

- Digital collections acquired for preservation and/or access that have archival/historical value
  - Files, opical discs, data tape, etc.
- Content generated by campus partners such as UCTV
- Licensed audio and video for general circulating collections
- Audio and video acquired digitally for general circulating collections
Different from Digitization?

“Digitization is an act of creation where an analog original is converted to the digital domain, while born-digital files are the original.” – T. Owens

Digitization creates surrogate of original’s significant characteristics, but the act itself is not preservation

Born-digital media is the original that archives preserve, not surrogates, therefore have significant characteristics inherent to their respective formats

“[…] the distinction between born digital and digitized objects can obfuscate as much as it illuminates.” – Trevor Owens, Head of Digital Content Management (Library of Congress)

Hosted by ALCTS, Association for Library Collections and Technical Services
Importance to Libraries and Archives

- Content is created in the digital domain and it must be preserved for access and re-use.
- Digital asset management systems are not necessarily tailored toward preservation.
- Knowing about properties of digital audio and video as well as the archiving process helps improve workflows.
- This knowledge will help communication between the archive and content producers/acquisitions.
• Do you digitize AV materials?
• Do you collect born-digital AV materials?
• Do you have a digital preservation policy for your AV?
• Do you have a repository with digital preservation services?
DIGITAL ARCHIVING AND PRESERVATION
• In context
  – Collection of digital files or objects
  – Objects can be born-digital or can be digital surrogates
  – Actively maintained for preservation and access purposes
  – NOT a hard drive that you put on a shelf and forget about
  – NOT a backup of your computer
  – NOT a networked drive (on its own)
• Having a specific and public collection development policy that addresses born digital material will help direct what you collect and how you make it available
• Digital objects and metadata should be stored together
• Open file formats/file normalization
  • https://www.archives.gov/preservation/products/definitions/filetypes.html
• Filenaming system
  – To keep or not to keep: Original file names
  – Creating a useful filenaming system
    • Unique IDs
    • Location pointers
    • Context that you can understand in 20 years
    • Consider adding a “readme” text file
DIGITAL ARCHIVING AND PRESERVATION

- Intellectual control
- Controlled metadata vocabularies
- Rights and provenance information

Ensures access over time

Hosted by ALCTS, Association for Library Collections and Technical Services
Digitization and Disk Imaging

- Digitization is the process of an analog signal changing to binary data (01010101)
- Disk Imaging creates a file containing the contents and structure of a disk volume (hard disk drive, tape drive, floppy disk, optical disc or USB flash drive)
DIGITAL ARCHIVING AND PRESERVATION

• Stabilize media, extract, image, etc.

Characterization
“Snapshot” activities, such as ISO-creation. Manifest creation. Checksum generation.

Compliance

Delivery
Quality Control, again.

Ingest
Fixity activities. Metadata wrangling, embedding, and parsing.

Hosted by ALCTS, Association for Library Collections and Technical Services
Creating and maintaining a repository of digital information objects and associated metadata for long-term preservation and access

Collection of digital files or objects
Actively maintained for preservation and access purposes

Open file formats/file normalization
Sustainability factors: Disclosure | Adoption | Transparency | Self-documentation | External Dependencies | Impact of Patents | Technical Protection Mechanisms

Dispersed Storage with Fixity Checking
Infrastructure, multiple copies, checksum verification

Having a specific collection development policy/preservation policy
Guide’s the archive’s work

Fileneing System
Intellectual control

Security and Auditing
Access control, Personally Identifiable Information, ISO 16363

Hosted by ALCTS, Association for Library Collections and Technical Services
DIGITAL ARCHIVING AND PRESERVATION

• Wrapper: MXF
• Codec: DV25
• Goal: Create a best-effort restoration
• Tools: Terminal on Mac, Quicktime, Hex Fiend
• Method: Identify bit-level difference, identify bad file section, replace with good file section
• Result:
CONSIDERATIONS FOR AV

CHALLENGES

ACQUIRING
• Subscription service vs. file custody
• Carrier and file readability
• Licensing

MANAGING
• Digital storage space
• Repository infrastructure
• Metadata
• Integration with course reserves and other systems

USING
• Format obsolescence
• Access control
• Streaming media
• Digital preservation

Hosted by ALCTS, Association for Library Collections and Technical Services
CONSIDERATIONS FOR AV IN DIGITAL PRESERVATION

• Digitization is key to the survival of your analog materials
• Migration is key to the survival of your born-digital materials
• “Master” files should be archived and preserved
• Access files should be used for production purposes and user access
  • Mezzanine files serve as intermediary
  • Access files serve as end user files
How do I choose a format?

- Non-proprietary, open standards are key
- What works in your institution (in terms of size and ease of use)
- Can be transcoded easily for future migrations
- Least amount of software and hardware dependencies
The Sustainability of Digital Formats Web site provides information about digital content formats. The analyses and resources presented here will increase and be updated over time. The compilers, Caroline R. Arms, Carl Fleischhauer, and Kate Murray invite feedback on the content.

**Introduction**

Background information and overview: What is a format? How shall we evaluate formats? What projects in other organizations are addressing these questions?

- Overview
- Formats, Evaluation Factors, and Relationships
- Papers and Presentations
- Related Resources

**Sustainability Factors**

What affects the ability of the Library to preserve content in a given format? These sustainability factors apply to all formats.

- Disclosure
- Adoption
- Transparency
- Self-documentation
- External Dependencies
- Impact of Patents
- Technical Protection Mechanisms

**Content Categories**

The evaluation of formats must take into account quality and functionality. These factors vary according to the type of content under consideration and the categories will be expanded as time passes.

- Still Image
- Sound
- Textual
- Moving Image
- Web Archive
- Datasets
- Geospatial
- Generic
- Browse All Formats

**Format Descriptions**

Documents with more information about specific formats.

- Browse categories
- Browse alphabetical list
- Format Descriptions as XML
## CONSIDERATIONS FOR AV IN DIGITAL PRESERVATION

<table>
<thead>
<tr>
<th>Common Wrappers/Codecs for Video Preservation</th>
<th>Common Format for Access</th>
<th>Common Wrapper/Format for Film Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Quicktime (MOV)</td>
<td>• MPEG-4 (mp4)</td>
<td>• 2k-4k DPX sequences</td>
</tr>
<tr>
<td>• AVI (Audio Video Interleaved)</td>
<td></td>
<td>• Conformed Audio/Video into Video Wrapper</td>
</tr>
<tr>
<td>• Matroska</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• V210 Uncompressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• FFv1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Archiving: Analog Video Source

<table>
<thead>
<tr>
<th>Video Compression: Uncompressed</th>
<th>Audio channels: same as original</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Size: Standard Definition: 720x4866</td>
<td>Audio Compression: Uncompressed, PCM</td>
</tr>
<tr>
<td>Frame Size: High Definition: Native</td>
<td>Audio Sample Rate: 48 kHz</td>
</tr>
<tr>
<td>Aspect Ratio: Native: 4:3 for SD; 16:9 for HD</td>
<td>Audio Resolution: 24-bit</td>
</tr>
<tr>
<td>Bit Depth: 10-bit</td>
<td></td>
</tr>
<tr>
<td>Color Space: YCbCr</td>
<td></td>
</tr>
<tr>
<td>Chroma subsampling: 4:2:2</td>
<td></td>
</tr>
<tr>
<td>Interlaced/Progressive: Native</td>
<td></td>
</tr>
<tr>
<td>Frame Rate: Native, 30 or 29.97</td>
<td></td>
</tr>
</tbody>
</table>
CONSIDERATIONS FOR AV IN DIGITAL PRESERVATION

Archiving: Digital Video Source

Video Compression: Native
Frame Size: Native
Aspect Ratio: Native: 4:3 for SD; 16:9 for HD
Bit Depth: Native, 8-bit or 10-bit
Color Space: Native, YCbCr
Chroma subsampling: Native, typically, 4:2:2, 4:1:1, or 4:2:0
Interlaced/Progressive: Native
Frame Rate: Native

Audio channels: Same as original
Audio Compression: Native, typically uncompressed, PCM
Audio Sample Rate: Native, typically, 48 kHz
Audio Resolution: Native, typically 24- or 16-bit
CONSIDERATIONS FOR AV IN DIGITAL PRESERVATION

• Metadata
  • Descriptive
  • Technical
    • File properties
  • Preservation
    • Administrative
    • Rights
    • e.g. PREMIS

• Standards
  • Dublin Core
  • PBCore
  • EBUCore
CASE STUDY - VHS MIGRATION PROJECT

• VHS tape collection currently numbers approximately 6,500 titles
• Physical deterioration, poor video quality, obsolescence, dwindling circulation

Hosted by ALCTS, Association for Library Collections and Technical Services
CASE STUDY - VHS MIGRATION PROJECT

Goals

• 90% of the VHS collection can be replaced title for title with DVD or Blu-ray formats
• 5% can be replaced with similar or new material on DVD or Blu-ray, or through online databases
• The remaining 5% will be reformatted to DVD or Blu-ray and re-cataloged
• Withdraw redundant/outdated titles

Hosted by ALCTS, Association for Library Collections and Technical Services
- Section 108(c) of the United States Copyright Act
  - "Limitations on Exclusive Rights: Reproduction by Libraries and Archives."
  - Allows libraries to make copies for preservation purposes.
This video is closed captioned. The video's dialogue will be displayed as words on your TV screen when you have a VideoCaption Reader connected to your TV and video recorder.
CASE STUDY - VHS MIGRATION PROJECT
CASE STUDY – UC TV

Mission Statement:

Showcasing the excellence and diversity of the nation's premier research university, UCTV embraces the core missions of the University of California -- teaching, research and public service -- through quality, in-depth television that informs, educates and enriches the lives of people around the globe.
Where to Watch UCTV

It is our aim to make UCTV available to everyone, anywhere, at anytime. Take your pick of these many ways to tune in:

- Live Webstream
- Cable Television
- YouTube
- Podcasting and iTunes U
- Roku
- Video On-Demand
- iPhone App
- Amazon Fire TV

Learn more about UCTV:
- Contact Us
- FAQs
- Retransmission Information
- UCTV Overview
CASE STUDY – UC TV

Hosted by ALCTS, Association for Library Collections and Technical Services
Holocaust Living History Workshop

About this collection

Organizers
- Hilman, Susanne
- UC San Diego Library
- University of California, San Diego. Division of Arts and Humanities. Jewish Studies Program

Project Directors
- Hertz, Deborah
- Schottlander, Britan

Creation Date
2012 to 2016

Description
The video recordings represent events organized by the Holocaust Living History Workshop (HLHW). Established in 2008 as a collaborative project of the UC San Diego Jewish Studies Program and the Library, the HLHW aims to preserve the memory of the victims and survivors of the Holocaust. Starting in January 2010, the HLHW has been hosting meetings and presentations that connect local survivors, their families, researchers, UCSD students, and San Diego communities. The collection includes the recordings of these events since 2012. Its content covers multiple topics of the history of the Jewish and Romani Holocaust, antisemitism, and Jewish migration mostly in the 1930's – 1940's in Europe and the United States. Besides that, the collection amply represents the intergenerational Holocaust impact and contemporary reflections on the Holocaust.

Extent
37 digital objects.

Topics
CASE STUDY – UC TV

• Initial overview:
  • Provide secure management/preservation access to Holocaust Living History Workshop (HLHW) videos. Currently these videos are held in a mix of locations, seven videos on the UC San Diego Library YouTube channel and four on local hard drive storage. In both instances the content is insufficiently described and not committed to preservation storage. Ideally, once these videos are described and ingested to the DAMS as objects, a link will be built to their DAMS collection page from the HLHW website; enabling the DAMS instance as the primary access copy for this content.

• Initial objectives:
  • Assess technical specifications for the files and develop metadata needed for DAMS ingest
  • Ingest to the DAMS, establish collection landing page for HLHW videos
  • HLHW Website will link to DAMS collection landing page
CASE STUDY – UC TV

Initial project successful, MOU created to establish relationship between Library and UCTV, now a “living archive”

- Accept files from producer
  - Ensure they are “masters”
- Identify files and verify they conform to our agreements
- Create a mezzanine master
  - Remove bars and tone
  - Remove other content irrelevant to user
- Describe content through metadata creation
- Build digital object/ingest into DAMS
- Quality control
- Go live

Hosted by ALCTS, Association for Library Collections and Technical Services
Lessons Learned

• UCTV files provided need close QC to ensure compatibility with online video player
  • Experienced issues with incorrect aspect ratio displaying in jwplayer
• Files from producer need conform to baseline standards with respect to file format properties and supplied metadata
  • Teach UCTV how to provide master files and basic metadata
• Rights review required us to collect appropriate releases from individuals documented in videos and acknowledge underlying rights of content
CONCLUSION

• AV media preservation in the digital domain requires knowledge of the analog originals, the files generated through digitization, and files acquired through born-digital means
• Archiving and preservation is part of a larger process that incorporates people, policies, and technology
• Decisions must be made based on available resources and what makes sense for the individual institution
Thank you!

**Stefan Elnabli**
Media Curation Librarian \ Manager, Digital Reformatting Operations
UC San Diego Library
stelnabli@ucsd.edu