

# DIY Video Archiving with the Community Media Archive

John Hauser, Access Humboldt  
Kayla Creamer, Amherst Media  
James Jones, Attelboro Access Cable System

Nov 20, 2015

ACM North East Regional Conference

<https://cutt.ly/StqMjLW>

# Context

- Access Center as the repository for a community's cultural history
- Rationale has evolved from VOD through archiving to distribution
- Do It Yourself Archiving
- It's a marathon not a sprint!
- You CAN do it!

# Scope of the Internet Archive

- nonprofit digital public library - founded 1996
- 15+ million items (books, videos, audio, live music, games, software)
- 445+ billion web pages in the Wayback Machine
- Goal is “Universal Access to All Knowledge”
- Documentary about the Internet Archive

# Community Media Archive

- a collection hosted by the Internet Archive
- attempt to solve VideoOnDemand for Access Humboldt
- 54,000+ videos, 50,000+ hours
- 44 Access Centers have contributed
- 5.4 million downloads
- 92TB of “original” video files (not “derivative” formats)

# Community Media Archive Vision

- A collection of broadcast quality, locally produced shows with sufficient descriptive information that are freely shareable
  - Good!
- Archive as the hub of a sharing/distribution system for broadcast quality video between access centers
  - DoubleGood!

# How do I get started?

- Register an email address and assign a password
  - <https://archive.org/account/login.php>

# How do I get a collection?

- email Collections group ([collections-service@archive.org](mailto:collections-service@archive.org))
  - title of your collection
  - logo and descriptive blurb about your center
  - request it be a sub-collection of the Community Media Archive
  - provide email address(es) registered with the archive

# Getting Started - Concepts

- Collection, item, identifier, file
- A file is uploaded to an item
- Item identifier must be unique across 15M+ items
- “details” page – item level
- Metadata – item level
- Getting organized is more important than technical knowledge



# How do I upload files?

- Use manual interactive interface to upload first several videos
  - <https://archive.org/upload/>
- Bulk uploader available
  - <https://github.com/kngenie/ias3upload>
  - Uses a Comma Separated Value file for metadata - [example](#)

# How do I upload my videos?

- Use the archive's interactive upload interface
  - <https://archive.org/upload/>
  - Double check that your collection name is selected
    - #I mistake!
- Solution: Preset your collection name in URL
  - [https://archive.org/upload?collection=access\\_humboldt](https://archive.org/upload?collection=access_humboldt)

# Metadata and the Archive

- Minimal required – identifier, title, description, subject
- Anything accepted (and retrievable through `_meta.xml` file on the item's “detail” page)
- subject(s) – what shows up under “Topics” link for collection
- Identifier must be unique across 15+ million items!

# Metadata and Archive - More

- File/Identifier naming restrictions
  - A-zA-Z0-9.\_ - no spaces, parenthesis, braces, pound signs, colons allowed in identifiers or file names
- Playback server is likely more permissive

# Pretty Good Metadata Practices

- consider including a “presenter” element
- include a “series” metadata element
- include a “runtime” in HH:MM:SS format
- use multiple “subject” elements
- put year in a separate “subject” element
- put station name, initials and state in separate “subject” elements

# Series found on the Internet Archive

- Big History Project
- Media Edge
- PopTech
- Peak Moment
- Rompeviento TV
- TedTalks
- The Royal Society for the Arts
- Veterans Helping Veterans

# How can I help?

- Improve the metadata of your items
- How will someone find this item? (in a collection of 54k items)
- Take the time to learn Internet Archive's interfaces
  - “Advanced Search”, “Edit Item”, “Item History”
  - “Item Manager”, [ias3upload.pl](#) (bulk uploader)
- Underwrite my work on the distribution aspect of the CMA

# AM Archives Case Study

- 2013 Community Preservation Act grant
- digitize 2,300 video tapes
- interns used for tape digitization process
- bulk upload to Internet Archive in batches
- lessons learned



# DoubleACS Case Study

- Operations Manager attended prior year's presentation
- had uploaded videos online but hosting co. said "too much"
- delivered 1,900 videos on a 3.7 TB external hard drive
- metadata in CSV format file
- uploaded in batches of 100, 4 simultaneous upload threads

# Thanks!

- To Brewster and the Internet Archive staff
- To Sean and Access Humboldt for underwriting the CMA effort
- To the Access Centers that have contributed video to the CMA
- Let's aim for "DoubleGood!" instead of settling for "Good!"

# More Information

- These slides: <https://cutt.ly/StqMjLW>
- Community Media Archive wiki: <https://cutt.ly/ztqMRCu>
- [john@accesshumboldt.net](mailto:john@accesshumboldt.net)
- [kayla@amherstmedia.org](mailto:kayla@amherstmedia.org)
- [jjones@doubleacs.com](mailto:jjones@doubleacs.com)