

# Metadata and the Future of Video in an OTT World

John Hauser, Community Media Archive

March 23, 2017

ACM Central States  
Regional Conference

<https://cutt.ly/Ftq0x8M>

# Context

- It's not about archiving!
- It's about metadata
- It's about designing your metadata at a collection level, not an item level
- so that you're ready for the future

# Key Points

- the future of video distribution is not TV, it's the internet
- you need consistently tagged video to get ready for the future
- metadata storage should be vendor neutral
- preparing for archiving/preservation can help prepare for the future

# Why Metadata Matters

- Key to “Discoverability” in an online world
- Bridge between Cablecasting and new forms of video distribution
- It’s what feeds Search Engines and Relevance Ranking algorithms

# Think about Metadata

- think about how you want to group your video programs
- think about how users interact with your video
  - how did they get here?
  - where do they want to go next?
- Don't lock up categories/genre in your website

# Metadata Makeovers

## Before

- Title
- No Subject Tags

## After

- Title
- Geographic and Organization
- Series
- Government Meeting?
- Name of Board or Commission
- Provenance (Youtube or Vimeo)
- Year

# Metadata Recommendations

- Consider metadata from the collection perspective vs. the item level
- Use Consistent Naming in Titles (spelling & CAPITALIZATION!)
- Add Dates or Episode numbering to Title - make it unique!
- Use something other than a common word or phrase for the title
- Include data only you would know and probably have - producer
- Use a vendor neutral solution

# Scope of the Internet Archive

- nonprofit digital public library - founded 1996
- 20+ million items (books, videos, audio, live music, games, software)
- 450+ 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
- 350,000+ videos, 250,000+ hours
- 44 Access Centers have contributed, 475 represented
- 12.4 million downloads
- 200TB 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 20M+ items
- “details” page – item level
- Metadata – item level
- Getting organized is more important than technical knowledge

# 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 20+ 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

# Civic Video Tagging Project

- started with request from DeeDee Halleck for a playlist of videos
- we requested the ability to play clips from an archived video
- archive.org added stop=nnn parameter
- [hyperlinked](#) PDF meeting agendas
- poor man's Granicus - meeting [video](#) with associated PDF agenda
- see Community Media Archive [wiki](#) for more details

# Non Commercial Series Available on the Internet Archive

- [Aspen Ideas Project](#)
- [Big History Project](#)
- [Global Connections TV](#)
- [Intelligence Squared U.S.](#)
- [The Laura Flanders Show](#)
- [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 350k 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

# Youtube —> Archive Project

- catalog use of Youtube/Vimeo by Access Center, Government, Schools and Libraries
- update metadata and thumbnails for new videos in channels weekly
- develop a semi automated way of downloading, enhancing metadata and archiving the Youtube/Vimeo channels
- easiest method of archiving video
- “Archiving as a Service” vs. “Do it Yourself” Archiving

# Takeaways

- Design metadata at the collection level before the item level
- Focus on Metadata Consistency
- Choose “DIY” Archiving or “Archiving as a Service”
- Get Started!
- Easiest way to get your video archived is upload to Youtube or Vimeo and let me know your channel URL

# 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
- To Eureka City Schools & HCOE for the 1 Gb/s network connection
- Go forth and Archive!

# More Information

- These slides: <https://cutt.ly/Ftq0x8M>
- Community Media Archive wiki: <https://cutt.ly/ztqMRCu>
- [Documentary video](#) of the Internet Archive
- [More photos](#) by Jason Scott of the Internet Archive
- [john@accesshumboldt.net](mailto:john@accesshumboldt.net)