

# Community Media Archive – How to Get Started

John Hauser – Access Humboldt

Wendy Blom – Somerville Community Access TV

2013 ACM National Conference

May 30, 2pm – 3:30pm

Slides: <http://goo.gl/Wx5m8>



# Vision

- A collection of
  - broadcast quality shows
  - locally produced
  - sufficient descriptive information
  - freely shareable
- Archive as the hub of a sharing system



# Scope of Internet Archive

- Nonprofit digital public library – 1996
- 7+ million items (books, videos, audio, live music)
- 280 billion web pages archived in the Wayback Machine
- Aim is “Universal Access to all Knowledge”
- Documentary about the Internet Archive



# Scope of Community Media Archive

- About 5 years old
- 24,000+ videos
- 31 access centers
- 1.3 million downloads
- 35TB storage



# Related Collections

- Access centers not part of CMA – CMAP, CTN5
- Access centers without their own collection
  - Grand Rapids, Fort Wayne, Maui
- Collections w ACM compliant MPEG2
  - [Tedtalks](#) – 1,350 videos
  - [Poptech](#) – 223 videos



# Community Media on the Roku

To add the private Internet Archive channel that contains the Community Media Archive to your Roku box, go to:

<https://owner.roku.com/add/NMJS5>

[Screenshots](#) of Roku interface



# Getting Started – Conceptual

- Collection, item, file
- A file is uploaded to an item
- Item identifier must be unique – 7M+ items
- [“details”](#) page – item level
- [Metadata](#) – item level
  
- Getting organized is more important than technical knowledge



# Getting Started – Procedural

- Register an email address
- Email the collections group to add a collection (collections-service@archive.org)
  - Include a descriptive blurb, graphic logo, registered email addresses for admin access
  - Ask that it be setup as a “subcollection” of the community\_media collection
- input formats they accept and what output formats they derive
- <http://archive.org/help/derivatives.php?mediateype=movies>





# Getting Started – More

- 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

# Metadata and the Archive

- Minimal required – identifier, “creator”, title, description
- Anything accepted (and retrievable through `_meta.xml` file on the item's “detail” page
- subject(s) – what shows up under “Browse by subject/keyword” link for collection
- Identifier must be unique across 7+ million items!



# Metadata and the Archive – More

- File naming restrictions
- A-zA-Z0-9.\_- no spaces, parenthesis, braces, pound signs, colons allowed in file names
- Playback server is likely more permissive
- File suffix matters for animated gif to appear in collection listings



# Internet Archive Interfaces

- [Advanced Search](#)
  - Fielded, boolean search (\*not\* Google style!)
- Metadata XML and JSON files
  - [http://archive.org/details/\\$identifier&output=json](http://archive.org/details/$identifier&output=json)
- Metadata cleanup after the fact
  - Single or few items - Edit Item interface
  - Many Items – Bulk Uploader in Metadata Merge Mode – no files uploaded, metadata in CSV format
- RSS feeds



# Internet Archive Admin Interfaces

- “Edit Item”
  - Useful for adding or correcting metadata for a single item
- “Item Manager”
  - “make dark” function
  - “re-derive” certain derived video formats
- “Item History”
  - Links to all prior logs for the item



# Finding Programs for Re-Broadcast

- Advanced Search “AND format:MPEG2” search qualifier
- TedTalks, PopTech collections
- Archive.org RSS feeds aren't good enough
  - No runtime reported
  - Subject/keywords not shown
    - Used MRSS 2.0 spec but no one else does

# Downloading Programs for Re-Broadcast

- Grab the MPEG2 version!
- Grab the \_meta.xml file or “screen scrape” title, description, runtime
- Grab the \_files.xml file; runtime (in seconds) available from the <length> element
- Want to be notified when a new episode becomes available?
  - Subscribe to RSS feed for the collection or create a custom one from advanced search interface



# Or Grab the \_archive.torrent file!

- Every item in the Community Media Archive has a torrent file!
- [http://archive.org/download/\\$identifier/\\$identifier\\_archive.torrent](http://archive.org/download/$identifier/$identifier_archive.torrent)
- Video formats + metadata in one torrent
  - Original + derived formats
  - Meta.xml file
  - Animated GIF from jpegs





## Somerville Case Study

- Their Director attended last year's presentation
- Collection of 1,500 shows on video tape
- Organized the collection using interns
- Filemaker database of metadata dumped to spreadsheet
- Videos digitized to MPEG4 format from tape
- First batch of 405 items - 630GB of MPEG2 files
- Shipped external hard drive via UPS
- modified spreadsheet for use by bulk uploader

