

# Video Archiving with the Community Media Archive

John Hauser, Access Humboldt

May 1, 2014

Central States ACM Regional Conference

<http://goo.gl/H9OoPQ>

# Context

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

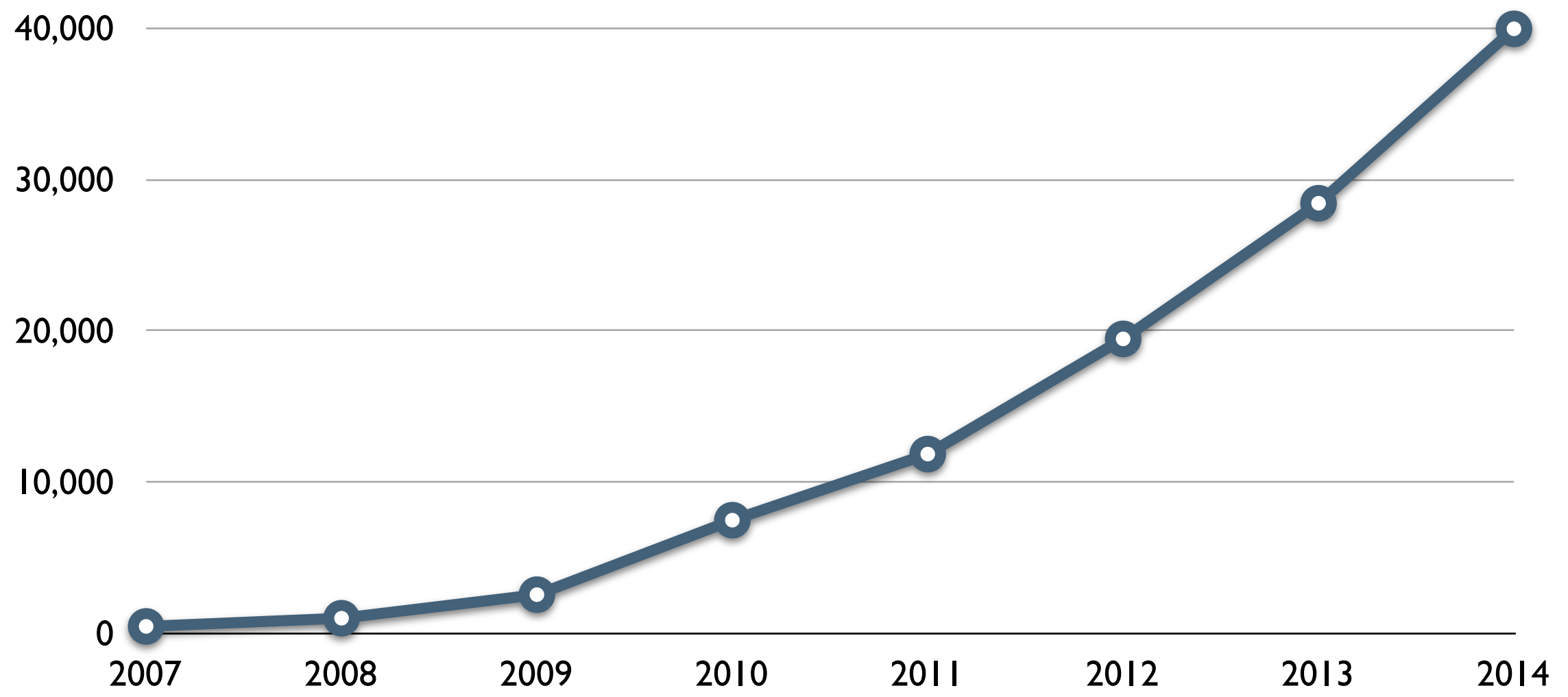
# Scope of the Internet Archive

- nonprofit digital public library - 1996
- 10+ million (books, videos, audio, live music)
- 400+ 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
- setup about 5 years ago - attempt to solve VOD for Access Humboldt
- 33,000+ videos, 29,000+ hours
- 39 Access Centers have contributed
- 1.3 million downloads
- 57TB of “original” video files (not “derivative” formats)

# ○ Total Community Media Archive Items



# 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
  - request the “MPEG2 derivative” - if original file not MPEG2



# Getting Started - Concepts

- Collection, item, identifier, file
- A file is uploaded to an item
- Item identifier must be unique across 10M+ items
- “details” page – item level
- Metadata – item level

# 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](#)

# Metadata and the Archive

- Minimal required – identifier, “creator”, title, description, subject
- 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 10+ 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
- File suffix matters for animated gif to appear in collection listings

# 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

# How can I help?

- Improve the metadata of your items
- How will someone find this item? (in a collection of 40k items)
- Take the time to learn Internet Archive's interfaces
  - “Advanced Search”, “Edit Item”, “Item History”
  - “Item Manager”, [ias3upload.pl](#) (bulk uploader)
- Help underwrite work on the distribution aspect of the CMA

# 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 the Digital Bicycle effort from ~ 10 years ago!
- Let's aim for "DoubleGood!" instead of settling for "Good!"

# More Information

- These slides
  - <http://goo.gl/H9OoPQ>
- Community Media Archive wiki
  - <http://goo.gl/Wx5m8>
- [john@accesshumboldt.net](mailto:john@accesshumboldt.net)