Archiving, YouTube & You

John Hauser Access Humboldt, James Jones DoubleACS

Thursday, August 18, 2016 9:00 AM

http://goo.gl/U2AaeZ
Archiving, Youtube and You

John Hauser, Access Humboldt
James Jones, DoubleACS

Aug 18, 2016 - 9:00am

ACM National Conference

http://goo.gl/U2AaeZ
Overview

• Introduction
• Themes
• Metadata matters
• Case Study - DoubleACS smartphone, Roku & Apple TV apps
• archive.org specifics
• Youtube —> Archive project specifics
Themes

• “DIY Archiving” or “Archiving as a Service”

• Consistency Matters; Organizational Implications

• Archiving = Preservation

• Archiving = Distribution

• Distribution = Archiving

• Youtube → Archive Project

• Vendor Agnostic

• Power of Partnerships
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
Metadata Exercise

• How would you group your videos?
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 “Topic” link in right hand col
- 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 / Youtube / Vimeo is more permissive
Pretty Good Metadata Practices

- consider including a “presenter” or “creator” 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
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 - should be 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)

• 500+ 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
- 190,000+ videos from 165 Access Centers and Gov. entities
- 44 Access Centers have contributed via “DIY” archiving
- More than 7 million downloads
- 192TB 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.createaccount.php

• Let John know what the email address is

• he’ll get it registered as an admin for your collection
How do I get a collection?

- email Collections group (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 files?

- Use manual interactive interface to upload first several videos
  - [https://archive.org/upload/](https://archive.org/upload/)

- Bulk uploader available
  - [https://github.com/kngenie/ias3upload](https://github.com/kngenie/ias3upload)

- Uses a Comma Separated Value file for metadata - [example](https://archive.org/upload/)

- Upload to Youtube and let John know you want it archived
How do I upload my videos?

• Use the archive’s interactive upload interface
  • [https://archive.org/upload/](https://archive.org/upload/)

• Double check that your collection name is selected
  • #1 mistake!

• Solution: Preset your collection name in URL
archive.org filename specifics

- filenaming restrictions: A-Za-z0-9_-.
- no special characters or spaces allowed ( ,’,”,%,@,#,:,{},)
- identifier length > 4 chars < 100 chars
archive.org UI specifics

• “show as list”, “show details”, views, “date archived” vs “date published”
• sorting and searching UI elements
• Topic “faceting” in right hand column
• Getting a derived “web version” 640x480 from your .mp4 original
• How to view an HD file in original format, not 640x480 web version
more archive.org UI specifics

- item history
- edit items, multi-item edit
- quick view of total download counts; “About” tab of collection
- requests to collections group — be respectful, be specific/actionable
The Archive adapts to us!

• security model allows for admin rights to sub collection without parent rights

• “broadcast_ready” collection - MPEG2 auto derived from your original

• stop parameter allows video clips

• RSS output of (advanced) search results

• runtime shown on items detail page
more Archive adaptations

• torrent file created for every video uploaded to the CMA
• RSS feed includes MPEG2 format and torrent file
• allows automated downloading of new items in a collection or advanced search
• metadata available in XML, CSV, JSON formats
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
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 190k items)

• Take the time to learn Internet Archive’s interfaces
  • “Topic” selection, “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
- statistics, metadata and thumbnails updated 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
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:  http://goo.gl/U2AaeZ

• Community Media Archive wiki: http://goo.gl/Wx5m8

• john@accesshumboldt.net

• jjones@doubleacs.org