Working with the Internet Archive

Topic Outline

Housekeeping
How to get a collection set up on the Community Media Archive
How to have your previously uploaded videos moved into your new collection
How to get admin access to your already established collection that I setup
How to monitor the health of the Internet Archive and the state of your queue
How to find out download/view stats of your videos?

Metadata Recommendations

Searching
How to use the Archive's search and Advanced Search features
How to use the "Play All" feature to preview search results

Downloading
How to use the Internet Archive as a source of video for your channel
How to automatically download new items in a search or collection

Uploading
How to upload your own videos to the Archive
How to upload MP4 originals so that you avoid "Error loading media: File could not be played"
How to have an SD MPEG2 video auto-derived from your HD wide-screen original
How to upload HD originals and have both an HD and and smaller sized MPEG4 video available

Interesting Random Bits
How to use the Archive's API (Application Programming Interface)
How to use start & stop points in Internet Archive video with hyperlinked meeting agendas
How to use the auto-derived captions for your video

Housekeeping

How to get a collection set up on the Community Media Archive

1.   Create an account by registering an email address:
http://www.archive.org/account/login.createaccount.php
 This email should be a "real" email address where you can receive email and that you check regularly. (The Internet Archive will only honor requests for modification/removal of an uploaded video if the request comes from the same email address that uploaded the video or that has admin access to the collection.)
At Access Humboldt we use a generic uploads@accesshumboldt.net account rather than a personal one, so that any of our 3 operations or programming staff can login to upload files, make changes to metadata or debug issues. It also prevents issues if someone leaves the organization.

Working with the Internet Archive

2.   To get a collection established for your organization, please email the Collections Service group (collections-service@archive.org) with the following information (they will use this information to establish your collection under the Community Media Archive):

* the title you want for your collection
* request to be a "sub collection" of the Community Media Archive (collection:"community_media")
* the email address you used to register for an archive account that should have admin access to the collection.
* a descriptive text blurb for your collection's detail page.
* a logo for your collection

How to have your items moved into your collection

Email collection-service@archive.org and provide the specific item identifiers that you need moved and the name of the collection you want them moved to. The email request must come from the same email address registered with archive.org that uploaded them.

How to get admin access to your collection that I setup

1.   Register an email address at: https://archive.org/account/login.createaccount.php
This email should be a "real" email address where you can receive email and that you check regularly. (The Internet Archive will only honor requests for modification/removal of an uploaded video if the request comes from the same email address that uploaded the video.)

2.   Let me (jhauser@pobox.com) know what the email address is.

3.   I'll get the IA collections service group to set it up with admin privileges for your collection. Most access requests are handled in the same day as received. Since my email address is the one with existing admin access to your collection of Youtube or Vimeo auto uploads, the request has to come through me, not directly from your email.

4.   After your email address is set up as an admin you can add new items, or edit metadata for existing items

You'll know you have admin access when you go to the your collection page and see "Edit" and "History" tabs in the Actions List on the upper right side of the page in addition to the "Share", "Favorite" and "Play All" actions.
On existing items, you'll now have "Edit", "Manage" and "History" Actions in the Actions List which you'll find under the filmstrip icon on the left hand side of the page under the video player area of the page.

John Hauser   jhauser@pobox.com   707-441-9401

Working with the Internet Archive

How to monitor the health of the Internet Archive and the state of your queue

your own queue this shows the status of any jobs related to items you've submitted.
Screenshot showing 1,015 of Community Media Archive video jobs being derived at once.

graphs for all of archive.org I check "Waiting Derives" and "Derives/hour" to make sure there's not a big backlog in the Archive's processing

How to find out download/view stats of your videos

The "About" tab of your collection's page
https://archive.org/details/pgctvmd&tab=about

Search by Collection, sort by descending views
https://archive.org/search.php?query=collection%3Apgctvmd&sort=-downloads

Same as Above, in a CSV file
https://archive.org/advancedsearch.php?q=collection%3Apgctvmd&fl%5B%5D=creator&fl%5B%5D=downloads&fl%5B%5D=identifier&fl%5B%5D=publicdate&sort%5B%5D=downloads+desc&sort%5B%5D=&sort%5B%5D=&rows=50000&page=1&callback=callback&save=yes&output=csv

Metadata Recommendations

See my 2 page Metadata "Cheat Sheet"
Basic metadata fields: **Title, Description, Subjects/Topics/Keywords**
Most important additional metadata fields are: **Runtime, Series, Producer, Genre/Category**

Searching

How to use the Archive's search and Advanced Search features

See lower half of https://archive.org/advancedsearch.php for syntax and examples
Explain/demo how I use search iteratively and then convert to advanced search and output the results in HTML table, CSV or RSS formats

How to use the "Play All" feature to preview search results

Upper right hand corner of the Search Results page.

John Hauser   jhauser@pobox.com   707-441-9401

Working with the Internet Archive

Downloading

How to use the archive as a source of video for your channel

[16 Series (17,800 videos) with Broadcast Ready MPEG2 versions downloadable from the Internet Archive](#)

Classical Arts Showcase replacement programming
      Search Query   - Search Results
[collection:community_media AND dance](#)   - 3,469 videos
[collection:community_media AND danc*](#) - 4,786 videos
[collection:community_media AND poe*](#) - 3,430 videos
[collection:community_media AND concert](#) - 8,329 videos
[collection:community_media AND concert*](#) - 9,375 videos
[collection:community_media AND symphony](#) - 864 videos
[collection:community_media AND art*](#) - 30,682 videos
[collection:community_media AND art* AND licenseurl:*creativecommons.org*](#) - 5,590 videos
[collection:community_media AND music*](#) - 23,010 videos
[collection:community_media AND paint*](#) - 2,748 videos
[collection:community_media AND pian*](#) - 784 videos
[collection:community_media AND culture](#) - 5,316 videos
[collection:community_media AND cultural](#) - 2,991 videos

How to automatically download new items in a search or collection

Use a Bittorrent client (uTorrent, qBittorrent) that can accept RSS feeds. Every collection and item in the Community Media Archive has a *.torrent file. Archive's RSS feeds include the *.torrent files. Advanced Search results can be returned as RSS feeds. Demo TEDTalks example of qBittorrent and the 16 [series](#) that have MPEG2 versions

Uploading

How to upload your own center's videos to the Archive

1.   For most users, use Archive's web interface: https://archive.org/upload
To make sure you don't upload to the wrong collection, use a bookmark or shortcut with your collection identifier appended. That way, you won't have to select your collection from the dropdown list in the web interface for every item uploaded. Example for Access Humboldt:
https://archive.org/upload/?collection=access_humboldt

John Hauser   jhauser@pobox.com   707-441-9401

2.   For advanced users only, use the batch uploader:
Mac version: mac_bulk_uploader.zip
Windows version (Portable Apps): windows_bulk_uploade..>
Kayla Creamer from Billerica Access TV, Joe Ameur from Amherst Media, James Jones from DoubleACS and Jason Daniels of ECAT, all have experience with the batch loader.

Be prepared to devote significant time to learning how things works with the batch uploader. The batch uploader pulls metadata from a CSV file that you create.

Updated suggestions for metadata

3.   For Telvue customers, use your Hypercaster's archiving function. See
https://telvue.com/knowledgebase/archive-org-integration/

4.   For Tightrope customers, ask Ray Tiley <ray.tiley@trms.com>
about what's involved in using Media Sherpa. It's a utility to upload video from Cablecast servers to Internet Archive. Plans to provide a supported "upload to Archive" feature have been added to Cablecast's 2018 product roadmap. Look for it towards the end of the year according to Steve Israelsky of Cablecast/Tightrope.

How to upload MP4 originals so that you avoid "Error loading media: File could not be played" in Archive's video player

Upload MPEG4 files with an .mpeg4 suffix, not an .mp4 suffix. The Archive will auto-derive a web- optimized MPEG4 file if and only if an original file with an mp4 suffix doesn't exist.
For a more detailed description of the problem, read
https://rigor.com/blog/2016/01/optimizing-mp4-video-for-fast-streaming

How to have a Broadcast Ready MPEG2 video auto-derived from your HD wide screen original

Add an extra "collection" metadata element with the value "broadcast_ready" to your metadata. If it's a new item, an NTSC MPEG2 version will be created as part of the "derive" step. If it's an existing video, you'll have to go into the Item Manager for the item and select "all" in the Re-derive section of the Item Manager page to force a re-derive of an existing item.

How to upload HD originals and have both an HD and and smaller sized MPEG4 video available

Make sure your HD video doesn't have an .mp4 suffix. I change the suffix to .mpeg4 or HD.mov. Archive will auto derive a smaller size "web viewable" .mp4 file as long as the original file doesn't have an .mp4 suffix. If it does, no smaller size .mp4 is created. If you want to view the

Working with the Internet Archive

HD version in Archive's jwplayer, click on the "Gear" icon and select 1080p or 720p instead of the default 360p. This is a member education issue.

Interesting Random Bits

How to use the Archive's API - Application Programming Interface

https://internetarchive.readthedocs.io/en/latest/cli.html

How to use start & stop points in Internet Archive video with hyperlinked meeting agendas

See Civic Video Tagging Project
Example Meeting: https://archive.org/details/AH-fcc_4-16-18
Hyperlinked Agenda:
https://archive.org/download/AH-fcc_4-16-18/Fcc_agenda_linked_4-16-18.pdf

How to use the auto-derived captions for your video

Look for a "SubRip" file in the downloads area of an item's page. That will be the text file that contains time code, sequence and captions.
https://archive.org/download/Dance_Ptown_Dance_Floor_Fundraiser/Dance_Ptown_Dance_Floor_Fundraiser.asr.srt

You can download that text file and import it into a program that edits captions, such as CADET. Then re-upload the corrected captions to the item.

John Hauser   jhauser@pobox.com   707-441-9401