

Archiving with the Internet Archive

John Hauser, Access Humboldt
Nick Ring, Amherst Media

Oct 11, 2013

ACM-NE Regional Conference

<http://goo.gl/limdzq>

Context

- Access Center as the repository for a community's cultural history
- State funding (MA - CPA) may be available
- Rationale has evolved from VOD through archiving to distribution
- It's a marathon not a sprint
- You **can** do it!

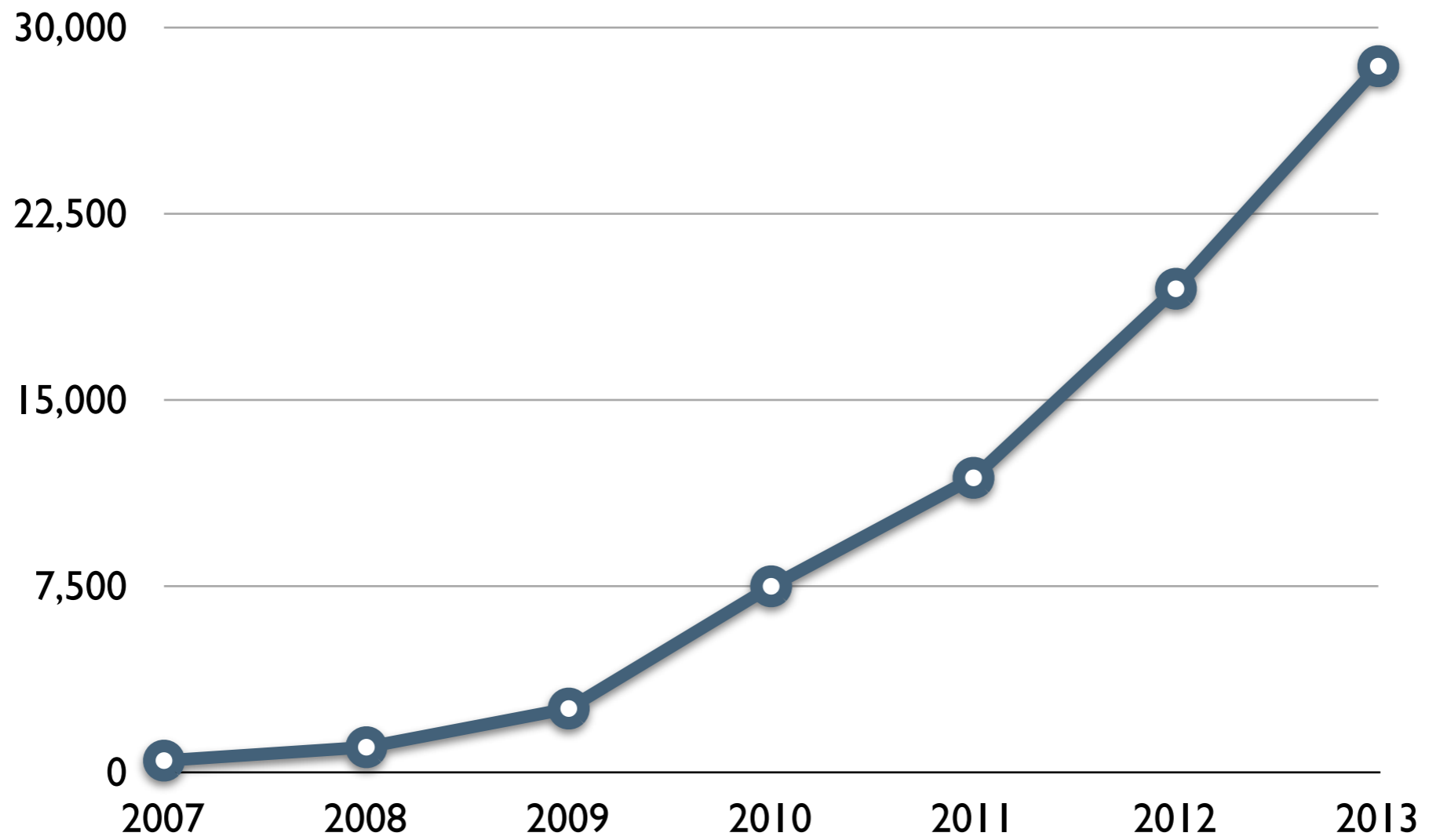
Internet Archive

- nonprofit digital public library - 1996
- 7+ million (books, videos, audio, live music)
- 280+ 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
- 26,000+ videos, 21,000+ hours
- 38 Access Centers have contributed
- 1.3 million downloads
- 42TB 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)
 - 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

Basic Concepts

- Collection, item, identifier, file
- A file is uploaded to an item
- Item identifier must be unique across 7M+ 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>

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 7+ million items!

Metadata and Archive

- 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

PG 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
- separate station name, initials and state

How can I help?

- Improve the metadata of your items
- How will someone find this item? (in a collection of 24k 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

Amherst Case Study

- access centers as repository of community's cultural history
- heard about the Community Preservation Act
- 2 years from start to pilot
- Nick will discuss their planning in more detail

Amherst Media

Local community cable access stations are repositories of local history and therefore have a fundamental responsibility to preserve that history and make it available to future generations.

– Diane LeBlanc

Regional Administrator,

National Archives & Records Administration, Northeast Region

Evaluating the Archiving Process

- **Evaluated our archives**
 - **media types, quantities, condition, metadata**
- Consulted with people who know
 - Massachusetts State Historic Records Advisory Board (Massachusetts SHRAB)
 - “Roving Archivist”
 - John Hauser / Archive.org
 - Other access centers
- **Equipment, Time, and Money**

Community Preservation Act

Purpose

- preserve open space
- historic preservation
- create affordable housing
- develop outdoor recreational facilities

Community Preservation Act

Definition

An historic resource is defined as a building, structure, vessel, real property, document or artifact that is either:

- listed on the State Register of Historic Places; or
- determined by the local Historic Commission to be significant in the history, archeology, architecture, or culture of the city or town.

Community Preservation Act

The Money:

- Property tax
- Community Preservation Trust Fund

The Deliberative Process:

- Community Preservation Committee
- Community's legislative body

Community Preservation Act

Funding and Projects

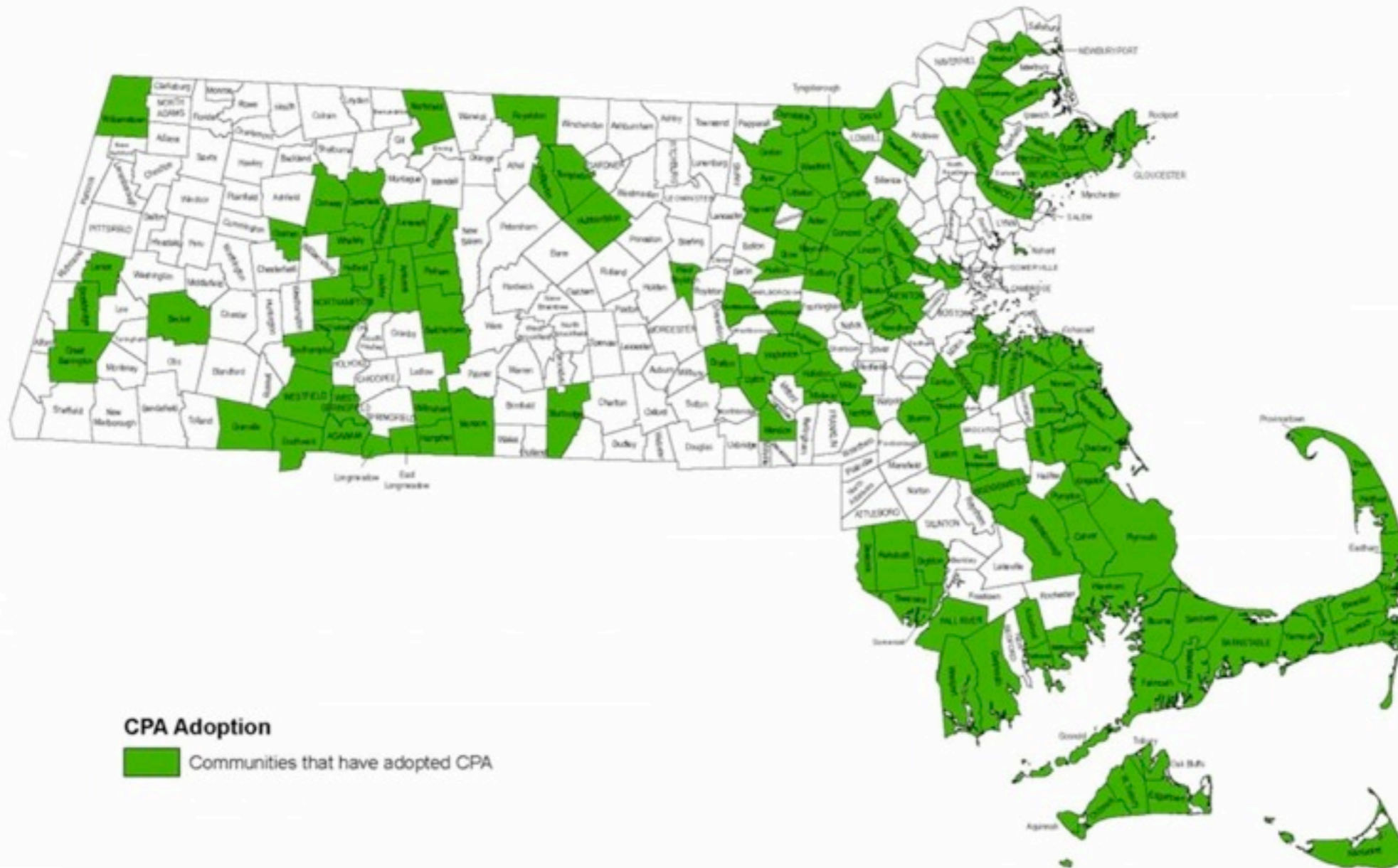
- 155 communities have adopted CPA (44% of the Commonwealth's cities and towns)
- Close to \$1.2 billion has been raised to date for community preservation funding statewide
- Over 6,600 projects approved
- Over 7,800 affordable housing units have been created or supported
- Nearly 19,700 acres of open space have been preserved
- Over 3,300 appropriations have been made for historic preservation projects
- Over 1,000 outdoor recreation projects have been initiated

CPA Funding: Why Community Media Centers

- We possess critical history
- We can provide free access to that history
- We can preserve that history with quality and longevity
- Media management is at the core of our expertise

Community Preservation Act Adoption

May 1, 2013



Source: [Community Preservation Coalition](#)

Amherst Media

Acton	Concord	Groveland	Maynard	Provincetown	Tewksbury
Acushnet	Conway	Hadley	Medway	Quincy	Tisbury
Agawam	Dartmouth	Hamilton	Mendon	Randolph	Truro
Amherst	Deerfield	Hampden	Middleborough	Rehoboth	Tyngsborough
Aquinnah	Dennis	Hanover	Middleton	Rockport	Upton
Ashland	Dighton	Hanson	Millis	Rowley	Waltham
Ayer	Dracut	Harvard	Monson	Royalston	Wareham
Barnstable	Dunstable	Harwich	Nahant	Salem	Wayland
BECKET	Duxbury	Hatfield	Nantucket	Sandwich	Wellesley
Bedford	East Longmeadow	Hingham	Needham	Scituate	Wellfleet
Belchertown	Eastham	Holliston	Newburyport	Seekonk	Wenham
Belmont	Easthampton	Hopkinton	Newton	Sharon	West Boylston
Beverly	Easton	Hubbardston	Norfolk	Shutesbury	West Bridgewater
Bourne	Edgartown	Hudson	North Andover	Somerset	West Newbury
Boxford	Essex	Kingston	Northampton	Somerville	West Springfield
Braintree	Fairhaven	Lenox	Northborough	Southampton	West Tisbury
Brewster	Fall River	Leverett	Northfield	Southborough	Westfield
Bridgewater	Falmouth	Lexington	Norwell	Southwick	Westford
Cambridge	Georgetown	Lincoln	Oak Bluffs	Stockbridge	Weston
Canton	Gloucester	Littleton	Orleans	Stoughton	Westport
Carlisle	Goshen	Longmeadow	Peabody	Stow	Weymouth
Carver	Gosnold	Manchester	Pelham	Sturbridge	Whately
Chatham	Grafton	Marion	Pembroke	Sudbury	Wilbraham
Chelmsford	Granville	Marshfield	Phillipston	Sunderland	Williamstown
Chilmark	Great Barrington	Mashpee	Plymouth	Swansea	Yarmouth
Cohasset	Groton	Mattapoissett	Plympton	Templeton	

Amherst Media Archives Contain:

- 2,306 tapes of various formats
- 2,577 estimated hours of material

Additionally, there are thousands of tapes and DVDs housed at the public library

Amherst Media CPA Proposal:

Hardware: \$13,278

Personnel: \$43,316

Estimated start of project: June, 2013

Estimated date of completion: July, 2014

Media Formats

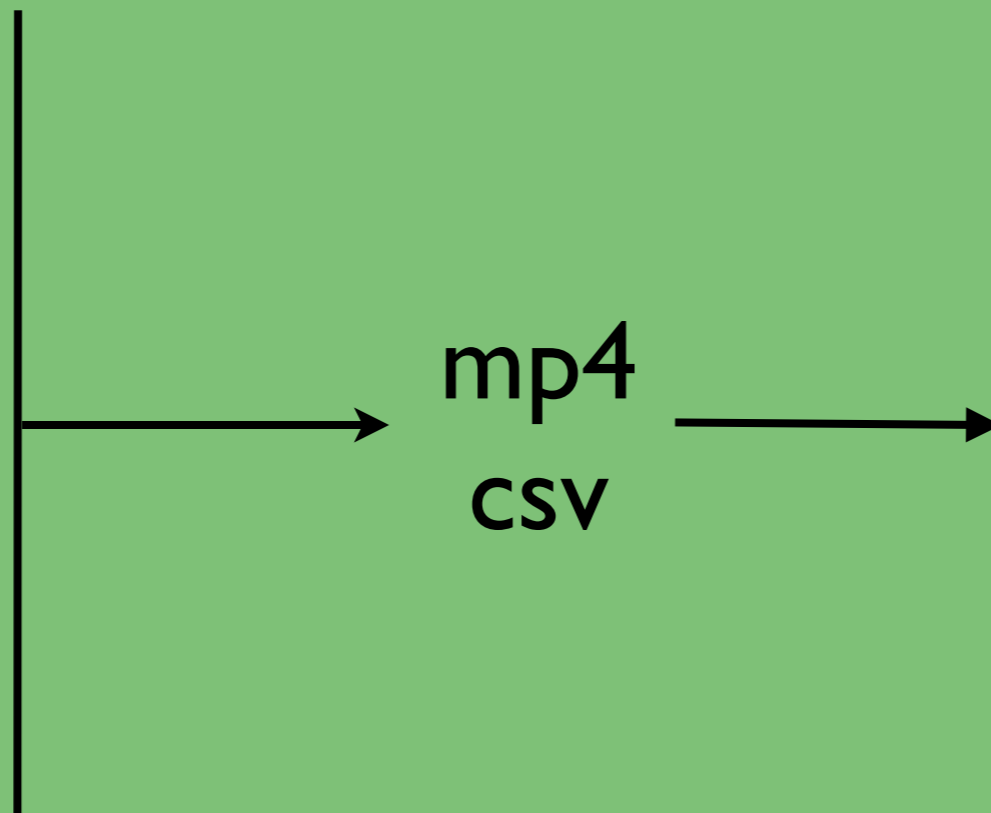
Ingest and Catalog

Archiving

miniDV
3/4" tape
Beta
Hi-8
VHS
SVHS
DVD
MPEG2

mp4
csv

archive.org
LTO tape
originals



More Information

- These slides
 - <http://goo.gl/LZWVTi3>
- Community Media Archive wiki
 - <http://goo.gl/Wx5m8>
- john@accesshumboldt.net
- nick@amherstmedia.org