WGBH Boston - media archives

"In progress"

⚠️

This page is a work in progress and is intended to provide resources for working with media assets in a repository environment.

Media archives interest group

Challenge Areas

These are areas of difficulty that we are all facing together. For each of these topics, there are at least a handful of people who are actively working to reduce these challenges to non-issues. Are you one of those people? If you are, make sure we all know about your work, and make sure it's reflected here on the wiki.

- Workflows
- Cataloging
- Preservation
- Digitization
- Encoding + Processing
- Metadata + Controlled Vocabularies
- Asset formats and delivery - streaming vs progressive download, quicktime vs flash
- Content Modeling and the CMA
- Managing Large Files
- Frontends, annotation environments, etc

Hot Topics

Digitization

There is a lot of discussion about digitization and the tradeoffs of an uncompressed format vs a lossless format.

Other communities interested in digitization and preservation include:

- Association of Moving Image Archivists (AMIA-L)
- Association for Recorded Sound Collections (ARSCLIST)

Helpful utility: AJA Data Rate Calculator Mac | Win

Content models

WGBH OpenVault model

WGBH has created a very atomistic content model to support their rich media archives repository.

Content types:

- Concept
- Video
- Image
- Document

Delivery

There are three basic approaches to delivering video files: progressive download, rtsp/rtmp streaming, and progressive download with HTTP range requests. Many media repositories are currently using Fedora redirect datastreams to reference content hosted on dedicated media servers (either locally, on a CDN, etc). The obvious advantages here are optimized delivery of media content and the ability to offload processing and storage to external processes. The biggest disadvantages are loss of integrated authn/authz from Fedora and additional effort for allowing Fedora to manage those assets.

One current limitation in Fedora is it uses the Java URL parsing library, which fails to recognize rtsp:// format URLs. This is further confounded by Quicktime's inability to follow HTTP redirects to rtsp:// streams. One possible work-around is to use the quicktime media link xml schema to redirect the Quicktime plugin appropriately. WGBH has developed this proof-of-concept.

Metadata formats

- Dublin Core
- PBCore
- EBUCore
- METS (e.g. LoC Video + Xscript)
- MPEG-21
- MXF
- Microformats

Resources

Archives with significant media content

- WGBH Media Library and Archives | http://openvault.wgbh.org/
- Jewish Women's Archive | http://jwa.org
- Spoken Word Services at Glasgow Caledonian | http://www.spokenword.ac.uk/
- Indiana University Digital Library Program | http://www.dlib.indiana.edu/
- Canadian Centre for Ethnomusicology at the University of Alberta
- SAFIR at University of York | http://music.york.ac.uk/atyork/resources/archives.php