

CLIR symposium paper January 2015

Accelerating Exposure of Audio Visual Collections, What's Next?

Authors: Karen Cariani, Sadie Roosa, Jack Brighton, Brian Graney

Abstract

The paper presents issues of exposing large audiovisual collections online with examples from Hidden Collections initiatives. User interest in the exposure of digitized objects must be constantly weighed against the cost of preparation, management and storage.

The Boston Local TV News Digital Library (<http://bostonlocaltv.org/>) project is a core example. As a collaborative project, it has been a signal success, exposing around 50,000 individual film and video news stories created between 1960 and 2000. The four institutional partners addressed issues faced by libraries and other institutions with large moving image collections: how to effectively and economically expose large collections for better access, solve rights issues that may effect online access and use, and build awareness and use for educational needs. The project showed that exposing minimal item-level descriptive data increases awareness and discoverability of collections.

The paper is enhanced by examples from other moving image Hidden Collections projects, with particular attention to overcoming technology barriers, funding structures, and staffing limitations. It will focus on the importance of sharing metadata standards for audiovisual works, in particular the use of PBCore, and will report from the PBCore user community. Lastly, presenters will address the effectiveness of metadata workshops in a period of transition.

Title: Exposing Large Collections of Media
Karen Cariani
Director WGBH Media Library and Archives

The challenge for analog audiovisual collections in the digital networked age is their relative invisibility compared to numerous online resources that can be easily searched and viewed through public search engines. If an archive record doesn't appear in a search, it may as well not exist. Any plan for a sustainable online archive must take into consideration this extremely competitive environment for the public's attention. Due to the volume of individual surviving TV news collections, in particular, and each collection containing millions of feet of film and tens of thousands of stories, it is not economically feasible to catalog and digitize the full extent of each collection to expose each archive's content. The solution is not to select narrow topics and abandon the volume of the holdings to dark archives, thereby restricting future use; instead, the challenge is to illuminate the breadth of the collections by determining the information necessary for research within entire collections. The Boston Local TV News Digital Library project demonstrated and tested how to organize widely disparate news stories from separate collections, expose the full range of content from 1960-2000, and contextualize the stories for public access, educational, and scholarly use.

Despite rapid advances in the world of digital libraries, moving images are usually the last to benefit from effective standards-based work—and television news is further deferred, offered in small topical selections, because of the size and technological challenges of large existing collections. Large media collections are hidden because they are on formats that are inaccessible – film, 2", 1", 3/4" – all now considered obsolete formats. The equipment alone to watch the content stored on analog formats is disappearing - no longer being manufactured, breaking down, and being discarded. In addition, with the deteriorating magnetic signal on the tapes, and disintegration of the film, often the next time a tape is played when one seeks to determine its content may be the last time the tape can ever be seen, or the signal on the tape ever capable of being captured to make a copy.

So how can we discover what is on these tapes and film and in these media collections? How can we let scholars and researchers know about this content? The dilemma for media material is that content on obsolete formats needs to be digitized to be accessible and then it needs to be described and cataloged to be identified and discovered. Videos are opaque and so the content is unknown without playing them. Titles alone don't necessarily tell you the content. You can't scan tape or film for searchable data. Most research and searching on the web occurs via text search engines. With no robust descriptive metadata for the media, having many hours of digital content available on the web does not make it discoverable. To increase use and improve findability, and access for scholars and researchers, the content needs better descriptive information that can be indexed by search engines.

There are a number of attributes and data that could be gathered without having to fully watch or listen to a program, which would greatly enhance accessibility. Collecting, posting and exposing a minimal amount of data can be helpful. Minimal information that can be gathered from the physical item – on tape labels and film cans, is better than nothing. Verifying or documenting the correct title(s) and title types from the title card and dates associated with the content, identifying

primary creators and contributors by just watching or listening to the credits, or only documenting the statement in the copyright notice would help with rights assessment. Additionally, noting the type or genre of program that might not be obvious from the title – like dramatic performance, news, instructional, public affairs, documentary, etc. would be extremely helpful and would aid researchers in their search for material. Having this data as part of the archive will greatly enhance discoverability of the collection for scholarly and humanities work.

To increase access for scholars at WGBH, we have been testing the value of publishing minimal data for our media archive. We launched our archive ‘catalog’ on-line. It has minimal data, often just a title, in the hopes that it is at least now discoverable. When possible, we have published transcripts of interviews, which are exposed to search engines, further enhancing discoverability. We have worked with our users to add information to our minimal records, thus enhancing the catalog even more. But most of these records are titles of full programs that may have been referenced somewhere at some time. Our news collection posed a more complex challenge. The first was its volume, and second was the variety of topics, which covered more than 10 years of daily news in Boston. Daily news doesn’t have a succinct title other than on a collection level that represents 10 years and thousands of items.

Television news has provided a daily chronicle of our communities since the late 1940s, especially through local TV news. Across the United States, collections of local television news on film and defunct analog video occupy shelves at TV and radio stations, colleges and universities, public and specialty libraries, local and state historical societies, and archives and museums. These records of local TV news represent an as yet largely untapped resource for study and use. Despite their potential value to the public and in-school audiences, few other records of contemporary culture have been as unavailable as 20th century television news. In its *Television and Video Preservation 1997* report (See Appendix 2), the Library of Congress stated, “The most devastating losses have already occurred among news film and videotape files of local television stations across the United States.” Compounding the access problem is the volume of those individual surviving collections and the overwhelming cost of making them accessible and useful. To preserve and enhance the value of audiovisual news collections, they must be organized, described, put into context, and made known to users.

Daily local TV news has been created and produced for over 50 years – 30 minutes to an hour, every day. It is a lot of content, a lot of tape, a lot of hours to view to catalog. For time-based media, the content needs to be seen or heard in real time for a human to be able to describe it and tag it. As an example, our American Archive of Public Broadcasting project, the 40,000 hours we have digitized would take one person working 8 hours, 7 days a week, 30 years to catalog. The resources necessary to fully describe and catalog hours of programming are overwhelming and unattainable.

The few local TV collections offered in digital libraries is relatively small scale and topically focused - chiefly on the civil rights era. These and other projects offer important documentation, yet they do not expose entire collections. This project sought to try and expose entire news collections. The WGBH Media Library and Archives with Northeast Historic Film, Cambridge Community Television, and the Boston Public Library, developed *The Boston TV News Digital Library: 1960-2000*, the first online resource offering a city’s commercial, noncommercial, and

community cable TV news heritage to educators and the public. The purpose of the collaboration was to use, test, and demonstrate open source tools to assist custodians of similar resources, while creating an online library offering 40 years of urban moving image materials, resulting in approximately 70,000 news records. *The Boston TV News Digital Library: 1960-2000* is designed to effectively and economically expose large collections for access. Each of the four partners brought a television collection to the digital library ranging from hundreds of analog videotapes to 4 million feet of 16mm film.

Online access to TV news collections poses complicated rights issues. Where users once expected to make research visits to collections, global exposure via the web changes the range of potential legal risks. Consequently, an examination of rights issues for online access to local TV news in public collections was also greatly needed. The ephemeral nature of TV news – it airs and never will be seen again -- made releases and rights issues less risky for broadcasters, so releases and permissions were generally not secured. As part of this project, we worked with our legal team and the Berkman Center at Harvard University to answer some definitive questions to allowing on-line access to the news collection. We wanted to develop a guide for archivists to identify and think about the risks to make decisions of what they can do with their own collections.

As a collaborative project, it has been a signal success, exposing around 50,000 individual film and video news stories from between 1960 and 2000. The four institutional partners addressed issues faced by libraries and other institutions with large moving image collections: how to effectively and economically expose large collections for better access, solve rights issues that may effect online access and use, and build awareness and use for educational needs. The project showed that exposing minimal item- level descriptive data increases awareness and discoverability of collections.

Title: An example of exposing large media collections: The Boston Local TV News project
Sadie Roosa
Project Assistant, WGBH Media Library and Archives

Using the Boston TV News Digital Library as an example of a large audiovisual collection, we can explore some of the obstacles and challenges one might face while processing, cataloging, and making any similar collection accessible. We can also use the successes of the project as guidelines for the work needing to be done on other large AV collections.

At the beginning of the project, what we had was essentially four separate AV collections. Each collection was too large for any of the individual holding institutions to handle as part of their everyday workflow. The Boston Public Library (BPL) holds the WHDH Channel Five collection containing more than 23,000 assets, and Northeast Historic Film (NHF) holds the WCVB Film Collection containing more than 15,000 assets. Both of these collections were completely hidden from users, with no electronic inventories or catalogs. Additionally, WGBH holds the Ten O’Clock News collection with nearly 8,000 assets, and Cambridge Community Television (CCTV) holds their own collection of nearly 5,000 assets. Both of these collections were partially cataloged in Filemaker databases but were only available onsite at the holding institution. The collections were not well advertise as a resource to potential researchers on the

collection level, let alone the item level. We decided that the best approach to handling such large collections would be to bring them together, sharing our technical and personnel resources.

With 51,000 assets, 38,000 of which did not have any form of electronic record, we needed to figure out how to make these collections useful. The initial focus of our energy was on the two completely hidden collections. We designed PBCore compliant Filemaker databases to capture and manage the new electronic records. We also took the existing Ten O'Clock News and CCTV Filemaker databases and mapped the existing fields to PBCore. While each collection had particularities in what data was available and relevant, by mapping it all to the same standard, we were able to ensure that these disparate collections could exist as one cohesive collection that could be searched as a whole. Bringing the collections together provided many benefits, including providing more context for each individual collection and allowing for easier comparison between different styles and eras of news reporting.

Rather than beginning by fully cataloging each item in the two hidden collections, which would take considerable time and require the use of playback devices that were not easily available, we created item level records based on what was available to us without any playback. Project assistants and library science interns transcribed a set of index cards that had served as a basic catalog for the WHDH collection. NHF's archivist created item level electronic records for each film reel, recording the can labels and the slugs written on the leaders of the films themselves. By the end of the transcription process, many of the electronic records had dates, and some had significant keywords and names of places and people, but they were remarkably sparse overall.

While we knew the data was far from perfect, and we expected many hours of grooming and normalization ahead, we decided not to wait to publish the records online. There is no clearly defined standard for what data is "good enough" for users to be able to search. We wanted to push the limits and see if the barest item level records, essentially just a date and a keyword or two, a full sentence if we were lucky, would be enough to catch anybody's interest. We discovered that for some users this really was enough. If they knew exactly what they were looking for, they could narrow down our collection based on date and then scan the sparse descriptions for the story they needed.

As not all researchers could conduct such narrow searches, we continued to improve the records by adding subject, name, and location headings that we authorized in a controlled vocabulary consisting of many Library of Congress subject headings and name authority files, as well as many locally authorized headings. At this point we still had not played back any of the items, so all of this cataloging was being done based on the sparse descriptions available by just looking at the index cards, film cans, tape cases, and other labels. Because we were not viewing or listening to any of the items yet, we were able to move through this minimal cataloging at a rapid pace. Using this method we were able to make all of the records slightly more searchable much more quickly than we would ever have been able to fully catalog even one of the individual collections, let alone the combined collection of all four. As we improved the records, we refreshed the data on the website, making it more searchable and user-friendly, not only because of the normalized names and topics, but because the website links the headings, which allows users to go from a record with a specific heading to a list of all the records with that heading.

As we worked on this minimal cataloging, we also had the project staff and interns pick out especially interesting sounding news stories that they came across. Based on the date and sparse description, they would research newspaper databases and other resources to determine what the news story was likely about. They would then write up a post for the project blog. By featuring these more in-depth looks at our content on the website, we were able to give the richer descriptions for at least a few items, as well as make our collection more visible and attractive to potential users. Users would find our blog posts in search engine results, read the posts containing richer descriptions, and then go on to search all of the sparser records for additional content of interest.

We decided that because of the fragile state of many of the items, especially the $\frac{3}{4}$ inch videotapes, and the scarcity of easily accessible playback equipment, especially for the BPL's film, we would wait to view and more fully catalog any of the content until it had been digitized. The risk we took in choosing to digitize before viewing was that we might end up digitizing content that had been mislabeled or was actually less significant than the labels indicated. However, we decided that issues with playing the analog formats and the amount of time it would take to view items before digitizing were enough of a deterrent. We also did not feel comfortable making the blanket judgment that a user would find certain content more valuable than other content.

It was clear from the beginning of the project that we would never secure the funding and other resources to digitize the entirety of any of these collections. Rather than deciding internally which items would be prioritized for digitization, we conducted an experiment to see if we could get enough public attention to have our users decide the priority through a process of user voting at the item level. With only the sparse descriptions and the blog posts highlighting a few of the items, we had a total of 1184 votes from the public, spread over 825 items. Based on these votes, and some input from our advisors, we began digitizing the prioritized items.

Once we had the items digitized, it was much easier to view them for cataloging. Project staff and interns viewed the digitized items, wrote short descriptions, and added additional subject, name, and location authority headings. Because we were viewing the electronic copies of the items, we were also able to manipulate the content to make cataloging easier. For example, we played content at two or four times the regular speed, so that speech was still understandable, but the viewing process took half or a quarter of the time. We refreshed the website with these richer descriptions and additional headings, and we added the digitized media to the records, so that users could stream the videos in the site.

When the project ended, we had 1,593 items digitized and streaming on the site, with enhanced descriptions and headings. The other 49,000 records still only have the sparse descriptions and minimal headings, but they are available to search along with the records for the digitized content. We implemented technology that allows users viewing the records to request digitization on an item by item basis. Without additional grant funding to support more digitization, we have to require the users to cover the costs. Overall we are pleased that all of the records, not just the ones we were able to view and fully catalog, are searchable by users, and available to view in some manner.

Title: Accelerating Exposure of Audio Visual Collections, What's Next?

Confessions of an Accidental Archivist

Jack Brighton

Director of New Media & Innovation, Illinois Public Media |College of Media

“My main theme is the extension of the nervous system in the electric age...”

Marshall McLuhan - *Letters of Marshall McLuhan* (1987), p. 300

The ability to record and play back moving images and sound has been part of the human technology repertoire for barely more than 100 years. Electricity wasn't commonplace anywhere until the turn of the twentieth century. The combination of electricity and media technology has brought an acceleration of change in our capacity to see and hear, paradoxically both extending and annihilating distance and time. It has also allowed us to create new forms of time-based narrative arts, and increasingly intervenes in every economic, political, and social interaction.

And most of this happened before the Internet.

If “all media are extensions of some human faculty,” as McLuhan has it, then archival media are an extension of our memory and imagination. For this to actually mean anything, preservation of moving images and sound is necessary but not sufficient: its value in human culture depends on ready access.

I began my professional life in public broadcasting as a journalist and creator of electronic media. It is toward the realization of “persistent access” to the vast store of our multimedia memory that my career has increasingly become focused.

Fortunately we have no resources

Exposure to archival practice came from my struggles as a media producer in the emerging digital age. I began designing websites with streaming and downloadable multimedia in 1997, and quickly realized that without an archival plan the situation was becoming hopeless. I saw how quickly technology was changing, and suspected that the media we published on the web at that time would be unplayable within a few years. I placed bets all over the board by publishing RealMedia, QuickTime, Windows Media, and Flash but I assumed all these formats would sooner or later be obsolete. I guessed wrong how quickly that would happen. But I got one or two things right: I saved the original physical formats, and high-resolution digital files derived from the masters. And I began to maintain a database of everything I saved.

Fortunately nobody saw me doing this. As a lowly producer, I had no budget. Therefore, no one could cut my budget or tell me to stop. I was the guy running the station website in his “spare time,” and no one complained so long as I did the magic. But I soon noticed something truly disturbing: the problem of persistent media was growing more difficult, more complex, and just plain larger every day. And the rate of change was accelerating.

And that's not counting our shelves and desk drawers everywhere filled with analog media in fragile formats with barely a peeling label to tell us what they are.

Toward content management and a life of data

You build a website with content and metadata. The presentation is just the access layer. A web page is accessed by humans by way of a browser that parses the html and presents a view we can understand and use. The same web page is accessed by machines (like web crawlers and other applications) by virtue of well-structured data they can parse and reuse. As I began to understand these things, I learned from my colleagues at other public media organizations that many of us were becoming deeply invested in technical solutions that could manage this data, and present it to humans and machines, at the cost of a very large contract and ongoing vendor dependency.

The solution I pursued seemed half-baked by comparison with these expensive, proprietary systems: a simple MySQL database tied to a file system. With this as a data core, we can build applications that speak the language of the Internet. We can present web pages to humans; RSS/podcast feeds to iTunes; Dublin Core and PBCore records to library and archival services. The most difficult challenge is getting good data to begin with. So that's what I focused on, and sought external funding for with some success.

Today we openly share well-structured data with a growing number of public institutions and archival services, including the American Archive, the Pop Up Archive, Collective Access, and the Internet Archive. We provide open access to high-resolution digital media files and all the metadata we can capture and share. In some cases our partners add to and enhance the existing metadata. For example, the Pop Up Archive provides a speech-to-text transcription service that feeds back into our core data. Most of this happens automatically using PBCore as an exchange format between systems.

The way the automation works is as follows:

- Content producers create media files, such as radio news stories, video productions, interview segments, etc. Most of these assets are finished products for broadcast or online distribution. They also might be products of oral history interviews, youth media projects, and special collections based on election cycles and other topic-based productions.
- Producers export media files to a watch folder, from which scripts route the files to software that encodes them to web-friendly formats. Source media files are then moved to network storage where they become subject to our preservation protocols.
- Producers upload the web-friendly media files to our website content management system, which stores the files in an NFS mount addressable by http. While uploading these files, producers add a variety of (mandatory) descriptive metadata. The location of the files and all other metadata then resides in our MySQL database.
- Our website content management system provides an html front end for public access to all uploaded media files and associated metadata. The assets are available for browsing and search on the Illinois Public Media website, and exposed to social media services via Open Graph and other metadata formats.

- The website CMS also provides full access to all content and metadata in the form of PBCore records using the pbcoreCollection element for exchange of multiple records as a “feed.”
- Services like the American Archive and the Pop Up Archive consume Illinois Public Media’s PBCore feed for automated ingestion and processing of the media essence files and associated metadata.
- Where external services add to our existing metadata, e.g. text transcripts and SRT files from the Pop Up Archive, the enhanced metadata is added to the MySQL store of data at Illinois Public Media where it is used for search optimization, video captioning, and progressive enhancement of our website user experience.

I would emphasize one key feature of this system: It requires no additional effort on the part of producers beyond publishing media content on the Illinois Public Media website. The system itself could be described as a system of shareable metadata and media files. It just happens to also produce a website.

At Illinois Public Media we don’t aspire to fully solve the archival problem. We simply try to be an awesome version of the best data source ever. I think it’s important to keep role and scope in mind as more public media and cultural heritage institutions take on the challenge of preserving and creating access to their media collections. Illinois Public Media will serve as the source and authoritative voice for our collections, and if we can’t stand up our own trusted repository it’s within our mission to share it with other public institutions that can.

If you don’t know what you have, you don’t really have it

The complexities of persistent access in the age of rapid change in media technology can seem overwhelming for public media and other cultural heritage institutions. The situation is more dire with the deterioration of our legacy analog and physical media collections. We are racing against both time and scope, and for some significant portion of our audiovisual heritage, we will lose this race.

If we have any remaining arguments about which system or tool is best for managing media content and metadata, we can enjoy that conversation over drinks. The most important thing is to get the data about what we have, and put it in structured form so it can be accessed and shared. Media census projects, like those recently completed at Indiana University and the University of Illinois at Urbana-Champaign, are needed to begin answering the question of what there is to preserve. We can then marshal that data to prioritize the work of preservation and “persistent access.”

Persistence of Vision

“The written symbol extends infinitely, as regards time and space, the range within which one mind can communicate with another.”

Samuel Butler, *Life and Habit*, London: Trubner & Co, 1978

We live at a time when all previous forms of media are potentially hyperlinked and accessible. We added moving images and sound to the written symbol as means to extend perception,

communication, knowledge, and imagination. It may seem lofty to claim that our media archives are to culture what memory is to the human brain. But to the extent that we have not yet embraced caring for these as means of extending our senses, our vision and reach are impaired.

The work still to be done is far from trivial, but my experience has been that it begins by focusing on the data. The technical means of handling data will change, but we can make the data accessible as systems and standards evolve. Small institutions like Illinois Public Media can't do everything required for preservation, but we can serve as a node in a larger preservation ecosystem. In terms of preserving an aggregate of potentially all institutional media collections, resources can be allocated at different levels depending on roles and scope. The thing that has brought such rapid change, the Internet, is also the means of connecting the levels.

We may have limited resources, but fortunately we have an architecture of collaboration we're just now beginning to understand.

Title: Hidden Collections at Northeast Historic Film

Brian Graney

Archivist, Black Film Center/Archive, Indiana University

Northeast Historic Film, a regional moving image archive based in Bucksport, Maine, was among WGBH's partners on the Boston Local TV News project, and was at the same time exploring related questions regarding the exposure of large audiovisual collections over the course of two sequential projects funded in 2009 and 2010 through CLIR's Cataloging Hidden Special Collections and Archives program: Intellectual Access to Images of Work Life, 1916-1960; and, Moving Images 1938-1940: Amateur Filmmakers Record the New York World's Fair and Its Period. Both projects were directed by Karan Sheldon, a co-founder of Northeast Historic Film.

The 2009 project, Work Life, described 50 unique hidden collections of nearly 800 amateur and industrial films documenting work and labor in the first half of the 20th century “to help advance discovery and inclusion of moving images as primary source materials in the universe of traditional library and archives collections.”

This mission had roots at NHF in cataloging efforts extending back to its 1995 Collections Guide, which made use of descriptive standards Archives, Personal Papers, and Manuscripts (APPM) and Archival Moving Image Materials (AMIM). More recently, the 2007 project, Finding and Using Moving Images Online, funded as a National Endowment for the Humanities Digital Humanities Start-up Grant, demonstrated the value of treating films with integrity as primary source materials for the humanities by selecting and presenting archival moving images explicitly as unedited texts within a hierarchical structure rooted in provenance.

With the Work Life project, NHF continued this approach by formally adopting an efficient metadata model integrating provenance-based archival description of moving image collections with descriptive and technical metadata at the item level, based on the widely-supported standards Describing Archives: A Content Standard (DACS) and Encoded Archival Description (EAD), along with the then-emerging audiovisual metadata standard, PBCore 1.3. NHF also worked with database developers to implement a custom profile of the open source CollectiveAccess platform, designed for web-enabled production and presentation of finding aids with embedded images and digital video; search and navigation across collections and items; and export of XML-based EAD and PBCore records. Elements of the open source profile developed for this project have since served in CollectiveAccess implementations at large and small institutions, including the Academy Film Archive and the Chicago Film Archives.

NHF's 2010 CLIR project set out to describe hidden collections of amateur film across three institutions, including NHF, the George Eastman House International Museum of Photography and Film, and the Queens Museum of Art. The emphasis on amateur film shot at the 1939-1940 New York World's Fair aimed to establish the subject within a broader research context of amateur filmmaking of the period. There was at the same time a strategic emphasis on promoting wider implementation of the descriptive practices and tools developed through the Work Life project and dissemination of related resources, tools, and strategies, primarily through a series of cataloging workshops held over the project year in Rochester, NY, with the L. Jeffrey Selznick School of Film Preservation; in Boston, MA, with the Northeast Document Conservation Center and the Simmons College School of Library and Information Science; and in Austin, TX, with the Association of Moving Image Archivists.

There was throughout both projects an underlying effort to interrogate the practices and priorities at the archive in order to establish efficient workflows in areas of support for project cataloging – such as inspection, conservation, and reformatting – and in cataloging itself. In evaluating this, we returned regularly to discussions of what might constitute efficient processing for moving images – particularly for home movies and amateur films.

In the case of the World's Fair project, we resolved this question situationally within the thematic framework of the project, which privileged above all films of the 1939-1940 New York World's Fair, examined them alongside other amateur films of the 1938-1940 period, and established all of the films as individual items within the context of their individual collections, which, on the whole, were broader in their subjects and dates of creation.

As in the earlier Work Life and Moving Images Online projects, DACS-based collection level records were created across the board, meeting the standard's single-level minimum element requirements and enhancing records with some additional elements recommended for optimal description, including administrative and biographical histories of the creators, and geographic,

subject, and name-based access points. Each finding aid also included a full collection inventory, corresponding to the individual PBCore records, based on minimal Intellectual Content and Instantiation element sets adopted for the project.

Within the PBCore Intellectual Content class, we determined the level of description to employ by considering the relevance of each item to the project themes. New York World's Fair films received highly detailed shot-level descriptions; 1938-1940 films received rich item-level descriptions; and other reels within the project collections which fell beyond the scope of its principal themes received minimal item-level descriptions, often based solely on notes drawn from the film cans, leaders, or inventories in the collection files. Among this last group, we made exceptions to the minimal approach for select films that were discovered to be of particular significance and potentially high user interest. Two examples of these are the mountaineering films of Charles and Oscar Houston, including film of the landmark first ascent of Alaska's Mount Foraker in 1934; and 1927 amateur footage shot by Simmons Brown of Charles Lindbergh and "The Spirit of St. Louis" at Orchard Beach, Maine.

PBCore Instantiation class elements, associated with each Intellectual Content record, were populated for all original film items and for video transfers derived from them. In the case of the thematically-prioritized films, the archive's technical staff provided highly detailed inspection reports, including information about the stock types, date codes, editing structure, and distinctive aperture marks which can help to determine the camera used to shoot the film. While far from minimal in practice, this strategy was directly responsive to stated user interest in aspects of the material film artifacts not otherwise captured during digital surrogation or content description.

We configured the CollectiveAccess database modules in our profile to correspond to our metadata schema, and provided mapping schemas to the developer for generating valid EAD and PBCore XML documents for harvesting from the CollectiveAccess records.

Finally, as an important supplement to the metadata strategy, we embedded visual and audiovisual digital media in the records at different levels, again according to the thematic priorities established above. Collection-level displays featured single frame enlargements and brief streaming video excerpts considered to be representative of the collection; 1938-1940 item record displays were each illustrated with a frame enlargement, which appeared as a thumbnail in item-level search results; and the item record displays for New York World's Fair films featured multiple frame enlargements and full reel transfers in streaming video. While this fell outside the parameters of the cataloging project, our rationale here was that, while researchers could expect a reasonable level of recall on their searches based solely on textual descriptions and access points at the collection and item levels, their subsequent determinations of relevance – particularly for researchers working with home movies and amateur films – were often most easily narrowed by primarily visual means.