

Cataloging and Preservation of Moving Images: A Survey of Organizations and Initiatives

Helen Sam

Helen Sam completed her MLIS at the San Jose State University School of Library and Information Science in 2009. She has worked in the serials section of the library at UC Santa Cruz. Before pursuing her MLIS degree, she worked in the film industry and became interested in digital asset management, digital archiving, and preservation, and she is pursuing a career as a media or film librarian. She can be reached at: gradschool_info@yahoo.com

Introduction

There is a large body of work and research being done in the field of cataloging and preservation of moving image (film and video) content. Much of the work is being spearheaded by organizations such as the National Film Preservation Foundation, Moving Image Collections (MIC), and the Corporation for Public Broadcasting (CPB). Many organizations such as WGBH/Boston, The Pacific Film Archive, and CNN are already digitizing their collections for public/private access, streamlining internal workflows, and participating in cooperative information sharing projects. This is all possible due to advancements of asset management systems and metadata standards being used to organize and retrieve moving image content. This paper offers an overview of the work that is currently being done in the area of metadata creation and implementation related to cataloging moving image content. Because the field and organizations involved with moving images is so vast and diverse, I have limited my survey to organizations developing metadata specifically for moving images. The paper begins by offering a brief background of moving image cataloging and preservation. It then highlights three major organizations (MIC, CPB, and MPEG) involved in creating metadata for moving images, problems and suggestions encountered in the field.

Since its inception in the late 19th century, moving images have played an important role in shaping our society through entertainment, education, and communication. Like books and photographs, moving images are also important cultural artifacts that warrant the need for adequate cataloging, archiving, and preservation. In 1994, the Library of Congress in consultation with the National Film Preservation Board, created a national film preservation plan which recommended:

- Redefining film preservation to include (1) low-temperature, low-humidity storage to retard film deterioration, (2) the copying of decaying film onto new, more stable film stock, and (3) use of video and other access technologies to share newly duplicated films with the public.

- Increasing the availability of films for education and exhibition.
- Developing cooperative projects to advance national preservation goals.
- Creating a new federally chartered foundation to help public and nonprofit organizations preserve American orphan films and share them with the public. (National Film Preservation Foundation, 2009)

The national film preservation plan not only focuses on the preservation and access of physical films, it is also engaged in digital preservation and development of cooperative projects (digital libraries and catalogs) that will give users access to shared moving image collections for research and education use. The increase in moving image content in digital format whether they are born digital or being digitized from its original analog format has created a need for systems and tools to help manage the content for current and long-term access. The advent of computer technology and networks has allowed many organizations and companies the ability to catalog their collections into online databases and easily share information without having to physically retrieve tapes from the library. While these technologies have created efficient asset management systems for digital content, they are only as useful as the information contained in it. Thus there is a need for quality metadata to support efficient information retrieval. According to NISO (2004), "metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource." For example, traditional cataloging using MARC 21 and AACR2 standards are a form of metadata creation. In the digital environment, metadata has become extremely important in facilitating the retrieval of information resources. If people cannot locate an item in a collection via searching an online catalog, then the catalog and the collection are useless.

Survey of Organizations

My awareness of moving image collections and metadata comes from working in the field of film and television. I worked at the [Independent Television Service](#) (ITVS) in San Francisco, where I was responsible for maintaining the tape library records using a FileMaker Pro database. I also had the opportunity to intern at the CNN Library in Los Angeles performing duties such as providing research assistance and creating records within the library database (MIRA). While working in the film and television field, I became aware of a shift in how film and video content were being managed and cataloged. Organizations that produced moving image content were implementing asset management systems. These organizations have been engaged in efforts of digitizing their film and tape libraries and implementing asset management systems as a way to efficiently manage assets and streamline workflow methods. The efforts of digitizing moving images into an asset management system also brought about the need for adequate description of these assets for retrieval and resource sharing amongst different organizations and systems. While working at ITVS and CNN, I noticed content description and guidelines were not standardized or enforced within both library databases. There was also no form of quality control to determine whether the information entered into the database fields were correct or adhered to a particular metadata or controlled vocabulary standard. Information entered into the fields varied from one record to the other because there was no guide explaining how information is to be entered and what type of information the field was asking for. Inconsistent record information caused confusion amongst staff in terms of not being able to locate footage within the library or the inability of the database to produce accurate reports because record information was inconsistent.

These inconsistencies made me aware of the need for libraries to use and build upon established metadata standards rather than create new, localized standards. It also prompted me to seek out organizations that were engaged in creating metadata standards for moving image description and resource sharing. From my survey of the field, I discovered that the major organizations involved in creating metadata for moving images included the Moving Image Collections, The Corporation for Public Broadcasting, and the Motion Picture Experts Group.

A major project involved in the creation of a moving image metadata schema is the Moving Image Collections (MIC), cosponsored by the Library of Congress and the Association of Moving Image Archivists (AMIA). The MIC was created in response to the 1994 national film preservation plan published by the Library of Congress in consultation with the National Film Preservation Board. The MIC:

- Documents moving image collections around the world through a catalog of titles and directory of repositories, providing a window to the world's moving image collections for discovery, access and preservation.
- Provides a technology base and informational resources to support research, collaboration, preservation, and education for archivists, exhibitors, educators, and the general public.
- Is a portal for integrating moving images into 21st Century education.
- Is a key access program of the Library of Congress' National Audio Visual Conservation Center. (MIC, 2009)

The MIC offers information resources for organizations and individuals interested in areas of cataloging, preservation, and outreach related to their moving image collections. Two main features of the MIC are the online catalogs, the Union Catalog and the Archive Directory, which are freely accessible through MIC's homepage: <http://mic.loc.gov>. The Union Catalog brings together catalog records for individual moving images collected and managed by individual organizations from all over the United States. Users can search for moving image records across multiple collections. Union Catalog records include information such as title, date, format, and subject information about a moving image. Organizations that contribute records to the Union Catalog include the CNN Library, MBRS, National Geographic Television and Film Library, and the Smithsonian Institution. The Archive Directory lists organizations and individuals involved in collecting moving images. Users can locate archive contact information or the type of services and collections within the archive. These catalogs allow users to seek out information and collaborate on describing and maintaining moving image resources without having to duplicate unnecessary cataloging work. The catalogs are two different metadata schemas. The Union Catalog is based on the MPEG-7 schema with 49 core elements and the Archive Directory schema was developed by the MIC, which has 99 core elements. The Union Catalog is able to import and export catalog records in MARC, MPEG-7, MODS, and Dublin Core formats (MIC, 2009). Organizations can import their own records in several different formats, which are then mapped according to the MIC Core Registry schema for inclusion into the Union Catalog. The MIC is continually working on educating and improving access to moving image collections. As of March 2008, there are 558,489 records listed in the Union Catalog and 250 archives listed in the Archive Directory.

Another organization engaged in creating a metadata standard for moving image content is the Corporation for Public Broadcasting (CPB). CPB is a private, non-profit created by Congress in 1967 to promote public telecommunications services (television, radio, and online) for the American people (CPB, 2009). While the CPB is not a government agency, it is funded by taxpayer money. The main goals and objectives of the CPB are to:

- Promote an educated and informed civil society through significant, high-quality content and services.
- Increase awareness of and appreciation for the essential contribution that public media makes to civil society.
- Recognizing the transformational change taking place between media and audience, foster innovation in public media by supporting projects that advance creative or resourceful ideas for improving content, service, diversity, and audience reach, including projects that employ collaboration as a tool for innovation. Help increase the resources available to public media. (CPB, 2009)

Not only is the CPB supporting the creation of diverse and innovative programming for public broadcasting, it is also engaged in the research and creation of standards to meet the information sharing needs of the public broadcasting community (PBS, NPR, PRI, local stations, and other partner organizations). As public broadcasting entities begin to acquire and implement asset management systems to organize their content, they must also implement a standard metadata schema. This has led to the development of the Public Broadcasting Metadata Dictionary (PBCore) funded by the CPB and administered by WGBH/Boston. PBCore is designed to provide public television, radio, and web activities a standard way of describing and using data that allows content to be more easily shared and retrieved among different user groups, systems, and organizations (PBCore, 2009). PBCore is based on the Dublin Core schema and currently contains 53 elements arranged in 15 containers and 3 sub-containers; all organized under 4 content classes (PBCore, 2009). PBCore is currently being used by local public television stations such as WGBH in Boston and WNET in New York. According to a recent announcement on PBCore's homepage, a new version of PBCore v1.2 is currently being developed (PBCore, 2009).

The final group in my survey engaged in creating a metadata schema for moving image content is the Motion Picture Experts Group (MPEG). Established in 1988, MPEG is a working group of ISO/IEC in charge of developing standards for coded representation of digital audio and video content (MPEG, 2009). Some of MPEG's mandates include the:

- Development of international standards for schemes that declare and describe digital items, multimedia content data structures and related information, enabling creation, exchange, distribution, transaction, storage, search, retrieval, browsing and filtering of digital items and multimedia content.
- Development of international standards for coded representation of moving pictures, visual information and associated metadata. The evaluation of coding techniques and description schemes is performed based on their performance (both objectively and by formal subjective testing), efficiency

with respect to software implementation, VLSI (programmable & dedicated) implementation and feasibility of systems architectures. (MPEG, 2009)

The MPEG has been credited for creating the MPEG-2 compression standard for the transmission of audiovisual content for digital broadcast television. Another standard is MPEG-7, which was specifically designed to describe, manage, and provide access to moving images in digital format (Agnew, Kniesner, and Weber, 2007). MPEG-7 is currently being used as a metadata schema in programs (Richoh MovieTool and IBM MPEG-7 Annotation Tool) designed to assist in the creation of descriptions for audiovisual content (Smith, 2002). Organizations such as the MIC are also working to integrate MPEG-7 standards within their Union Catalog.

Issues and Recommendations

Through my survey, I noticed that there is no one size fits all metadata standard for describing moving image content. The organizations I surveyed have created their own metadata schema to address the specific needs of the organization and community they fund or support. While some level of local customization is necessary, I worry that the long-term effects of creating multiple specialized metadata schemas will create an inability for archives to cooperatively share and retrieve information. A multitude of metadata schemas can also be overwhelming to individuals and organizations deciding on a viable schema to adopt or adapt to that will ensure future migration and retrieval of information. To alleviate confusion and ambiguities between metadata schemas, it is best to create a crosswalk or metadata map showing the relationships, equivalencies, and gaps between different schemas and their elements. An example of a crosswalk is the Metadata Standards Crosswalk created by the Getty Research Institute. Crosswalks not only support semantic interoperability, they are also instrumental for converting data from one format to another (Woodley, 2008). It would be helpful if a publicly available crosswalk similar to the Getty's Metadata Standards Crosswalk was designed specifically for metadata schemas related to moving images existed.

As more organizations implement asset management systems, guidelines and standards must be set up to evaluate the quality of the metadata being used to catalog moving image content. Quality and consistent metadata across collections can be implemented with the use of a common controlled vocabulary or thesauri. Not only is it important to test whether the metadata schema and its elements are correct and appropriate, it is also necessary to test the functionality of how well the schema can find, identify, select, and obtain information from a search query based on information entered for each field. A study done by Zhang and Li (2008) tested the usefulness of the individual metadata fields within the MIC Union Catalog and Archive Directory based on the four "generic tasks" based on the International Association of Library Associations' Functional Requirement for Bibliographic Records (IFLA FRBR) as the framework for the study (Zhang and Li, 2008). The four tasks were:

- *Find* bibliographic records that correspond to stated search criteria on a topic.
- *Identify* potentially relevant records from a retrieved set through interpreting the information in each retrieved record.
- *Select* records that the users would like to get the corresponding physical items by comparing the information in multiple records.

- *Obtain* the selected physical items from an organization/ archive based on the information provided in the metadata records.

The results of the study determined that the IFLA FRBR framework could be applied to other user-centered functional metadata evaluations. This example suggests that more user-centered studies need to be done to test the functionality of metadata information. By using standards such as controlled vocabularies and gathering user feedback, the organization can better improve the quality of metadata being used and ensure efficient retrieval of information.

Conclusion

This survey brings insight into the work being done in cataloging and preservation of moving image content. While efforts to create and implement standards have been in place for print and art materials, the same should be taking place for moving images. In order to ensure long-term access for moving images, organizations should continue to work together to develop cooperative standards such as metadata schemas, controlled vocabularies, and user feedback methods within the field.

The survey brings to light many areas of further research. These topics include:

- Can the metadata schemas highlighted in the survey be used to create a general metadata crosswalk standard for moving image cataloging? Some work on mapping metadata schemas for moving images has already been done.
- How many organizations are creating metadata schemas or cataloging moving images based on already established schemas (Dublin Core, PBCore, or MPEG-7).
- Determining whether the PBCore schema can be mapped and used by nonpublic broadcasting organizations as a standard for cataloging moving image content.

References

Agnew, G., Kniesner, D., & Weber, M.B. (2007). Integrating MPEG-7 into the Moving image Collections portal. *Journal of the American Society for Information Science and Technology*, 58, 1357-1363.

Baca, M. (2008). Introduction to metadata version 3.0. Crosswalks, metadata harvesting, federated searching, metasearching: Using metadata to connect users and information. Available: http://www.getty.edu/research/conducting_research/standards/intrometadata/path.html

CPB (Corporation for Public Broadcasting). (n.d.). *Goals and objectives*. Available: http://www.cpb.org/aboutcpb/goals/goalsandobjectives/goalsandobjectives_full.html

CPB (Corporation for Public Broadcasting). (n.d.). *What is the Corporation for Public Broadcasting?* Available: <http://www.cpb.org/aboutcpb/whatis.html>

MPEG (Motion Picture Experts Group). (n.d.). *The MPEG home page*. Available: <http://www.chiariglione.org/mpeg>

MPEG (Motion Picture Experts Group). (n.d.). *Terms of Reference*. Available: http://www.chiariglione.org/mpeg/terms_of_reference.htm

MIC (Moving Image Collections). (2006). *About the Union Catalog*. Available: http://mic.loc.gov/public_portal/pub_unicatlg.htm

MIC (Moving Image Collections). (2005). *About us*. Available: http://mic.loc.gov/about_1.htm

National Film Preservation Foundation. (n.d.). *Why the NFPF was created*. Available: <http://www.filmpreservation.org>

NISO (National Information Standards Organization). (2004). *Understanding metadata*. Available: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

PBCore (PBCore Public Broadcasting Metadata Dictionary Project). (2005). *About the PBCore metadata dictionary elements*. Available: <http://www.pbcore.org/PBCore/index.html>

PBCore (PBCore Public Broadcasting Metadata Dictionary Project). (2005). *Background of the PBCore Public Broadcasting Metadata Dictionary Project*. Available: http://www.pbcore.org/PBCore/PBCore_background.html

PBCore (PBCore Public Broadcasting Metadata Dictionary Project) (December, 2008). *Announcements and news about the PBCore Public Broadcasting Metadata Dictionary: PBCore v1.2 is on the Horizon!!!* Available: <http://www.pbcore.org/announcements.html#pbcore1-2soon>

Smith, J.R. (2002). *MPEG-7 Industry Use v.2.0* [Electronic version]. International Organization for Standardization/ Organization Internationale Normalisation ISO/IEC JTC 1/SC 29/WG 11/ Coding of Moving Pictures and Audio, 3-4.

Zhang, Y., & Li. Y. (2008). A User-Centered Functional Metadata Evaluation of Moving Image Collections. *Journal of the American Society for Information Science and Technology* 59: 1331-1346.