

---

# MIC metadata strategies: Thinking beyond asset management

## Jane D. Johnson

is MIC Project Manager for the Library of Congress, and has worked on the initiative since its inception. Previously, Jane was a UCLA film and television archive cataloger for nearly two decades. She serves on the AMIA Board of Directors and has given numerous presentations on cataloging principles, media asset management and MIC. She extends her gratitude to MIC architect Grace Agnew, whose vision and clear articulation of key concepts inform this writing.

**Keywords:** *moving images, metadata, standards, interoperability, collaboration, preservation*

**Abstract** This paper examines the metadata strategies of Moving Image Collections (MIC), a collaboration of the Library of Congress and the Association of Moving Image Archivists (AMIA). MIC offers a union catalog, archive directory and informational resources about moving images and their preservation, through a portal structure delivering customized information to diverse constituencies. The paper discusses three key principles underlying MIC's strategy and how they have been applied to meet the respective missions of the two co-sponsors. The key principles are promotion of standards, acceptance of diversity in collections and schemas, and the need to extend effective metadata use to all repositories. This visionary metadata strategy takes MIC's functionalities beyond asset management. MIC creates a framework for collaborative community building and an R&D platform to explore issues of digital rights management, low-level indexing and the intersection of public and private sector goals, in an effort to advance the Library of Congress mission to preserve and educate.

## INTRODUCTION

*Shattered Glass. Control Room. Outfoxed.* Three provocative films, all recently released within a span of a few short months, and all pointing to ethical issues in news coverage. As a profession and a discipline, journalism is broadly acknowledged to be in a state of crisis. Traditional outlets for journalism products, such as newspapers and magazines, are experiencing shrinking audiences even as standards for ethics and practice are eroding. Purchases and mergers within the industry have arbitrarily changed the scope, focus and

acceptable practices of many news resources in ways that are often subtle or hidden from the consumer. A greater understanding of convergences and forces that are shaping television news requires the ability to compare practice and performance across many news sources and over many time periods. But where is the news?

There is little access to yesterday's news programming, and news is not alone. A dearth of online access to moving images of all kinds threatens to shrink the public's knowledge base and diminish collective intelligence at this

Jane D. Johnson  
MIC Project Manager,  
Library of Congress  
Visiting Scholar, Rutgers,  
The State University of  
New Jersey,  
Library Technical  
Services Building,  
47 Davidson Road  
Piscataway,  
NJ 08854-5603, USA.  
Tel: +1 732 445 5904  
Fax: +1 732 445 5888  
Email: [jjohnson@loc.gov](mailto:jjohnson@loc.gov)

critical juncture when nations vie for leadership and competitive edge in an increasingly global labor market. Moving images of all types remain largely outside of the education mainstream. Experts allude to media's far-reaching influence, yet most of the world's media output remains largely unavailable for study to students, scholars and lifelong learners.

Beginning in 1994, the Library of Congress published two documents detailing a different crisis: the crisis in film preservation. *Redefining Film Preservation* (1994),<sup>1</sup> mandated by the US Congress as part of the National Film Preservation Act of 1992, was the first national moving image preservation plan. Three years later it was complemented by a follow-up plan entitled *Television and Video Preservation* (1997).<sup>2</sup>

Somewhere between the Library's search to save deteriorating film and a world with instant access to its digital assets, Moving Image Collections (MICs) emerged. Pronounced "Mike," MIC is a collaboration between the Library of Congress and the Association of Moving Image Archivists (AMIA).<sup>3</sup> MIC offers a union catalog of moving images, an international directory of archives and informational resources about moving images and their preservation. MIC's mission is to immerse moving images into the education mainstream, recognizing that what society uses, it values, and what it values it preserves.

MIC began when the Library of Congress turned to AMIA for assistance in developing strategies for implementing the numerous recommendations included in the two preservation plans. AMIA identified the first and most crucial step in any

preservation solution as a standardized way to identify holdings, particularly unique titles, so that strategic planning and collaborative decision-making could occur.<sup>4</sup> MIC brings together in one place for discovery the wealth of the world's moving images in all its variety of genres, forms, subject areas and formats, from a number of organizations of vastly divergent types, sizes, missions, financial resources and metadata standards. For the first time students, educators, researchers, archivists and the general public can now search across multiple repositories to find current, detailed descriptions of moving images, and the images themselves.

MIC's innovative design employs a flexible, metadata-driven strategy to simultaneously address multiple goals of expanding education, outreach, access, preservation and research in culture and information technology. A look at this strategy and a review of metadata for moving images can help readers understand the metadata landscape for these materials, evaluate and select metadata schemas for their collections, and discover how the MIC model might be used or modified for their own digital assets.

Three key principles drive MIC's approach to asset management:

- to promote established cataloging principles and standards-based metadata use;
- to embrace the inherent diversity in repository types, missions, holdings, financial resources, metadata schemas and content standards; and
- to extend metadata for digital collection description, management and access to any moving image archive, regardless of

technical readiness or cataloging and archiving expertise.

### **PRINCIPLE 1: PROMOTE STANDARDS**

Following standards reaps many benefits. Basic principles of cataloging, as reflected in standards, have existed and evolved for over a hundred years. Use of standards allows record sharing and record migration. Even organizations with no mission or intent to exchange their records benefit from standards when the time comes to migrate to a new or upgraded system. When specific needs demand development of new standards or best practices, existing standards can provide a reality check. Certainly there is a greater choice of off-the-shelf tools if one conforms to standards, and field-wide use of standards creates a critical mass of customers, the profit motive, for system developers. The emergence of better, more powerful systems is dependent on consistency of practice. Finally, organizations seeking grant funding will achieve far greater success if standards are followed; rarely will grant monies fall to those reinventing the wheel.

MIC promotes standards in a number of ways. First, it illustrates by example the benefits of best practices, providing a working tool that accommodates a variety of schema, demonstrating the ease and value of standards through practical application. As MIC grows in size and stature (the current record count stands at about 350,000), it will facilitate research and development in emerging technologies by making available a sizeable set of bibliographic records representing a cross section of the moving image community. Computer

science researchers can partner with this community to explore issues such as digital rights management, active privacy policies and fair use, or low-level indexing applications.

Secondly, MIC is committed to the use of standards-based interoperability protocols to provide an integrated and easy-to-use portal for digital educational resources. MIC's union catalog incorporates both open-source Z39.50 capabilities and support for the Open Archives Initiative (OAI). MIC's architecture includes an LDAP database for the archive directory, a PostgreSQL database for the union catalog, and a Zebra/Yaz XML-based search engine. In short, MIC is usable, flexible, extensible, open source-based and interoperable.

Thirdly, MIC promotes standards by educating asset managers in their use, through a wealth of information resources gathered by AMIA members as part of the organization's educational mission. This innovative partnership between the Library of Congress and a professional association maximizes the strengths of both organizations to make a total contribution to the field beyond the sum of its parts. The Library of Congress will serve as the permanent host site for MIC, providing the technology infrastructure and management staffing to support MIC's long-term maintenance and ongoing development. AMIA, in accordance with its organizational mission, is responsible for developing MIC's educational content. AMIA committees, interest groups, task forces and other voluntary participants develop the site's informational resources through the working groups of the MIC Education and Outreach Committee.

## PRINCIPLE 2: EMBRACE DIVERSITY

MIC's core registry of data elements for moving image description is the key to accommodating the myriad of metadata schemas employed in this diverse field. In the archival moving image world, the most common metadata schemas — excluding in-house schemas — are probably MARC21<sup>5</sup> and Dublin Core.<sup>6</sup> While there are few MPEG-7<sup>7</sup> implementations, this schema shows great promise, particularly for non-textual digital video indexing applications. Other prominent standards include MODS<sup>8</sup>, PBCore<sup>9</sup>, and SMPTE DMS-1.<sup>10</sup>

MARC21 (MAchine Readable Cataloging) is a set of standards developed by the Library of Congress in the 1960s as a way structuring and tagging data to facilitate the interchange of records between databases or files. The cornerstone of the set is the MARC21 Format for Bibliographic Data, which historically was used by libraries and archives primarily for descriptions of textual materials. The format is extremely detailed and includes redundancies attributable to the technical shortcomings of its early development period. With almost 200 fields, most with multiple subfields, MARC is generally considered too complex for the untrained professional. On the other hand, it has been so widely adopted, and is so integrated with other standards, that it lends itself to interoperability, especially with print materials, and for retrieval via Z39.50 in bibliographic utilities and other remote databases.

In the mid-1990s, the explosion of information on the web, and a dearth of tools to organize it for access, called for a cataloging standard that was easier to

use than MARC. MARC cataloging was not only labor-intensive, it generally required highly trained professional cataloging librarians for effective application. The learning curve was steep. Thus was born the Dublin Core metadata schema. Dublin Core consists of 15 data elements likely to be useful across a broad range of disciplines and within any type of organization needing to organize and classify information. Although Dublin Core can theoretically be used to describe any kind of resource, its development was driven by a desire to facilitate discovery of web-based resources. It was not intended for collection management, and is not very accommodating of administrative, preservation, technical and rights metadata. Further, it was not designed to handle multiple manifestations or expressions of a particular work. Dublin Core's use in asset management is therefore limited. On the upside, Dublin Core is easy to use; it facilitates data sharing and allows OAI harvesting.

MPEG-7 is one of the few metadata schemas developed specifically to describe, manage and provide access to moving images. Designed primarily by digital video engineers, it is intended to support a wide range of users including commercial production studios, educational agencies and individual users.<sup>11</sup> Unlike MARC and Dublin Core, MPEG-7 was designed to support multiple manifestations and accommodates technical and administrative metadata. Moreover, MPEG-7 supports the automatic generation of segments for storyboards and summary clips that can be combined into learning objects. By providing schemas that process the digital bitstream for non-textual digital video indexing

applications, it allows retrieval of digital objects by properties like color, shape and texture. This is in contrast to the high-level indexing traditionally employed in libraries, wherein a cataloger manually assigns index terms from a predefined subject heading list or thesaurus.

MIC's goal is to ensure that these metadata schemas, and others to come, could be mapped into — and out of — its core registry for ingest into the MIC union catalog and subsequent export. The registry, a rigorously maintained and standardized application, consists of 48 fields.<sup>12</sup> Data elements from virtually any schema can be mapped, or translated, to and from these 48 fields. Currently MIC's extensible format-independent metadata design accommodates searching, export and display in MARC21, Dublin Core, MPEG-7, and several organizations' own formats. Maps to additional metadata schemas, such as MODS, PBCore and SMPTE DMS-1, are forthcoming.

The MIC core registry is detailed enough to retain the richness of extensive archival descriptions, but simple enough to provide users with readable, succinct displays of heterogeneous metadata from widely divergent sources. The MIC Core Registry supports the core user needs as defined by IFLA's Functional Requirements of Bibliographic Records (FRBR): to *find* entities that correspond to the user's stated search criteria, to *identify* an entity, to *select* an entity appropriate to the user's needs, and to *obtain* a copy of a selected entity.<sup>13</sup>

MIC has now broken new ground by creating a mapping utility that enables any moving image repository, regardless

of metadata schema used, to easily map records itself for sharing globally through the MIC union catalog. The organization simply submits an application, followed by a set of sample records and its own list of data elements (field list). MIC populates an online form with the field list. The user is then led through the list of MIC data elements, selecting from a pull-down menu the organization's own equivalents for the MIC data elements.

MIC's strategy to accommodate multiple schemas accomplishes two goals. First, it encourages participation because organizations aren't asked to conform to a particular schema imposed from above and unsuited to their needs.<sup>14</sup> Secondly, the ability to map to multiple schemas allows archives to make their records available through a number of avenues, maximizing exposure to their collections. MARC allows interoperability, especially with print materials. Dublin Core enables OAI harvesting. MPEG-7 facilitates non-textual indexing. Each archive's own schema is important for local ownership. Using the MIC export utility, an archive employing Dublin Core for ease of use in-house can make its records available in MARC for Z39.50 access or in MPEG-7 for low-level indexing. Figure 1 shows how data are mapped from a variety of schema, through the MIC core registry, for export in different schema, which can then be distributed via numerous venues for different purposes.

Finally, MIC embraces diversity through its portal structure, which allows delivery of customized information to a number of distinct constituencies, including educators, archivists, and the general public.

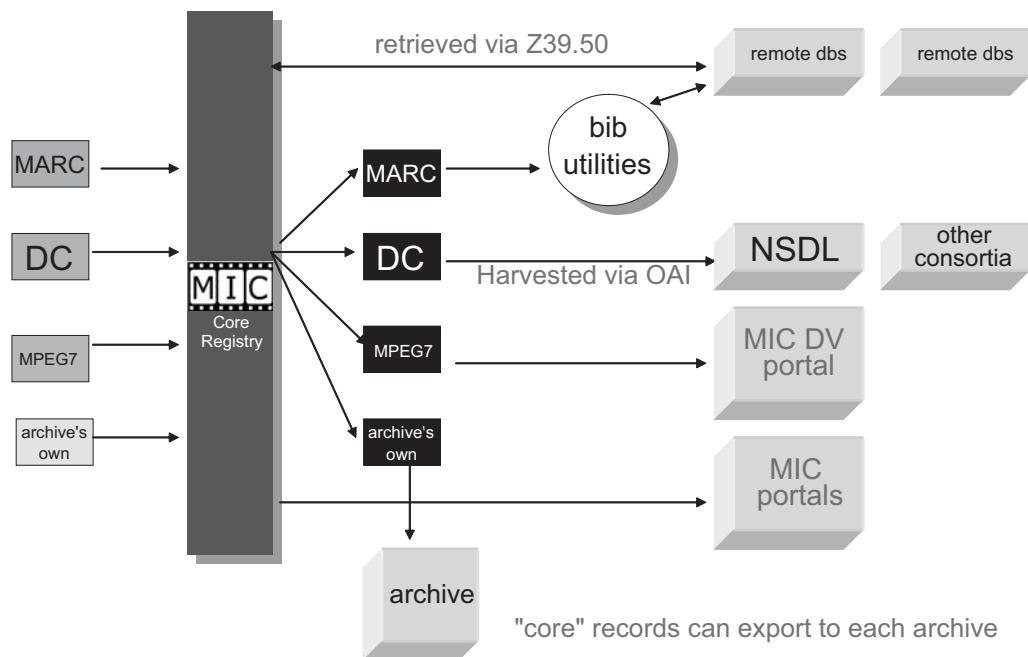


Figure 1: MIC mappings

Within the archivists' portal are subportals for preservation, cataloging and metadata, and exhibition programming. Portals customize MIC information in a variety of ways, in the union catalog, the archive directory and in the educational resources. (It is the PortalID data element, found in both the archive directory and the union catalog, that enables retrieval of portal-specific information in the catalog and directory databases.)

For example, in the science educators' portal ("Science Goes to the Movies"), a union catalog subject search on "whales" will retrieve a documentary on the topic, held by the National Library of Medicine, but not the fictional feature film *Free Willy*. Also in the union catalog, bibliographic record displays are standard across portals, but in the archivists' portal, the user can opt to display a record in any of several additional formats, including MIC

XML, MARC HTML, MPEG-7 XML, Dublin Core XML and the original record format. Archive directory displays are also customized by constituency. A directory entry displayed in the general users' portal offers organization and collection profiles, description of services, guidance on how to obtain moving images, and a link to the repository's own catalog. The same entry in the archivists' portal adds detailed information about cataloging and preservation practices.

Directory data also enables MIC's sponsors, the Library of Congress and AMIA, to identify and target particular educational needs, potential collaborations, and emerging trends, in order to focus community training and support.

The last portal feature promotes a community-building environment that takes MIC's functionality beyond discovery and distinguishes MIC from

other similar initiatives. Through the directory archivists can obtain the information they need to build collaborations, evaluate cataloging and preservation activities in similar organizations, and self-identify organizations with common interests. They can then exploit MIC's portal capabilities to sponsor research and education portals, and offer active collaborative training and development in areas of mutual interest. For example, local news organizations might have a portal to implement common record displays and shared subject heading lists for indexing television station logbooks. Legal and medical moving image organizations could use a subject-specific portal to explore privacy and rights metadata issues particular to their clientele by concatenating their collections for shared rights metadata development and testing. A digital video portal might be used by organizations with linked video for download or streaming, to provide a research space for exploring digital access strategies such as MPEG-7 non-textual indexing. Thus, the portal architecture can be used not just to customize information delivery; it can serve as a technology platform to support any number of collaborative activities.

### **PRINCIPLE 3: EXTEND METADATA TO ANY MOVING IMAGE ARCHIVE**

This paper has previously referenced several kinds of metadata. NISO defines three main types: descriptive, structural and administrative.<sup>15</sup> Descriptive metadata are those which allow discovery and identification of a work, or an expression of that work, such as title, date, director or subject. The MIC

core registry covers descriptive metadata. Structural metadata indicate how compound objects are put together, for example, how pages are ordered to form chapters. Administrative metadata provide information to help manage a resource, and include rights metadata, technical metadata and preservation metadata.

MIC's future plans include a cataloging utility with front-end input form enabling organizations to easily input descriptions of their holdings directly into MIC, but not limited to descriptive metadata. The MIC cataloging utility will be a Metadata Encoding and Transmission Standard (METS) implementation<sup>16</sup> incorporating the administrative, technical, preservation and rights metadata required for the effective management of any moving image asset through the course of its lifecycle. METS provides a structure for encoding all types of metadata, expressed in XML. METS is being developed as an initiative of the Digital Library Federation, and is maintained by the Library of Congress.

This utility will enable individual archives or collectors to describe their materials for preservation, management and access purposes, without the extensive expertise, technical infrastructure and funding previously required for this activity. This will empower small under-funded organizations to make national standard records available to a much broader audience. At the same time it will allow larger institutions employing multiple or legacy metadata schemas to map different collections through MIC's mapping utility and export records in a single metadata schema. The cataloging utility exemplifies the Library of

Congress' commitment to providing help with tools and standards to under-supported archives, while building consensus and shared solutions for broader constituencies.

MIC developers are laying the groundwork for this functionality by defining comprehensive source metadata for analog moving images, as well as technical metadata for digital video and audio data streams.

## **BENEFITS BEYOND DISCOVERY**

The ideal metadata strategy extends beyond immediate goals to describe and manage assets, and so MIC's strategy reaps benefits beyond discovery of moving images, repositories and informational resources. MIC promotes strategic preservation decision-making and management at national and international consortial levels through its union catalog and archive directory. It documents unique and best extant materials for preservation purposes, enabling archivists to accurately identify past preservation work and emerging critical need, reducing duplication of effort, preventing loss through deterioration, and ensuring that titles are preserved from the best surviving footage. Through its informational resources, including its remembrance wall of lost and found films, preservation timelines, and links to film preservation fundraising initiatives, MIC educates the public on the value of moving images and the importance of moving image preservation, both for the world's cultural heritage, and for individuals' own private holdings, such as home movies.

MIC contributes to the continuing education of archival professionals

through creation and gathering of informational resources about moving images and their preservation. It aids archivists in their day-to-day work by answering questions about nitrate storage, donor policies, and digital asset management systems. (MIC also intends to guide archivists and individuals to providers of products and services for archival moving image collections, through its service providers' directory, currently in development.)

MIC's educational mission reaches beyond archivists and the general public. The MIC union catalog provides the framework for an integrated and comprehensive search system combining phoneme-based search tools for audio streams with metadata, closed captions, transcripts and shot and segment logbooks to enable end users to identify and compile digital video segments for use in scholarly publications or as learning objects. Resources about a film can be linked to the asset itself, collocating various expressions of a work, such as foreign releases or director's cuts, with works about the film. Linking assets to a body of contextual works will advance film studies and will have broader applications beyond academia.

Educational and commercial interests are not mutually exclusive. MIC's cataloging utility, incorporating rights metadata, would lay the groundwork for developing advanced digital rights management strategies to support widespread, equitable access to digital moving images and active privacy for users. It is possible and necessary to negotiate and implement access portals that represent a common ground between the needs of educators, the public and commercial entities. For

example, MIC could develop a digital news portal for journalism studies incorporating applications and resources to serve commercial stock footage interests while improving literacy skills for using digital video resources in education. The news is out there. MIC provides the framework to make it available.

Finally, MIC can serve as a model to facilitate parallel projects in allied fields, for example, for sound recordings, allowing the immersion of all audiovisual materials into the information landscape, in support of education. MIC's metadata strategy serves the Library of Congress mission to sustain and preserve a universal collection of knowledge and creativity for future generations.

## CONCLUSION

MIC was originally designed as a tool to address the crisis in film preservation. As this paper demonstrates, recommendations rooted in the practical requirements of preserving analog artifacts, when implemented within a sound and visionary metadata strategy, can evolve into a research and development platform that serves a clientele beyond archivists and explores the leading edge of non-textual indexing, digital rights management, and educational use, all the while continuing to meet the daily needs of archivists through continuing education and support of collaborative preservation, access, digitization, education and metadata initiatives.

## References

- 1 Melville, A. and Simmon, S. (1997) *Redefining Film Preservation, A National Plan: Recommendations of the Librarian of*

*Congress in Consultation with the National Film Preservation Board.* Library of Congress, Washington, DC.

- 2 Murphy, W. (1997) *Television and Video Preservation.* Library of Congress, Washington, DC.
- 3 The Association of Moving Image Archivists (<http://amianet.org/>) is a non-profit professional association established to advance the field of moving image archiving by fostering cooperation among individuals and organizations concerned with the acquisition, preservation, exhibition and use of moving image materials.
- 4 This was a finding of the Cataloging Task Force under AMIA's Committee on US National Moving Image Preservation Plans (CoUSNMIPP); for more background information, see <http://www.amianet.org/committees/PastProj/USNational/mandate.html> (Accessed: 24 October 2005).
- 5 The MARC Standards page of the Library of Congress Network Development and MARC Standards Office may be found at <http://www.loc.gov/marc/> (Accessed: 24 October 2005).
- 6 The Dublin Core Metadata Initiative home page may be found at <http://dublincore.org/> (Accessed: 24 October 2005).
- 7 An overview of the MPEG-7 standard may be found on the Motion Picture Experts Group website: <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm> (Accessed: 24 October 2005).
- 8 The MODS: Metadata Object Description Schema official website may be found at <http://www.loc.gov/standards/mods/> (Accessed: 24 October 2005).
- 9 The Corporation for Public Broadcasting's PBCore Public Broadcasting Metadata Dictionary Project website may be found at <http://>

- www.utah.edu/cpbmetadata/ (Accessed: 24 October 2005).
- 10 A summary of the DMS-1 standard (SMPTE 380M-2004) may be found at [http://smpte.org/smpete\\_store/standards/index.cfm?scope=1&CurrentPage=17&stdtype=smpete](http://smpte.org/smpete_store/standards/index.cfm?scope=1&CurrentPage=17&stdtype=smpete) (Accessed: 24 October 2005).
  - 11 Agnew, G., Kniesner, D. and Weber, M. B. (2005) "Integrating MPEG-7 into the Moving Image Collections Portal," *Journal of the American Society for Information Science and Technology* (in press)
  - 12 The MIC Core Registry is documented online at [http://gondolin.rutgers.edu/MIC/text/how/unioncat\\_registry\\_table\\_01\\_05.htm](http://gondolin.rutgers.edu/MIC/text/how/unioncat_registry_table_01_05.htm) (Accessed: 24 October 2005).
  - 13 IFLA Study Group on the Functional Requirements for Bibliographic Records (1998) *Functional Requirements for Bibliographic Records. Final Report*; p. 82. G. Saur, Munich.
  - 14 Compare, for example, the National Moving Image Database (NAMID) initiative to centralize information on film and television holdings, begun in 1984 under the National Center for Film and Video Preservation of the American Film Institute (AFI), and which required conformance to the MARC standard; described in Green, D. (1997) "Beyond word and image: Networking moving images — More than just the 'movies'" *D-Lib Magazine* (July/August), Available at <http://www.dlib.org/dlib/july97/07green.html> (Accessed: 24 October 2005).
  - 15 NISO (2004) *Understanding Metadata*. p. 1. NISO Press, Bethesda, MD: p. 1.
  - 16 The METS: Metadata Encoding & Transmission Standard official website may be found at <http://www.loc.gov/standards/mets/> (Accessed: 24 October 2005).