A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories

Citation to Publisher Version: Otto, Jane. (2010). A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories. *Cataloging & Classification Quarterly 48*(5), 403-422.http://www.tandfonline.com/doi/abs/10.1080/01639371003757012#.VNztNGMhB8E.

Citation to *this* Version: Otto, Jane. (2010). A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories. *Cataloging & Classification Quarterly 48*(5), 403-422. Retrieved from doi:10.7282/T31G0NWR.

A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories

Jane Johnson Otto[1]

Abstract

Preservation of our audiovisual heritage is critical, and technical metadata is at the heart of any effective preservation program. This paper documents the efforts of Rutgers University Libraries to implement the Audio Engineering Society's draft "Audio Object" schema, AES-X098B, extend it for moving images, align it with existing standards, and integrate it with technical metadata for text, three-dimensional objects, and graphics. The paper compares several existing and emerging technical metadata standards, provides a description and assessment of the AES schema, and concludes with an application profile for several New Jersey repositories.

Keywords: technical metadata; moving images; sound recordings; preservation; metadata standards; Audio Engineering Society standards.

Introduction

Preservation of our moving image and sound heritage is critical. Our society's collective memory for the last hundred-plus years is inextricably tied to the visual and audio record. Much of this heritage is already lost, endangered, or inaccessible: deteriorating, underdocumented, and embedded in obsolete or soon-to-be obsolete mediums such as film, disc, and tape, or on transitory digital file formats whose future is uncertain.

Increasingly the focus for preservation is on digital surrogates, as digital formats become more faithful to the original, more stable, and less expensive, and as more and more resources are born digital. Documentation of the digital object's technical characteristics--i.e., technical metadata--is central to a strategic and robust preservation program. For the moving image

---

[1] Jane Johnson Otto is Multimedia Metadata Librarian, Rutgers University Libraries, 47 Davidson Road, Piscataway, NJ 08854 (E-mail: jjotto@rci.rutgers.edu).

The author wishes to thank Grace Agnew for her vision and guidance, and Chris Lacinak, Brad McCoy, and Carl Fleischhauer for advice, assistance, and technical expertise.

archivist, it is axiomatic that preservation includes documentation of the object's own characteristics *and* those of its source. The archivist must record the technical characteristics of the digital resource (attributes of the file, its size, coding format, etc.), the hardware and software used to create it and play it back, *and* the characteristics of the analog source from which the digital surrogate was created.

Most archivists would agree that the more technical information that is known about a resource, the more effectively it can be managed. However, technical metadata, particularly for moving image and audio materials, can be extremely complex, as well as voluminous, and the value added must be weighed against the significant cost of its creation and maintenance. The complexities are numerous: a moving image may have multiple layers or tracks of images and sound, and require linking between multiple generations of an analog source and the digital surrogate. Metadata accumulates throughout the object's life cycle from the moment of production, but is gathered by different personnel with complementary expertise and no mechanism to aggregate the data and communicate it down the line. As for cost, much metadata for digital objects must be manually input, although some can be extracted from the file header. No one is entirely sure what metadata is critical. The descriptive metadata community, guided by user experiences in search and retrieval, has developed core descriptive metadata over a number of decades for a variety of contexts. In contrast, technical metadata is still in its infancy, and we lack the user studies that would inform a core technical metadata standard, as well as the tools to make it useful. The great diversity in holding organizations, with differing missions, users, resources, and business models, also impedes consensus. For all of these reasons, technical metadata standards for moving image and audio materials have been slow to develop and difficult to adopt.

This paper looks at one emerging technical metadata standard with great potential for multimedia archives: the Audio Engineering Society's draft "Audio Object" schema, AES-X098B.[1] This paper documents our efforts to implement this complex schema for audio, including audio elements of motion pictures, and to extend it for moving images. Our objective was to develop a practical application of the schema for use in a busy academic setting, and also by a statewide digital video initiative, NJVid, which includes New Jersey libraries and archives of all sizes and types. The paper begins with a needs assessment and comparison of several existing and emerging technical metadata standards, followed by a description and assessment of the AES standard, and concluding with our application profile for the Rutgers Community Repository (RUcore) and NJVid. The bibliographic utility incorporating this application profile is available as part of the OpenMIC open source bibliographic utility now available from the Rutgers University Libraries.[2]

A Vanishing Heritage

Movies have been called the twentieth century's "most vital social memory and its most distinctive art form"[3] Fifty percent of all films produced before 1950 have been lost[4]; perhaps more alarming, in a society reliant on television for coverage of current events, some 25 years (covering approximately 1950-75) of American state and local history were destroyed in the switch from film to video.[5] The video produced since that time survives on obsolete formats or in evolving digital file formats. As home movies, itinerate filmmaking, and other independent and orphan works gain importance as social documents, and moving images increase with the proliferation of mobile devices and social networking spaces, preservation becomes still more important.

Collaborative long-term preservation requires a complete picture of a work, as provided by descriptive, preservation, rights, and technical metadata for all extant manifestations. Shared databases documenting "who has what" enable archivists to identify past preservation work and emerging critical need, thus reducing duplication of effort, preventing loss through deterioration, and ensuring titles are preserved from the best surviving footage.[6]

The Role of Technical Metadata in Moving Image Preservation

Extensive technical metadata is essential for the long-term preservation and management of audiovisual resources. A moving image consists of a series of sequential momentary images, perceived as continuous motion through the optical effect known as persistence of vision. The basic technical features that characterize a moving image are the size and shape of each image, the detail in the image is recorded, the speed at which the images follow each other in sequence when presented, any accompanying audio, and the overall duration of the sequence. Without metadata describing resolution, gamut (color and brightness of the pixel), pixel aspect ratio, frame rate, and audio aspects,[7] a digital image cannot be authentically reproduced.

Although technical metadata is commonly described as metadata documenting the creation and characteristics of digital files[8], the properties of analog carriers are equally important to preservation. In the case of moving images, particularly film, the digital file is not a true equivalent of the analog source. Photographs and text may be digitally reproduced with astonishing fidelity to the original, but the encoding of a moving image that can play back on a computer still results in noticeable information loss. Restoring and assembling negatives to produce preservation masters and quality viewing copies from pre-print materials also requires technical metadata: emulsion depth, color timing (color correction) data, key codes (the analog equivalent of the time code), film stock, etc. In the digital realm, standards for the moving image itself (the "essence") continue to emerge, but both analog and digital moving images suffer from a lack of standardization and backward compatibility. In any case, the transitory nature and constantly improving quality of physical carrier and file formats require archives to retain the film elements long after they are mastered to a preservation format. The technical metadata for these originals should reside alongside that of any digital surrogates produced in the preservation process.

Broadly defined, and for the purposes of this paper then, technical metadata is the data describing the physical properties and/or digital characteristics of any library material, be it analog or digital, physical object or set of bits. Thus, technical metadata might apply here to a roll of film, a digital videotape, an analog videotape, or a digital file.

Ideally, technical metadata should be useful in multiple contexts, and transferable from one environment to another. For example, metadata created on the production end should be accessible and comprehensible to preservationists and end users. It should be comprehensive enough, and granular enough, to facilitate retrieval, playback, end user displays, and management of the resource throughout its lifetime, supporting the collocation of all generations of a title, whether on film or videotape, or born digital. At the same time, the ideal schema would provide a clear migration path from the source copy that produced the digital surrogate, since archives often hold multiple analog copies. Above all, technical metadata should be standardized to ensure interoperability between systems, shared understanding of the information, and to avoid reinventing the wheel.

A Dearth of Technical Metadata Standards

A proliferation of overlapping standards over the last few decades has not obscured the fact that for audio and video, we lack technical metadata element sets and schemas.[9]

Nonetheless, much groundwork has been laid. Perhaps most importantly, the METS Metadata Encoding & Transmission Standard provided the means to incorporate preservation metadata with descriptive and rights metadata in a single package with the digital object itself and metadata pertaining to the analog source instantiation from which that object was derived.[10]

For digital still images, ANSI/NISO standard Z39.87 defined a set of metadata elements to enable file exchange, long-term management, and access.[11] PREMIS' data dictionary defines a preservation metadata set broadly applicable to digital preservation repositories.[12] Neither of these standards fully address technical metadata for AV materials, however. With ANSI, the images must be still, and digital. PREMIS' core elements are applicable to objects in all formats, but it lacks the detailed, format-specific technical metadata that is "clearly necessary for implementing most preservation strategies."[13]

PBCore, the Public Broadcasting Metadata Dictionary, was created by the U.S. public broadcasting community to describe its collections, and places greater emphasis on moving images and audio. PBCore is more comprehensive than other schemas, incorporating descriptive, rights, and technical metadata for moving images, sound recordings, still images, and text. One strength of PBCore is the number of controlled vocabularies provided for many of its data elements. In many cases, however, it lacks granularity; for example, physical format terms combine dimension, carrier (videotape), brand names, popular names, and sometimes, the defining standard. It also lacks some important technical metadata, particularly for film. There is no place to record layer characteristics, including composition of the base (nitrate or safety) or recording layer (for tapes). PBCore also lacks elements and vocabularies for time code characteristics, thickness (for example, of a film's emulsion), stock brand, speed, or specific color processes. Several key elements take only free-text values, rather than controlled vocabularies.

MPEG-7, one of the few metadata schemas developed specifically to describe, manage and provide access to moving images, incorporates significant technical metadata, but carries with it a steep learning curve, few examples or implementations, and limited documentation.[14] To some degree, MPEG-7's complexity is understandable, given its support of functionalities such as segmenting and non-textual navigation and indexing, but its difficulties are compounded by an obscure nomenclature and an unusual data model that subordinates descriptive metadata to technical. A further limitation is that MPEG-7 is intended exclusively for digital files.

The Society of Motion Picture and Television Engineers (SMPTE) Metadata Dictionary RP210 was developed by the SMPTE 30 Metadata and Registers Committee, an ANSI-accredited due process standards body. While often cited as a technical metadata standard, it is not actually a standard, but rather a registry of metadata elements *used* in SMPTE standards. The registry in its spreadsheet format numbers nearly 2000 rows and 38 columns of data elements and specifications.[15] The registry is useful for identifying the types of data historically recorded for video, but SMPTE RP210 excludes film from its scope and its organizational scheme limits its applicability to other environments. (In fact, its elements were not designed for re-use.)[16] Instead of following the library/archive model of descriptive, technical, rights, and structural metadata, SMPTE RP210 recognizes ten data classes, or nodes: Identifiers and Locators, Administration (business and security data), Interpretive (how a value is to be interpreted; thesauri), Parametric (technical data for coding and compression), Process (identification of processes and device or processor settings), Relational (identification of how data relates to other data), and Spatio-Temporal (time data and geo-spatial coordinates),[17] along with a few others.[18]

Like other SMPTE standards, RP210 is part of a complex web of interdependent standards. For example, a full explanation of RP210 is contained in its "defining document," SMPTE 335M (Television - Metadata Dictionary Structure). A full understanding of leaves and nodes as used in RP210 requires SMPTE EG-37-2001, Node Structure for the SMPTE Metadata Dictionary, or SMPTE's Labels Register (SMPTE 224.9), which is defined by SMPTE 400M. The Labels Register standard references SMPTE 298M-1997 (Universal Labels for Unique Identification of Digital Data), SMPTE 336M-2007 (Data Encoding Protocol Using Key-Length-Value), and SMPTE 359M-2001 (Television and Motion Pictures — Dynamic Documents). And so on. Identifying, obtaining, and determining the usefulness locally of all the related standards is something beyond the means, expertise, and will of most libraries and archives.

Possibly the most broadly useful source for moving image technical metadata "published" to date are the video metadata (VMD Schema) data dictionary and extension schema for the Metadata Encoding and Transmission Standard developed as part of the Library of Congress' AV Prototyping Project,[19] begun in 1999. According to the Library of Congress, a number of digital projects report using it in the absence of a true standard, and there is interest from the field in formalizing it, along with the other AV Prototype schemas.[20] Interestingly, the AV prototype, which offers audio as well as video metadata elements, drew heavily from an early draft of the Audio Engineering Society's "Audio Object" schema, or AES-X098B.[21] It stands to reason then, since the AES schema expands on the audio metadata prototype, that any update or expansion of the Prototype's parallel video metadata should align with AES-X098B, insofar as possible.

The AES Audio Object Schema

AES-X098B (still a draft at press time), is a set of data elements and corresponding vocabularies, expressed as an XML schema and designed to advance the long-term archival storage and preservation of sound recordings. Numerous AES characteristics make it a compelling choice, even a model, for technical metadata. Its preservation metadata is comprehensive enough for end-to-end management of a resource, yet granular enough to be manipulated for a multitude of purposes. It accommodates both analog and digital materials, whether on physical carriers or stored as streams of bits,[22] as well as segmenting. A rigorous delineation between technical and other metadata means it can readily be plugged into the METS environment without overlap with rights or descriptive metadata. About the only major downside to AES-X098B is the limited scope: it applies only to audio recordings.

At Rutgers University Libraries, we encountered AES-X098B in developing OpenMIC, an open source, web-based METS-compliant cataloging tool used for RUcore (Rutgers' digital repository), the New Jersey Digital Highway (the state repository), NJVid, a statewide digital video repository, the Video Mosaic Collaborative, an NSF-funded portal of mathematics education videos created over a 20-year period, and MIC, the Moving Image Collections initiative sponsored by the Association of Moving Image Archivists (AMIA) and the Library of Congress.[23] In addition to metadata import/export services, OpenMIC provides a complete metadata creation system for analog and digital materials in all formats: moving images, sound recordings, graphic materials, text, and three dimensional objects.

As OpenMIC metadata development progressed, we realized that its moving image and audio elements and vocabularies were inadequate for the breadth and scope of moving images we were supporting, from analog source objects in obsolete formats to born-digital mini-DV files. We were directed by several parties to the AES schema as a potentially useful source for audio metadata, even though it still was in draft form. The schema had been developed over the course of several years by audio archiving experts with unparalleled knowledge of the technical characteristics of audio resources. Although it is fairly complex, it possesses a number of noteworthy characteristics and conventions and showed considerable promise.

Several features of the AES standard warrant further discussion and will be explained in turn. First, the schema defines four primary levels of hierarchy (linked by ID attributes), and assumes a strict one-to-one mapping. It is based in large part on structure type (tape, disc, cylinder), and distinguishes between the media itself (e.g., the tape) and its shell (e.g., the cassette or cartridge). It also uses W3C data types, and additional custom data types defined in accordance with W3C XML schema recommendations.

Four levels of hierarchy. The AES Audio Object schema specifies four levels of hierarchy which together describe the audio object. The audioObject is the top level of the document, and is described in terms of repeatable face, region, and stream sections.

Audio object

--Face

--Region

--Stream

The *face* represents a single side of a two-sided object, for example, the B side of a 45 rpm record. Many objects are single-sided and consequently would be described in terms of a single face. The *region* represents a section within a face, for example, a section of a tape where the speed changes from the norm. Generally one delineates and describes object regions only where necessitated by changes of format; consequently, most objects would be described in terms of a single region. The *stream* represents an individual channel of audio information.

Each object has one or more faces; each face has one or more regions, and each region has one or more streams. Even a wire recording, which has just one stream of data, no changes of format, and a single face, is described in terms of (a single) face, region, and stream. Elements for describing these four sections are described in greater detail below.

Strict one-to-one mapping. The AES Audio Object schema specifies that "each audio object is described by a single instance document in a strict one-to-one mapping." In other words, an audio object is defined as either a single physical item or a single digital file.[24] This means that two audiocassettes comprising a single work are described as two separate audio objects. Even for two files meant to be played synchronously for stereo output, "each file shall have its own audio object instance document populated with values scoped to that file only."[25] This is where the AES standard deviates from traditional library practice. AES relies on structural metadata (outside the scope of the standard) to associate multiple files comprising a single intellectual work. While structural metadata allows one to link the audio object instance documents and the files themselves, the question remains: How does one synthesize the technical details to describe the resource in its entirety to users?

Structure type. The AES schema ties physical description elements to structure types, and defines four such types: tapeStructure, opticalDiscStructure, analogDiscStructure, and cylinderStructure. Wire recordings are described under tapeStructure, presumably because tapeStructure subelements are adequate to describe them. Once the structure type is declared, only values corresponding to that type are recorded. For example, an audiotape is described in terms of its four layers and measured in length and width; an analog disc is described in terms of its three layers, and measured in diameter.

Media and shell. In contrast to the description of physical items in AACR2 and other standards, AES delineates descriptions of the media and its shell. Description of the shell is limited to its dimensions: either length, width, and depth, or diameter and depth.

W3C data types. As part of its XML Schema language specifications, W3C recommends parameters for types of data, to enable data validation and facilitate data interchange.[26] For example, a phone number or a date can be formatted in any number of ways. A specified datatype for telephone number lays out parameters for the data which can trigger an error report when a data string meant to be a phone number does not conform to the rules for valid phone numbers.

AES uses nine W3C data types: dateTime, ID, IDREF, IDREFS, integer, long, nonNegativeInteger, positiveInteger, and string. For example, the objectCreationDate is of the dateTime data type, so would take the form 2002-10-10T12:00:00+05:00. In addition, AES defines several custom data types, following the rules of the W3C recommendations. For example, AES defines a dimensionsType data type for expressing measurements, comprised of seven sub-elements: width, height, depth, shape, diameter, length, and thickness. Each of these subelements is of the measurementType data type. MeasurementType calls for storing the value as a decimal number and unit of measure. Valid units of measure include micrometres, millimetres, centimetres, mils, inches, feet, and threads_per_inch. (AES further specifies that "metric units should be used where practical," with U.S. units used to accommodate large backlogs of legacy data.[27]) The use of W3C data types and dataType rules are meant to ensure consistency of data values across implementations.

AES elements and attributes. Each hierarchical level has its own set of elements. The face element[28] is described in terms of

- start time and duration ("timeline," expressed as timecode)
- region
- direction (e.g., front and back, for a 78 rpm record)

Under face, the region is described in terms of

- start time and duration ("timerange," expressed as timecode)
- number of audio streams ("numChannels"[29])
- condition note
- security note (security measures embedded within the object)
- stream (individual channel of audio)

Under region, the stream is described in terms of

- channel assignment (channel number with pan positions within the left-right and front-rear axes)
- condition note

AES also accommodates format regions, tied to structure types. The formatRegion data for analog discs, for example, include subelements like speed, groove width, and noise reduction. Optical disc format region data includes bit depth, sample rate, and word size, among others. Also, if an object changes format several times, but there are only two total formats (format A changes to format B, then back to A, and back to B again), just the two formats need be described, with each linked to the appropriate region on the tape.

Format regions are described, as applicable, in terms of

- physical properties

- speed
- bitDepth
- sampleRate
- wordSize
- trackLayout
- grooveOrientation
- grooveWidth
- grooveCreationMethod
- soundField
- noiseReduction
- equalization
- bitrateReduction

Many of these elements include their own subelements, many with their own data types. For example, bitrateReduction includes subelements codecName, codecNameVersion, codecCreatorApplication, codecCreatorApplicationVersion, etc.

In addition to these categories, the audio object is described in terms of

- identifier
- checksums (one for the sound data and one for the entire file)
- format
- physical and digital properties
- objectCreationDate
- title (to link back to descriptive metadata in human-readable fashion)
- analogDigitalFlag

- generation
- disposition (indicating the object has been filed, discarded, etc.)

SchemaVersion, technically "metametadata," is also included.

The final section of the schema provides detailed specifications for timecode formatting. Annex A is the Audio Object XML Schema; Annex B is the Time Code Format XML schema; and Annex C is a bibliography (forthcoming).

Creating an Application Profile

Although the AES Audio Object schema is fairly complex, its assets were abundant and clear. Our challenge was to reduce the complexity to a workable application profile for use by busy archivists and librarians in a number of different settings, and for a number of different types of resources. We needed to integrate the AES structure, insofar as possible, into our existing OpenMIC tool, and incorporate individual AES elements into our existing schema. We would have to identify which elements were already in our schema (often under a different name), which were new, and which had application beyond sound recordings and moving images.

From the outset, our objective in using the AES schema in OpenMIC was to adopt it, not adapt it, but in the end this was not entirely feasible. AES-X098B and the original Rutgers Workflow Management System (WMS) which has evolved into OpenMIC were developing at about the same time, but in different spheres. AES was developed by engineers for sound recordings, whereas WMS was developed by librarians for numerous formats, including sound recordings, moving images, text, three-dimensional objects, and still images. One difficulty was that OpenMIC already followed ANSI/NISO Z39.87[30] for still images, requiring a balance between aligning with AES and achieving consistency of elements and vocabularies across formats. A pre-established data model coupled with our development timelines further complicated the retrofit and forced some compromises. In some cases OpenMIC would have benefited from an earlier alignment with AES; in other cases OpenMIC's path was arguably the better one. Ultimately we were unable to fully achieve our objective to adopt rather than adapt. Nonetheless, because AES is still a draft, it's possible closer conformance can be achieved through discussion during AES's public "call for comment" period.

The following describes some of the issues we encountered in incorporating the AES Audio Object schema into OpenMIC, and how they were addressed. Because the AES standard is still a draft, OpenMIC developers were working with a moving target. For the most part, we referenced an undated version of the draft which was downloaded from the AES website in February, 2007, although we consulted later iterations as the standard was revised.

Element equivalence. From the outset, AES and OpenMIC shared many of the same elements, although they were not always called by the same name. Because of programming time limitations and the difficulty of making global changes to element names across our suite of applications,[31] we made the decision to retain the existing OpenMIC element name in these cases, and use the AES definition where applicable. For the most part, AES data elements could be readily incorporated into OpenMIC.

Structure type. In our work we found the AES "structure type" construct to be the most elucidative, elegant, and extensible feature of the schema. Technical metadata is voluminous, highly detailed, and at times arcane. The logical framework provided by structure type gave us the facility to contain the web of complex relationships between elements and detailed terminologies, organize vocabularies, and, in a word, *manage* the metadata for the greatest possible comprehensibility and usability by our end users, the metadata inputters.

OpenMIC is designed for management of an object through its life cycle, so we expect a number of people to be editing any one metadata record: a cataloger enhancing descriptive metadata, a preservationist adding technical metadata, a student keying condition data from inspection reports, or a curator adding donor information (rights metadata). One obstacle to wide use of a metadata tool like OpenMIC is the vast array of elements and vocabularies, many of them irrelevant to one or another workflow. One of our challenges is to create dependencies to narrow the number of elements displayed to any one inputter.

For example, OpenMIC includes several hundred element-vocabulary combinations. By declaring the object ‘analog’ in AES’ analogDigitalFlag element, all elements pertaining to digital characteristics can be suppressed from the display. For this reason the analogDigitalFlag element is one of the first values selected by the inputter. Also right up front is sourceType, where the inputter chooses a structure type (audiotape, optical disc, etc.). Like the analogDigitalFlag element, SourceType drives data element displays. For example, if the inputter declares the object analog and selects ‘audiotape’ as the source type, only the elements applicable to analog audiotape are displayed. This eliminates two sizable sets of elements from the display: digital characteristics (codec, sampling rate, etc.) and image characteristics (color, aspect ratio, etc.). The number of displayed elements is reduced to about 25, most of them optional.

Another key benefit of the structure type model, perhaps unforeseen by the standard’s authors, is the ability to define vocabulary subsets based on the structure type value. In describing a videotape, for example, the person inputting gauge will choose from a pulldown offering only those dimensions applicable to videotape. The choices will be 1/4 inch, 1/2 inch, 3/4 inch, 1 inch, 2 inch, 6.35mm, and 8mm. The pulldown will not include film gauges, which number over a dozen (e.g., 16mm, 35mm, 70mm, etc.).

This last feature prompted us to include an additional structure type for audio materials: wire recording. (Although the draft standard is not explicit on this point, AES includes wire recordings under the tape structure type.[32]) If one were to use this feature as justification for new structure types, others might be added to accommodate additional materials such as music box discs, paper prints, player piano rolls, and early cinema formats on glass. At this point we have not fully assessed the need for additional structure types, but we would like to explore the possibilities across all formats, including text, graphics, and three-dimensional objects.

Face-Region-Stream. The AES construct of defining face, region, and stream for each audio object is a useful one, but these entities were not incorporated into OpenMIC due to time constraints and an existing data model which did not identify entities at this level. It was felt that most objects, initially at least, would not be described at this level of detail by our end users. Nonetheless, description at the region and stream level are important; it is frequently necessary to break description of the essence into segments and the region entity allows this. Therefore the face-region-stream model has been tagged as a possible area for future development.

Strict one-to-one mapping. As described above, AES utilizes a strict one-to-one mapping wherein each audio object is described by a single record. A double LP, such as The Beatles' *White Album*, would require two technical metadata instantiations, linked to one another by structural metadata outside the scope of the AES Audio Object standard. This is not in keeping with the METS or FRBR[33] models, and is complicated by the possibility of multiple manifestations of the two-record set, any or all of which, or even parts of which, might be linked to one or more source instantiations. In film preservation, it is quite common to derive printing elements from multiple source objects, for example, an interpositive from three different negatives. It is a source of frustration to archivists that integrated library systems cannot graphically represent relationships between the sources and their derivatives. We have identified a way in OpenMIC to link multiple instantiations within the METS document and would like to see graphic representations of the "migration trail" in a future release. AES does not mention any mechanism for linking multiple instantiations of technical and/or source metadata to a single instance of descriptive metadata, nor a means of tracking derivatives through multiple duplication processes. The principle of strict one-on-one mapping will require greater attention to this issue.

Another issue with the one-to-one mapping relates to the overall physical description of a title. In the typical library catalog, some data in the record, probably the physical description in descriptive metadata, would describe the collective whole, e.g., 2 audiocassettes. Few would argue the utility of such a statement, but it's unclear where this would fit in an AES description, or how it could be linked to the AES technical metadata. This fundamental contradiction between the AES one-to-one map and traditional library practice was probably the primary factor in our decision to adapt rather than adopt.

To address this issue, we added two data elements designed to allow a standard AACR2 physical description: an extent element, with positive integer value, to concatenate with a "type" value for each structure type, to produce a simple extent statement such as "2 analog disc(s)" or "1 DVD audio(s)." These type values were largely based on AACR2's specific and general material designations. A strict AES implementation would not need the extent element, because the numeric value would, by definition, always be '1.'

Media-Shell-Container. Much of the beauty of the AES Audio Object schema lies in its granularity, which gives it the flexibility to serve many management, preservation, and access functions. One example is the AES delineation between the media itself, and its shell. For example, AES describes an audiocassette in terms of width (1/4 in.) and length, *then* describes the dimensions of the shell (cassette) encasing it.

We liked the AES delineation between media and shell, and carried it a step further, adding elements to describe any non-integral container, as well as the mount (hub for audiotape, reel or core for film materials).[34] The storage of archival objects, in terms of size and physical composition of container and mount have significant preservation ramifications. In theory, container dimensions could also help determine linear shelving requirements.

Extending AES to moving images

Although we initially consulted AES as a source for audio metadata, we soon realized it offered a useful model for moving images too, and might be adapted for these materials. After all, motion picture audio tracks have been recorded on most of the formats AES was meant to describe: audiotape, optical disc, and analog disc (e.g., Vitaphone discs). We believed we could extend the audio metadata to include not just audio tracks recorded on film, but the moving images themselves, whether on film, videotape, or videodisc. The schema's logical organization and structure type-based model certainly facilitated this.

To adapt AES for moving images, we began with our existing OpenMIC data elements for audio and moving images. These had been developed largely by Grace Agnew based on her PBCore work. In accordance with METS, we had defined both source metadata (for analog objects) and technical metadata (for digital files). Our "technical metadata" (in this narrow METS sense of the word) included many of the source metadata elements (e.g., duration, aspect ratio), but included additional elements to describe digital characteristics. Elements pertaining to digital videotape had not been included in the source metadata, probably due to a common assumption, implicit in METS documentation, that a physical object would by definition be analog. In building out the moving image and audio metadata, we made the decision to include digital audiotape, digital videotape, and digital discs in source metadata, reserving technical metadata for digital files.

Next, we took the AES structure type model and applied it to moving images, establishing three new structure types: film, videotape, and videodisc. At this point we established three new "bucket" elements in OpenMIC: digital characteristics, picture qualities, and sound qualities. These were useful in organizing the totality of elements. We then created a "superset" of possible new elements, by combining video-specific elements from the Library of Congress AV Prototype with the AES data elements. In consultation with AES developers and Carl Fleischhauer (a principal author of the LC Prototype), we eliminated duplication between the two, and aligned element names and definitions. This streamlined superset was then compared with our existing OpenMIC elements.

For each candidate element, we asked these questions:

1) Is the element useful for moving images, audio, or both?

   1a) If moving images, for the picture characteristics, sound characteristics, or both?

   1b) If moving images, to which of the three new structure types (film, videotape, videodisc)?

2) Does the element apply to digital materials, analog, or both? If digital, does it apply to digital materials on physical carriers, or just to digital files?

3) Is there a match with OpenMIC? If so, we generally retained the existing OpenMIC element name, to minimize programming changes, and with the knowledge that the OpenMIC schema could be mapped to AES.

Some AES elements, such as analogDigitalFlag, apply to all materials. Others, such as gauge, could apply to audio or video, although extension to video required a new controlled vocabulary. Still others are audio-specific, but would apply to the audio portion of a video object (e.g., noiseReduction). In some cases, an element would apply to *both* audio and video, and was needed twice under the same structure type. For example, a single digital video object could have one codec for the video image and a different codec for the accompanying audio. Some AES elements had no application to moving image materials.

These decisions, while time-consuming, were usually straightforward, but not always. For example, after some debate we decided to use the timecode element for film, to record timecode's analog equivalent (arguably): the keykode.[35] Often extensive research was required. While most of us don't think of videodiscs issued in integral cases, quite a few were produced that way; JVC's 25cm VHD videodisc, introduced in Japan in the early 1980s, was housed in a caddy and sold in cardboard sleeves. RCA's CED discs were similar.

Occasionally an AES element was of little practical use for a particular structure type. For example, optical discs are technically comprised of layers, yet few archivists would take the time to describe them in detail. In general, we erred on the side of including the element, particularly if it had already been established under a different structure type. However, in these marginal cases we did not go the extra step of developing corresponding controlled vocabularies.

As this decision-making proceeded, we were also identifying which new elements should be added under our other three format types (text, graphic materials, and three-dimensional objects), and developing new controlled vocabularies for each element/structure type combination.

Vocabularies

A truly useful technical metadata schema includes not just data elements, but the corresponding vocabularies. One guiding principle behind the OpenMIC metadata is to offer vocabulary selection from pulldowns, to enable data input by non-catalogers and reduce typographical errors. While AES-X098B includes controlled vocabularies for some elements, most elements lacked vocabularies. For example, in the AES metadata schema, there is a 'speed' data element for analog discs, but no associated controlled vocabulary, which should include 16-2/3 rpm, 33-1/3 rpm, 45 rpm, 78 rpm, 120 rpm, and other.

The final part of our metadata development, and probably the most labor-intensive, was to identify authoritative sources of controlled vocabularies for each data element, correlated to structure type. Where no authoritative vocabularies were found, they were compiled or created in consultation with preservationists, librarians, archivists and audio engineers.

In some cases we have tinkered with the AES vocabularies where they do not conform to traditional cataloging content standards. For example, AACR2 relies heavily on abbreviations and measures objects in U.S. units of measure, rather than metric.

Working with Experts

All of the work on elements and vocabularies was undertaken in consultation with experts within diverse communities of librarians, archivists, and sound engineers, particularly those within the Association of Moving Image Archivists, Audio Engineering Society, Association for Recorded Sound Collections, and the Library of Congress' National Audio-Visual Conservation Center in Culpeper, Virginia. Active members of the AES Working Group were particularly helpful in elucidating the complex schema and pointing us to related projects: Chris Lacinak (Audiovisual Preservation Solutions), David Ackerman (Harvard College Library), and Bill Casey (Indiana University and the Sound Directions project). Development of the technical vocabularies keyed to individual data elements, with definitions, was a monumental task akin to compiling an encyclopedia in the form of a jigsaw puzzle. We consulted numerous existing standards and tools, including the LC AV Prototype, SMPTE RP210, PBCore, MAVIS (Merged AudioVisual Information System), the collection management tool used by the Library of Congress and Australia's National Film and Sound Archive, as well as documentation produced NFSA. Although many of these resources were highly detailed, each had its own application and no one was comprehensive; some covered sound but not moving images (or vice versa), analog but not digital (or vice versa), or offered terms but no definitions. Wikipedia proved invaluable for definitions and pointers to obscure formats. Throughout the process our work in progress was disseminated through diverse venues, including the LITA National Forum preconference, AMIA annual meetings, the Digital Library Federation Fall Forum, and for the Library of Congress' Digital Future and You webcast.[36]

Conclusion

Tools for managing audiovisual materials throughout their life cycles are essential to a strategic and robust collaborative preservation program, and technical metadata is at the heart of the process. Yet a dearth of technical metadata standards for both moving image and audio materials (analog or digital) has hindered progress. As an emerging standard, the Audio Engineering Society's draft "Audio Object" schema, AES-X098B, is a valuable model for technical metadata, and can be adapted for moving images. Its structure type-based model and rigorous organization make it potentially extensible to other formats as well. Rutgers University has successfully adapted the AES schema to align it with existing standards and integrate it with technical metadata for text, three-dimensional objects, and graphics. We have created an application profile for the OpenMIC open source bibliographic utility which is currently in production for the Rutgers Community Repository and the statewide digital video collaborative, NJVid, and available for use from the Rutgers open source applications website.

With the benefit of tremendous support from moving image and sound archivists--experts in the field and working professionals at home --we feel very confident that we have captured the range of elements needed to successfully document the management information needed for the preservation and use of analog and digital moving images and sound files. We have not only developed a technical metadata schema for every format, we have integrated extensive technical metadata into our current workflow and shared it with others via the OpenMIC platform.[37] However, the ink is not yet dry on our application profile and its use at Rutgers University Libraries and beyond. We are already investigating the next big question: can we develop sustainable practices that can be successfully applied in busy libraries and by the next generation of metadata professionals? We will be performing workflow analyses with our own staff, the staffs of libraries participating in collaborative initiative, and the metadata students in the Rutgers MLIS course, to answer this question.

---

1. Audio Engineering Society. Working Group on Digital Library and Archive Systems (SC-03-06). AES-X098B, DRAFT "AES standard for audio metadata : audio object structures for preservation and restoration." An earlier title was "Standard for audio preservation and restoration : administrative and structural metadata for audio objects"; members of the Working Group tend to refer to the standard as simply the "Audio Object Schema." Retrieved from AES

website for registered members of AES standards working groups: https://secure.aes.org/standards, February 7, 2007 and June 25, 2008. General information on AES and its standards may be found at http://www.aes.org/standards.

2. Rutgers University Libraries, "OpenMIC Overview," http://rucore.libraries.rutgers.edu/open/projects/openmic (accessed January 8, 2010).

3. *Film Preservation 1993 : A Study of the Current State of American Film Preservation: Report of the Librarian of Congress* (Washington, D.C.: Library of Congress, 1993), http://www.loc.gov/film/study.html (accessed January 8, 2010).

4. *Film Preservation 1993.*

5. *Television and Video Preservation 1997: A Report on the Current State of American Television and Video Preservation: Report of the Librarian of Congress* (Washington, D.C.: Library of Congress, 1993), http://www.loc.gov/film/tvstudy.html (accessed January 8, 2010).

6. For example, see MIC (Moving Image Collections): http://mic.loc.gov (accessed February 21, 2010).

7. Joint Information Systems Committee. *The Significant Properties of Moving Images* (March 26, 2008), pp. 37-40, http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf (accessed February 22, 2010).

8. See, for example, ODLIS Online Dictionary for Library and Information Science (http://lu.com/odlis/) and the Dublin Core Metadata Initiative's DCMI Glossary (http://dublincore.org/documents/usageguide/glossary.shtml). METS defines technical metadata as that metadata regarding files' creation, format, and use characteristics, and distinguishes it from "source metadata" describing "the analog source from which a digital library object

derives" (http://www.loc.gov/standards/mets/METSOverview.v2.html). ANSI/NISO Standard Z39.87, the data dictionary of technical metadata for digital still images, defines metadata elements enabling users "to develop, exchange, and interpret digital image files." Harvard University Library, on its Metadata Standards page (http://hul.harvard.edu/ois/digproj/metadata-standards.html), describes technical metadata as that focusing on "how a digital object was created, its format, format-specific technical characteristics, storage and location, etc. Accurate technical metadata helps a repository manage digital objects over time and keep them usable."

9. Library of Congress, *Library Services Strategic Plan, 2008-2013, Performance Goal 4.C.1 and 4.C.2* (August, 2007). This internal report documents the findings of a working group chaired by Sally McCallum of the Network Development and MARC Standards Office, of which this author was a member. The group's three-point charge was to "describe current standards activities and the Library's role nationally and internationally, identify gaps between current activities and needs in the information field, [and] propose a future Library role."

10. METS Metadata Encoding & Transmission Standard, http://www.loc.gov/standards/mets.

11. ANSI/NISO Z39.87 - Data Dictionary - Technical Metadata for Digital Still Images, 2006, http://www.niso.org (accessed January 8, 2010).

12. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* (Dublin, Ohio: OCLC Online Computer Library Center, May 2005), vii (accessed January 8, 2010).

13. *PREMIS Data Dictionary for Preservation Metadata, version 2.0* (March, 2008), 24 (accessed January 8, 2010).

14. Grace Agnew et al., "Integrating MPEG-7 into the Moving Image Collections portal," *Journal of the American Society for Information Science and Technology* 58 (2007): 1357-1363.

Perhaps the best documentation is found in B.S. Manjunath et al., *Introduction to MPEG-7, Multimedia Content Description Interface* (John Wiley & Sons, 2002).

15. *SMPTE Metadata Dictionary as specified in RP210-11-2008*, http://www.smpte-ra.org/mdd (accessed January 8, 2010).

16. The origins and organization of SMPTE RP210 and its relation to other SMPTE standards, as set out in this paper, were derived from the standards themselves, with input from Karen Broome, Senior Staff, Technology Standards Office at Sony Electronics, and Chair of SMPTE's 30.10 Metadata Definition working group, which manages the registration of SMPTE metadata.

17. Jim Wilkinson, *SMPTE Object Identification* (paper presented at Digital Objects and the Management of Information, June 20, 2006), http://www.doi.org/doi_presentations/seminar_20Jun06/Wilkinson-SMPTE.pdf (accessed January 8, 2010).

18. Class 13, "User Organization Registered for Public Use," allows specific user organizations, such as the Public Broadcasting Service, the Audio Engineering Society, or the Association of Moving Image Archivists, to register individual metadata elements. These would be reserved and managed separately from other classes of metadata by the SMPTE Registration Authority in accordance with SMPTE Administrative Practices. Class 14 provides a similar space for private use. Class 15 is for Experimental Metadata, which has been described as a sandpit for testing, and not interoperable.

19. Library of Congress, "AV Prototype Project Working Documents - Extension Schemas for the Metadata Encoding and Transmission Standard," http://www.loc.gov/rr/mopic/avprot/metsmenu2.html (accessed January 9, 2010).

20. Library of Congress, *Library Services Strategic Plan, 2008-2013, Performance Goal 4.C.1 and 4.C.2* (August, 2007).

21. The Prototype web page acknowledges the schema's debt to Harvard University and the Audio Engineering Society: http://www.loc.gov/rr/mopic/avprot/metsmenu2.html. Dave Ackerman, a sound engineer at Harvard, chairs the AES Working Group on Digital Library and Archive Systems (SC-03-06) and is primary author of AES-X098B. LC Prototype author Carl Fleischhauer, in conversations with this author (2008-2009), has stated that the audio metadata portion has been superseded by the draft AES schema.

22. AES-X098B explains that "the term physical audio carriers shall be understood to mean only those carriers to which an audio data essence is intrinsically bound to the physical object. For example, an audio DAT tape is a physical digital audio object. Wave files on the other hand are file-based digital audio objects. They can be moved, copied, and renamed without the need to play the audio essence. In the later case, the file exists on a storage device, but it is not intrinsically bound to that device. Therefore there is no need to describe the physical storage given its transient nature." (p. 7)

23. Rutgers University Libraries, "OpenMIC Overview," http://rucore.libraries.rutgers.edu/open/projects/openmic (accessed January 8, 2010).

24. AES-X098B, p. 7.

25. AES-X098B, p. 7.

26. W3C, "XML Schema Part 2: Datatypes," Second edition, W3C Recommendation 28 October 2004, http://www.w3.org/TR/2004/REC-xmlschema-2-20041028 (accessed January 10, 2009).

27. AES-X098B, p. 16.

28. Each section (face, region, and stream) includes elements for identifiers, human-readable labels, and pointers. (For example, the region element points to the enclosing face element and any included stream elements.) These are the 'ID,' 'label,' and 'ref' subelements.

29. It is beyond the scope of this paper to fully define the AES data elements. The AES schema is not publicly available, but the OpenMIC data dictionary defines most of these terms: http://rucore.libraries.rutgers.edu/open/projects/openmic/index.php?sec=guides&sub=metadata.

30. ANSI/NISO Z39.87 - Data Dictionary - Technical Metadata for Digital Still Images, 2006, http://www.niso.org.

31. OpenMIC is just one part of a larger architecture that includes Fedora and OpenWMS, the file handling module.

32. For example, 4.4.2.1.2., Dimensions, specifies how a wire should be measured, and some wire-specific data is defined in a wireFormatRegion element set "for the convenience of the implementer."

33. International Federation of Library Associations and Institutions, *Functional Requirements for Bibliographic Records: Final Report* (IFLA, September, 1997 as amended and corrected through February, 2009), http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf (accessed January 9, 2010).

34. Technically, AES includes some types of "mounts" in the shell element, which "may be used to provide physical measurements of the audio object's enclosure … this element may be used to describe for example, a tape reel, a tape cassette shell, or a non-removable optical disc caddy. However, because shell dimensions are governed by the measurementType data type, the terms don't match natural language terminology, and are limited to dimensions. The preservationist, however, would like to know whether that 2-inch core is metal, plastic, or wood. Audiotape is

mounted on a large, or NAB hub, a small hub, or a pancake. A film reel is generally described not in terms of its diameter (as called for by AES), but in terms of the number of feet of film it holds, e.g., 800-foot reel or 1200-foot reel.

35. Wikipedia describes keykode as a variation of time code designed to uniquely identify frames in a film stock: Wikipedia, "Keykode," http://en.wikipedia.org/wiki/Keykode (accessed January 8, 2010).

36. May 12, 2008: http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4321.

37. Rutgers University Libraries, "OpenMIC Overview," http://rucore.libraries.rutgers.edu/open/projects/openmic (accessed January 8, 2010).