

# Federal Agencies AV Working Group, Recorded Sound Subgroup

Notes from meetings about embedding metadata, held at the National Archives (College Park) July 9, 2009, and Library of Congress Packard Campus for Audio-Visual Conservation, July 10, 2009.

*Abstract: Four agencies represented, discussion of proposed guideline for embedding metadata in the bext and INFO chunks of WAVE/BWF files, and discussion of plan to develop a software tool to support such embedding.*

## Introductory

**Round robin.** Some members highlighted their activities in a round robin.

- a) One representative from the Library of Congress gave an update reported on IRENE, the optical-capture device for discs and cylinders; progress was being made on the workup of the 3D scanning approach.
- b) One representative from NARA reported on the exploration of an enterprise-wide system for identifiers and metadata; the ultimate goal is develop a plan for enterprise-wide non-mnemonic identifiers for content.
- c) One member said that she would be organizing a subgroup to compare ideas about technical metadata. Interested parties should signal their interest in participation.

## Regarding The Elements To Embed And Related Matters

**Element list.** The group was led through the list of elements that a special subcommittee has drafted, a proposal for a set of elements to embed in the <bext> and INFO (List INFO) chunks in Broadcast WAVE files. The draft will be posted for public comment at the Federal Agencies Web site very soon.

**UTC?** During the discussion, one member asked about the rendering of date and time of day in the <bext> OriginationDate and OriginationTime elements. Should these be expressed as UTC (Coordinated Universal Time, from the French Temps Universel Coordonné)? This could be important if files contain digital recordings made of broadcast content made off-air (off-cable, off-Internet, etc.) capture of broadcast content. The <bext> chunk elements do not permit using ISO 8601 standardized structures to provide the datetime information. The Working Group welcomes comments from outside readers on this topic.

**TimeReference element.** The consultants to the group explained some details of the <bext> TimeReference element. This metadata element records what is sometimes called a *timestamp*. This is the start time for a given file, in terms of a timeline, e.g., the timeline that is provided by digital audio workstation software. Files can be placed on the timeline in terms of their sequence using TimeReference. In <bext>, the time value is provided in terms of sample count for timeline location for the first sample in the file.

**Timecode for synchronization.** Some members asked about the native ability in <bext> to express timecode information, e.g., the kind of information needed to synchronize picture and

sound. This seems not to be possible; synchronizing timecode information will depend on other packaging structures, e.g., MXF.

## **Regarding A Tool To Support Embedding**

**Source of data to embed.** Discussion of the *conformance point data*, i.e., the data needed by an embedding tool, to embed. One consultant showed how an Excel spreadsheet (in effect, comma separated values) could be used by many archives as a landing zone for data extracted from databases or for the provision of boilerplate data. The format would make it easy in many settings to get an overview of the data to be embedded and to edit or massage it if needed. Some members said that the Excel form of the conformance point data set seemed useful. At least one group, however, highlighted the importance of XML data exchange in their operation, saying that they would prefer to see conformance point data in XML form.

**What tools are out there now?** The consultants offered a review of some of the classes of tools on the market, comparing them with what we want our tool to do. One class is *audio-centric*; examples often lack batch processing, provide inefficient ways to enter metadata, may fail to work in conformance with metadata specifications, may rewrite files when any element is changed, and often fail to provide a log file produced. A second class is *metadata-centric*; examples of these also tend to lack batch processing, fail to include “undo” features, and do not offer validation. One well regarded tool has a number of valuable features but it is expensive, requires users to master a relatively complex set of elements, and does not offer an API, thus presenting difficulties in interfacing with external applications.

**Features we would like to have.** The tool we seek ought to be easy for technician-level workers to use, run batches, permit undo actions, log the results, and also extract metadata from files. The general design ideas that the consultant are recommending include the user’s ability to “switch on and off” changes for specific elements, e.g., “keep CodingHistory untouched, substitute our agency’s name for the vendor’s, in the Originator element.”

**Sketch for a roll out plan.** The plan to roll out the tool project will employ a contractor and include such elements as identifying the pilot organizations; communicating with them to determine (i) a mechanism to find the files in a batch, (ii) review of available data to embed (conformance data), (iii) building tool for data transformation, (iv) identify storage location, (v) defining workflow and tools for generating new files, and (vi) write up the pilot activities as case studies. Meanwhile the contractor will set up a code repository to communicate milestones, get bug reports, store demos, etc. The tool design will be cross platform, written in C with wxWidgets or Java, and will start with a command-line version, then work up to a GUI. If contracting can take place promptly, the work might start in August or September.