SDR Team : Audio - MJF

This page last changed on Sep 03, 2008 by hfrost@stanford.edu.

The following is a description and analysis of an audio object packaged for preservation in the SDR. In this case, the object is part of the Monterey Jazz Festival collection.

Sample TM XML file: 0688.xml

Constituent Files of the Object – all tarred up together

1. Content files -
    a. Preservation Master WAVE file (*pm.wav)
        i. May take the form of split channel files (*pm.L.wav, and *pm.R.wav) or four multi-track files (*-TR#pm.wav)
    b. Service High WAVE file (*sh.wav)
2. Auxiliary files - Auxiliary application files are created by WaveLab software in the course of audio quality control and content description processes. They do not contain audio data and as such do not represent "content." However, they are useful, convenient, and meaningful. For these reasons, they shall be retained to facilitate future access events. They are relatively small in size (measured in KB), and so are a drop in the bucket in terms of storage. They may or may not be present for preservation master files, but should almost always be present for service files.
    a. (*.cd)
    b. (*.MRK)
    c. (*.wav.gpk)
    d. (*.wav.mem)
3. Transfer Manifest (SIP form)

Other Related Objects (to be analyzed under separate cover)

1. SDR Agreement (e.g., MJF-DepositAgreementV01.tar)
2. Collection-level Record (e.g., MJFCollection.tar)

Transfer Manifest

metsHdr

- agent = conversion software
- altRecordID = the data provider ID, in this case a string of 20 characters representing a MJF collection "proprietary" scheme that uniquely identifies an audio object by the item ID (first 8 characters), followed by a semi-colon, and the item face ID (last 11 characters).

dmdSec

- declares that this is the schema for describing an MJF audio object using MODS
- includes description information:
    - name of the object (tape number and side)
    - who created it
    - resource type
    - table of contents and performers (lists of all song titles performed and all performers contributing to the content embodied in the resource)
    - date originally recorded
    - date digitized
    - MIME type
    - recording extent (duration, total running time)
    - rights/access/copyright statement
    - constituent parts of the recording as "related Items" – segment-level description (i.e., each tune performed in a concert is described individually

amdSec (Object Level)

The first instance of an amdSec contains:

- rightsMD - expressed in PREMIS. identifies the agreement by file name (without file extension) and UUID; specifies the access permissions/restrictions outlined in the agreement

- sourceMD - data about the object source. namespace is declared as "http://library.stanford.edu/sulair/sdr/raw"; elements are based on local practices (to be codified in AES standardized schema for future collections).

amdSec (File Level; Content File)

Subsequent instances of an amdSec describing each preservation master and service high WAVE file contain:

- status of the file is indicated by filename (name ending in "pm" or "sh") and format (extension = .wav)
- techMD (instance 1) - premisObject metadata for the content file
  - names the file (original file name)
  - preservation service level
  - MD5, file size, file format
- techMD (instance 2) - JHOVE output - generated by conversion code
- techMD (instance 3) - other technical metadata for the draft AES object schema - imported from collection database

amdSec (File Level; Auxiliary File)

Subsequent instances of an amdSec pertinent to auxiliary application files contain:

- techMD - premisObject metadata for the auxiliary file
  - names the file (original file name)
  - preservation service level
  - MD5, file size, file format

fileSec

- fileGrp (instance 1) - grouped by Use as the "Archive Masters" files
  - declares the fileID for each file
  - includes MD5, create date, MIME type, original file name ("Owner ID"), file size
  - FLocat shows where it found within the object structure
- fileGrp (instance 2) - grouped by Use as the "Auxiliary Application" files
- fileGrp (instance 3) - grouped by Use as the "Service High" files
- fileGrp (instance 4) - grouped by Use as the "Data Container" file (i.e., .tar)

structMap

One "div" containing several nested divs, showing the relationships between files and metadata.

- div (sub-instance 1) - for the "object aggregate"
  - mptr - denotes UUID ("SURI") for the entire object (confirm)
  - div (sub-instance 1) - "Level02_Aux". Pointers to each auxiliary files by file ID
  - div (sub-instance 2) - "Level02" for the "documentEntire". Contains a sub div ("Level03") for each "related item" indicated in the dmdSec where content files are correlated to the descriptive metadata by way of the DMDID (which is based on time stamp).
- div (sub-instance 2) - points to the fileID for the data container