Getting it together!

Automating Standardized Technical Metadata for Images and Audio

Jody L. DeRidder
University of Alabama Libraries
DLF 2015
October 27, 2015

Why do we do this?

- Standardized technical metadata
- Future migrations, emulations, access

But why automate the process?

- Lower the bar for necessary expertise
- Reduce costs
- High production environment

Overview

- FITS for validation
- MIX for images
- FITS -> valid MIX for images
  - *Software*
- AES57 for audio
- Spreadsheet & FITS -> AES57 for audio
- Database entries
  - *More Software*

FITS for Validation

File Information Tool Set (FITS)

- A wrapper for multiple open source tools
- Compares their results
- Identifies, validates, extracts technical metadata
- Provides a single XML output
- Can generate partial MIX file

Harvard: http://projects.iq.harvard.edu/fits

FITS for Validation

Per JHOVE developer Gary McGath:

- **Invalid** = has errors that reduce functionality
- **Not well-formed** = is unusable

http://www.garymcgath.com/jhovenote.html

FITS Errors

- Unknown TIFF IFT tag
- Type mismatch for tag 700, expecting 1, saw 7.
- Unknown data type Type = 0, Tag = 0; and Tag 0 out of sequence
- File format conflict and FocalPlaneResolutionUnit value out of range: 4.
- Value offset not word-aligned; Tag out of sequence, Unknown TIFF IFD tag.

FITS Errors

Common causes:

- Camera Raw files with TIFF extensions
- Captured with UMAX Magic Scan
- Viewed in Windows Photo Viewer
- Renamed in Bridge
- Problems with Adobe Raw Converter 7.1
- Thumbnails embedded in TIFF

MIX for Images

NISO **M**etadata for **I**mages in **X**ML Schema

- Basic Digital Object Information
- Basic Image Information
- Image Capture Metadata
- Image Assessment Metadata

http://www.loc.gov/standards/mix/

FITS -> Valid MIX

Additions to FITS MIX:

- dateTime Created
- Image Producer
- Capture Device
- ID Type & Value
- iccProfile (if in color)
- Format name & version
- Checksum information
- File Size
- Analog source type (optional)

Software

http://bit.ly/1UdfFFi

fits2mix.pl

-- Requires FITS, TIFF files, name of producer
-- Generates FITS and MIX files
-- OR if not valid/well-formed/TIFF,
copies TIFF to repairMe directory

Creative Commons Attribution 3.0 Unported License

AES57 for Audio

- Physical properties of analog
- Signal characteristics
- Digital file characteristics
- Condition comments

One AES57 file per analog object

http://bit.ly/1J1oAtA

AES57 for Audio

- **Audio Object:** format, byteOrder, block size, use, checksum, etc.
  - **Faces:** startTime, duration, direction, label
    - **Regions:** startTime, duration, channels, condition, security note, label
      - **Streams:** channel number, position, condition, label

Spreadsheet & FITS -> AES57

- identity@format
- identity&@mimetype
- byteOrder
- blockAlign
- audioDataEncoding
- md5checksum

From FITS

- offset
- lastModified
- wordSize
- channel
- bitDepth
- sampleRate
- duration

Spreadsheet & FITS -> AES57

- File name
- Title
- Format
- Clip begin
- Clip end
- Notes: region, streams
- Prior/end streams
- Direction
- Label
- Speed correction

From spreadsheet

Spreadsheet & FITS -> AES57

Type of original assigned based on RDA format

Example:

1 sound tape reel : analog (41:04 min); 7 inches

-> analogTapeFormatRegionType

Other region types for:

Digital tape, analog disc, optical disc, cylinder, wire

Spreadsheet & FITS -> AES57

Change in direction or speed correction = new "face"

ID	Direction	Speed Correction
Item4_01	FORWARD	
Item4_02	REVERSE	
Item4_03	REVERSE	1.5

Spreadsheet & FITS -> AES57

Omitted regions calculated from duration and time codes, and added

ID	Clip Begin	Clip End	Prior Streams	Ending Streams
Item4_01	02:23	05:48	Silence	
Item4_02	06:16	08:02	Applause	
Item4_03	08:36	10:12	Applause	Applause

3 Entries: 7 regions

Spreadsheet & FITS -> AES57

Comments and condition notes

… down to the second

Example in a single region entry:

[12:23-12:30] crackly; Security: phone number; [13:01-14:12] background whine

- Security note for entire region
- Separate condition notes for specified ranges

Spreadsheet & FITS -> AES57

Seconds * sampleRate * speedCorrection = duration

Example, with sample rate 44100 Hertz

ID	Clip Begin	Clip End	Speed Correction
Item4_03	08:36	10:12	1.5

((8*60) + 36) * 41000 * 1.5 = startTime

(((10*60) + 12) * 41000 * 1.5) – startTime = Duration

Database Entries

- Identifier
- Mime type
- Format
- Format version
- Contains thumb *or* Channels (audio)
- Format Registry (PRONOM)
- Format Registry Key

And conflicts!!

More Software

http://bit.ly/1PbdxwQ

fits2aes.pl

-- Requires FITS, spreadsheet, WAV files

-- Generates FITS and AES57 files

See Code4Lib article October 2015

Creative Commons Attribution 3.0 Unported License

Thank you!

Questions?

fits2mix.pl http://bit.ly/1UdfFFi
fits2aes.pl http://bit.ly/1PbdxwQ

Jody L. DeRidder
Head of Metadata & Digital Services
University of Alabama Libraries
jlderidder@ua.edu