Media Archiving IT Challenges at the Library of Congress

Photo by Bieberdorf

LIBRARY OF CONGRESS

National Audio Visual Conservation Center
Packard Campus for Audio Visual Conservation
http://www.loc.gov/avconservation/packard/

Media-related Archiving IT challenges

- Designing & implementing media-related IT technologies that are as standardized as possible
- Designing & implementing media production/migration systems that can work on a mass scale (10s to 100s of thousands of units per year)
- Designing & implementing data structures (files) that will last at least as long as the physical media they came from
- Begin to ingest Born Digital media files from media producers
  - They start as files, why shouldn't they come to us as files?
- Extending the life of physical data media: can we make at least one of our copies one that lasts decades longer?
  - Film lasts more than a century and can still be played in standardized machines: Can we do a version of that with data storage?
  - Lower the TCO of at least one of the copies?

Working On Technical Challenges

How do we maintain the benefits of IT economies of scale while meeting the long-term needs of media archiving?

MXF: Media eXchange Format

- ISO standard by SMPTE (Society of Motion Picture & Television Engineers) via the AMWA (Advanced Media Workflow Association)
- Standardized media file definition to allow a common file design to be used across multiple vendor platforms & usage designs
  - Promotes interoperability between vendors
  - Allows metadata & other essential associated data to be included within file
  - Not vendor-specific, lowering licensing costs & obsolescence issues (unlike .avi [Microsoft] & .mov [Apple Computer] file formats)
- 7 major flavors adapted for specific functional requirements: Operational Patterns (labeled “OP-##”)
- Use-specific focused file definitions for common use requirements: Application Specifications (labeled “AS-#”)
- Tightly defined definitions for specific implementations: ‘shims’
  - For specific needs or implementations

MXF: Media eXchange Format

- Media production adopters:
  - CBS/Paramount TV
  - Disney/ABC
  - NBC Universal
  - Public Broadcasting Service
  - BBC
  - CBC
  - Deutsche Welle
  - Digital Cinema Initiative
    - AMC Entertainment
    - Regal Entertainment
    - 20th Century-Fox Studios
    - Paramount Pictures
    - Universal Pictures
    - Sony Pictures/Columbia Pictures
    - Walt Disney Company
- Media equipment vendor adopters:
  - Omneon
  - Harris
  - Front Porch Digital
  - AmberFin
  - OpenCube/EVS
  - DVS (Digital Video Solutions)
  - Dolby
  - Christie
  - Barco
  - Doremi
  - Qube Cinema

MXF: Media eXchange Format

- Archive-specific Application Specification (AS) being worked on currently by LOC, NARA and other partners
- Intended to address archives & libraries' needs to have enhanced metadata fields, associated essences and greater versioning abilities

AXF: Archive eXchange Format

AXF: Archive eXchange Format

- Candidate standard being worked on by SMPTE
- Develop a standardized file organization structure that is media agnostic and can be written across any type of data recording media without being adapted for media-specific requirements
  - Standardized metadata fields
  - Forward error correction (redundant data inserted in files to recover a complete copy without accessing an external 2nd or 3rd copy)
  - Media materials allow for error concealment & error correction under defined conditions
  - Cryptographic hash checksums: entire files and internal file elements
  - Agnostic to storage media: disk, tape, solid state, holographic, others????
- Now in drafting stages
- More details: www.smpte.org/standards
  - Look for TC-31FS30 committee

Media Longevity Studies

Media Longevity Studies

- How do we store data for longer periods of time?
- How do we recover data from media that are dozens if not hundreds of years old?
- Can we engineer better longevity of the media?
- Can we engineer better machines that last longer?
- Can we maintain designs that can be easily reproduced decades or centuries in the future?

Media Longevity Studies

- Methods:
  - Study current materials – media and machinery
  - Research new materials via materials sciences
  - Look at previous experiences in error coding (IE ‘entropy coding’) and forward error correction (FEC) science
    - Well studied in the media production & transmission fields:
      - Used since the 1980s in digital audio and videotape
      - Version is built into MPEG-2 and DVB broadcast & satellite transport standards

        How to apply it to non-transmission data storage technologies?
- We’re at the beginning….stay tuned!

Standardized Metadata Schemas

Standardized Metadata Schemas

- Develop standardized metadata schemas that can serve the specific needs of media archiving of all types
- Document them well and publish as open standards without licensing agreements
- Create a clear versioning scheme to allow for future growth and change without orphaning previous schemas

Questions?

Thank you