Audio and Video at Scale: Indiana University's Media Digitization and Preservation Initiative
Monday, October 27, 3:45am-5:15pm
Salons 4,5,6, Georgia Tech Hotel and Conference Center

Colocated with: Managing the Digitization of Large Press Archives (Group notes: https://docs.google.com/document/d/1fOSWBzMQ5PXQoKisJAEhyDUM8JMkc88SMC5v_zon9sc/edit#heading=h.v5ec1r31vuuf)

Colocated with: Building a Ten-Campus Digital Library Collection at the University of California (Group notes: https://docs.google.com/document/d/1tOcC09wNylyaRXhd68egnSLp5Nwqg5-l8oVDRudVk/edit#)

Session Leaders
Juliet Hardesty, Indiana University
Jon Dunn, Indiana University

Slides
http://www.slideshare.net/DLFCLIR/audio-and-video-at-scale-dunn-hardesty

Notes
We have a 15-20 year window of opportunity to digitally preserve audio and video holdings. Digitization of a large quantity of audio and video materials. In 2009 IU-Bloomington initiated a media survey; part of larger study of special collections at university, particularly those outside of library and archives locations. Span over 50 different media formats. Both magnetic and optical; actively degrading; formats that are rapidly becoming obsolete. “degralescence” (degradation and obsolescence) High risk of loss of content. Scarcity of equipment and parts for that equipment.

2011 report: “Meeting the Challenges of Media Preservation” Includes creation of facility to digitize these materials; AV in libraries heavily influenced by AV in production world; opens opportunities for access. Collaborative process.

2013: Media Digitization and Preservation Initiative. IU President announces project to digitize it all by 2020 (accelerating the time frame); undertaking as public-private partnership with Memnon Archiving Services in Brussels, Belgium. Ability to do rapid, massive transfer…
4 phases: (1) pre-digitization, prioritization, inventorying, forming content into batches for high throughput digitization process; (2) IU is also setting up digitization lab on IU Bloomington campus focused on unique, one-to-one formats that need attention rather than mass approach Collaboration among staff from IT, Libraries, units holding materials

“Bins, Boxes, Barcodes, Batches”
everything that goes out to Memnon gets a barcode for tracking; groups of barcoded items in boxes go into bins; bins get barcoded; bins gathered into groups that go over in groups that make up batches. Batches are format-based. One bin might have nothing but vinyl LPs or nothing but CDs, for example.
Starting with Music library and Archives of Traditional Music. Open reel tapes, DATs, CD-Rs. Starting with collections that can fill batches to get process well underway. Sources of descriptive metadata include library catalog and media score survey (?)
Starting with audio formats.
Information technology challenges - tight timeline, must have enough material to keep digitization process underway.
Developed in-house system to manage project: physical object database (Ruby app with MySQL backend). Different statuses can be applied to whole batches of items. Used Agile methodology to create tracking database. Started with user stories to determine functionality requirements.
Memnon sending to IU:
Preservation master (audio-broadcast wav, 96/24; video-10-bit uncompressed QT.mov);
Mezzanine file; access file; metadata (checksums and process history)
Post-digitization workflow: MDPI item - scholarly data archives (mass storage system)-initial validation of files-manual QC for portion of items-simultaneous SIP/AIP and at same time mezzanine prod file will be used to create derivatives for access repository.
Using different types of access technology - IUCAT, Avalon, Archives Online, etc

Preservation repository - don’t have one yet; generating over 9PT of data in total; as much as 10TB per day...Anticipate placing directly in UITS Scholarly Data Archive, still missing management layer (DAMS). Also keeping tabs on APTrust and DPN development.

Challenges:
- rights issues at scale
- descriptive metadata and discovery
- quality ctrl strategies for mass digitization
- strategies for born digital
- out-of-region preservation storage
- approach for film (right now doing audio and video) film considered more stable as an analog, and carries larger storage challenges.

http://mdpi.iu.edu
Reports and background:
www.indiana.edu/~medpres/

Q - how much does this cost?
  - orig press release 15M; but that's not the whole story - in-kind time/effort is much higher.