2009 Media Preservation Survey Findings

- More than 560,000 audio, video, and film objects are owned by the Bloomington campus on more than 50 formats housed in more than 80 units
- Actively degrading, obsolete formats, high risk of loss of content over the next decade
- We have a 15- to 20-year window of opportunity to digitally preserve audio and video holdings.
2011 “Meeting the Challenges of Media Preservation”

Key Recommendations
- Preservation planning
- Facility Development
- Facility operation and workflow development
- Prioritization
- Strategies for Film
- Technology infrastructure needs
- Access
- Collaboration
2013 Announcement: Media Digitization and Preservation Initiative

Goal: Digitize, preserve, provide access to rare and unique audio and video by 2020 – all IU campuses
MDPI Leadership

MDPI Co-Chairs

Carolyn Walters
Ruth Lilly Dean of University Libraries

Brad Wheeler
Vice President for Information Technology and Chief Information Officer

MDPI Executive Director

Laurie G. Antolovic
Associate Vice President and Deputy Chief Information Officer

PASIG 2015
March 12, 2015
Public-Private Partnership

Memnon Archiving Services
Brussels, Belgium

Solutions for small and institutional archives
Digitization and Preservation: The Phases

Pre-Digitization

- Inventory
  - Catalog
  - Prioritize
  - Batch & Queue

Digitization

Memnon
Massive parallel digitization

IU Operation
Digitization of selected unique and highly vulnerable formats
Quality control

Digitization and Access

- Metadata
  - Rights Issues
- Technical aspects

Discovery and Access

- Metadata
  - Technical infrastructure
- Ongoing monitoring and migration

Digital Preservation and Storage
Organizational Structure

- IU Operations
- Memnon Archiving Services, USA
- IT Team
- MDPI Operational Task Force
Batches: Format Based
Strategy

• Memnon – parallel transfer workflows
• IU – 1:1 workflows for fragile formats and items
• Digitally preserve approximately 300,000 recordings in four years
IU Facility Functions

- Audio Digitization
- Video Digitization
- QC of Memnon and IU output
- Clean, Repair, Restore, Metadata, Etc.
Timeline

• Renovation complete mid-March?
• IU audio digitization begins late-April
• Memnon audio digitization begins May
• Starting formats: open reel tape, LPs
• Next formats: DAT, Betacam SP
Preparations

• Starting with:
  • William and Gayle Cook Music Library
  • Archives of Traditional Music
  • Open reel tapes, LP’s, DATs, and Betacam SP
What is Memnon sending to us

- **Preservation Master File**
  - Audio: Broadcast WAV (96/24)
  - Video: 10-bit uncompressed (QT .mov)
- **Mezzanine File**
  - Audio: Broadcast WAV (96/24)
  - Video: 50 Mbps Iframe-only MPEG-2
- **Throw-away Access File (for QC)**
  - Audio: MPEG-4 AAC
  - Video: MPEG-4 H.264
- **Metadata**
  - checksums, process history
Workflow Support: POD

- Physical Object Database (POD)
- Support for tracking of objects and construction of batches
- Populated from MediaSCORE/MediaRIVERS and IUCAT
- Ruby-on-Rails based web application using MySql database
- Agile development process
<table>
<thead>
<tr>
<th>Unit</th>
<th>Format Type</th>
<th>2009 Total</th>
<th>2014 Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropology</td>
<td>CD-R</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Archives of African American Music and Culture (AAAMC)</td>
<td>CD-R</td>
<td>6</td>
<td>160</td>
</tr>
<tr>
<td>Archives of Traditional Music (ATM)</td>
<td>CD-R</td>
<td>59</td>
<td>437</td>
</tr>
<tr>
<td>Bands, Department of (Jacobs)</td>
<td>CD-R</td>
<td>282</td>
<td>12</td>
</tr>
<tr>
<td>Black Film Center Archive</td>
<td>CD-R</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Center for Language Technology and Instructional Enrichment</td>
<td>CD-R</td>
<td>3613</td>
<td>6272</td>
</tr>
<tr>
<td>Center for the Study of Global Change</td>
<td>CD-R</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Center of International Education and Development Assistance</td>
<td>CD-R</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
High Level Process Overview

Original Media Batches

Mass Digitization

Digitized Media Mass Storage

New File Detection

Automated File Processing

QC Checks
QC File Generate
Create Derivatives
Metadata Build
SIP Generation
Update Item Status

Manual QC Check on % of files

Digitized Media Mass Storage

Fedora Preservation System

Avalon Media Player
High Level Systems Overview

- Physical Object Database (POD)
- Batch Processing system
- FFMpeg processing module
- QCTools analysis system
- AudioInspector
- Scholarly Data Archive (SDA)
Digital Preservation and Access

- Media Files and Metadata
  - Masters, Mezzanines
  - Transcodes

- Digital Preservation Repository
- Access Repository
Digital Preservation Repository

- Bit Storage:
  - IU Scholarly Data Archive (SDA)
  - Mirrored between IUB and IUPUI

- File and Metadata Management:
  - HydraDAM2
HydraDAM2

- NEH-funded collaboration with WGBH/Boston, 2015-2016
- Goal: Preservation repository system for time-based media
- Based on:
  - Fedora 4 digital repository software
  - Hydra framework
Access Repository: Avalon Media System

- Co-developed by IU and Northwestern University
- Funding from IMLS, 2011-2015
- New funding from Mellon, 2015-2016
- In production at multiple institutions
- Uses Hydra/Fedora technologies
- Leverages institutional streaming infrastructure

PASIG 2015
March 12, 2015
Avalon Functionality for MDPI

- Basic searching and browsing
- Item-level access control
- Linking and embedding
  - Omeka plugin for exhibits
- Coming Soon:
  - Structural metadata
  - Captioning
  - Transcripts
Long-Term Preservation Technology

• 9 petabytes+ to be preserved
• Local storage
  • UITS Scholarly Data Archive
  • Fedora 4 repository layer
• Out-of-region storage
  • APTrust, DPN
• Data swap agreements
MDPI Challenges

- Dealing with rights issues at scale
- Descriptive metadata and discovery
- Quality control strategies for mass digitization
- Strategies for born-digital media
- Out-of-region preservation storage
- Approach for film
Storage Growth

• Chart goes here