From Mass Digitization to Description: Indiana University’s Strategy to Overcome the Next Great Challenge

Chris Lacinak, AVPreserve  
Jon Dunn, Indiana University

AMIA Conference - Pittsburgh  
November 10, 2016
MDPI:
Media Digitization and Preservation Initiative

- Goal: “To digitize, preserve and make universally available by IU’s Bicentennial—subject to copyright or other legal restrictions—all of the time-based media objects on all campuses of IU judged important by experts.”
- 280,000+ audio and video items
- ~7PB over 4 years
- 9TB per day peak
- 80 different units
- 20+ different physical formats
Public-Private Partnership for Digitization

Memnon Archiving Services, A Sony Company
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

Discovery and Access
- Metadata
- Rights Issues
- Technical aspects

Digital Preservation and Storage
- Metadata
- Technical infrastructure
- Ongoing monitoring and migration
Progress to Date

Indiana University's comprehensive work to preserve historical and cultural time-based media for the research, education, and enrichment of future generations.

Why we're doing this

154,398 of 280,000 ITEMS PRESERVED TO DATE

mdpi.iu.edu
Many Types of Units... Some Examples:

Libraries
- Herman B Wells Library
- Cook Music Library
- Lilly Library

Archives
- Archives of African American Music and Culture
- Archives of Traditional Music
- Black Film Center/Archive
- IUL Moving Image Archive
- University Archives

Museums
- Sage Costume Collection
- Mathers Museum

Production Units
- Radio and Television Services
- Media Design and Production

Performing Groups
- African American Arts Institute
- Department of Bands

Academic Units
- School of Journalism
- Department of Folklore and Ethnomusicology
- Jacobs School of Music

Research Centers and Institutes
- Kinsey Institute
- Russian and East European Institute

Athletics!
Rights

- Goal: “make universally available”
- Many rights situations:
  - Copyright held by IU
  - Copyright held by others
  - Copyright held by others with license or deed of gift to IU
  - Public domain
  - Copyrighted orphaned work
  - Copyright status and ownership unknown
- Also, ethical and cultural considerations
- Decision trees being developed to guide rights and access determinations
Existing metadata sources: IUCAT
## Existing metadata sources: POD

**Physical Object Database**

### Physical Object [132330]

- **Assignment**: University Archives Betacam SP
- **Bin**: 40000002330068
- **Box**: Not Assigned
- **Batch**: IU-0001-BSP

### Betacam Technical Metadata

<table>
<thead>
<tr>
<th>Pack Deformation</th>
<th>Preservation problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format Version</th>
<th>Cassette Size</th>
<th>Recording Standard</th>
<th>Format Duration (min)</th>
<th>Tape Stock Brand</th>
<th>Image Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>large</td>
<td>NTSC</td>
<td>60</td>
<td>Sony BCT 60MLA</td>
<td>4:3</td>
</tr>
</tbody>
</table>

- **MDPI barcode**: 4000000063815
- **Title**: Interviews: From John Ryan’s Orli
- **Call number**: 2014/001 35
- **IUCAT barcode**: GR00100653
- **Year**: 1
- **Group key**: GR00100653
- **Group position**: 1
- **Carrier stream index**: GR00100653_1.1
- **Imported from spreadsheet**: Binned >> Batched >> Shipped
- **Workflow Status**: Binned >> Batched >> Shipped
Access: Avalon Media System

The image shows a screenshot of the Avalon Media System interface. The interface is designed for managing and accessing media content. It includes a search bar and options to browse by various categories such as format, sound recording, moving image, main contributor, date, genres, collection, and unit. The interface also allows for searching by relevance and displaying a limited number of items per page. The main content lists media items, including titles, the number of contributors, dates, and other details. The background includes logos for INDIANA UNIVERSITY and other institutions like Northwestern and the Andrew W. Mellon Foundation.
Digital Preservation: HydraDAM2

- Media Files and Metadata
  - Masters, Mezzanines
- Digital Preservation Repository (HydraDAM2)
- Access Repository (Avalon)
- SDA
  - Out-of-Region Storage
Bit Storage: Scholarly Data Archive
Collection and Metadata Examples
Program, 1996-1997, no. 384

Corporate name  Indiana University Concert Orchestra.
Title  [Program, 1996-1997, no. 384] [sound recording]
Format  Musical Recording
Published/Produced  1996.
Physical description  1 sound cassette : digital, stereo
Performer  Edwin Anderson, bass trombone (1st work) ; Atar Arad, viola (3rd work) ; Indiana University Concert Orchestra ; Keith Brown, conductor.
Recording info  Recorded on November 20, 1996, Musical Arts Center, Indiana University, Bloomington.
Contents  Concerto for bass trombone, piano and strings (1966) / David Anderson (24:27)  Symphony no. 8 in F major, op. 93 / Ludwig van Beethoven (28:08)  Harold in Italy, op. 16 / Hector Berlioz (47:30).
Subject  Concertos (Bass trombone with string orchestra)  Symphonies  Symphonies (Viola with orchestra)
Program, 1996-1997, no. 1

Corporate name: Indiana University School of Music
Title: [Program, 1996-1997, no. 1]
Format: [Musical recording]
Published/Produced: 1996
Physical description: 1 sound cassette
Performer: Edwin Anderson, bass trombone; Atar Arad, viola; Indiana University Concert Orchestra; Keith Brown, conductor
Recording Info: Recorded on November 20, 1996, Musical Arts Center, Indiana University, Bloomington
Contents: Concerto for bass trombone, piano and strings (1996) / David Anderson (24:27) -- Symphony no. 8 in F major, op. 93 / Ludwig van Beethoven (28:06) -- Harold in Italy, op. 16 / Hector Berlioz (47:30)
Subject: Concertos (Bass trombone, piano and strings) / Symphonies (Viola and orchestra)
Related name: Brown, Keith, d. 1933 October 21- / Anderson, Edwin / Arad, Atar
Includes: Anderson, David, d. 1962- / Concertos, m. trombone, string orchestra
Uniform series: Indiana University, Bloomington. b: School of Music. l: Program; v: 1996-1997, no. 384

[Program, 1996-1997, no. 384]

Date
1996

Main contributor
Indiana University Concert Orchestra.

Contributors
Indiana University Concert Orchestra, Brown, Keith, 1933 October 21.; Anderson, Edwin; Arad, Atar; Anderson, David, 1962; Concertos, trombone, string orchestra.; Beethoven, Ludwig van, 1770-1827. Symphonies, no. 8, op. 93, F major.; Berlioz, Hector, 1803-1869. Harold en Italie.

Subjects
Concertos (Bass trombone with string orchestra); Symphonies; Symphonies (Viola with orchestra)

Collection
William and Gayle Cook Music Library

Unit
William and Gayle Cook Music Library

Physical Description
1 sound cassette: digital, stereo.

Notes
PERFORMERS
Edwin Anderson, bass trombone (1st work), Atar Arad, viola (3rd work), Indiana University Concert Orchestra; Keith Brown, conductor.

CONTENTS
Concerto for bass trombone, piano and strings (1996) / David Anderson (24:27) -- Symphony no. 8 in F major, op. 93 / Ludwig van Beethoven (28:08) -- Harold in Italy, op. 16 / Hector Berlioz (47:30).
Dan Burton congressional papers, 1983-2012

Author: Burton, Dan, 1938-
Title: Dan Burton congressional papers, 1983-2012.
Format: Mixed Materials
Description: 123 linear feet (123 boxes)
video and audio 1049 tapes
electronic records 6.97 gigabytes
URL: <Finding aid for collection> http://purl.dlib.indiana.edu/iudi/findingaids/politicalpapers/VAD2039
Variant title: Burton papers
Notes: Select files closed to researchers.
Advance notice is required for access to the collection.
Danny Lee Burton (b. 1938) represented Indiana in the United States House of Representatives from 1983-2012. Republican Dan Burton was elected to represent the 5th and 6th congressional districts of Indiana throughout his time in Congress.
Materials are in English.
Abstract: The collection consists of papers, audio-visual materials, and electronic records generated and received by the office of Congressman Dan Burton.
Subject headings: Burton, Dan, 1938---Archives.
Burton, Dan, 1938---Correspondence.
United States. Congress--Records and correspondence.
Legislators--Indiana--History--20th century--Sources.
Legislators--Indiana--History--21st century--Sources.
Dan Burton Congressional Papers, 1983-2012


Finding aid prepared by Hope Grebner and Nicole Biegansi, Project Archivists

Summary Information

Repository
Modern Political Papers Collection, Indiana University Libraries
Herman B Wells Library E480
1320 East Tenth Street
Bloomington, IN 47405-7000
Phone: 812-855-4601
Email: congpol@indiana.edu
http://www.libraries.indiana.edu/modern-political-papers

Creator
Burton, Dan, 1936-

Title
Dan Burton Congressional Papers, 1983-2012

Collection No.
MPP 21

Extent
123 linear feet (123 cartons), 1049 video and audio tapes, and 6.97 gigabytes of electronic records

Language
Materials are in English.

Abstract
Danny Lee Burton (b. 1936) represented Indiana in the United States House of Representatives from 1983-2012. Republican Dan Burton was elected to represent the 6th and 8th congressional districts of Indiana throughout his time in Congress. The collection consists of papers, audio-visual materials, and electronic records generated and received by the office of Congressman Dan Burton.

Access Restrictions
This collection is open for research, with the exception of the unpublished records of the House of Representatives Committees on Government Reform and Oversight and Foreign Affairs (formerly the Committee on International Relations). Select files from these series may be available to researchers after review by political papers archivists.
<table>
<thead>
<tr>
<th>Box 96</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frontline Special (VHS)</strong>, Oct 6 1998</td>
</tr>
<tr>
<td><strong>Sunday Morning Shows NBC Meet the Press, McLaughlin (VHS)</strong>, Nov 18 1998</td>
</tr>
<tr>
<td><strong>Crossfire (VHS)</strong>, Dec 1 1998</td>
</tr>
<tr>
<td><strong>ABC Nightly News, O'Reilly Factor - Monica Lewinsky and Other General News (VHS)</strong>, 1999</td>
</tr>
<tr>
<td><strong>Open 30 Seconds, Close 30 Seconds (Betacam SP)</strong>, Feb 23 1999</td>
</tr>
<tr>
<td><strong>NBC - Fleecing of America: Housing/HUD Story (3:30) (VHS)</strong>, Feb 25 1999</td>
</tr>
<tr>
<td><strong>Dateline - Juanita Broaddrick (VHS)</strong>, Mar 1999</td>
</tr>
<tr>
<td><strong>Oprah - Child/Men Abuse Women (VHS)</strong>, Mar 1 1999</td>
</tr>
<tr>
<td><strong>O'Reilly Factor (VHS)</strong>, Mar 16 1999</td>
</tr>
<tr>
<td><strong>O'Reilly Factor (VHS)</strong>, Mar 16 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton with Representative Stephen Horn (20:54) (Betacam SP)</strong>, Mar 24 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton with Representative Stephen Horn (20:54) (VHS)</strong>, Mar 24 1999</td>
</tr>
<tr>
<td><strong>Fox News - Carl Cameron Report on Chinese Espionage and Campaign Funds (VHS)</strong>, Apr 8 1999</td>
</tr>
<tr>
<td><strong>Yugoslavia - Death of a Nation (VHS)</strong>, Apr 9 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton Guest Jane Seymour VR#17 (29:33) (VHS)</strong>, Apr 19 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton Guest Stephen Horn Recorded Mar 24 1999 VR#12 (20:54) (VHS)</strong>, Apr 19 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton Guest Jane Seymour VR #16 (29:33) (VHS)</strong>, Apr 19 1999</td>
</tr>
<tr>
<td><strong>Representative Burton with Governor Tommy Thompson VR#15 (30:30) (VHS)</strong>, Apr 22 1999</td>
</tr>
<tr>
<td><strong>Representative Dan Burton Two Promos Each 30 Seconds 1st Promo with Pete Sessions 2nd with Tommy Thompson VR#18 (.60) (VHS)</strong>, Apr 25 1999</td>
</tr>
<tr>
<td><strong>Evening News - Chung, ABC, CBS, NBC, Lehrer (VHS)</strong>, May 11 1999</td>
</tr>
<tr>
<td><strong>Moneyline - Chung (VHS)</strong>, May 11 1999</td>
</tr>
<tr>
<td><strong>House Committee Hearing - Chung (Tape 1 of 3) (VHS)</strong>, May 11 1999</td>
</tr>
<tr>
<td><strong>House Committee Hearing - Chung (Tape 2 of 3) (VHS)</strong>, May 11 1999</td>
</tr>
<tr>
<td><strong>House Committee Hearing - Chung (Tape 3 of 3) (VHS)</strong>, May 11 1999</td>
</tr>
</tbody>
</table>

Date
unknown/unknown

Main contributor
Unknown

Collection
Office of University Archives and Records Management

Unit
Office of University Archives and Records Management

Other identifier
Other: GR00157239

Sections
1. Betacam 1/1 Part 1 (40000000659635) (31:33)
2. Segment 1 (00:36)
Colombia, Atlántico, Bolívar, Córdoba and Magdalena, 1965

Author/Composer, etc: List, George, 1911-2008.

Format: Non Musical Recording

Title: [Colombia, Atlántico, Bolívar, Córdoba and Magdalena, 1965] [sound recording] / collected by George List.

Physical description: 30 sound tape reels: analog, 7 1/2 ips, full track; 7 in. 4 slides and documentation.

Language: Sung and spoken in Spanish.

Performer: Various native informants and musicians.

Recording info: Recorded by List from February 19-April 6, 1965 in the Departamentos de Atlántico, Bolivar, Cordoba, and Magdalena.

Notes: Interviews, tales and folk music. Accompanied by a complete catalog of the tapes in Spanish with some English translations, and biographical information on the performers and informants.

Local note: Slides housed in slide file cabinet in vault.

Summary: Interviews with musicians concerning various musical genres, instruments, velorios for children, folktales and prayers. Includes demonstrations of musical instruments and performances by various musical groups.

Terms of use: Archival use and reproduction for non-profit educational or research purposes. Archives of Traditional Music.

Subject: Folk music--Colombia.
Tales--Colombia.
Funeral rites and ceremonies--Colombia.
**Collection title:** Colombia. George List, 1965.

<table>
<thead>
<tr>
<th>shelf no. / item no.</th>
<th>duration</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>4:05</td>
<td>“Vicente se fue a bañar” (Juan Lara). Porro. Ensamble completo: gaita hembra, gaita macho con maraca, tambor mayor y tambor llamador.</td>
</tr>
<tr>
<td>05</td>
<td>3:12</td>
<td>“Las cinco notas” (Juan Lara). Porro. Ensamble completo.</td>
</tr>
<tr>
<td>06</td>
<td>1:27</td>
<td>Conversación sobre cuando se usa letra en la canciones.</td>
</tr>
<tr>
<td>07</td>
<td>2:02</td>
<td>“Las cinco notas” (Juan Lara). Porro. Solo las dos gaitas, sin percusión.</td>
</tr>
<tr>
<td>08</td>
<td>0:24</td>
<td>“Las cinco notas” (Juan Lara). Porro. Solo gaita macho.</td>
</tr>
<tr>
<td>09</td>
<td>3:15</td>
<td>Conversación sobre el nombre de la canción.</td>
</tr>
<tr>
<td>10</td>
<td>1:53</td>
<td>Conversación sobre el compositor de las tres canciones que se tocaron.</td>
</tr>
</tbody>
</table>
## Physical Object [138450]

<table>
<thead>
<tr>
<th>Assignment</th>
<th>RTVS Betacam SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picklist</td>
<td>400000000003378</td>
</tr>
<tr>
<td>Bin</td>
<td>Not Assigned</td>
</tr>
<tr>
<td>Box</td>
<td>M-0004-BSP</td>
</tr>
</tbody>
</table>

### MDPI barcode

| MDPI barcode        | 40000000784936         |

### Format

| Format      | Betacam               |

### Has ephemera

| Has ephemera | False                  |

### Ephemera returned

| Ephemera returned | False                  |

### Unit

| Unit | B-RTVS                |

### Title

| Title | "A Conversation with Herman B W" |

### Call number

| Call number |                                   |

### IUCAT barcode

| IUCAT barcode |                                   |

### Year

| Year |                                   |

### Group key

| Group key | GR00106221                         |

### Group position

| Group position | 1                                   |

### Carrier stream index

| Carrier stream index | GR00106221_1_1                   |

### Imported from spreadsheet

| Imported from spreadsheet |                                   |

### Workflow Status

| Workflow Status | Unpacked                           |

### General Note

| General Note |                                   |

### Include in Shipping Manifest

| Include in Shipping Manifest | False                           |

### Notes

| Notes | dub of one inch #1356 apx. 29 mins |

## Betacam Technical Metadata

### Pack Deformation

| Pack Deformation |                                   |

### Preservation problems

| Preservation problems | SP                                  |

### Format Version

| Format Version | small                              |

### Cassette Size

| Cassette Size | small                              |

### Recording Standard

| Recording Standard | NTSC                               |

### Format Duration (min)

| Format Duration (min) | 30                                  |

### Tape Stock Brand

| Tape Stock Brand | 3M                                  |

### Image Format

| Image Format | 4:3                                 |
"A Conversation with Herman B. Wells"

Date
unknown/unknown

Main contributor
Unknown

Collection
Radio and Television Services

Unit
Radio and Television Services

Other Identifier
Other: GR00106221

---

Sections

1. Betacam 1/1 Part 1 (40000000784936) (31:10)
2. Segment 1 (00:31)
3. Segment 2 (00:18)
4. Segment 3 (28:42)
Hours

280k
Core Principles

• Technological Heroism
• Human Assisted Computing
• People First
• Phased and Iterative
• Service Oriented Architecture
Methodology
<table>
<thead>
<tr>
<th>Applied permission</th>
<th>Notes</th>
<th>Date other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied perm note</td>
<td>Parts/components</td>
<td>Event</td>
</tr>
<tr>
<td>Collection</td>
<td>Physical description</td>
<td>Geographic other</td>
</tr>
<tr>
<td>*Date created</td>
<td>*Publication status</td>
<td>Keywords</td>
</tr>
<tr>
<td>*Default permissions</td>
<td>Relationship</td>
<td>Language</td>
</tr>
<tr>
<td>*Department/Unit</td>
<td>*Rights holder</td>
<td>Publisher</td>
</tr>
<tr>
<td>Duration</td>
<td>*Rights status</td>
<td>Beats per minute</td>
</tr>
<tr>
<td>Full text</td>
<td>*Role (for names)</td>
<td>Color info</td>
</tr>
<tr>
<td>Geographic Identifiers</td>
<td>Sound/silent</td>
<td>Color/B&amp;W</td>
</tr>
<tr>
<td>Links &amp; relationships</td>
<td>Source</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>*MDPI barcode</td>
<td>Subjects</td>
<td>Frequency info</td>
</tr>
<tr>
<td>Music/Speech</td>
<td>Target audience</td>
<td>Gender</td>
</tr>
<tr>
<td>*Names</td>
<td>*Title</td>
<td>Genre</td>
</tr>
<tr>
<td>Type of Resource</td>
<td>Type of title</td>
<td>Music present</td>
</tr>
<tr>
<td>Year of renewal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata Generation Mechanisms
<table>
<thead>
<tr>
<th>Phase One MGMs</th>
<th>Phase Two MGMs</th>
<th>Phase Three MGMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial content ID</td>
<td>Facial recognition</td>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>Human generation</td>
<td>Linked data</td>
<td>Audio frequency analysis</td>
</tr>
<tr>
<td>Human transcription</td>
<td>Named entity recognition</td>
<td>Beats per minute detection</td>
</tr>
<tr>
<td>Legacy closed caption</td>
<td>Natural language processing</td>
<td>Chroma analysis</td>
</tr>
<tr>
<td>Legacy date &amp; timecode</td>
<td>Paper &amp; video OCR</td>
<td>Content matching</td>
</tr>
<tr>
<td>Memnon output</td>
<td>Phoneme analysis</td>
<td>Genre detection</td>
</tr>
<tr>
<td>Music vs speech detection</td>
<td>Speech to text</td>
<td>Speaker identification</td>
</tr>
<tr>
<td>Queried/Imported data</td>
<td>Subtitle recovery</td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td>Speech-to-text comparison &amp; weighting</td>
<td></td>
</tr>
<tr>
<td>Shipping manifest (POD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silence detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech-embedded metadata</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Phase One
1) Staffing & Training
2) Acquire/Develop & implement core infrastructure & MGMs
3) Rights analysis
4) Associated materials
5) Phase 2 Prep

Phase Two
1) Develop/Acquire & implement phase 2 MGMs
2) Additional system dev
3) Phase 3 prep

Phase Three
1) Develop/Acquire & implement phase 3 MGMs
2) Additional system dev
3) Focus on AI

- Required & prioritized
- Manual
- Low hanging fruit
- Semi/Automated
- Time-based MD
- Linked data
- Synthesized/Al
- Specialized/Domain specific
Without the Project
Possible Searches

• “Wells”

• “Radio and Television Services Collection”
Phase One

Title – A Conversation with HB Wells

Title (Type of Title – Alternative) – A Conversation with Herman B. Wells

Collection – Radio and Television Services

Department/Unit – Radio and Television Services

Physical Description – Betacam

Duration – 31:07

MDPI Barcode – 40000000784944

Type of Resource – moving image

Date Created – 1970/1979

Music/speech – Speech

Sound/silent – Sound

Subjects –

Indiana University (http://id.loc.gov/authorities/names/n80126369)

Universities and colleges—United States—Administration (http://id.loc.gov/authorities/subjects/sh85141150)

Universities and colleges—United States—History—20th century (http://id.loc.gov/authorities/subjects/sh2010117482)

Wells, Herman B. (http://id.loc.gov/authorities/names/n81012520)
Phase One

Names –
Mundt, Bruce (Producer)
Petranoff, Robert (Producer)
Office of University Relations (Production company)
Wells, Herman B. (Interviewee)
Clark, Thomas D. (Interviewer)
WTIU Bloomington (Production place)

Geographic (Recording location) –
United States

Parts/components – 1 of 1

Notes – Betacam is transfer from film original based on film leader at start of content. Date range is based on main library building being complete and having electricity.

Rights holder – Indiana University

Rights status – In Copyright

Year of Renewal – 2067

Publication status – Unpublished

Applied permissions – Campus only

Default permissions – Campus only
Phase One Search

• Unpublished videos
• Produced by Radio and TV
• Audio type speech
• Herman B. Wells is interviewee
• IU is rights holder
Phase Two (Additional)

**Language** – English

**Keywords** –
- Board of Trustees
- Faculty
- University administration
- University president
- Music
- Arts
- Midwest universities
- President's spouses
- 20<sup>th</sup>-century leaders

**Names** –
- Bryan, William Lowe
- Kinsey, Alfred
- Roosevelt, Eleanor
- Rockefeller Foundation

**Date other** – 1940/1970

**Full Text**
- OCR from text within video
- Speech to Text

**Geographic other**
- Bloomington (http://www.geonames.org/4254679)
- Indiana (http://www.geonames.org/4921868)
Phase Two Search

• Every place within the video
• Face of Herman B. Wells appears
• “Eleanor Roosevelt” exists in text
Phase Three (Additional)

**Color/bw** – Color

**Color Information** – [time stamp and values representing predominant colors being represented]

**Music present** – [time stamp at points at which music is present]

**Genre** – Interview

**Speaker Identification** – [time stamp and name each time a speaker changes, combined with the names field]

**Relation** – [Content matching will identify where any portions of content in this video also exist in any portions of videos within the repository]
Phase Three Searches

• All repository content with any portions of the audio or video in this video

• Information for each piece of commercial music
<table>
<thead>
<tr>
<th>Service</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Staff Hours</td>
<td>Avalon Dev</td>
</tr>
<tr>
<td>Human Generation MGMs</td>
<td>Workflow System Dev</td>
</tr>
<tr>
<td>Query and Import Hours</td>
<td>RMS Dev</td>
</tr>
<tr>
<td>SMART Assessment Hours</td>
<td>Consulting</td>
</tr>
<tr>
<td>MGM Products &amp; Services</td>
<td>Scanning &amp; Imaging Costs</td>
</tr>
<tr>
<td>MGM Dev</td>
<td>Hardware Costs</td>
</tr>
<tr>
<td></td>
<td>Phase 1</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>Cost Per Item (All In)</td>
<td>$16.26</td>
</tr>
</tbody>
</table>
A Few Takeaways

- Organizational effort and preparation
- Segmentation vs timestamping
- Lower than anticipated costs
Next Steps

1. Access Workflows task force formed
   - Develop workflows for items/collections with good metadata and clear rights

2. Continued internal messaging
   - Digitization ≠ access

3. Longer-term planning
   - Implementation plan and budget for full metadata project
   - Development of platform for workflow engine / data warehouse
   - Grant funding, collaborators?
Questions?

- mdpi.iu.edu
- avpreserve.com
- jwd@iu.edu
- chris@avpreserve.com