Why Media Preservation Can’t Wait
The Gathering Storm

Mike Casey
Director of Technical Operations
Media Digitization and Preservation Initiative
Indiana University
May 24, 2017

Website: https://mdpi.iu.edu

Slide design by Mike Lee
Why Media Preservation Can’t Wait
The Gathering Storm

• The time-based media preservation problem
  • Analog and physical digital
  • Source content for digital libraries
  • Degradation, obsolescence, and time frame
Why Media Preservation Can’t Wait
The Gathering Storm

- IU as case study
  - Survey, planning project, implementation (MDPI)
  - General to specific
Why Media Preservation Can’t Wait
The Gathering Storm

• Progress at other US institutions
Once Upon A Time...
The King & Queen of Media
“...it is alarming to realize that nearly all recorded sound is in peril of disappearing or becoming inaccessible within a few generations.”

--National Recording Preservation Board in "Capturing Analog Sound"
“in the mid- to long-term there is a major risk that carrier degradation combined with playback obsolescence will defeat the efforts of archivists...”

--International Association of Sound and Audiovisual Archives
“Audiovisual materials are the fastest growing segment of our nation’s archives and special collections.”

--The Library of Congress National Recording Preservation Plan, page 6
What is the problem?

• Large numbers
• Degradation
• Obsolescence
• High research value
• Short time window
Indiana Numbers

• More than 560,000 audio, video, film objects
  – 364,000 audio (64%)
  – 125,000 video (22%)
  – 78,000 film (14%)

• 80 units

• 50+ formats
Numbers

Study by AVPreserve and NEDCC

250 million audio recordings in the US are not digitized and are considered preservation-worthy

• Does not include video
• Does not include outside the US
Degradation

- All analog and physical digital media objects actively degrading
  - some catastrophically
Digital Audio Tape (DAT)

- Degradation?
- Mis-alignment of transport?
- 30% of DATs at IU Music Library
Obsolescence

- Media formats
- Equipment (playback machines, test devices)
- Repair parts
- Playback expertise
- Repair expertise
- Tools
- Supplies
I'm spinning like an old turntable
Three speeds going nowhere fast
I hesitate at the door to the future
Holding on to my bitter-sweet past
Me and all of the percolators
Me and all of the rotary phones
Me and all of my vinyl records
Warped and scratched and out of date
33, 45, 78

Performed by Kathy Mattea
Written by Steve Key
Obsolescence

- Technics SP-15 turntable main bearing unavailable at any price
- Styli made by one small company
- Sony PCM-70x0 series DAT machine capstan motors unavailable at any price
- No new open reel audio machines
- Audio alignment tapes—one company
- Last run of playback heads for Studer tape machines—one company left

--information courtesy of George Blood
Obsolescence

• Sony ends sales of new videotape machines
• 1” video machines, parts, playback expertise scarce
  – Type C head assembly $16,000
• Not enough working audio and video playback machines to digitize everything currently in vaults
The Non-Linearity of Obsolescence

• Technics SP-15 bearing made by Dave Cawley
• Revox open reel playback machine—semi-pro, 2017, target price $4500
• 3D printing—Tascam part—partly working
Obsolescence in Action

• 12 Panasonic AG-DS840
• Free!
• 200 dried out capacitors on each machine
Obsolescence

- Quad (video) machines, parts, playback expertise very scarce
- Standard industry professional format
- Estimated 100-200 machines left worldwide
- Ampex AVR-1 (1970s) compressed air diaphragm unavailable at any cost
- Playback heads must be replaced every 3-4 months
Obsolescence

• One supplier - technician is age 65
• Company appealing for money to continue refurbishing program
• Trouble getting supplies
• 2009 cost for new head = $3500
• 2012 cost = $5200

--information courtesy of George Blood
Obsolescence

- Obtaining Studer open reel tape machines at the British Library

---information courtesy of Andrew Pearson
The slate is blank, the day is new
The past is over, future's in view
I take my strength and apply it here
Obsolescence: that was what you feared
Don't look back at what you know
It's over and this chapter has been closed

Performed by Tsunami Bomb
Written by David Alderman Ball, Richard Graham Thomas Norris, Vanessa Quinones
Obsolescence

“...the old persist[s] alongside and despite the new, surviving as echoes and shadows...”

The Old Ways: A Journey on Foot
Robert Macfarlane
Obsolescence

How should we situate ourselves in relation to obsolescence?

• Progress and modernity?
• Stability and expertise?
• New and old together?
Obsolescence

How do we deal with “echoes and shadows”?

• Obsolescence budget (maintenance, repairs, “new” machines)
• Network of contacts
• Persistence
• Ingenuity and creativity
Obsolescence

All analog audio and video formats are at different points on a similar obsolescence slope
Degradation/Obsolescence

“For video the problem is even sharper: complete disappearance of an (affordable) ability to transfer.”

--Richard Wright, AMIA list, 2/17/2013
75% of the analogue video held in Europe in 2006 will be lost by 2023 when video digitisation will simply have "ceased to be."

--Richard Wright, PrestoCentre Answers, 2/5/2013
Degradation/Obsolescence

“There is a massive disaster happening here.”

-- Clifford Lynch, Coalition for Networked Information, 2011
Degralescence Attacks the Kingdom of Media
Prince Codec Brandishes the Sword of Migration
The Sands of Time Slip Away
Degradation/Obsolescence

How much time do we have?

• 15-20 years
“…many analog audio recordings must be digitized within the next 15 to 20 years – before sound carrier degradation and the challenges of acquiring and maintaining playback equipment make the success of these efforts too expensive or unattainable.”

--The Library of Congress National Recording Preservation Plan, page 7
Degradation/Obsolescence

How much time do we have?

• 15-20 years (analog audio and video)
• Less for some formats
• Degradation + Obsolescence = Impossible/Too Expensive
• Vendors and a few institutions stockpiling
Degradation/Obsolescence

How much time do we have?

10-15 Years
Degradation/Obsolescence

Audio and Video
How long will it take?

At our current pace...

Archives of Traditional Music: 58 years
Music Library: 120 years

Key Words: Massive, rapid, considered
The Battle for Media Commences
What is Indiana University Doing to Weather this Storm?
Numbers

More than 560,000 audio, video, film objects

- 364,000 audio (64%)
- 125,000 video (22%)
- 78,000 film (14%)
44% Unique or Rare
(248,000)

56% Commercial
Lesson 1

Highlight research and instructional value

(along with Degralescence)
Meeting the Challenge of Media Preservation:
Strategies and Solutions

INDIANA UNIVERSITY
BLOOMINGTON
Lesson 2
Preserverance
(Preservation perseverance)
2013 State of the University Announcement

Media Digitization and Preservation Initiative (MDPI)
2013 State of the University Announcement

Media Digitization and Preservation Initiative (MDPI)

Three fundamental missions of universities:

1. The creation of knowledge (research and innovation)
2. The dissemination of knowledge (education and learning)
3. The preservation of knowledge
MDPI Overview

• Digitally preserve all significant audio and video
• Complete by IU Bicentennial in 2020
• University-wide initiative
MDPI Goal

Digitally preserve 280,000 audio and video recordings in 3-4 years

Progress:
213,501 digitized as of Monday
MDPI Funding

- Office of the President
- Office of the Provost
- Office of the Vice President for Research
- Additional funding and in-kind support: UITS, Libraries
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

Dennis Cromwell
Executive Director
MDPI Personnel

Management and Coordination

• Mike Casey, MDPI Director of Technical Operations
• Brian McGough, Director, Enterprise Integration, UITS
• Julie Bobay, Executive Associate Dean, Libraries
MDPI Personnel

IU Media Digitization Studios staff

• Mike Casey
• Melissa Widzinski, Audio Engineer
• Dan Figurelli, Audio Engineer
• Rob Mobley, Video Engineer
• Adam Nickel, QC Specialist
• Glenn Hicks, QC Specialist
• Jonathan Richardson, AV Tech
MDPI Personnel

Strategic Media Access Resource Team (SMART)

• Patrick Feaster, Media Preservation Specialist
• Sherri Michaels, Head of Collection Management, IU Libraries
• 30-35 graduate students
MDPI Personnel

Software Development

• Brent Moberly
• Andrew Albrecht
• Adam Ploshay
• Will Cowan
• Sherri Michaels
• Brian Wheeler
• Patrick Feaster
• Brian McGough
MDPI Personnel

Library and IU Units Operations Team

- Julie Hardesty
- Thomas Whitaker
- Erika Dowell
- Gary Charbonneau
- Alan Burdette
- Phil Ponella
- Tony Tadey
- Jon Dunn
- 10 others previously named above
MDPI Personnel

IT Team

- Kurt Siefert
- David Hunter
- Kristi Kallback-Rose
- Danko Antolovic’
- John Wright
- Brian McGough
MDPI Personnel

IT Communications

• Keith Danielson
• Alan Mauro
• Amanda Chambliss
• Joe Stone
• Ralph Zuzolo
• Brian Hawkins
• Madeline Grdina
• Daphne Siefert-Herron
MDPI Personnel

Finance, facilities, and administrative assistance

- Misty Smith
- Doug Mayo
- Heather Pawluk
- Don Brock
- Doug Chambers
- Chris Hayden
MDPI Personnel

SIP Working Group

• Ronda Sewald
• Julie Hardesty
• Susan Hooyenga
• Brian Wheeler
• Brian McGough
• Jon Dunn
MDPI Personnel

Access Advisory Working Group

• Naz Pantaloni
• Dina Kellams
• Barbara Truesdell
• Heidi Dowding
• Nick Homenda
• Rachael Stoeltje
• 13 others previously named above
MDPI Personnel

Media-holding Unit Staff

• Numerous staff from many units
MDPI Personnel

Audio and Video Technical Working Groups

- Konrad Strauss
- Mark Hood
- Tony Tadey
- John Wright
MDPI Personnel

Consultant
• AVPreserve

Colleagues
• Michael Angeletti, Stanford University
• Chris Lacinak, AVPreserve
• Dave Rice, CUNY
• Library of Congress Packard campus
• Danny Sbardella, New York Public Library
• National Library of Norway
• NARA
• Tom de Smet, Sound & Vision
• Brecht Declercq, VIAA
• Many others...
Lesson 3
It takes a village...
MDPI Digitization Strategy

- Paradigm shift: partnership with a private company
- Sony Memnon
- Key to realizing goals around cost and timeframe without sacrificing quality
MDPI Digitization Strategy

• Memnon - parallel transfer (industrial-scale) workflows
• IU - 1:1 workflows for fragile formats and problem items
• 6.5 PB in 4 years
Lesson 4

We don’t have to do it all ourselves
MDPI Digitization Strategy

Project file formats

- Audio preservation master - BWF, 24/96
- Audio production master - same
- Video preservation master -
- FFV1/Matroska
- Video mezzanine - 50 Mbps MPEG-2
MDPI Digitization Facility

- Innovation Center, Tech Park, IU Bloomington campus
- 8,400 square feet (780 sq m)
- 21 rooms
- Open for non-IU business
MDPI Pre-Digitization Workflow

- Feeding the beast
  At peak -
  - 9 TB per day
  - 320 hours digitized per day
  - 616 physical objects digitized per day
  - 2,700 files created per day
MDPI Pre-Digitization Workflow

SMART - led by Sherri Michaels and Patrick Feaster

- Sort by technical characteristics
- Gather technical metadata
- Barcode
- Place in batch
- Transport to facility
Lesson 5

Prep workflow requires more time and resources than expected

Bottleneck in the system
### Physical Object [114455]

<table>
<thead>
<tr>
<th>Assignment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Picklist</td>
<td>ATM ORT for IU side</td>
</tr>
<tr>
<td>Bin</td>
<td>Bin: 400000000014847</td>
</tr>
<tr>
<td>Box</td>
<td>Not Assigned</td>
</tr>
<tr>
<td>Batch</td>
<td>IU-0004-OR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MDPI barcode</th>
<th>40000001211863</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>Open Reel Audio Tape</td>
</tr>
<tr>
<td>Has ephemera</td>
<td>No</td>
</tr>
<tr>
<td>Ephemeran returned</td>
<td>No</td>
</tr>
<tr>
<td>Unit</td>
<td>B-ATM</td>
</tr>
<tr>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>Call number</td>
<td>OT 5160</td>
</tr>
<tr>
<td>IUCAT barcode</td>
<td>0</td>
</tr>
<tr>
<td>Year</td>
<td>1976</td>
</tr>
<tr>
<td>Group key</td>
<td>GR00083515</td>
</tr>
<tr>
<td>Group position</td>
<td>1</td>
</tr>
<tr>
<td>Carrier stream index</td>
<td>GR00083515_1_1</td>
</tr>
<tr>
<td>Imported from spreadsheet</td>
<td>ATM_OPENREEL_MIXEDSPEEDS_OBJECTS_INHOUSE_SlowUnknownSpeeds.osv</td>
</tr>
<tr>
<td>Workflow Status</td>
<td>Binned &gt;&gt; Batched &gt;&gt; Shipped</td>
</tr>
</tbody>
</table>

#### Hidden Fields

**Open Reel Audio Tape Technical Metadata**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pack deformation</td>
<td>Moderate</td>
</tr>
<tr>
<td>Preservation problems</td>
<td></td>
</tr>
<tr>
<td>Reel size</td>
<td>5 in.</td>
</tr>
<tr>
<td>Playback speed</td>
<td>3.75 ips, 7.5 ips</td>
</tr>
</tbody>
</table>
One batch
Memnon Digitization

Receiving and ingest
Import metadata

<table>
<thead>
<tr>
<th>Batch number</th>
<th>Batch description</th>
<th>Unit</th>
<th>Object ID</th>
<th>Cell number</th>
<th>Container barcode</th>
<th>Container name</th>
<th>Container description</th>
<th>Output file name prefix</th>
<th>Year</th>
<th>Tape type</th>
<th>Reel size</th>
<th>Track configuration</th>
<th>Soundfield</th>
<th>Playback speed</th>
<th>Tape thickness</th>
<th>Tape type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234567890</td>
<td>Sample batch</td>
<td>1</td>
<td>A123456</td>
<td>78901234</td>
<td>ABCD123456</td>
<td>XYZ123456</td>
<td>Sample container</td>
<td>SAMPLE00000000</td>
<td>2019</td>
<td>Polyester</td>
<td>7 in</td>
<td>Half track</td>
<td>Stereo</td>
<td>75 psi</td>
<td>15 ml</td>
<td>Scotch</td>
</tr>
<tr>
<td>0987654321</td>
<td>Another batch</td>
<td>2</td>
<td>12345678</td>
<td>98765432</td>
<td>ABCD789012</td>
<td>XYZ987654</td>
<td>Another container</td>
<td>ANOTHER00000000</td>
<td>2019</td>
<td>Polyester</td>
<td>7 in</td>
<td>Stereo</td>
<td>75 psi</td>
<td>15 ml</td>
<td>Scotch</td>
<td>Scotch</td>
</tr>
<tr>
<td>1122334455</td>
<td>Third batch</td>
<td>3</td>
<td>23456789</td>
<td>34567892</td>
<td>ABCD234567</td>
<td>XYZ234567</td>
<td>Yet another container</td>
<td>YET_ANOTHER00000000</td>
<td>2019</td>
<td>Polyester</td>
<td>7 in</td>
<td>Half track</td>
<td>75 psi</td>
<td>15 ml</td>
<td>Scotch</td>
<td>Scotch</td>
</tr>
</tbody>
</table>

Memnon Digitization
Memnon Digitization

Ultrasonic cleaning of LPs
Memnon Digitization

4:1 LP digitization
Memnon Digitization

4:1 Tape digitization
Memnon Digitization
Video digitization
Memnon Digitization

Quality control
Memnon Digitization

Full time maintenance engineer
Audio preservation
• 7,000 field cylinders, lacquer discs, mixed speed tapes, wire recordings
• Sound Directions 1:1 workflow
Video preservation

- Hi 8/8mm, Betamax, ½” EIAJ
- Problem VHS, Umatic, Betacam SP
Productivity
- Theory of Constraints
- Scripts
- Scrum methodology
  - Jira backlog
  - Two week sprints
MDPI Quality Control

- QC preservation master files and derivatives
- Average package size
  - Betamax 78.6 GB
  - 8 mm 72.4 GB
  - VHS 95.2 GB
- QC workstations - 10 Gb connection over 50 micron multi mode fiber
MDPI Quality Control

- Random sample
- Visual/aural/metadata inspection
MDPI Quality Control

- 100% QC Specialist
- ~50% Processing and QC Specialist
- ~25% AV Specialist
- Hourly students--2
MDPI Post-Processing Workflow

- Transcode—create access derivatives
- Embed metadata
- Collect metadata for SIP
- Structural QC
- Make files available for QC
- Push to long-term storage
- Push to access system
- Fully automated
MDPI Storage

- **Bit Storage:**
  - IU Scholarly Data Archive (SDA)
  - Mirrored between IUB and IUPUI
    - Dual write over fiber
  - 42 PB tape capacity
  - Disk cache front end—1800 TB
Preservation Repository

- HydraDAM2 (Phydo)
- IU Libraries NEH-funded collaboration with WGBH/Boston, 2015-2016
- For time-based media
- Based on:
  - Fedora 4 digital repository software
  - Hydra framework
Preservation Repository

- Metadata and preservation event management tool
- Sits on top of storage
Access: Avalon Media System

- Co-developed by IU and Northwestern University Libraries
- Currently in production at:
  - IU
  - NW
  - University of Virginia
  - Yale
  - Washington University, St. Louis
  - University of Alberta
  - Calvin College
  - Qatar National Library
Access: Avalon Media System

- Open source software system
- Enables libraries and archives to provide access to audio and video collections
Access: Avalon Media System

- 2010-2011: IMLS Planning Grant
- 2011-2015: IMLS National Leadership Grant
- 2015-2018: Andrew W. Mellon Foundation
- 2017-2019: IMLS National Leadership Grant
Access: Avalon Media System

- SaaS—pilot hosted service
- LYRASIS
- Fall 2017
Unsolved Challenges

• Metadata at scale
  • Discoverability
  • AVPreserve consultation
• Rights at scale
  • IU Libraries Copyright Program Librarian
  • IU legal counsel engaged
Unsolved Challenges

Metadata at scale + Rights at scale = Access at scale!

- IU Libraries Access Task Force
- Out of region storage for PB of data
  - DPN
What has been done to date?

- Quick, non-scientific survey
- Help from IASA TC
- 42 projects
Large-Scale Media Digitization

What has been done to date?

- 90% undertaken outside the US
- 55% by broadcast organizations
- 19% by cultural heritage organizations
- Remainder by organizations holding both broadcast and cultural heritage materials
What has been done to date?

- 43% include video
- 75% include audio
- 10% include film
Large-Scale Media Digitization

Implications?

The future *media* documentary record might emphasize:

- Audio holdings
- Broadcast collections
- Materials from outside the US
Large-Scale Media Digitization

Implications?

What has been digitized in quantity:

- Books
- Still images
- Manuscripts
Large-Scale Media Digitization

Implications?

If, indeed, film, video, and audio collections are not preserved....the digital libraries of the future will contain embarrassing gaps.

-- David J. Francis, former Chief, MBRS Division, Library of Congress
Survey of Progress in the US
Stanford University Libraries

- Stanford Media Preservation Lab—2007
- 4 full time staff
- High quality playback and capture of most common formats at Stanford
University of North Carolina Libraries

- Large-scale digitization supported by Mellon
- Smaller scale digitization underway for many years
University of Illinois Libraries

- Media preservation program—2011
- Campus-wide media preservation census completed 2014
New York Public Library

• Completed comprehensive assessment of media holdings with support from Mellon
• Reallocated existing resources to support media preservation
• Beginning large-scale digitization phase
Syracuse University Libraries

- Completed media preservation survey of special collections
- Planning for expanded digitization program
ARSC Guide to Audio Preservation

Sam Brylowski, Maya Lerman, Robin Pike, Kathlin Smith, editors

SOUND DIRECTIONS

Best Practices for Audio Preservation

By Mike Casey, Indiana University and Bruce Gordon, Harvard University
Media Preservation

10 things we have learned...
Media Preservation

1. Degradation may make digitization impossible or less accurate
2. Obsolescence may make digitization prohibitively expensive
Media Preservation

3. Time frame is short
Media Preservation

4. Waiting longer will make the work more difficult and more expensive
5. Foregrounding research and instructional value is critical
6. Time-based media is underrepresented in digital libraries
7. Standards and practices are ready for audio
   --somewhat ready for video
   --barely ready for film
Media Preservation

8. An increasing number of institutions are beginning to tackle these problems
9. We don’t have to do it all ourselves
Media Preservation

Hope is like a road in the country; there was never a road, but when many people walk on it, the road comes into existence.

---Lin Yutang

- Recognition of the value of AV for research
- New generation of media preservationists
- Increase in AV specialist positions in libraries
- University libraries taking action
- Open source movement
- New machines
- New software
- Skilled vendors
Degralescence is Brought under Control
The End
Why Media Preservation Can’t Wait
The Gathering Storm

Mike Casey
Director of Technical Operations
Media Digitization and Preservation Initiative
Indiana University
May 24, 2017

Website: https://mdpi.iu.edu