

Audio and Video

Repositories at Scale

**Indiana University's Media Digitization
and Preservation Initiative**

Jon Dunn
Indiana University Libraries

Open Repositories 2014
Helsinki, Finland
June 12, 2014





(for physical time-based media objects)



Degradation

All analog and physical digital media objects actively degrading

– some catastrophically

Vinegar Syndrome
Sticky Shed Syndrome
Fungus
Delamination
Windowing
Plasticizer Exudation
Curling
Hydrolysis
Color Fading
Shedding
Scratches
Undiagnosed
Unplayability
Efflorescence
Crystalline Residue
Oxidation
Shrinkage
Breaking
Mechanical Issues
Binder Breakdown
Cupping



Degradation



Obsolescence

- Media formats
- Equipment (playback machines, test devices)
- Repair parts
- Playback expertise
- Repair expertise
- Tools
- Supplies



How long do we have?

- Short time window in which to migrate or lose forever
- 10-15 years?



Indiana University

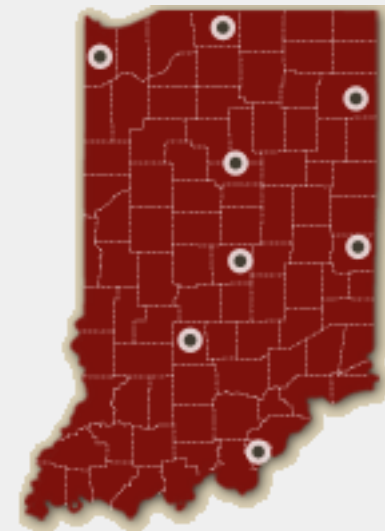
State university

Founded in 1820

8 campuses

110,000 students

18,000 faculty and staff



Indiana University Bloomington

Flagship IU campus

Major research university

Strong arts and humanities focus

46,000 students

8500 faculty and staff



IU Repository and Digital Library Environment

- Started with audio, 1992
- Repository infrastructure:
 - Fedora, Hydra, Dspace
- Strong collaboration between Libraries and University IT Services
- Strong enterprise and research storage infrastructure

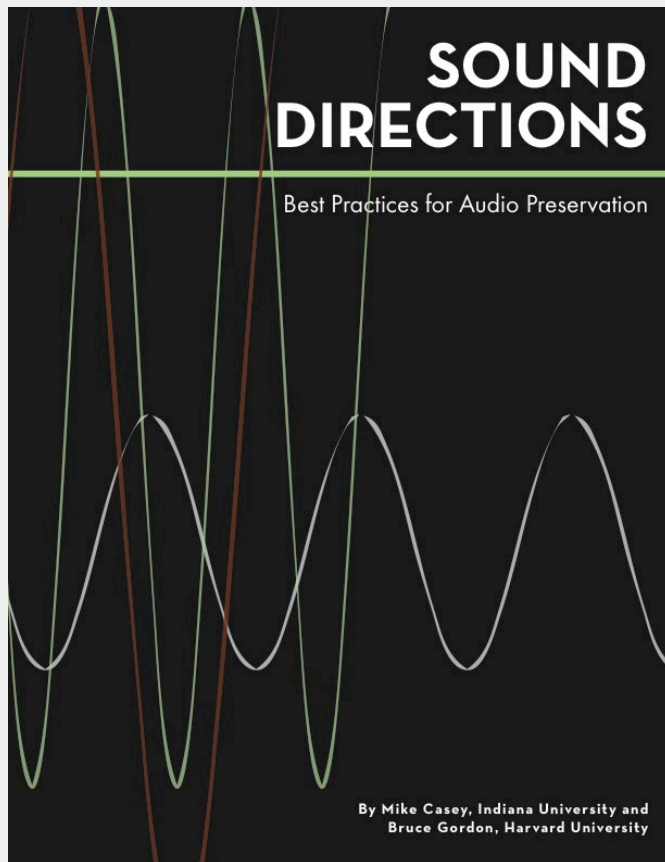


Context: IU Collections

- School of Music
- Archives of Traditional Music
- Kinsey Institute
- Moving Image Archive
- Lilly Library
- University Archives
- Athletics
- International collections



Sound Directions, 2007

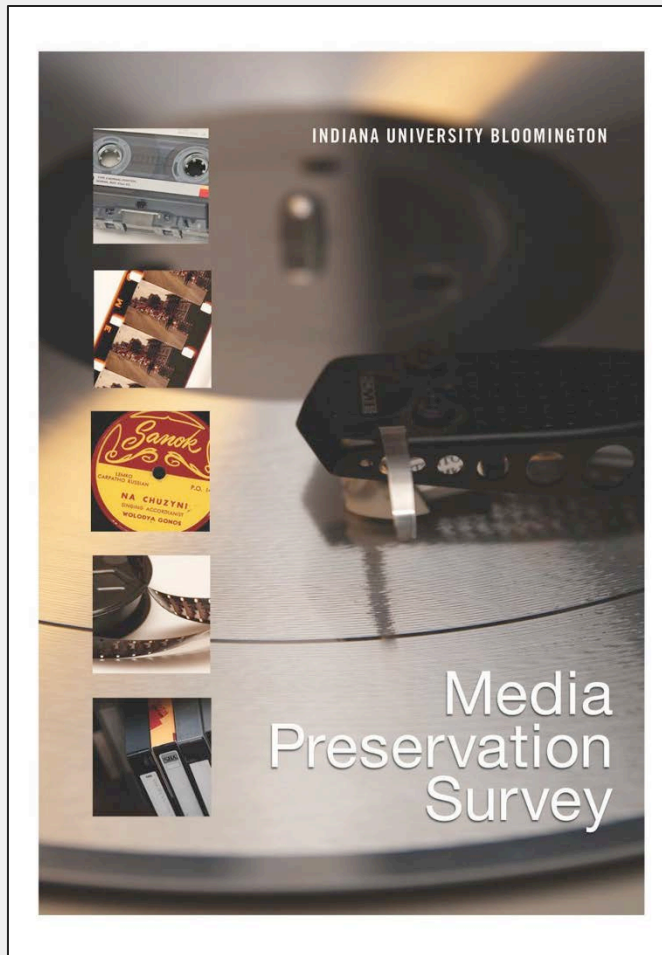


IU, Harvard

Best practices for
audio preservation
digitization and
storage



Media Preservation Survey, 2008



Findings:

- 569,000 items
- Unique and valuable collections
- 51 physical formats
- 80 units at IUB

Recommendations:

- Digitize within 15-20 years
- Form task force to plan



Meeting the Challenge Report, 2012



Recommendations:

- Create center to digitize 300K rare or unique objects within 15 years
- Audio, video for preservation
- Film for access



Media Digitization and Preservation Initiative (MDPI), 2013

- Announced by President McRobbie in October 2013
- Goal: Digitize, preserve, provide access to rare and unique audio and video **by 2020** – all IU campuses
- Expect to start in 2014

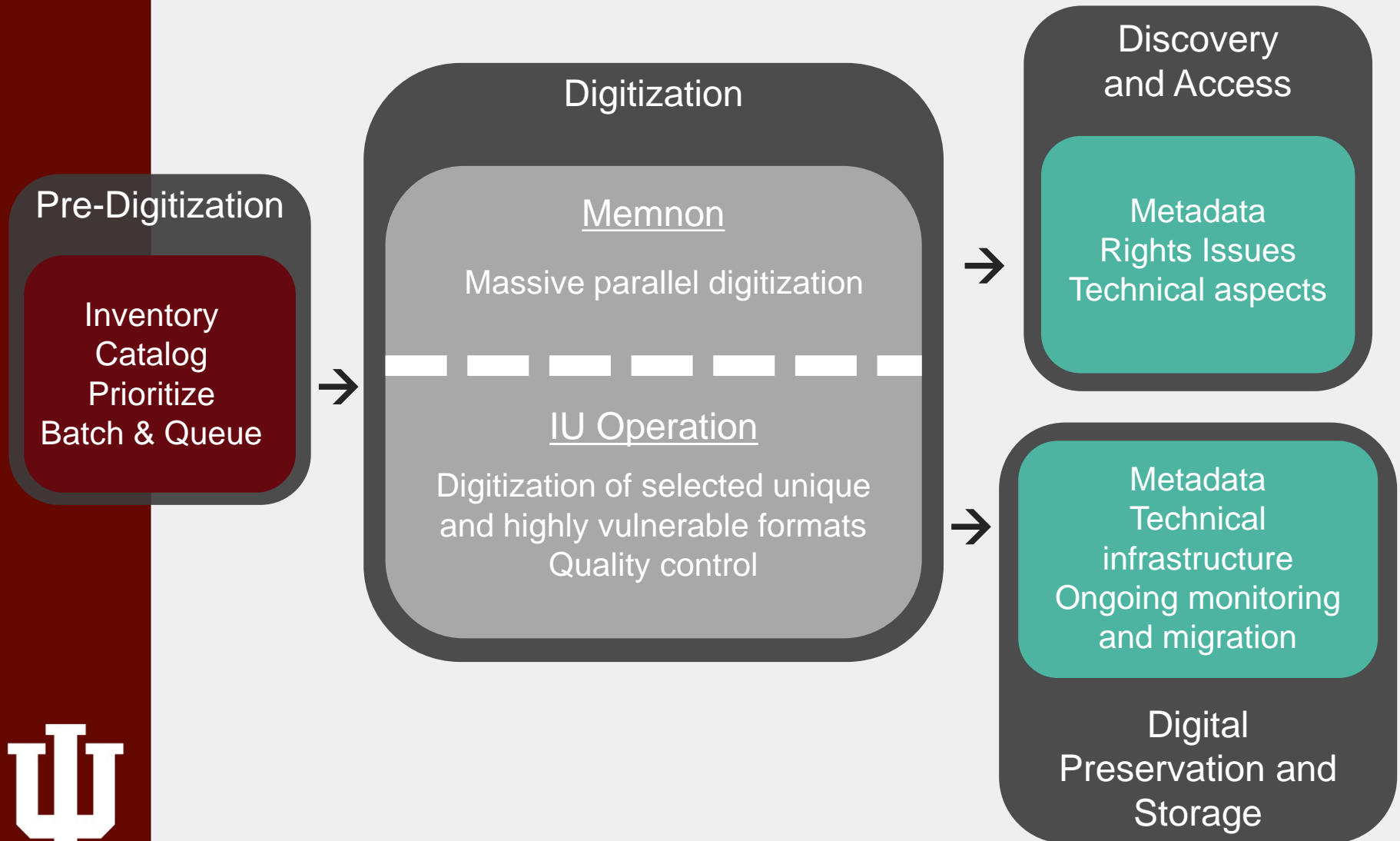


Public-Private Partnership

Memnon Archiving Services Brussels, Belgium



Digitization and Preservation Phases



Bins, Boxes, Barcodes, Batches

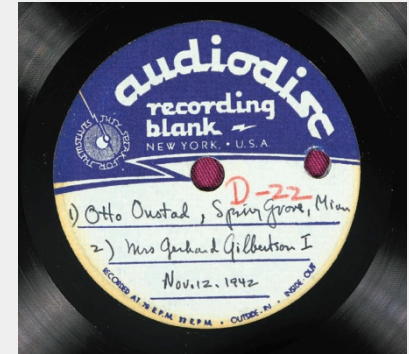


Workflow for Picking and Packing

Number	Item ID	MSP# Barcodes	Bin Barcodes	Print Item Removed	Print Item Returned	Title	Author	Library	Shelving Key
4640 G7.1	3200000222000					000020TH CENTURY MUSIC	PERUCHETTI, B. MUSIC		TP 4640 G7
5222 78.4.26 v.1	3200000229286					PROGRAM 001977 001978 NO 000799 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 4522 78
5222 78.4.26 v.2	3200000229294					PROGRAM 001977 001978 NO 000799 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 4522 78
3856 D7.1e v.1	3200000247478					CHRISTMAS MUSIC OF H. HENRICH SCHUTZ	SCHUTZ, H. H. B. MUSIC		TP 3386 D7
3856 D7.1e v.2	3200000247486					CHRISTMAS MUSIC OF H. HENRICH SCHUTZ	SCHUTZ, H. H. B. MUSIC		TP 3386 D7
5286 76.11.30 v.2	3200000482959					PROGRAM 001976 001977 NO 000849 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 3586 76
6131 77.2.14 v.1	1474686-2001					PROGRAM 001976 001977 NO 000762 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 3631 77
6131 77.2.14 v.2	1474686-2001					PROGRAM 001976 001977 NO 000762 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 3631 77
1294 77.4.26 v.1	1473965-2001					PROGRAM 001976 001977 NO 000825 ARTIST DIPLOMA B INDIANA UNIV B-MUSIC			TP 3124 77
1294 77.4.26 v.2	1473965-2001					PROGRAM 001976 001977 NO 000825 ARTIST DIPLOMA B INDIANA UNIV B-MUSIC			TP 3124 77
7643 76.11.30 v.3	32000007410063					GRADUATE RECITAL	ACKMANN, R. B. MUSIC		TP 5 A328 8
2643 76.11.30 v.2	32000007409834					GRADUATE RECITAL	ACKMANN, R. B. MUSIC		TP 5 A328 8
A1828 81.4.17 v.1	3200000968172	40000000086767				GRADUATE RECITAL	ADAM, ADOLF B. MUSIC		TP 5 A318 1
A1828 81.4.17 v.2	3200000968172	40000000086775				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A318 1
A1828 81.4.17 v.3	3200000968172	40000000086783				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A318 1
A19318 05.1.3 v.1	3200000968172	40000000086791				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A245 21
A19318 05.1.3 v.2	3200000968172	40000000086809				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A245 21
A19318 05.1.3 v.3	3200000968172	40000000086817				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A245 21
A19318 05.1.3 v.4	3200000968172	40000000086825				GRIELLE BALLEFPANTOMME WITH PROLOGUE AND TWO ADAM ADOLF B. MUSIC			TP 5 A245 21
A2145 76.3.12 v.1	3200000968172	40000000086833				GRADUATE RECITAL	ADAMS, JOHN B. MUSIC		TP 5 A218 7
A2145 76.3.12 v.2	3200000968172	40000000086841				GRADUATE RECITAL	ADAMS, JOHN B. MUSIC		TP 5 A218 7
A2145 76.3.12 v.3	3200000968172	40000000086849				GRADUATE RECITAL	ADAMS, JOHN B. MUSIC		TP 5 A218 7
A2145 76.3.12 v.4	3200000968172	40000000086857				GRADUATE RECITAL	ADAMS, JOHN B. MUSIC		TP 5 A218 7
A2328 77.12.9 v.1	3200000968172	40000000086865				PROGRAM 001977 001978 NO 000991 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 5 A328 7
A2328 77.12.9 v.2	3200000968172	40000000086873				PROGRAM 001977 001978 NO 000991 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 5 A328 7
A2328 77.12.9 v.3	3200000968172	40000000086881				PROGRAM 001977 001978 NO 000991 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 5 A328 7
A2328 77.12.9 v.4	3200000968172	40000000086889				PROGRAM 001977 001978 NO 000991 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 5 A328 7
A2328 78.12.2 v.1	3200000968172	40000000086897				PROGRAM 001978 001979 NO 000355 GRADUATE RECITA INDIANA UNIV B-MUSIC			TP 5 A328 7



Batches: Format Based



Workflow Support: The POD

Media Digitization and Preservation Initiative
Physical Object Database

Physical Objects **Batches** **Bins** **Pick Lists** **Statuses** **Advanced Search** Barcode

Edit Delete Split Record Create New Physical Object Upload Spreadsheet

Physical Object	Open Reel Tape Technical Metadata	Digital Files
Title PROGRAM 001978 00197	Pack Deformation <input type="text"/>	
Title Control Number BAH0423BM	Preservation Problems <input type="text"/>	
Call Number TP-S_A9238 78-8-5 v. 1	Reel Size 7"	
Unit BL-MUS	Playback Speed 7.5 ips	
Home Location _MUIUTP	Track Configuration Half Track	
Shelf Location <input type="text"/>	Tape Thickness 1.5 mil (standard play)	
IUCAT Barcode 30000036034449	Sound Field Stereo	
Format Open Reel Tape ▾	Tape Stock Brand Scotch 208	
Carrier Stream Index 1	Tape Base Polyester	
Collection Identifier <input type="text"/>	Year of Recording <input type="text"/>	
MDPI Barcode 40000000089468	Directions Recorded 1	
Format Duration <input type="text"/>		
Content Duration <input type="text"/>		
Has Ephemera false		
Created At 2014-04-21 15:42:52 UTC		
Updated At 2014-04-21 19:18:04 UTC		
Workflow Status Barcoded ▾		
Picklist 5" Open Reel Tapes		
Bin/Box Box: 40000000089500		
Batch Not Assigned		

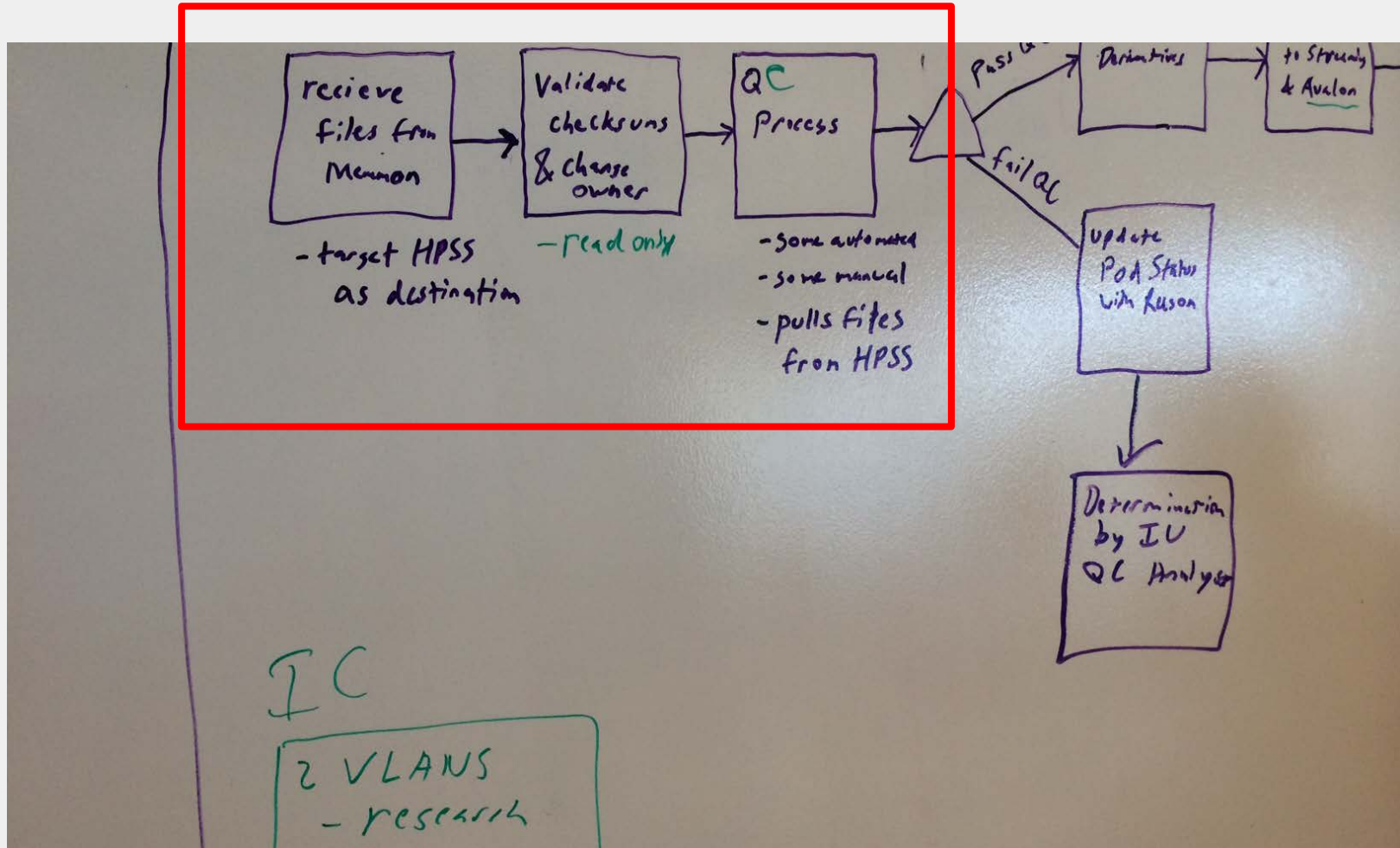


High Volume

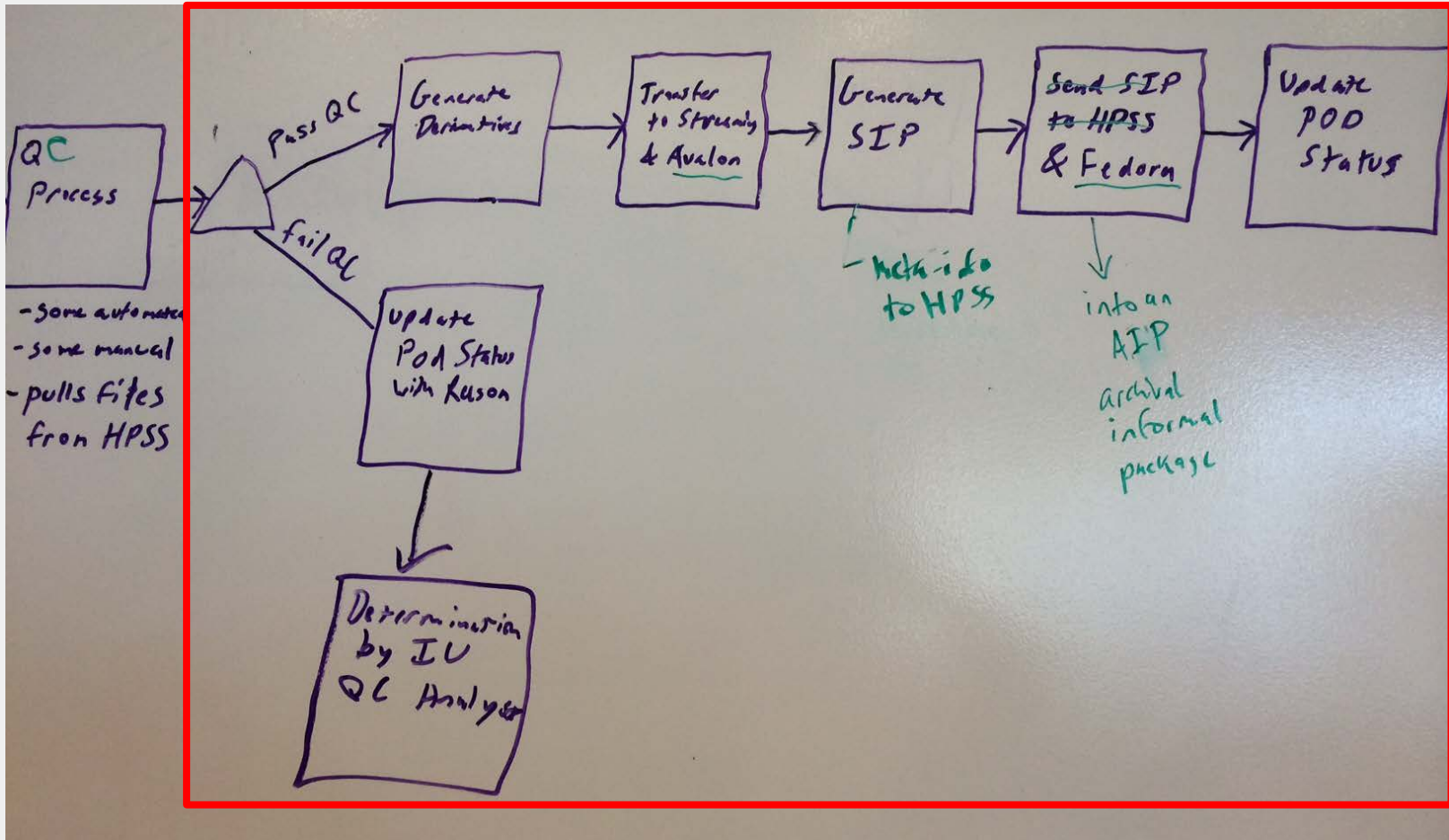
- 1500+ objects per week
- Peak of ~12 TB per day of digitization
- Total storage requirement: 9 PB



Post-Digitization Workflow



Post-Digitization Workflow



Metadata for Preservation and Access

Descriptive

- From MARC when available
- EAD finding aids, local inventories

Technical

- File and original object characteristics
- Checksums

Process History

- Digitization and preservation process

Structural

- Support relationships between and navigation within objects



IU Scholarly Data Archive (SDA)

- Central storage resource for IU since 1999
- Data replicated between Indianapolis and Bloomington
- Implemented using IBM HPSS software
- 40PB IBM TS1140 based capacity
- Available from desktop, web, high speed transfer protocols
- Currently holds ~6PB of data in 36 million files



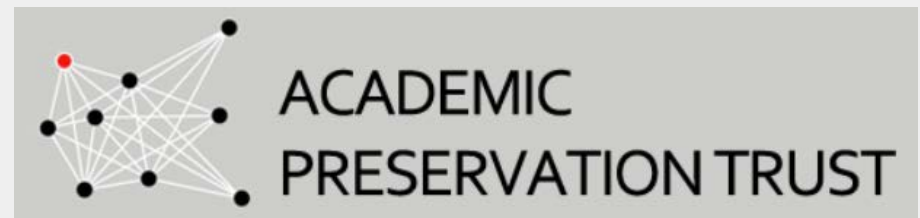
Access Technology

- Discovery
 - IUCAT, Archives Online (EAD Finding Aids), other environments?
- Delivery
 - Key requirements: usability, reusability, access control, performance
 - Avalon Media System



Long-Term Preservation Technology

- 12 petabytes+ to be preserved
- Local storage
 - UITS Scholarly Data Archive
 - Fedora 4 repository layer
- Out-of-region storage
 - APTTrust, DPN
 - Data swap agreements



MDPI Assets and Opportunities

- Collaborative planning process
 - Libraries, UITS, archives, faculty
- Research storage and IT resources
 - 40 petabyte mirrored HSM system
 - Experience with data-intensive workflows
 - High-performance networks
- Expertise in digitization, repositories, access systems
 - Variations, Avalon Media System
- Potential for external services



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



Stay Tuned...

mdpi.iu.edu

avalonmediasystem.org

