An Audiovisual Metadata Platform to Support Mass Description
Outline

1. Background and context - Jon
2. Workshop outcomes - Tanya
3. Next steps - Chris
Background

- **AMIA 2016:**
  - Chris Lacinak and Jon Dunn. “From Mass Digitization to Mass Description: Indiana University’s Strategy To Overcome The Next Great Challenge.”
  - [go.iu.edu/1Pvj](http://go.iu.edu/1Pvj)
The Challenge

- Growing AV collections
  - Digitization
  - Explosion of born-digital

- Increased expectations for access
The Challenge

- Many AV collections lack metadata
  - Discovery
  - Identification
  - Navigation
  - Rights
- Institutions lack resources for large cataloging/inventory/rights clearance projects
The Opportunity

- Mass digitization approach extended to AV
  - “Digitize first”
- Emergence and continued improvement of machine learning and other automated tools
- How can we leverage the best of automated tools and human expertise?
Existing Work in the context of AV Archives

- Application of specific machine learning tools
  - e.g. speech-to-text, named entity recognition
- “Black box” systems
  - One size fits all, brute force approach to automated metadata generation
- Customized workflows
  - e.g. MICO Platform
Context: Indiana University

- MDPI: Media Digitization and Preservation Initiative
  - 280,000+ AV items; 25,000+ films
- 80+ different units
- 20+ different physical formats
- Partnership with Memnon
- Variety of existing (or nonexisting) metadata
- Avalon Media System access platform
Consulting engagement with AVPreserve in 2016 to identify metadata and rights workflows
- Phased approach
- Identification of MGMs: Metadata Generation Mechanisms
- Need for platform to support workflows, metadata warehouse
Context: UT/HiPSTAS

- *High Performance Sound Technologies for Access and Scholarship*
- An assessment of scholarly requirements for analyzing sound
- An assessment of technological infrastructures needed to support discovery
- Preliminary tests that demonstrate the efficacy of using such tools in humanities scholarship
- Developing a freely available, open-source, API-driven application for general use
AMP: Audiovisual Metadata Platform

- Audiovisual Metadata Platform
- Planning grant from Andrew W. Mellon Foundation (July 2017 - January 2018)
  - Focus on technical architecture
- In-person workshop (September 2017)
- Planned deliverables:
  - White paper
  - Draft proposal for implementation and pilot test
AMP: Audiovisual Metadata Platform

- Open source software platform to support metadata creation for AV collections
- Design and execute workflows combining automated and human steps
- Integrate multiple MGMs
  - Automated, manual
  - Local, HPC, cloud
Core AMP Team

- Indiana University Libraries
  - Jon Dunn
  - Julie Hardesty
- University of Texas at Austin School of Information
  - Tanya Clement
- AVPreserve
  - Adeel Ahmad
  - Chris Lacinak
  - Amy Rudersdorf
Mellon-funded workshop

3 Days
16 People
1 Tech Platform Plan
Stakeholders, User Requirements, & Personas
Workshop Participants

- Kristian Allen, UCLA Library
- Jon Cameron, IU Libraries
- Maria Esteva, Texas Advanced Computing Center, UT at Austin
- Mike Giarlo, Stanford University Libraries
- Brian McFee, Music & Audio Research Laboratory, NYU
- Scott Rife, Packard Campus for AV Conservation, Library of Congress
- Sadie Roosa, WGBH Media Library & Archives
- Felix Saurbier, TIB/German National Library of Science & Technology
- Brian Wheeler, IU Libraries
- Maria Whitaker, IU Libraries
Workshop logistics

- Day one:
  - overview with framing and meeting goals
  - review of requirements, user personas (actors); current technical landscape
  - focus on non-technical criteria

- Day two:
  - focus on technical component candidate system identification and ranking

- Day three: presented four metadata generation scenarios
Actors

- System administrator
- Content owner
- Target system
- Target user
Business Requirements

- Automate analysis of AV content & metadata
- Provide an intuitive interface
- Leverage best-of-breed tools in a single workflow
- Generate metadata with minimal errors
- Offer a variety of metadata output
- Build a community of developers
- Offer easy-to-use APIs
- Process multi-TB batches of content at a time
- Support collaborative efforts
- Control what metadata is made public
Functional requirements

- Asset management
- Copyright and security
- Storage
- Metadata standards
- Modularity
- Multi-tenancy
- Scalability
- Usability
Technical Requirements

- Data Model (42)
- APIs (42)
- Open source (41)
- Scalable (41)

Queueing | 38
Data output | 38
Access to metadata | 36
Authentication | 36
Versioning | 36
Modularity | 35
SEO | 33
GUI/User Interface | 32
Ease of configuration | 31
Multi-tenancy | 30
Non-functional Requirements

- Clear service/cost model
- Wide adoption
- Robust documentation
- Active dev community

Also:
- Proven stability of the product
- Open licensing
- Established governance model
- Robust outreach & marketing program
- Training available
- Semantic versioning
- Support for internationalization
Metadata generation scenarios

Given actual media and asked to map out how metadata would be generated from media and related documentation ingest / creation to export.
Metadata Generation Mechanisms (MGMs)

Identified systems that could perform NLP, content matching, facial recognition, OCR from video, music/speech detection, and more
Metadata generation scenarios

- audio music concert performance
- audio oral history - Italian and Finnish interviews and songs
- video oral history - Yiddish interviews
- video promo segments aired during halftime on college basketball TV broadcasts
Next Steps
Example Workflow Scenario

ffmpeg

Video file with audio

AV Demuxer

Audio file

Video file with no audio

Embedded metadata extraction
Example Workflow Scenario

Audio file → Silence detection → Music Detection → MusicBrainz Gracenote

Speech detection:
- Yes → Language detection → Watson ARLO
- No → Fraunhofer AV Analyzer

Fraunhofer AV Analyzer

Watson ARLO → Speaker identification

Verbit.ai Kaldi Watson → Automated speech-to-text → Speech-to-text human refinement

Verbit.ai Custom internal

Mufin AudioID

Commercial? → Commercial music database query

Commercial match query
Example Workflow Scenario

Language detection → Speaker identification

Automated speech-to-text → Speech-to-text human refinement

Verbit.ai Custom internal

Verbit.ai

Kaldi

Watson

Microsoft NLP

Stanford NLP

Named entity extraction

Named entity extraction

Rosette

Data Recon

Sentiment Analysis

GPS coordinate identification

dBPedia URI identification

Music Detection

Yes

MusicBrainz

Gracenote

Mufin AudioID

Commercial music match query

Commercial music database query

VAMP Suite

JAMS

Beats per minute

Genre detection

Instrument detection

No
Example Workflow Scenario

Fraunhofer AV Analyzer
VU Digital

Video file with no audio
→ Scene detection
→ Object recognition
→ Facial recognition
→ Age analysis
→ Expression analysis

MediaInfo

Embedded metadata extraction
→ GPS coordinate resolution
→ Date and time identification

Script

Script
Borrow, build, buy considerations

- Ownership of “the machine”
- Black boxes
- Limited capabilities demonstrated
- Exploring building on top of MiCO
- Very limited options
- Metadata cultivation concept
- Focal point
Anticipated challenges & considerations
Stay Tuned!
Thanks!

Jon Dunn
jwd@iu.edu | @jwdunn

Tanya Clement
tclement@utexas.edu | @tanyaclement

Chris Lacinak
chris@avpreserve.com | @avpreserve