AMP Project Update:

Leveraging Machine Learning and Human Expertise for AV Collections Access

CNI Spring 2020

May 13, 2020

Jon Dunn / Indiana University / @jwdunn
Shawn Averkamp / AVP / @WeAreAVP

Project website: https://go.iu.edu/amppd

AUDIOVISUAL METADATA PLATFORM

Outline

1. Project background and overview
2. Collections
3. Tool Evaluation and Selection
4. Platform Architecture and Workflows
5. Next Steps

AMP

AUDIOVISUAL METADATA PLATFORM

AMPPD Project Team

Project Management:
Jon Dunn / IU
Amy Rudersdorf / AVP
Jack Sutton / IU

Development Team:
Vinita Boolchandani / IU
Ying Feng / IU
Dan Fischer / AVP
Brian Wheeler / IU
Maria Whitaker / IU

Collections Team:
Jon Cameron / IU
Brad Cook / IU
Alex Duryee / NYPL
Michelle Hahn / IU
Dina Kellams / IU
Chuck Peters / IU
Thomas Whittaker / IU
Sara Rubinow / NYPL
Melanie Yolles / NYPL

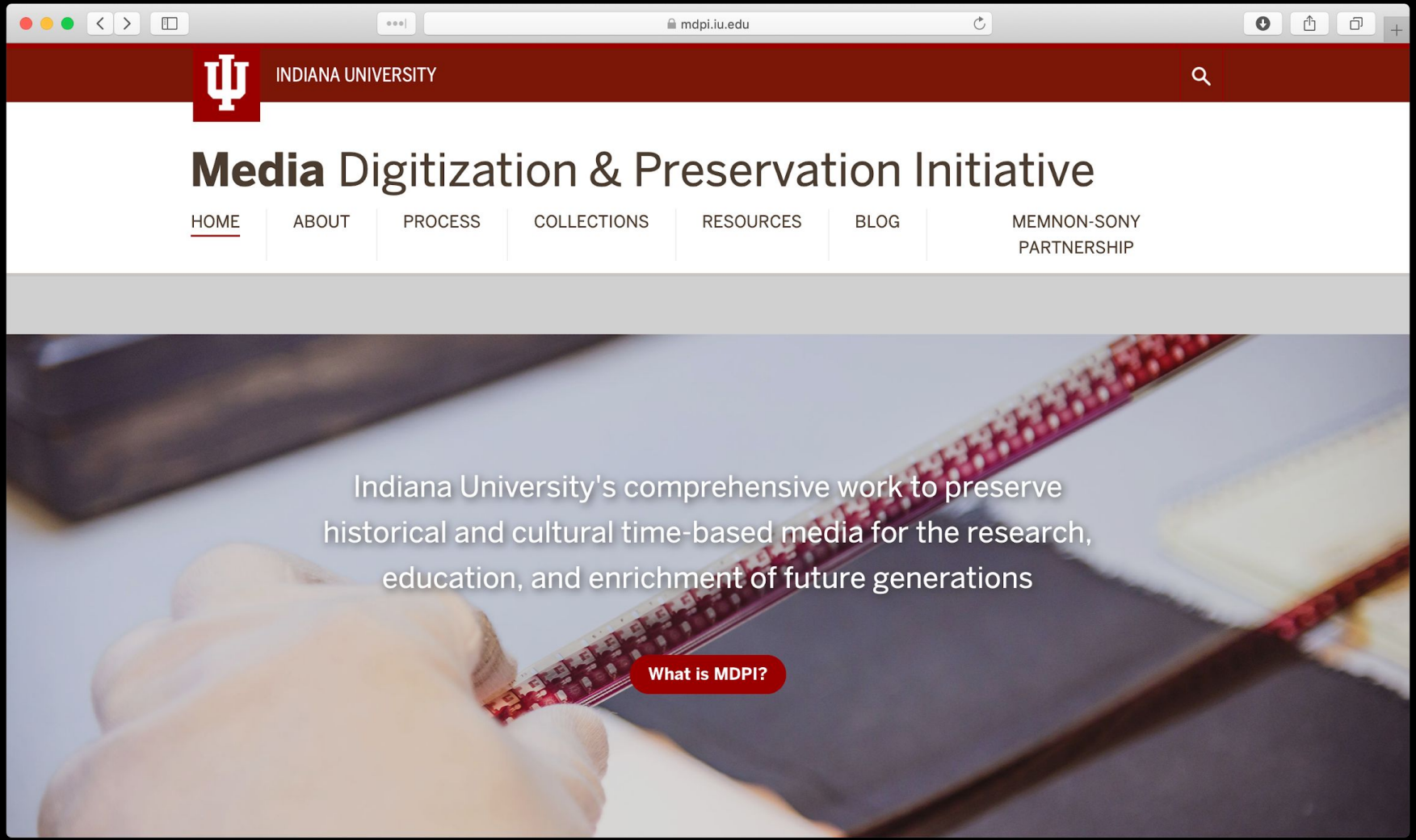
Metadata Generation Mechanisms (MGM) Team:
Shawn Averkamp / AVP
Tanya Clement / UT
Liz Fischer / UT
Julie Hardesty / IU
Bert Lyons / AVP

AMP
AUDIOVISUAL METADATA PLATFORM

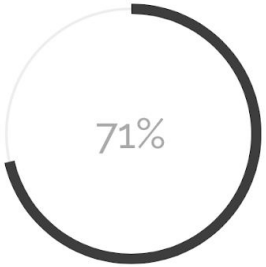
Project Background and Overview

AMP

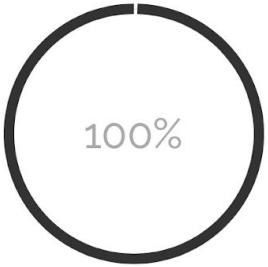
AUDIOVISUAL METADATA PLATFORM


mdpi.iu.edu
INDIANA UNIVERSITY
Media Digitization & Preservation Initiative
HOME
ABOUT
PROCESS
COLLECTIONS
RESOURCES
BLOG
MEMNON-SONY PARTNERSHIP
Indiana University's comprehensive work to preserve historical and cultural time-based media for the research, education, and enrichment of future generations
What is MDPI?

Digitization Progress to Date



17,829 out of 25000
Film Reels



326,424 out of 325000
Audio and Video

Featured collections

Edward S. Curtis collection

In 1906, Edward S. Curtis received funding from J.P. Morgan to photograph what remained of traditional Native American life. Over the next thirty years, Curtis captured the precolonial culture of tribes that included leaders such as Geronimo, Red Cloud, Sitting Bull, Chief Joseph, and Medicine Crow.

Curtis and his crew used wax cylinders to collect more than 10,000

The rake's progress a fable / by W.H. Auden and Chester Kallman ; music by Igor Stravinsky.

Date
2000

Main contributor
Weigel, Johan.

Contributors
Weigel, Johan.; Reich, Diane Thueson.; Oakden, Andrew.; Burchett, Christopher.; Valentine, Rebeckah.; Hauxwell-Hammond, Janice, 1962-; Mihalka, Toffer.; Sigmon, David.; Palló, Imre, 1941-; Liotta, Vincent.; Röthlisberger, Max.; Schwandt, Michael.; Stravinsky, Igor, 1882-1971. Rake's progress.; Indiana University Opera Theater.

Subject
Operas

Collection
William and Gayle Cook Music Library

Unit
William and Gayle Cook Music Library

Language
English

Share

Sections

1. VHS 1/1 Part 1 (40000001778846) (2:09:06)

1. Segment 1 (02:08)
2. Segment 2 (10:29)
3. Segment 3 (24:20)
4. Segment 4 (12:28)
5. Segment 5 (02:01)
6. **Segment 6 (18:16)**
7. Segment 7 (02:29)
8. Segment 8 (02:02)

Language
English

Physical Description
2 videocassettes : sd., col. ; 1/2 in.

Notes
Program bound separately as part 2.

CREATION/PRODUCTION CREDITS
Conductor, Imre Palló ; stage director, Vincent Liotta ; designer, Max Röthlisberger ; lighting designer, Michael Schwandt.

PERFORMERS
Johan Weigel, as Tom Rakewell ; Diane Thueson-Reich, as Anne ; Andrew Oakden, as Trulove ; Christopher Burchett, as Nick Shadow ; Rebeckah Valentine, as Mother Goose ; Janice Hauxwell, as Baba the Turk ; Kristoffer Mihalka, as Sellem ; David Sigmon, as Keeper of the madhouse ; Indiana University Opera Theater.

STATEMENT OF RESPONSIBILITY
by W.H. Auden and Chester Kallman ; music by Igor Stravinsky.

VENUE/EVENT DATE
Recorded on March 3, 2000, Musical Arts Center, Indiana University, Bloomington.

Other Identifiers
Catalog Key: 2812308; Other: GR00277701; Other: ARCHIVE VTP-S .S9126 A.3-20 v. 2; Other: ocm44875003; OCLC: ocm44875003; Catalog Key: CAK8960BM; MDPI Barcode: 40000001778846

View Raw Metadata | View Fedora Objects

Unpublish | Edit | Delete | Push to MCO

Tape B - Contains Digital Tapes 7-11

Date
unknown/unknown

Main contributor
Unknown

Collection
Office of University Archives and Records Management

Unit
Office of University Archives and Records Management

Other Identifiers
Other: GR00429921; Other: AV Box 32; MDPI Barcode: 40000003743053

Comments
[VHS 1/1 Part 1 (40000003743053)] Ingest: Signal - No HiFi Audio;

View Raw Metadata | View Fedora Objects

Share

Sections

1. VHS 1/1 Part 1 (40000001778846) (2:09:06)

1. Segment 1 (02:08)
2. Segment 2 (10:29)
3. Segment 3 (24:20)
4. Segment 4 (12:28)
5. Segment 5 (02:01)
6. **Segment 6 (18:16)**
7. Segment 7 (02:29)
8. Segment 8 (02:02)

Language
English

Physical Description
2 videocassettes : sd., col. ; 1/2 in.

Notes
Program bound separately as part 2.

CREATION/PRODUCTION CREDITS
Conductor, Imre Palló ; stage director, Vincent Liotta ; designer, Max Röthlisberger ; lighting designer, Michael Schwandt.

PERFORMERS
Johan Weigel, as Tom Rakewell ; Diane Thueson-Reich, as Anne ; Andrew Oakden, as Trulove ; Christopher Burchett, as Nick Shadow ; Rebeckah Valentine, as Mother Goose ; Janice Hauxwell, as Baba the Turk ; Kristoffer Mihalka, as Sellem ; David Sigmon, as Keeper of the madhouse ; Indiana University Opera Theater.

STATEMENT OF RESPONSIBILITY
by W.H. Auden and Chester Kallman ; music by Igor Stravinsky.

VENUE/EVENT DATE
Recorded on March 3, 2000, Musical Arts Center, Indiana University, Bloomington.

Other Identifiers
Catalog Key: 2812308; Other: GR00277701; Other: ARCHIVE VTP-S .S9126 A.3-20 v. 2; Other: ocm44875003; OCLC: ocm44875003; Catalog Key: CAK8960BM; MDPI Barcode: 40000001778846

View Raw Metadata | View Fedora Objects

The Challenge

- Growing digital AV collections
  - Digitization
  - Explosion of born-digital
- Increased expectations for access
- Insufficient metadata
- Limited resources for cataloging

AMP
AUDIOVISUAL METADATA PLATFORM

The Opportunity

- Emergence and continued improvement of machine learning and other automated tools
- New access tools (e.g. Avalon, Aviary) and standards (e.g. IIIF) that can leverage time-based metadata
- How can we leverage the best of automated tools and human expertise in flexible and configurable ways?
    - Diverse collections demand diverse workflows

AUDIOVISUAL METADATA PLATFORM

AMP: Audiovisual Metadata Platform

- Open source software platform to support metadata creation for AV collections
- Design and execute workflows combining automated and human steps: *Metadata Generation Mechanisms (MGMs)*
- Delivery of metadata to variety of target systems, e.g. online access systems, library catalogs, etc.

AMP
AUDIOVISUAL METADATA PLATFORM




New York Public Library

THE ANDREW W. MELLON FOUNDATION

2017-2018: AMP Planning Project

↓

2018-2020: AMP Pilot Development

↓

Future: AMP Implementation

Collections

AMP

AUDIOVISUAL METADATA PLATFORM

AMPPD Pilot Collections

University Archives, Indiana University Bloomington

William & Gayle Cook Music Library, Indiana University Bloomington

Gay Men’s Health Crisis Collection, New York Public Library

AUDIOVISUAL METADATA PLATFORM

Collections Use cases

IU Archives

- Bloomington Faculty Council Minutes
- Russian and East European Institute
- Focus Black America
- Herman B Wells

75hrs video | 25hrs audio

IU Music Library

- Jazz Ensemble
- Solo Recitals
- Orchestra Performances
- Opera Performances

80hrs video | 20hrs audio

New York Public Library

Gay Men's Health Crisis

- Protests
- Speaker Events
- Interviews
- Focus Groups

80hrs video | 20hrs audio

Collections Use cases

Common needs: Transcripts & Subject terms

Archives:
People & locations

Music Library:
Works performed

NYPL:
Content boundaries

MGM prioritization

First Priority

- Silence/speech/music detection
  - Content boundaries for processing/description
- Speech-to-text / Speaker diarization
  - Transcripts for discovery/accessibility
- Natural Language Processing (NLP) / Named Entity Recognition (NER)
  - Potential subject terms (people, corporate names, geographic locations, events, topics)

Next Priority

- Video OCR
  - Metadata from credits
- Structured OCR for supplementary materials
  - More subject terms
- Scene & shot detection
  - Copyright review
  - More content boundaries
- Music instrumentation detection
- Face detection

Evaluation and Selection

AMP

AUDIOVISUAL METADATA PLATFORM

Evaluation criteria

Accuracy	How does the MGM output compares to the expected value (or human-generated value). This should be a consideration of both quantitative and qualitative measures.
Input formats	Filetypes, encodings, compressions, etc. allowed by the MGM. Assess the level of difficulty involved in converting your files to the formats required for the tool.
Output formats	File types or data formats output by the MGM. Assess the level of difficulty involved in converting available output formats to the desired format. How will this impact automation?
Growth rate	Rate of increase of time and computing resources as volume/file size increases. Compare processing time between small, average, and large sized files to estimate time required as scale increases.
Processing time	Time required for the MGM to process the file. How will processing time affect your production workflows? Can processing time be improved by optimizing computing hardware, software, or networks?
Computing resources	Amount of computing resources, including processing power, memory, network connections, and bandwidth required to process the file. How will computing resources affect your production workflows?

https://wiki.dlib.indiana.edu/display/AMP/Evaluation+Criteria

Evaluation criteria (continued)

Social impact	The potential unintended consequences of an unmediated MGM's output. How could the MGM express hidden biases? What are the possible unintended negative impacts that could come from the output of this MGM? What measures can be taken to mitigate them? See FAT/ML's Principles for Accountable Algorithms for more information.
Cost	The cost of the MGM which could include paid services, file transfer and computing costs if running in the cloud, or local hardware and staff costs.
Support	Available human support, documentation, or logs output by the MGM which can help with learning or troubleshooting the MGM.
Integration capabilities	The ability of an MGM to fit into a workflow design or technical infrastructure or the ability to supply functionality for other computational needs, such as a speech-to-text tool that also provides segmentation and speaker diarization.
Training	Whether or not a model should be trained to utilize the MGM. Consider the costs, time, and social impact of training a model or using a model out-of-the-box.

https://wiki.dlib.indiana.edu/display/AMP/Evaluation+Criteria

FAT/ML Accuracy Principle

(Fairness, Accountability and Transparency in Machine Learning)

Guiding Questions

- What sources of error do you have and how will you mitigate their effect?
- How confident are the decisions output by your algorithmic system?
- What are realistic worst case scenarios in terms of how errors might impact society, individuals, and stakeholders?
- Have you evaluated the provenance and veracity of data and considered alternative data sources?

Initial Steps to Take

- Assess the potential for errors in your system and the resulting potential for harm to users.
- Undertake a sensitivity analysis to assess how uncertainty in the output of the algorithm relates to uncertainty in the inputs.
- Develop a process by which people can correct errors in input data, training data, or in output decisions.
- Perform a validity check by randomly sampling a portion of your data (e.g., input and/or training data) and manually checking its correctness. This check should be performed early in your development process before derived information is used. Report the overall data error rate on this random sample publicly.
- Determine how to communicate the uncertainty / margin of error for each decision.

From “FAT/ML Principles for Accountable Algorithms and a Social Impact Statement for Algorithms”
https://www.fatml.org/resources/principles-for-accountable-algorithms

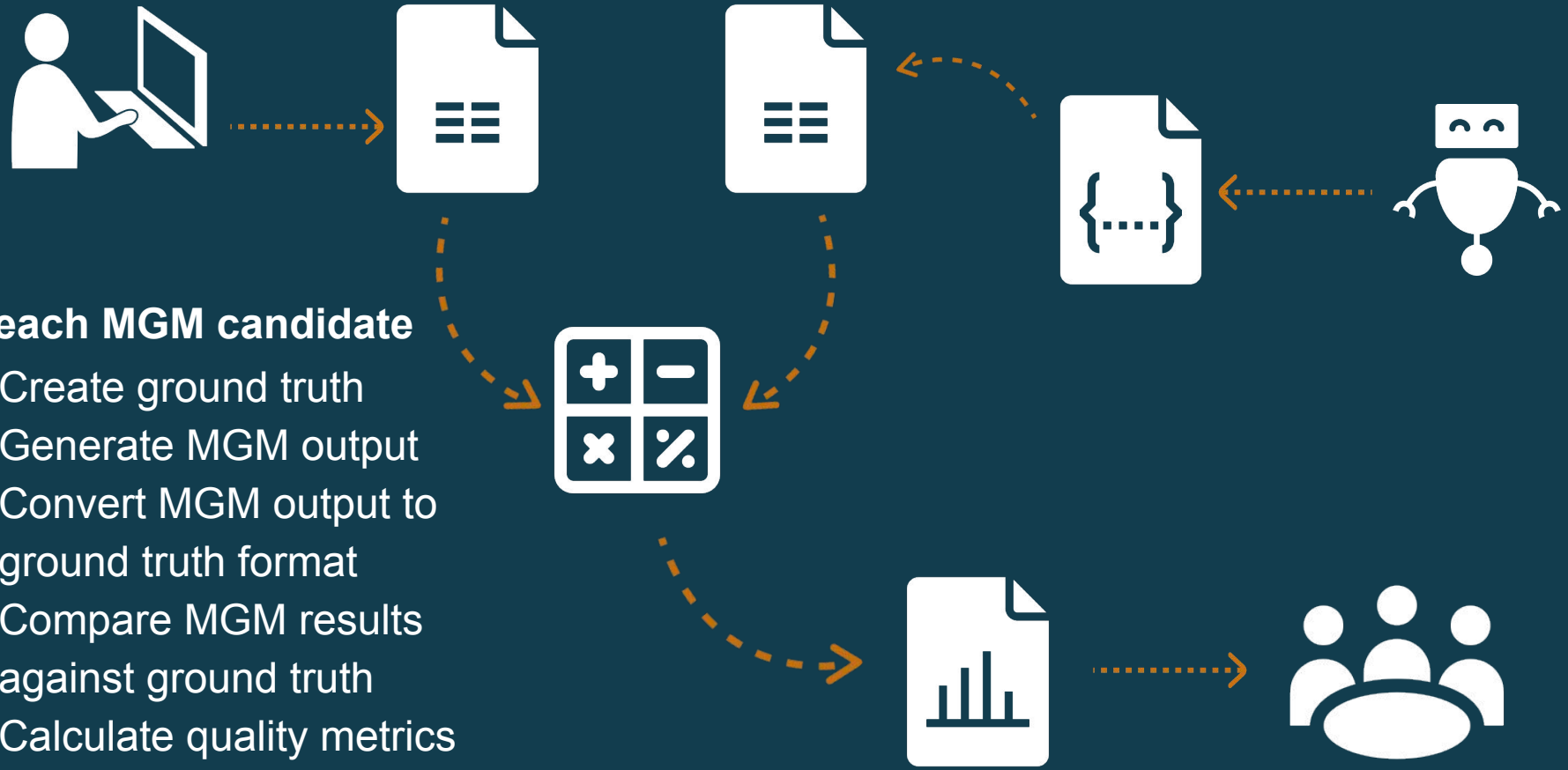
AMPPD metadata must be good enough to . . .

- Help (and not hinder) catalogers describing audiovisual assets
- Increase access to audiovisual assets through search
- Help provide greater context to audiovisual collections
- Increase navigability of audiovisual assets through provision of basic segmentation or structure

AMP
AUDIOVISUAL METADATA PLATFORM

Preparing to test

1. Worked with collection managers to select samples representing a diverse range of the content.
    - *What are possible errors the MGM could make? How could we select samples that can address our concerns?*
    - *Selected some samples that are “representative” of the collection*
2. Determined the quality measurement method and ground truth format
    - *Let our use cases guide us. What does success look like?*
    - *Kept it as simple as possible. For example, if use cases only require a list of words detected in video OCR, we didn’t worry about accuracy of timestamps or bounding boxes*



For each MGM candidate

3. Create ground truth
4. Generate MGM output
5. Convert MGM output to ground truth format
6. Compare MGM results against ground truth
7. Calculate quality metrics
8. Review scores and outputs

https://github.com/AudiovisualMetadataPlatform/mgm-assessment-scripts

MGM evaluations to date

Evaluated *(- and selected)*

Speech/music/silence segmentation
- *INA Speech Segmenter*

Speech-to-text transcription
- *AWS Transcribe*
- *Kaldi*

Speaker diarization
- *AWS Transcribe*

Named Entity Recognition (NER)
- *AWS Comprehend*
- *SpaCy*

Video OCR
- *MS Azure VideoIndexer*
- *FFMpeg + Tesseract*

In progress

Scene/shot detection

Structured OCR of supplemental materials

Facial recognition (of selected individuals)

Applause detection in musical performances

Instrumentation detection in musical performances

Evaluation summaries at https://go.iu.edu/amppd
under Documentation > MGMs (Metadata Generation Mechanisms)

Initial goals

BPM (beats per minute)
Color information (chroma values)
Color/BW
Date (other)
Duration
Ethnicity
Event (e.g., basketball game)
Frequency information (audio frequency)
Full text
Gender
Geographic (other)
Keyword
Language
Linked relationship (URL/URI)
Music present (binary: present/not)
Music/Speech (binary)
Note
Part/component
Phonemes (phonetic transcript)
Publisher
Relationship
Sound/silent (binary)
Source (e.g., provenance, donor)
Type of resource

AMP
AUDIOVISUAL METADATA PLATFORM

Revised goals

Produce data that is "good enough."

- May still require human intervention, either through Human MGMs or work performed after the data is processed.
- Researchers may be able to identify which (if any) names are creators, dates represent the time of creation, or keywords are subject headings.

AMP
AUDIOVISUAL METADATA PLATFORM

Metadata extracted with AMP

- Keyword *(NER, Video OCR)*
- Name *(NER, Video OCR, Program OCR)*
- Geographic location *(NER, Video OCR)*
- Date *(NER, Video OCR)*
- Full text *(Speech-to-text, Video OCR)*

AMP
AUDIOVISUAL METADATA PLATFORM

Metadata extracted with AMP

- Structural metadata for navigation *(Applause detection, Shot detection)*
- Information to support more efficient archival processing, copyright review *(Shot detection, silence detection, Program OCR)*

AMP
AUDIOVISUAL METADATA PLATFORM

Customizing/training MGMs

- NER custom vocabularies

  (find relevant keywords in NYPL transcripts with Homosaurus vocabulary and in IU Archives transcripts with IU Archives browse terms)
- Structured OCR

  (identify performers/instruments and composers/works in IU concert programs)
- Facial recognition

  (identify former IU president Herman B. Wells in IU Archives videos)

Bias and privacy issues

Social impact

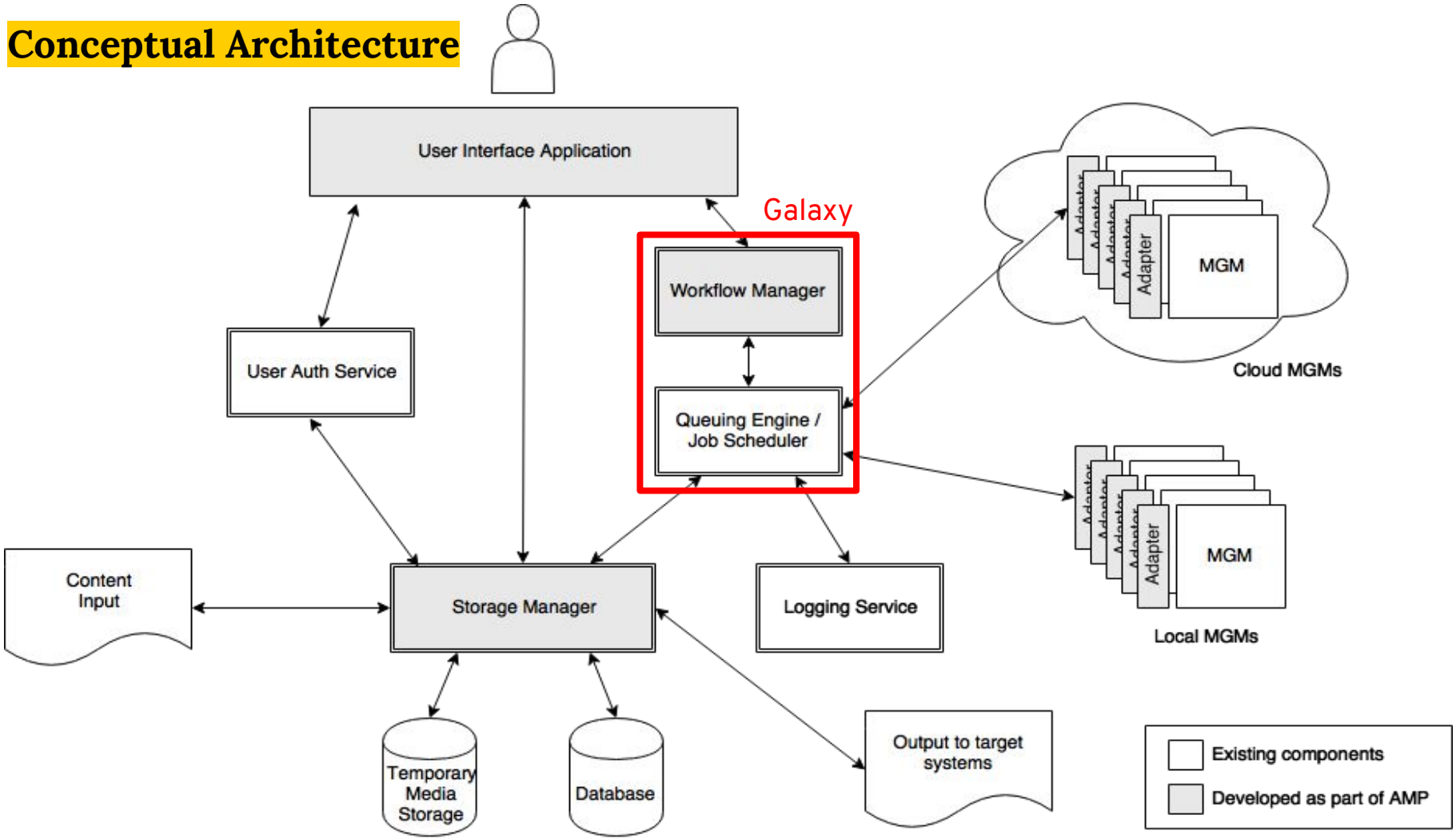
The potential unintended consequences of an unmediated MGM's output. How could the MGM express hidden biases? What are the possible unintended negative impacts that could come from the output of this MGM? What measures can be taken to mitigate them? See FAT/ML's Principles for Accountable Algorithms for more information.

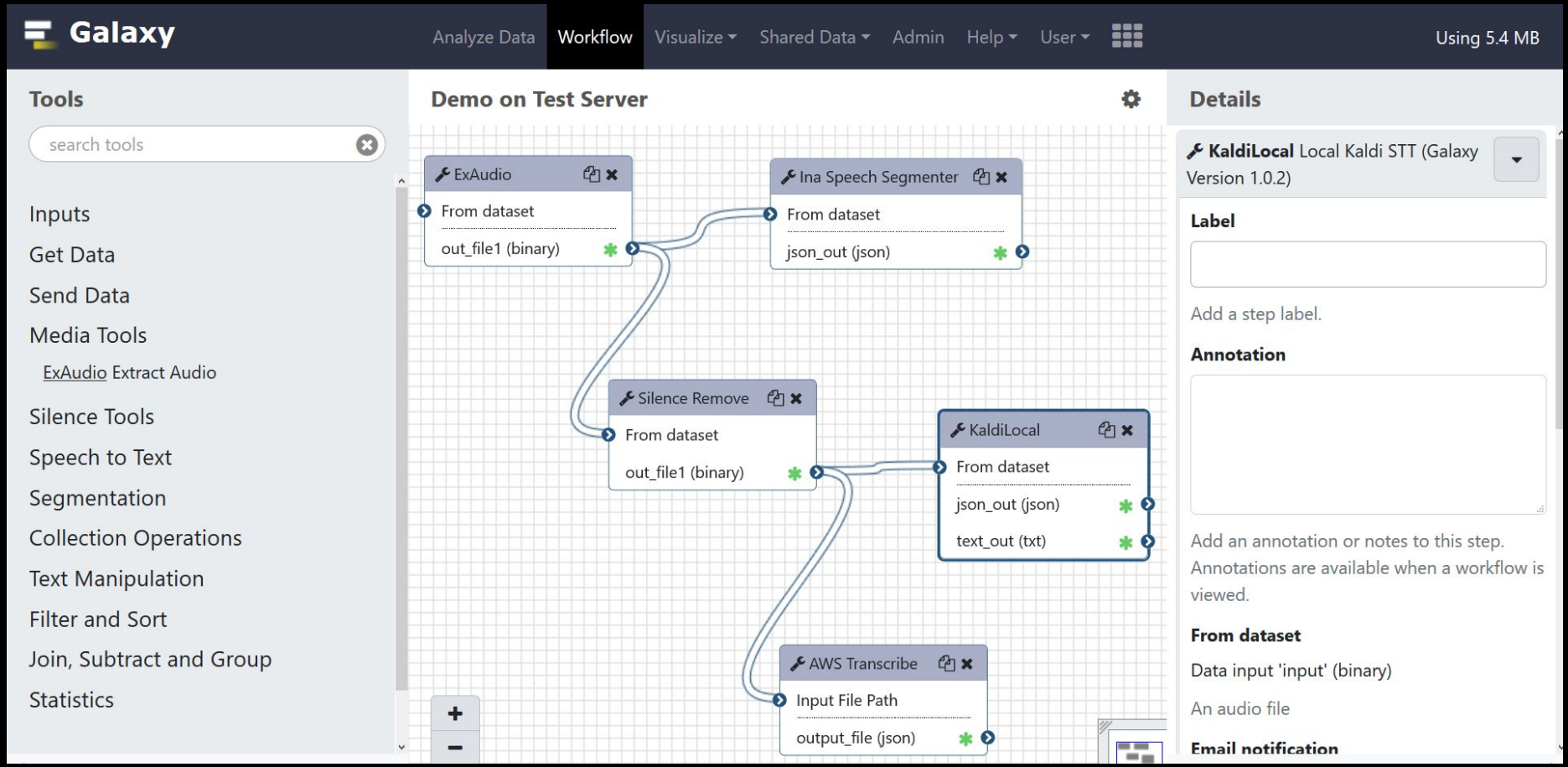
Platform Architecture and Workflows

AMP

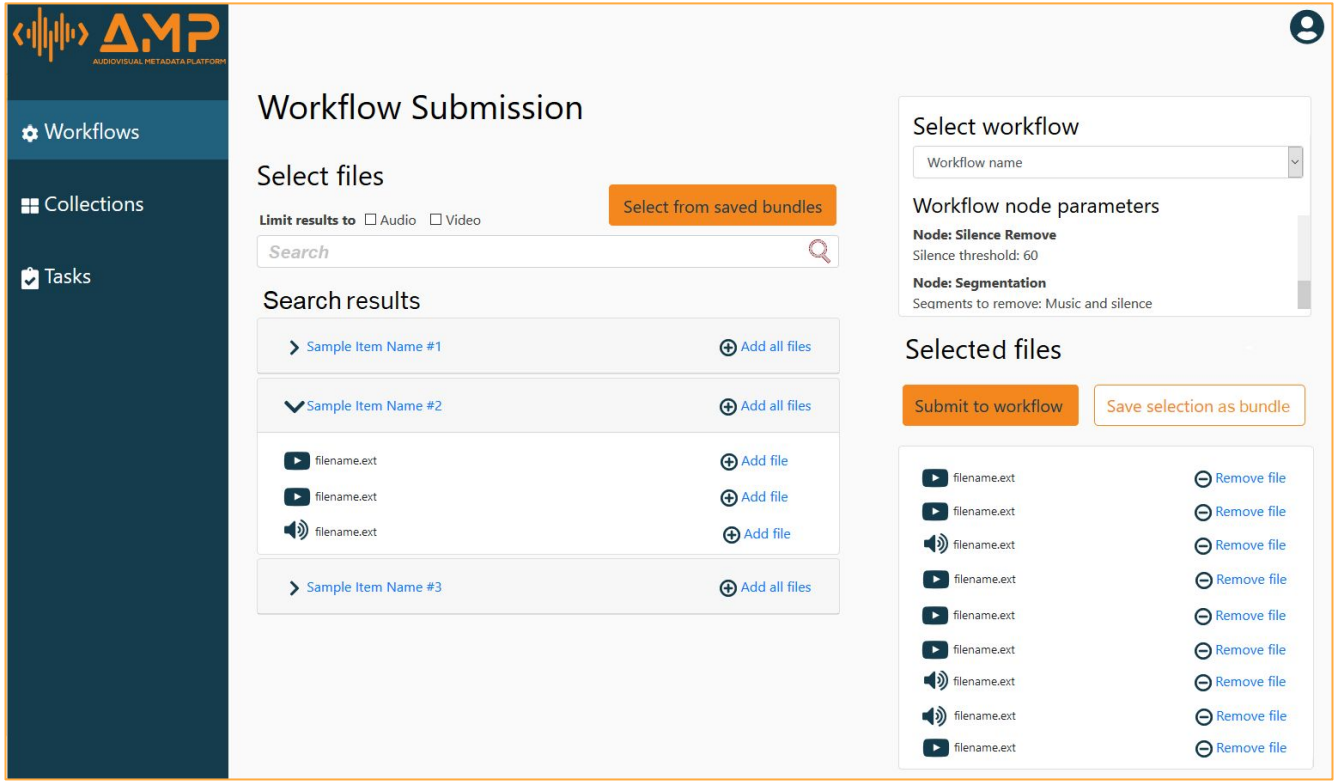
AUDIOVISUAL METADATA PLATFORM

Conceptual Architecture

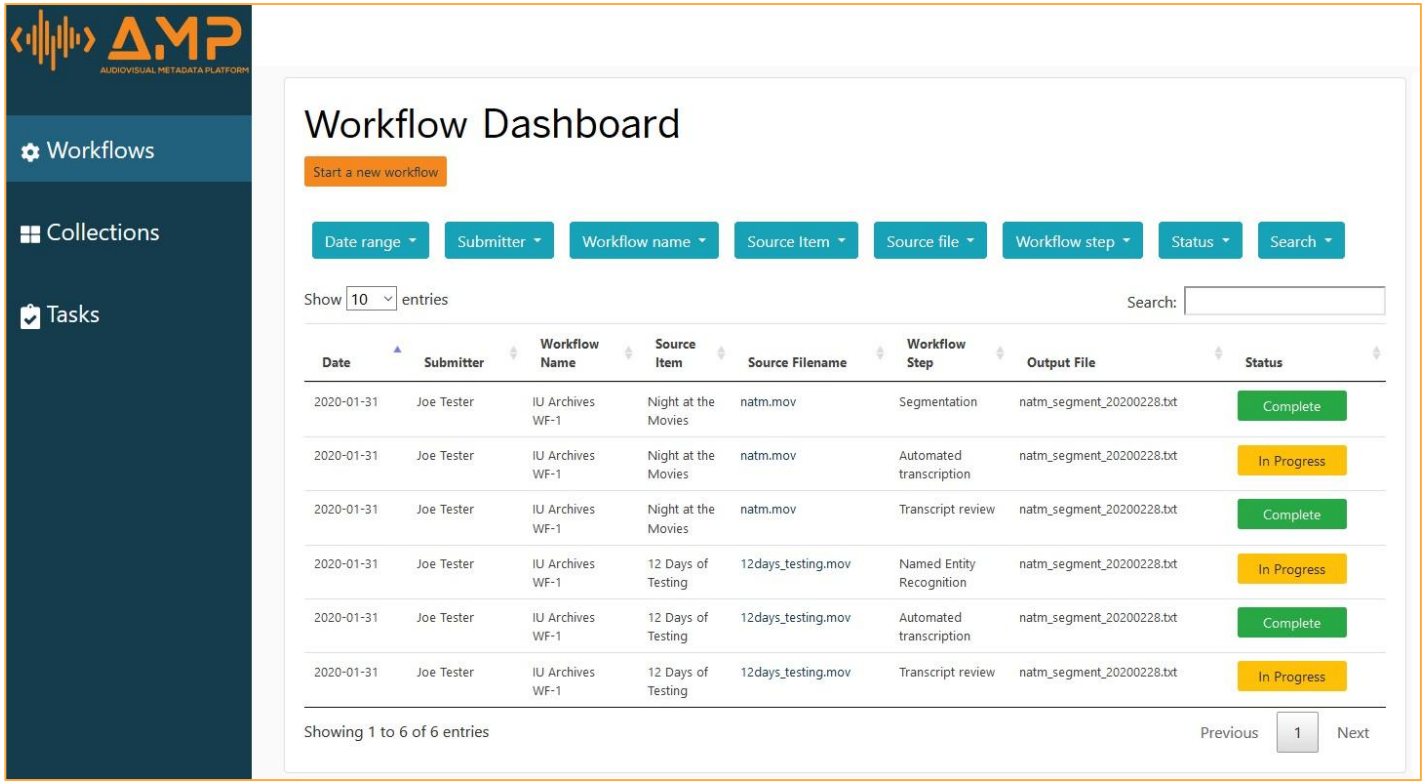



Galaxy
Analyze Data
Workflow
Visualize
Shared Data
Admin
Help
User
Using 5.4 MB
Tools
search tools
Inputs
Get Data
Send Data
Media Tools
ExAudio Extract Audio
Silence Tools
Speech to Text
Segmentation
Collection Operations
Text Manipulation
Filter and Sort
Join, Subtract and Group
Statistics
Demo on Test Server
ExAudio
From dataset
out_file1 (binary)
Ina Speech Segmenter
From dataset
json_out (json)
Silence Remove
From dataset
out_file1 (binary)
KaldiLocal
From dataset
json_out (json)
text_out (txt)
AWS Transcribe
Input File Path
output_file (json)
Details
KaldiLocal Local Kaldi STT (Galaxy Version 1.0.2)
Label
Add a step label.
Annotation
Add an annotation or notes to this step. Annotations are available when a workflow is viewed.
From dataset
Data input 'input' (binary)
An audio file
Email notification

AMP User Interface



AMP User Interface



Human MGMs

- Workflow steps that require human intervention rather than automated processing

- Examples in AMP:
  - Transcript correction (BBC Transcript Editor)
  - Named Entity Recognition revision (Avalon Timeliner)
  - Segmentation validation/editing (Avalon Structural Metadata Editor)

- Integrating with workflow tool (Jira) to support task management

AMP

AUDIOVISUAL METADATA PLATFORM

Next Steps

AMP

AUDIOVISUAL METADATA PLATFORM

Next Steps: AMPPD Current Phase

- Complete workflow design
- Integrate additional automated and human MGMs
- Workflow execution/evaluation against test collections
- Continue to refine application UI and functionality

AUDIOVISUAL METADATA PLATFORM

Next Steps: Future Work Phase

- Production implementation
- More focused MGM implementation/training
  - Integration of research-based tools, e.g. for music IR
  - Explore training of new models and additional training of existing models (transfer learning)
  - Requires additional data science expertise within the project
- Use of generated metadata for discovery and access

AUDIOVISUAL METADATA PLATFORM

Thank You!

More information on AMP:

- go.iu.edu/amppd
- twitter.com/AVMetadata

(And more information on IU's MDPI: mdpi.iu.edu)