AUDIOVISUAL METADATA PLATFORM (AMP) PLANNING PROJECT

PROGRESS REPORT & NEXT STEPS

avp

INDIANA UNIVERSITY

TEXAS
The University of Texas at Austin

This publication was made possible through a generous grant from The Andrew W. Mellon Foundation.

Report Publication Date

March 27, 2018

Report Authors

Jon W. Dunn and Juliet L. Hardesty, Indiana University Libraries
Tanya Clement, School of Information, University of Texas at Austin
Chris Lacinak and Amy Rudersdorf, AVP

Table of Contents

Introduction & Project Background

This report documents the outcomes of a workshop funded by the Andrew W. Mellon Foundation and hosted by Indiana University as part of a planning project for design and development of an audiovisual metadata platform (AMP). The platform will perform mass description of audiovisual content utilizing automated mechanisms linked together with human labor in a recursive and reflexive workflow to generate and manage metadata at scale for libraries and archives. The partners leading this planning project were the Indiana University (IU) Libraries, University of Texas at Austin (UT) School of Information, and AVP.

The AMP workshop was specifically focused on determining the technical details necessary to build the system, and bridging the gap between prior work of the project partners and future implementation. The workshop brought together individuals from within and outside the partner organizations, all of whom have relevant expertise and experience to assist the partners in analyzing the needs for the system and identifying the best technologies and approaches to building a functioning prototype. The workshop participants were:

- Adeel Ahmad, AVP (AMP Project Team Member)
- Kristian Allen, UCLA Library
- Jon Cameron, Indiana University
- Tanya Clement, University of Texas at Austin (AMP Project Team Member)
- Jon Dunn, Indiana University (AMP Project Team Member)
- Maria Esteva, Texas Advanced Computing Center, University of Texas at Austin
- Michael Giarlo, Stanford University
- Juliet Hardesty, Indiana University (AMP Project Team Member)
- Chris Lacinak, AVP (AMP Project Team Member)
- Brian McFee, Music and Audio Research Laboratory, New York University
- Scott Rife, Library of Congress
- Sadie Roosa, WGBH Media Library and Archives
- Amy Rudersdorf, AVP (AMP Project Team Member)
- Felix Saurbier, German National Library of Science and Technology
- Brian Wheeler, Indiana University
- Maria Whitaker, Indiana University

In the years leading up to this workshop, the project partners had embarked upon various initiatives investigating audiovisual description. In 2015, IU and AVP investigated models and developed a strategy for high-throughput description of audiovisual materials that are being digitized as part of IU's Media Digitization Preservation Initiative (MDPI).[1] AVP gathered information through interviews with collection managers at IU and users of MDPI content to understand whether metadata exists (it often does not), and if so, in which formats (video, audio, handwritten documents), applications (.xlsx, databases), and/or structures (.xml, .csv, .txt) it resides. Collection managers also identified optimal output formats and potential uses for the metadata, and considered related rights and permissions issues for the digitized objects and their metadata. These interviews resulted in (a) the establishment of a set of metadata fields for optimized discovery of audiovisual assets in IU's audiovisual access system called Avalon,[2] (b) identification of the metadata fields' value for discovery beyond Avalon, and (c) the values of those fields in the generation of other or subsequent metadata (e.g., general keywords can be analyzed to produce specific names, subject terms, and dates).

AVP then identified nearly 30 existing metadata generation mechanisms (MGMs) for populating the proposed metadata fields. These include, for example, natural language processing, facial recognition, legacy closed caption recovery, as well as human generated metadata and OCR of images and transcription, which have the potential for capturing and producing metadata at a massive scale when unified in the modular AMP architecture.[3]

AVP's initial research led to a proposal for an iterative approach to metadata capture, generation, and enhanced re-generation, wherein the full suite of envisioned MGMs would be deployed in three phases.

First-phase MGMs would produce sets of data that could be analyzed by second- and third-phase MGMs. By phase three, MGMs would begin to integrate various outputs from early processes to augment granular and topical description, ultimately increasing discoverability and usability. Throughout the three phases, AMP would act as the workflow engine, pushing data from one MGM to the next, as well as:

- serving as a decision engine,
- storing metadata for processing,
- providing a metadata warehouse for longer-term storage of all metadata generated, and,
- serving as a metadata source for target systems such as Avalon.[4]

As part of their initial study, AVP analyzed costs, staffing allocations, technology, and services required to implement AMP at IU. This project offered IU:

- an architecture and strategy for AMP,
- a realistic view of the resources, staffing, etc., required to implement AMP, and

---

[1] https://mdpi.iu.edu/
[2] Funded in part by grants from the Andrew W. Mellon Foundation and Institute of Museum and Library Services (http://avalonmediasystem.org/)
[3] See Framing Statement Appendix - MGMs and Descriptive Metadata for a diagram of possible MGMs and metadata fields potentially supported by them.
[4] See Appendix B for a diagram of the high-level data flow architecture.

- the opportunity for vast improvements to discoverability of and access to their audiovisual collections.

The MDPI metadata strategy project, then, provided a strong foundation for the 2017 AMP workshop discussions.

Parallel to the work performed at IU, Tanya Clement's High Performance Sound Technologies for Access and Scholarship (HiPSTAS) project[5] at the University of Texas is conducting research on how users can better access and analyze spoken word collections of interest to humanists through:

- an assessment of scholarly requirements for analyzing sound,
- an assessment of technological infrastructures needed to support discovery, and
- preliminary tests that demonstrate the efficacy of using such tools in humanities scholarship.

The HIPSTAS project has produced and documented workflows to show the movement and organization of files in "jobs" for the analysis of large collections of media. The workflows have been tested on collections of cultural heritage audio recordings including field recordings, oral histories, poetry performances, radio programs, and speeches at the UT Austin's School of Information, as well as several other institutions. Output metadata about these files includes genre and speaker identification, among other features.

Taken together, the work of IU and AVP begins to outline a high-level conceptual design of a system like AMP. The work of UT, specifically, presents a highly specialized and thoroughly researched example of an MGM. The cumulative work of all project partners (IU, UT, AVP) has informed their deep understanding of the metadata requirements of curators, collection managers, researchers, and end users. With a high-level model and clear knowledge of user requirements, it became obvious to project partners that in order to build a mass-scale metadata generation platform, convening a diverse group of specialists to discuss and advise on the technical components and architecture would be a valuable next step. Thus, in September 2017, a meeting of the project team and specialists listed above was held on the campus of Indiana University in Bloomington, Indiana.

---

[5] Funded by National Endowment for the Humanities and the Institute of Museum and Library Services (https://blogs.ischool.utexas.edu/hipstas/).

AMP Rationale, Goals, & Purpose

Libraries and archives hold massive collections of audiovisual recordings from a diverse range of timeframes, cultures, and contexts that are of great interest across many disciplines and communities.[6]

In recent years, increased concern over the longevity of physical audiovisual formats due to issues of media degradation and obsolescence,[7] combined with the decreasing cost of digital storage, have led institutions to embark on projects to digitize recordings for purposes of long-term preservation and improved access. Simultaneously, the growth of born-digital audiovisual content, which struggles with its own issues of stability and imminent obsolescence, has skyrocketed and continues to grow exponentially.

In 2010, the Council on Libraries and Information Resources (CLIR) and the Library of Congress reported in "The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age" that the complexity of preserving and accessing physical audiovisual collections goes far beyond digital reformatting. This complexity, which includes factors such as the cost to digitize the originals and manage the digital surrogates, is evidenced by the fact that large audiovisual collections are not well represented in our national and international digital platforms. The relative paucity of audiovisual content in Europeana and the Digital Public Library of America is testament to the difficulties that the GLAM (Galleries, Libraries, Archives, and Museums) community faces in creating access to their audiovisual collections. Europeana comprises 53% images and 43% text objects, but only 1.4% sound objects and 2.25% video objects.[8] DPLA is comprised of 50% images and 44% text, with only 0.34% sound objects, and 0.34% video objects.[9]

Another reason, beyond cost, that audiovisual recordings are not widely accessible is the lack of sufficiently granular metadata to support discovery, identification, and use, or to support informed rights and access decisions on the part of collections staff and users. Unlike textual materials—for which some degree of discovery may be provided through full-text

[6] See for example, *Quantifying The Need: A Survey Of Existing Sound Recordings In Collections In The United States.* AVP and the Northeast Document Conservation Center. https://www.weareavp.com/quantifying-the-need-a-survey-of-existing-sound-recordings-in-collections-in-the-united-states/
[7] See Casey, Mike (2015). "Why Media Preservation Can't Wait: The Gathering Storm." *IASA Journal* 44, 14-22. Available at https://www.weareavp.com/mike-casey-why-media-preservation-cant-wait-the-gathering-storm/
[8] Europeana. http://www.europeana.eu/portal/en/search?q=
[9] DPLA. https://dp.la/search

indexing—without metadata detailing the *content* of the dynamic files, audiovisual materials cannot be located, used, and ultimately, understood.

Traditional approaches to metadata generation for audiovisual recordings rely almost entirely on manual description performed by experts—either by writing identifying information on a carrier, typing bibliographic information into a database or spreadsheet, or creating collection- or series-level finding aids. The resource requirements and the lack of scalability to transfer even this limited information to a useful digital format that supports discovery presents an intractable problem. Lack of robust description stands in the way of access, ultimately resulting in the inability to truly derive value from collections of audiovisual content, which in turn can lead to lack of interest, use, and potential loss of a collection entirely to obsolescence and media degradation.

What is required for full descriptive access to audiovisual objects at scale is a variety of mechanisms (both automated and manual) working together to perform analysis of media and their associated materials (such as transcripts or transcribed information on carriers) in order to generate usable and meaningful metadata. These mechanisms might include natural language processing, speech to text, facial recognition, silence detection, scene detection, music detection, language recognition, manual description, optical character recognition, object recognition, and more. It is not exclusive to automated mechanisms, however. For the greatest success, automated mechanisms must work in concert with human labor managed by a recursive and reflexive workflow engine that supports an ecosystem of open source and proprietary tools and services, in local and cloud-based systems. The metadata must be compiled, refined, and delivered to a metadata warehouse where it can be harvested by target systems. At the same time, it must remain available to the MGMs for continued and ongoing "cultivation" by the evolving mechanisms' technologies and machine learning. In this way, the metadata remains in a constant and active state of refinement.

An intuitive system that is easy for non-developers and non-technical caretakers of collections to use would profoundly change the prospect for future access to hundreds of millions of hours of audiovisual content and open up collections in extraordinarily meaningful ways. The overwhelming resources required to meet even a fractional level of description using traditional workflows is incomparable to what might be achieved with AMP. The eventual goal, then, would be to make AMP available to libraries and archives to maximize their audiovisual assets' findability and usability by all users who require them.

Usage Scenario

Figure 1 below provides a high-level view of the AMP process from a content owner perspective. It demonstrates a user logging in, loading or creating a job, and running a job through to completion. For the purpose of understanding the workflow diagram, it is important to define a few terms in the context of AMP:

MGM: A *metadata generation mechanism* used in a workflow which performs a task that results in the production of metadata

Workflow: One or more MGMs and optional workflow tools linked together in order to process and analyze audio, video, still image, and/or text

Workflow Tool: A utility used in a workflow which performs a task that does not directly result in the production of metadata

Job: The assignment of a workflow to a specific set of files

For the sake of simplicity, this figure below demonstrates a single job, although it is envisioned that multiple jobs will run simultaneously.

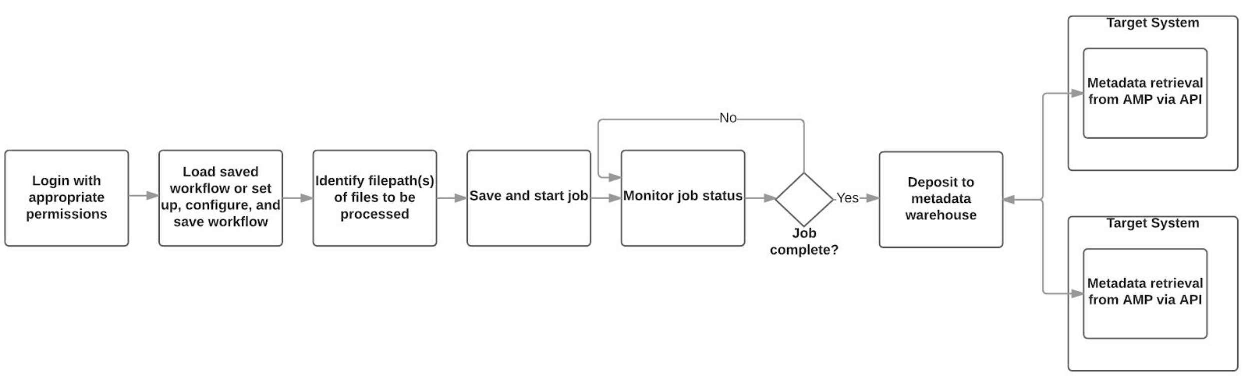


Figure 1. Example content owner interaction

Appendix C illustrates an example workflow. This shows the linking of MGMs and workflow tools. This figure in Appendix C does not show the inputs and outputs of each MGM but is intended to offer the reader insights into the vision for the type of MGMs that may be included in a workflow.

Prior Work

In addition to the work performed by the project partners described in the introduction, there have been several open source and commercial efforts to date that demonstrate the possibilities for computationally assisted metadata generation and improved discovery. Our research found that in many cases these systems are fixed, one-way pipelines that are designed to address a single type of content. For instance, an oral history application may employ two MGMs—a speech-to-text tool and natural language processing tool—in a workflow to process audio files and output a text-based document as a result. While a workflow like this serves as a proof-of-concept for AMP, it fails to meet the needs defined by the use cases and requirements for AMP. The needs call for a system with the following characteristics:

- Dynamic and flexible
- Two-way synchronous communication throughout
- Extensible, scalable, and modular
- Inclusive of human generation of metadata
- Contains conditional logic
- Outputs to a variety of data formats
- Topic or subject area agnostic

There are other systems that specialize in a particular type of content or subject, including the multi-institution tool MALACH (Multilingual Access to Large spoken ArCHives)[10], Cornell Lab of Ornithology's Raven[11], and BBC's Comma.[12] While the level of description these systems generate is extensive and deep for the subjects and content types for which they were built, the focus is narrow relative to the goals of AMP. Also, these systems are not built to handle the scale demanded by AMP. Additionally, these systems typically do not meet the rights and permissions requirements that are fundamental to AMP.[13]

Most closely approximating the goals of AMP are the commercial platform GrayMeta[14] and the open European Union project MiCO.[15] Both of these systems share many attributes of an envisioned AMP system, including the ability to deploy a variety of MGMs, analysis of multiple

[10] https://malach.umiacs.umd.edu/
[11] http://www.birds.cornell.edu/brp/raven/RavenOverview.html
[12] http://www.bbc.co.uk/rd/projects/comma
[13] These comments are not meant to disparage the great work behind these tools or the important work they perform. They are only meant to demonstrate why they are not the right fit for the vision and requirements laid out for AMP.
[14] https://www.graymeta.com/
[15] https://www.mico-project.eu/

media types, and storage of resulting metadata in a metadata warehouse. The MiCO project has extensive documentation that is publicly available. However, there is little public documentation available about GrayMeta, which is a licensed product used primarily by two domains (broadcasting and advertising).Thus, the findings of our GrayMeta research are based only on system demonstrations and review of the limited marketing materials available, but it was clear from that research that the goals of GrayMeta differ considerably from those of both the MiCO and AMP projects.

A core tenet of AMP is the inclusion of MGMs involving human interaction when cataloging or subject matter expertise is required, for the refinement of automatically generated metadata, and when human feedback supports or drives machine learning. Neither MiCO nor GrayMeta currently incorporates human-interaction MGMs in the integrated way that this is envisioned for AMP.

Additionally, AMP aims to utilize related supplementary documents (e.g., catalog records and transcripts) to augment a media file's metadata. This concept is not represented in either system. Instead, each object in MiCO and GrayMeta is treated as a single information package distinct from all other sources of data.

The target market for AMP is libraries and archives. Based on GrayMeta's public literature and some of the demonstrated features, they are highly focused on broadcast and advertising. The MiCO project is focused on video production environments and animal identification and analysis. This does not mean that their use cannot be extended to other disciplines, but their immediate target markets do influence their current implementation of MGMs and choice of media types and system features. In both cases, the media content on which they focus is contemporary and consistent (broadcast quality), and for that reason is high quality and relatively easy for automated MGMs to work with. Library and archival content, by contrast, is challenging due to the extreme variation across collections in content type, recording quality, recording specifications, and subject and discipline areas. The variability that exists in libraries and archives, then, requires specialized system design and performance and is a significant reason for AMP's emphasis on generation and refinement by humans.

Based on workshop discussions, there is a strong preference for the use of open source technologies for AMP. AMP will support an ecosystem that marries commercial, proprietary MGMs with open source MGMs (although open source components are strongly preferred where possible). For this reason, use of a commercial, proprietary, "black box" system to fulfill many, or all, of AMP's functions was deemed undesirable by workshop participants.

This is particularly true when considering the paradigm shift that AMP, and platforms like AMP, require in thinking about ownership. Traditionally, the value derived from the work to describe content has been the text output itself. Whether the description was performed in-house, through a service, or was performed by humans or automated processes, value and ownership remained on the data that was produced through the descriptive process. The deliverables have always been static text, in .txt, .doc, .docx, or .pdf documents, or structured in .xml, .csv, .json, MARC or other formats. AMP shifts this paradigm by storing the metadata outputs in a metadata warehouse where it will continue to be cultivated, groomed, and refined as MGMs evolve or as additional MGMs are integrated into AMP.

Machine learning will be leveraged over time as new algorithms are incorporated into AMP and humans continue to refine and provide feedback to improve the accuracy and performance of the MGMs. Amazon Web Services summarizes machine learning (ML) training as "…providing an ML algorithm (that is, the *learning algorithm*) with training data to learn from. The term *ML model* refers to the model artifact that is created by the training process."[16] The evolving machine learning algorithms and models and the human resources that help to create a smarter machine will become one of AMP's most valuable assets. With this in mind, ownership of the "machine" and the transparency of its functions becomes as important as the outputs themselves. It is critical that a library or archive own and benefit from the algorithms, models, and human resources that continuously build a smarter machine. AMP functionality must be transparent to support institutions' roles as stewards of collections they are charged with preserving, and to ensure the data they are producing is authentic and trustworthy to the greatest extent possible. Such transparency would not necessarily benefit a commercial entity, as it would expose the core of the system's value. Institutions that value—and indeed trade—on openness, trust, and neutrality, would not have access to the inner workings of the commercial system, making it difficult to fulfill their missions. This would effectively minimize the returns on their own human and financial investments and metadata quality, and the trust in its authenticity over time would suffer.

[16] Amazon Web Services. "Training ML Models."
http://docs.aws.amazon.com/machine-learning/latest/dg/training-ml-models.html

Meeting Logistics

Project partners recognized early on that to architect the best system possible, input from a larger group of experts would be necessary. The six members of the core team conceived of a workshop to explore the technical possibilities and requirements of AMP. Ten experts were invited from eight different institutions from as far away as Germany, and met over the course of two and a half days on September 13-15, 2017, at Indiana University in Bloomington, Indiana.

Day one of the workshop set the stage for the rest of the meeting. It included an overview of the project by the project team, framing and meeting goal setting, and a review of requirements, personas (end users who will make use of the generated metadata for access and discovery), and the current technical landscape. Technical requirements, non-technical requirements, and component types and instances for AMP were then identified through participant activities and discussion. The day ended with a group diagramming exercise to synthesize the discussions and begin a visualization of AMP.

The second day of the workshop involved clarifying what was meant by “workflow manager” as discussed on day one. The group then proceeded to identify component candidates by name for the various functional components of AMP and determined potential MGM candidates for different types of media processing and metadata extraction.

Day three of the workshop brought all of these discussions together by presenting four different system diagramming and metadata generation scenarios using example items from Indiana University’s collections (e.g. oral histories in Italian, basketball halftime segments from the 1980s). Each example was mapped using system components and MGMs identified in the previous two days to visualize how processing various types and quantities of media and existing description could work, and how each of the different examples affected the order and processing of MGMs.

AMP Requirements & Architecture

Introduction

Following the workshop, the notes that were taken by project partners were analyzed in-depth and color coded to identify business requirements, functional requirements, non-functional requirements, technical requirements, format requirements, and actors. After the statements were categorized and identified they were extracted, grouped accordingly, and placed into a spreadsheet along with related information. Once they were consolidated and organized, the core project team reviewed and vetted them further in order to come up with a refined set of data representing a path forward. The results of this process can be found in the Appendices and in the narrative that follows.

Actors

Actors represent the primary external applications and users that perform an action or actions to which AMP responds. Actors have a goal that must be satisfied by AMP. Four actors were identified in the workshop as the main users of AMP or applications that interact with AMP:

AMP Administrator (User)
The AMP Administrator is the AMP power user. This user is responsible for technical and functional administration of AMP, system configuration, workflow management, management of user permissions and access, and provision of user support.

Content Owner (User)
The Content Owner interacts with the AMP user interface to input and generate metadata content. They are typically the owners, creators, or caretakers of the media and related documentation that are ingested into AMP and processed by the MGMs to create metadata that will be used by them with or without the assistance of a Target System Application or by another Target User. Examples of Content Owners include archivists, librarians, collection curators, or creators of the source media.

Target System (Application)
The Target System is an application that uses the metadata produced from AMP via APIs and/or structured data (e.g., xml) download. Examples of Target Systems include digital asset management systems, content management systems, and discovery platforms.

Target User (User)

Target Users obtain outputs from AMP via a direct download of structured metadata (e.g., .csv or .xml files). Examples of Target Users are professors and researchers, librarians, archivists, students, and others in need of this metadata. As part of this project, a set of personas were developed to illustrate each of the main Target User types. These personas are available in Appendix A.

Business Requirements

Business requirements were identified and documented through analysis of the workshop notes and the relevant statements and themes that surfaced in discussions throughout the two and a half days. These ideas, along with business requirements created by the core team prior to the workshop, and which were confirmed by participants, generated the eleven business requirements below.

Business requirements define the overarching needs and goals of the organization that AMP is meant to address. These inform the definition of functional requirements and use cases that typically follow their development. Business requirements define the goals of the new technology.

The list of AMP business requirements and a brief description of each is included below.

1. Automate analysis of audiovisual content and human-generated metadata in a variety of formats to efficiently generate a rich set of searchable, textual attributes

A significant goal of the AMP initiative is to make the metadata generation process as efficient as possible, and to support that process with human generation where necessary. Inputs may be media only or related structured or unstructured textual documentation that remain associated within the system. Together, the media and related data and the MGMs will create robust and searchable metadata.

2. Provide an intuitive interface for metadata generation for novice and infrequent users

It was agreed upon by all participants that a user interface simple enough to run by general users—such as Content Owners, librarians, and archivists—was imperative for the project to be practical for its user base and increase the possibility of wider adoption.

3. Offer streamlined metadata creation by leveraging multiple, integrated, best-of-breed software tools in a single workflow

In order to produce the richest metadata possible, the automated MGMs must be the best, most up-to-date possible. They must remain current to ensure that AMP can utilize the greatest and most robust functionality available for metadata generation and support the evolution of the metadata over time.

4. Produce and format metadata with minimal errors

Along with richness, metadata with the least errors possible is imperative. We acknowledge this will be a challenge, but also recognize that metadata will continue to improve over time as the systems that produce it evolve. As such, metadata quality will evolve alongside the systems that will generate it.

5. Offer a variety of metadata outputs for consumption by multiple discovery and collection management systems

Metadata will be structured in a system-agnostic data model that enables the widest use. Metadata will be available through APIs, as well as by download via standard formats such as csv and xml.

6. Build a community of developers in the cultural heritage community who can develop and support AMP on an ongoing basis

The workshop participants stated on a number of occasions, most specifically during the non-technical requirements discussions, the need to build a community around AMP. Without a community of developers, AMP will struggle to gain wide adoption. The cultural heritage community already has a track record of successes in this area, including Fedora and Samvera (of which Avalon is a component), and it is the hope of this group that AMP will find the same sort of adoption and ongoing support.

7. Offer APIs that can be used with little training or knowledge of AMP

All workshop participants agreed that to ensure wide adoption, AMP must be intuitive and easy to use. Straightforward user interfaces for Content Owners (see #2) and equally intuitive APIs for Target Users are key. APIs must be straightforward and well documented.

8. Scale to efficiently process multi-terabyte batches of content

Digital audiovisual media are sometimes multi-gigabyte (or even terabyte) files. AMP must be able to ingest, manage, and process multiple large-format files at a time.

9. Support collaborative efforts with similar initiatives (e.g., MiCO).

AMP is not the only platform that exists in this domain. Others have very different purposes (specific domains or audiovisual formats) or motivations (profit driven). There may be potential areas of current or future functional overlap, so the initiative must remain open to the possibility of working with or using some of these parallel technologies.

10. Enable straightforward deployment of AMP for system administrators

Ease of use for content owners and metadata users aside, AMP must be easy to stand up and manage for system administrators. Not only should the system be intuitive, but also accompanied by documentation and/or robust online help channels to ensure administrators' success.

11. Automatically control what metadata is made publicly available, and what is not

A significant concern for content owners is the rights of reuse of the media and related files that will be processed via AMP. For some content, intellectual property concerns restrict the metadata from use beyond that of the Content Owner's home institution (or even department). Additional concerns, including personal privacy, must be addressed to ensure that personal data is exposed only to users with permission. Rights and permissions security are therefore a major requirement for AMP.

Functional Requirements

Functional requirements express characteristics and functions of a system designed to meet the needs of AMP. The needs are reflected in general requirements concerning issues of asset management, copyright and security, storage, metadata standards, modularity, tenancy, scalability, and usability. Through this process we generated fifty-four functional requirements (listed in detail in Appendix D) that include requirements for APIs, databases, file management, ingest, metadata, MGMs, multi-tenancy, organization, queueing, reporting, resources, scaling, search, user administration, user interface, versioning, and workflow.

Format Requirements

Format requirements refer to the file formats that would be processable through AMP, e.g., media files, transcripts, and other already existing supplementary metadata. Throughout the workshop discussions, file formats were identified that AMP needs to be able to manipulate to produce metadata for Target Systems and Target Users. The formats included both audio or video media and text and image files, as well as package formats. The following formats are possible sources for AMP processing:

- Audio (.mp3, .wav)
- Image (.eps, .jpg, .pdf, .png, .tif)
- Data (.xlsx, .csv, .ttl, .json)
- Presentation (.key, .pptx)
- Video (.mov, .mp4, .mkv, .mts, .mxf)
- Structured text (.xml, with or without defined schemas, such as TEI, MODS, EAD, MARCXML)
- Unstructured text (.txt, .docx)

AMP will need to process audio and video files to extract metadata, and additionally accept and incorporate metadata from supplementary documents about the content or technical or structural information within or about those files. Two scenarios from day three of the workshop involved media with extensive library catalog records that would be processed via MGMs similarly to the media files. The outputs of both would be associated in a single metadata record for access by Target Systems or Users, and also as data to be utilized in subsequent MGM processing.

Non-Functional Requirements

Non-functional requirements specify criteria that can be used to judge the overall operation of AMP, rather than its specific behaviors. The proposed system is designed to be robust, intuitive, and capable of processing media at scale, efficiently outputting structured metadata as correctly as possible. In addition to the technology and specific functionality necessary to support this, the workshop participants identified the following non-functional requirements to ensure its successful (i.e. wide) adoption:

- Robust documentation
- Clear service and cost model
- Active community of developers
- Proven stability of the product
- Open licensing
- Established governance model
- Robust outreach and marketing program
- Training opportunities
- Semantic versioning to communicate changes in each release
- Support for internationalization

Of these, the first three bullet points were considered the highest priority by workshop participants. First, detailed documentation will be required for AMP Administrators, as well as for Content Owners. In addition, APIs must be well documented for Target Users to integrate their systems with or otherwise utilize exports from AMP. Second, service and cost models must also be clearly articulated and documented. Finally, for AMP, as with all open source applications, one measure of its success will be the broadness and activeness of its developer community.

Technical Requirements

Technical requirements define the specifications for application development, hosting, database, system interfaces, authentication, and security functionality required by AMP. The intention is to build an application that is versatile and can respond to multiple kinds of users, user communities, and use environments with easily adoptable and sustainable solutions. Thus, the main suggestions for the technical system reflect the same general concerns expressed in the functional requirements.

First, AMP will be easy to install, configure, and deploy. AMP will be a hardware-agnostic application that is written in widely-used programming languages, can be run either locally or in cloud environments (e.g., Amazon Web Services), and is easy to install through command-line or GUI configurations and to deploy on Linux through containers or a bare metal operating system installation. The AMP GUI will be web-based and operate on a wide variety of browsers. As an API-driven system, AMP will support RESTful services for integration with other systems, to the metadata warehouse, between AMP and MGM components, and for data import and export.

Second, AMP should be sustainable. Using standard, commonly used technology components whenever possible for functions such as database, queuing, authentication/authorization, and logging, AMP will be easy to configure for established metadata standards and a wide range of data formats. Tracking versions of MGMs and other components and using standard ways of exchanging information across AMP instances will aid in intelligence gathering about the engine and, thus, in continued development and improvements.

Proposed System Architecture

From discussions in the workshop, the assembled participants began to envision a conceptual technical architecture for AMP that would meet desired requirements for configurability, ease of use, and flexibility in adapting to new MGM implementations, as shown in figure 2.

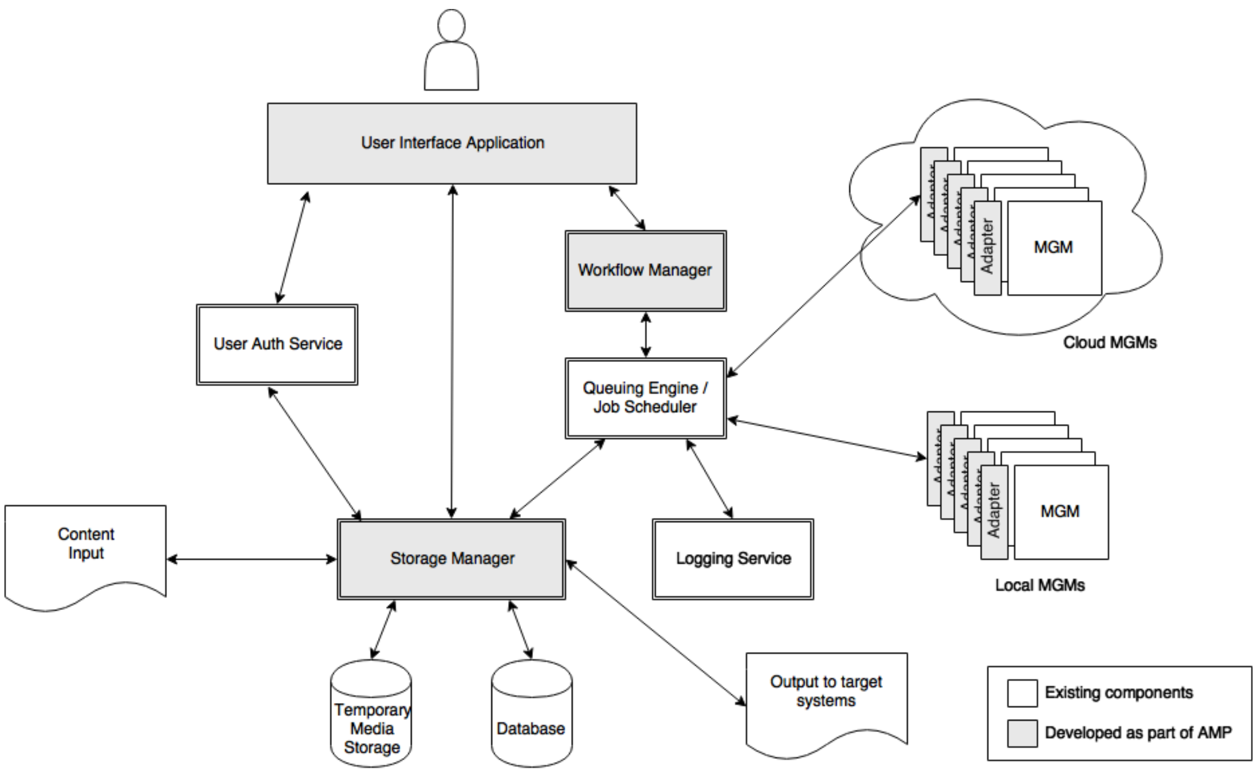


Figure 2. Proposed high-level system architecture

Based on the technical and functional requirements described above, options for existing implementations of a number of the envisioned technical components were discussed. However, the group felt that final selection of specific component candidates and detailed architectural design would need to be undertaken by the team developing the system rather than prematurely determined by the workshop participants, themselves.

Database

A database is required to serve as a data warehouse for storage of both intermediate and final metadata outputs of MGMs and MGM workflows. Given the variability of input-output format structures, a database structure that can flexibly accommodate both key-value pair and hierarchical data is needed, making a NoSQL database likely preferable to a traditional relational database. Potential NoSQL options identified include Mongo DB, Cassandra, Couch DB, Redis, and Level DB.

Storage Infrastructure

Because of the large size of audio and video media files, a design goal of the system is to minimize network movement and copying of AV data as much as possible. The system will support submission of content via URIs referencing Web-accessible content or content available on a filesystem shared with the machine on which the AMP platform components (and possibly MGMs) are running. Individual MGM steps may generate intermediate transformations of metadata that require temporary storage.

Storage Manager

Storage and retrieval functions for content and metadata will be mediated through a storage management API to enable adaptation to various mechanisms for content input and temporary storage (e.g., filesystem, S3, HTTP) and to different database systems in the future.

Job Queuing

A key component of the system architecture is an engine that can queue, execute, and track tasks performed by MGMs when executing workflows across items in a collection. Potential queuing and messaging systems identified include RabbitMQ, ActiveMQ, MQTT, Apache Kafka, AWS SNS, ZeroMQ, and Apache Thrift.

MGMs

Depending on the functional and throughput needs of a given installation of AMP, a variety of MGMs may be used. These MGMs may be either automated or manual (i.e. requiring human intervention), and automated MGMs may operate locally on the same server as AMP, in high-performance computing environments, or in commercial cloud services. See Appendix E for a list of identified possible MGMs. Because the mechanisms for calling particular MGMs, as well as input and output formats supported by MGMs, may vary, each new MGM will require a small amount of code as an “adapter” to support translation between AMP and the expected inputs and outputs of the MGM.

Future Work

Next steps

The September 2017 workshop outcomes refined and clarified concepts that existed previously, confirmed important decision points, and identified some new areas for consideration. The AMP project is now well positioned for implementation to begin. Given the scale of the platform and effort, the AMP project planning team recommends to Indiana University and the Andrew W. Mellon Foundation that a pilot project be undertaken, to serve as a proof of concept, establishing a core project team for implementation, benchmarking critical metrics, and stress-testing platform components.

Following the workshop, the AMP and MiCO teams held a meeting to further explore the possibility of leveraging the outcomes of the MiCO project. Technologies employed, intellectual property concerns, possible deployment options, and other relevant topics were discussed by the teams, offering both a clearer understanding of the two platforms, their overlaps, and their distinctions. The meeting concluded with an understanding by both parties that further collaboration would benefit both MiCO and AMP and that there were no apparent obstacles that would impede collaboration and building off of the work of the MiCO project. With this in mind, the AMP team recommends that the pilot project leverage the MiCO project architecture, using components such as RabbitMQ for messaging, Apache Camel for service Orchestration, and the MiCO Broker used for workflow management. The AMP team also envisions collaborating with the MiCO team during the pilot for assistance in building out further functionality.

The pilot project would consist of building an AMP system for processing representative sample sets from three digitized collections containing different content types (e.g., oral history and song, from different time periods and quality of sound), media types, and metadata extraction requirements.

In order to prioritize allocation of available resources within the pilot project, the full envisioned user experience and range of functionality for AMP would not be an outcome of the pilot. Rather, the project would address and demonstrate:

- User authentication
- Use of the workflow manager to configure workflows and connections between MGMs
- Error reporting and handling
- Workflow status tracking and reporting

- Cloud based MGMs
- Local MGMs
- Open source MGMs
- Commercial MGMs, including setup and tracking of billing costs
- Cost metrics for licensing commercial MGMs
- Automated MGMs
- Human refinement MGMs
- Human generation MGMs
- Classification of MGMs and MGM inputs and outputs
- Use of the messaging component
- Use of the queuing system component
- Use of the logging service component
- Use of the temporary storage component
- Importing existing metadata
- Use of supplementary materials to describe an audiovisual object
- Use of a data warehouse
- Metadata retrieval from AMP via APIs
- Generation of metadata to be used to guide rights, permissions, and access determinations
- Throughput metrics for specific MGMs and the overall system
- Quality metrics on use of varying resolutions of source items for analysis
- Value and utility of AMP metadata in target system

These areas are seen as the essential elements of AMP and should be the points of focus for resource allocation in the pilot. Accordingly, the pilot project should not focus resources on aspects of AMP that are important for full implementation, but which are not new or novel. Assuming successful completion of the pilot project and advancement to fuller implementation, improved user experience and full functionality would be addressed in future implementation phases. Examples of items that would not likely be fully addressed in the pilot, include:

- A comprehensive user interface and user experience design process. The pilot would likely use basic wireframes implemented existing UI frameworks such as Bootstrap as opposed to a fully design process.
- A robust GUI for the workflow manager. The pilot should focus on fulfilling core functionality required to demonstrate the proof of concept. More sophisticated features that improve user experience but do not add core functionality would not be addressed in the pilot.
- Full featured user management. The pilot should incorporate users and demonstrate user authentication, but we do not recommend putting pilot resources into building out extended features for the administration and management of users on the backend.
- AMP API. Within the scope of the pilot we would build out the API enough to demonstrate the functionality for a single target system, namely Avalon.

The proposed pilot is envisioned as a two-year project and would be led by the Indiana University Libraries with AVP serving as technical lead, carrying out development work in collaboration with developers at IU, and with consultation from the University of Texas at Austin School of Information. External financial resources would be required to support software development and testing, server and storage infrastructure, and use of commercial cloud-based MGMs. The pilot would focus on use cases involving three collections: two from IU and one from a third-party partner. Specific collections and partner are still to be defined.

This pilot would serve to validate the core architecture of AMP as well as its feasibility from perspectives of functionality, cost, and the value of the metadata generated. If successful, the pilot implementation would need to be followed by another project to support full implementation of a production-ready AMP, along with development of a business model for its support and sustainability.

Potential Challenges

There are several anticipated challenges in moving forward in the pilot and fuller implementation of AMP. Discussed below, these challenges are expected to be encountered and/or addressed to varying degrees throughout a pilot. During the pilot, every opportunity should be taken to better understand and gain insights into challenges and prospective solutions for planning and implementing AMP in future phases.

Bandwidth

With audiovisual files (which in digital form are typically quite large) as the primary type of data being processed within AMP, storage and bandwidth requirements will be significant. The expense of moving large files around, both in terms of time and, when applicable, bandwidth costs, will make processing these large files a primary challenge in the pilot. Exploring lossless compression techniques for sending and receiving data is one approach. Another area of research includes analyzing the cost-benefit ratio of processing files at different resolutions (e.g., preservation master or access copy) and the quality of the MGM output from each (e.g., accuracy of speech to text or facial recognition).

Computational demands

Along with substantial demands on network bandwidth and storage, machine learning-based MGMs can require significant, and potentially expensive, computational resources, particularly if executed within local infrastructure. Research will be necessary to determine the appropriate balance of MGM deployment across local compute infrastructure, national-level supercomputing resources such as XSEDE, and commercial cloud-based services.

Versioning and synchronization

A core concept of AMP is the continual enrichment and refinement of metadata over time. Implicit in this concept is the ability of target systems to pull updated metadata from AMP and to determine what has changed. The likely solution to this is versioning of metadata generated for each object. The initial set of metadata generated from the processing of a file through an MGM workflow would produce the first version of metadata for that file. Any successive processing through additional workflows would yield modified sets of metadata for the file and therefore a new version of metadata. This process is iterative, and will continue as long as the metadata is stored within AMP. A mechanism for AMP to identify the most current version of a metadata set, and another mechanism to trigger a notice to the target system, will be necessary, along with retention of prior versions in case any problems are introduced in successive processing.

Local vs cloud implementation of the platform

There are a number of competing factors that make it difficult to predict whether organizations will prefer an implementation of the AMP platform that is local or cloud based. While many collection caretakers tend to favor local storage of their collections, the trend in many IT departments is to moving these functions to the cloud. This is especially true for specialized infrastructure, and of organizations with fewer IT resources. It is also the case that organizations who already have their content stored in the cloud will likely favor a cloud implementation and those who have their content stored locally will likely favor an on-premise implementation. It is likely that that we will need to be able to offer a solution for both.

Service vs internally managed

While great emphasis will be placed on developing a platform that is as easy as possible to deploy and use, the reality is that AMP is complex. Making it available to organizations to adopt, install, and configure on their own will innately restrict the system to organizations with technically robust and sophisticated IT departments. Smaller organizations will likely need AMP offered as a hosted, or pay-to-play, system. It is feasible that both options can be accommodated and as AMP develops both options should remain open.

Maintaining an ecosystem with many dependencies

AMP will be constructed from multiple components and MGMs, which inevitably creates a number of dependencies. For example, updates to components and MGMs by third parties will require updates to AMP. With so many dependencies to manage, it will be important to build in a robust dependency tracking, reporting, and management system to mitigate the risk of downtime, poor user experience, and diminished quality of output.

Personally identifiable information

AMP is expected to be useful to, and used by, holders of collections of all subjects, dates, and quality. Many Content Owners will use AMP on collections that have no risks to personally identifiable information (PII).[17] Some will use AMP with collections that contain PII, and for those users, the semi-automated large-scale nature of AMP increases the responsibility of

---

[17] "PII is information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context." (https://en.wikipedia.org/wiki/Personally_identifiable_information)

Content Owners to thoroughly vet their collections prior to making them available in AMP. Use of MGMs focused on identifying PII may be one solution, but it will also be necessary from AMP administrators to understand the risks and then communicate them to ensure that Content Owners proceed accordingly.

The AMP workshop successfully gathered together a group of experts to talk about what would be needed to perform mass description of audiovisual content utilizing automated mechanisms linked together with human labor in a recursive and reflexive workflow to generate and manage metadata at scale for libraries and archives. The workshop generated technical details regarding the software and computational components needed and ideas for tools to use and workflows to implement to make this platform a reality. The initial concept of AMP has been extended with a solid architecture outline, list of MGM possibilities, and identified requirements. Challenges remain, but we have increased our knowledge about what is needed and our confidence in producing an audiovisual metadata platform to be able to move forward to building a pilot implementation, should resources be available.

Appendix A. Personas

AUDIOVISUAL METADATA PLATFORM PLANNING

Professor persona

Dr. Steven Shida is a professor of film history with a focus on 20th-century African film. He works at a small, private liberal arts college. Because the film collection at his library is relatively small, he is reliant on the collections of other academic institutions to supplement his courses and support his own research. He is an adept researcher, but does not like complex web interfaces and is impatient learning new systems. To make research more efficient, he would like as much information about a film available up front as possible.

NAME	Steven Shida
TITLE	Professor of film history
AFFILIATION	Small liberal arts college

MOTIVATIONS

He has a deep knowledge of his research area and is often looking for particular films and videos he knows. He searches for clips, transcripts, and related materials relevant to undergraduate courses he develops and teaches. He is particularly interested in finding resources by specific titles, filmmakers, genres, and dates.

Film production in his research area is prolific, so he is always looking for new film and video that may be useful to his undergraduate courses, as well as his research. Instead of specific topics, he's interested in browsing by general themes and keywords, as well as country of origin, language, and ethnic group.

He wants to understand the relationship between films and videos within an institution's collection and at other institutions, as well.

"The more comprehensive the information about a film is, the quicker and easier it is for me to find what I need. This means more time I can spend on my research and teaching."

GOALS

Perform Google-type searches

Browse by topics of interest to him

See full descriptions of the films he's interested in

Find what he needs in physically remote institutions

Access the *content*, even when videos are not available to view online

Understand permissions associated with showing films in and out of the classroom

METADATA ACCESS METHOD

Library catalog, list on website, digital repository

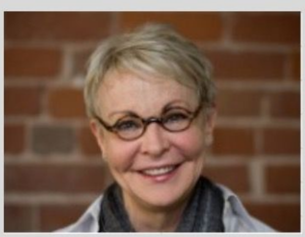
TECHNICAL PROFICIENCY

Low to moderate

Draft August 3, 2017

AUDIOVISUAL METADATA PLATFORM PLANNING

Archivist/collection manager persona



Emily Larsson is the manager of an oral history archive with more than 20 years of experience. The archive has limited funding, so she has to be creative in collections processing. The archive creates collection-level inventories, but she knows in order to make content discoverable and useful to researchers they need more granular data. She assists researchers and often performs searches for them, since she knows the collections well and is aware that searching for specific content is not easy. She is aware of intellectual property issues and cares about sensitive content.

NAME	Emily Larsson
TITLE	Manager, Oral history archive
AFFILIATION	Large nonprofit organization

MOTIVATIONS

Emily wants to be able to manage the collection efficiently in order to optimize resources, so she would welcome technologies that help create more complete records, including metadata about the content (names, descriptions) and context (ownership, acquisition info), and actual transcriptions.

Her efforts focus on making content discoverable, as she knows the research value of the collections. Although she and the staff are extremely familiar with the collections, they can't provide personalized assistance for every user, so she wants them to be as discoverable as possible in their online database.

Emily wants to integrate existing metadata from internal records and inventories with more granular data in a single record in the online database. But while she appreciates the value of automated metadata generation she is hesitant about more complex tasks like subject heading identification being performed by machines.

"We have incredibly rich content and as we process collections we do our best to provide access to our users. The collection continues to grow, though, and it's hard to keep up!"

GOALS

Perform searches by names and keywords to locate records.

Integrate human-created and automated metadata in the records.

Create more complete and granular item level records.

Make complete transcriptions of all oral histories available online

Make valuable hidden content discoverable.

Provide access where copyright and content sensitivities allow.

METADATA ACCESS METHOD

Excel spreadsheets, digital asset management system

TECHNICAL PROFICIENCY

Low to moderate

Draft August 3, 2017

AUDIOVISUAL METADATA PLATFORM PLANNING

Producer for sports broadcaster persona

Danni Jamison lives in Los Angeles, and is responsible for locating video for a variety of uses at the cable network for which she works. She uses Google to start her searches, but ends up having to write emails or call institutions to see if they have clips on a particular player or coach, a specific game, or even a snippet of a game containing a single event or play. She wishes that she could search by these topics and scrub through video herself without having to ask staff to search on her behalf because her production deadlines are often extremely short.

NAME Danni Jamison

TITLE Producer

AFFILIATION Major sports cable network

MOTIVATIONS

The fast-paced environment in which Danni works means she needs more control over how long it takes to search and locate the content she needs. Waiting for people to respond to emails and phone calls and to search collections on her behalf often takes too long, meaning the content doesn't make it to production.

Her work is all about the details. Often, she needs clips of a particular play by a specific athlete within a specific game. Finding footage on college games is hard because so little is available online. She really needs down-to-the-second details to make quick decisions, so watching entire games is out of the question.

Danni needs to understand the copyright restrictions on videos, so she knows if she can use them in her broadcasts. Clear statements that are readily available with the video descriptions is paramount.

"I need down-to-the-second details and I need them ASAP."

GOALS

- Perform Google-type searches
- Jump right to a specific point in a video
- Understand the copyright status of the content she needs
- Quickly locate content based on very specific topics
- Find what she needs in physical remote locations
- Download just the clip she needs

METADATA ACCESS METHOD

Google, digital repository, athletics department website

TECHNICAL PROFICIENCY

Moderate

Draft August 3, 2017

AUDIOVISUAL METADATA PLATFORM PLANNING

Metadata librarian persona

Melinda Farias is a librarian who works in the metadata department in a large academic university library. She wants to be able to capture lots of details about the audiovisual collections (both physical and digital) she is responsible for describing. She is often frustrated because so little information is available about the physical materials, in particular, and it is difficult for her to view or listen to them (due to lack of playback machines and workflow ROI). Full description of digitized av is also extremely time consuming, so she tends to create basic descriptions in MARC or Dublin Core.

NAME Melinda Farias

TITLE Metadata librarian for an access repository

AFFILIATION Large academic institution

MOTIVATIONS

Melinda is highly motivated by quality, but understands the need to balance this with efficiency and cost considerations. For that reason, automation is intriguing to her—she is willing to work with imperfect data and clean it up, if it means that she can provide more access points for her patrons.

Metadata that adheres to standards and taxonomies (e.g., LCSH) is important to her. PBCore, MARC, PREMIS, and rightsstatements.org are the types of standards she follows to create metadata. These are also used by the library's discovery and preservation systems, so when data conforms to them her job is easier.

Melinda sees the value in finding relationships among assets in the collection. She wants metadata to help make those connections easier and more obvious so patrons' find exactly what they need as quickly as possible.

"Metadata is costly and time consuming to produce, so I'm interested in finding ways to provide users with the greatest access to audiovisual material as possible, for the lowest cost and greatest accuracy."

GOALS

- Produce accurate metadata records for audiovisual collections
- Provide the most granular description of audiovisual content as possible
- Easily ingest assets and metadata into library systems
- Make patrons' search experience as efficient and successful as possible
- Enable clear communication of intellectual property and copyright restrictions and use
- Share metadata with data aggregators, partners, or researchers in various formats

METADATA ACCESS METHOD

Structured data formats (XML, json, csv, Excel)

TECHNICAL PROFICIENCY

High

Draft August 3, 2017

AUDIOVISUAL METADATA PLATFORM PLANNING

Music student persona

Josh Lakovic is a classical guitarist, taking courses on music composition, history, and performance towards a BFA. He uses his computer to write papers, browse the web, compose, listen to music, and share info with other musicians. He wants easy access to early guitar recordings but also modern performances (20th C.) because he often compares interpretation styles and trends. Josh took a class as a freshman about how to use the library catalog, but he is not acquainted with alternative search methods such as finding aids and mostly uses Google. He likes to watch performances on YouTube and doesn't really understand intellectual property so mostly ignores it.

NAME	Josh Lakovic
TITLE	Undergraduate music student
AFFILIATION	Conservatory, private liberal arts college

"I need information that will lead me to specific musical performances, artists, and stylistic interpetations, and I want to listen only to the parts that are pertinant to my research."

MOTIVATIONS

Josh takes a holistic approach to studying performances. When watching one, he pays attention to all aspects of it. So, when performing searches, he expects to find detailed information about it (venue, date, interpreters, type of ensemble, arrangements). He carefully watches the techniques of particular performers to improve his own playing.

For his composition courses he likes to listen to the music while watching a score and often jumps from one part of the recording to another.

His research papers are often comparative analyses of interpretations of specific pieces. He references different types of media, from text and images to specific recordings. He uses library materials and resources he finds online. To find works for his research, he searches by style, composer, time period, and performers.

GOALS

- Perform Google-type searches
- Integration of different types of media and platforms.
- Find complete contextual information about content.
- Export sections of media content for reference.
- Non-linear access to content so he can study specific sections.
- Wants access to all content—doesn't care about copyright

METADATA ACCESS METHOD

Google, Music Library database, commercial music vendor

TECHNICAL PROFICIENCY

Moderate

Draft August 3, 2017

AUDIOVISUAL METADATA PLATFORM PLANNING

Sound researcher persona

Dr. Amil Mehta is an Assistant Professor in a large research-1 university in the Radio-Television-Film department. He is an expert in the Civil Rights Movement. In particular, he is interested in how styles of speech can have a rhetorical influence in different contexts.

NAME	Amil Mehta
TITLE	Assistant Professor
AFFILIATION	Large Research Academic Institution

"Metadata should be sufficiently specific and complete to point me directly where I need to go to find the resources to support my research."

MOTIVATIONS

Amil's research topic concerns the recordings of Martin Luther King, Jr., across different audio contexts such as radio and speeches (as in recordings in the American Archive of Public Broadcasting and the Southern Christian Leadership Conference).

Dr. Mehta wants to find particular individuals in recordings and to identify where and when they were produced. He is frustrated by the limited metadata about recordings because he sometimes cannot identify when there is a speaker of interest or what kind of venue might be the context for a speech or radio interview.

Dr. Mehta is motivated by a need to produce scholarship for publication in academic journals. He needs ready access to digital files that he can download along with their correlating metadata.

GOALS

- Perform Google-type searching
- Quickly locate content based on very specific topics and personal names
- Find what he needs in local and physically remote locations
- Understand relationships between content from different collections
- Understand copyright and intellectual property issues

METADATA ACCESS METHOD

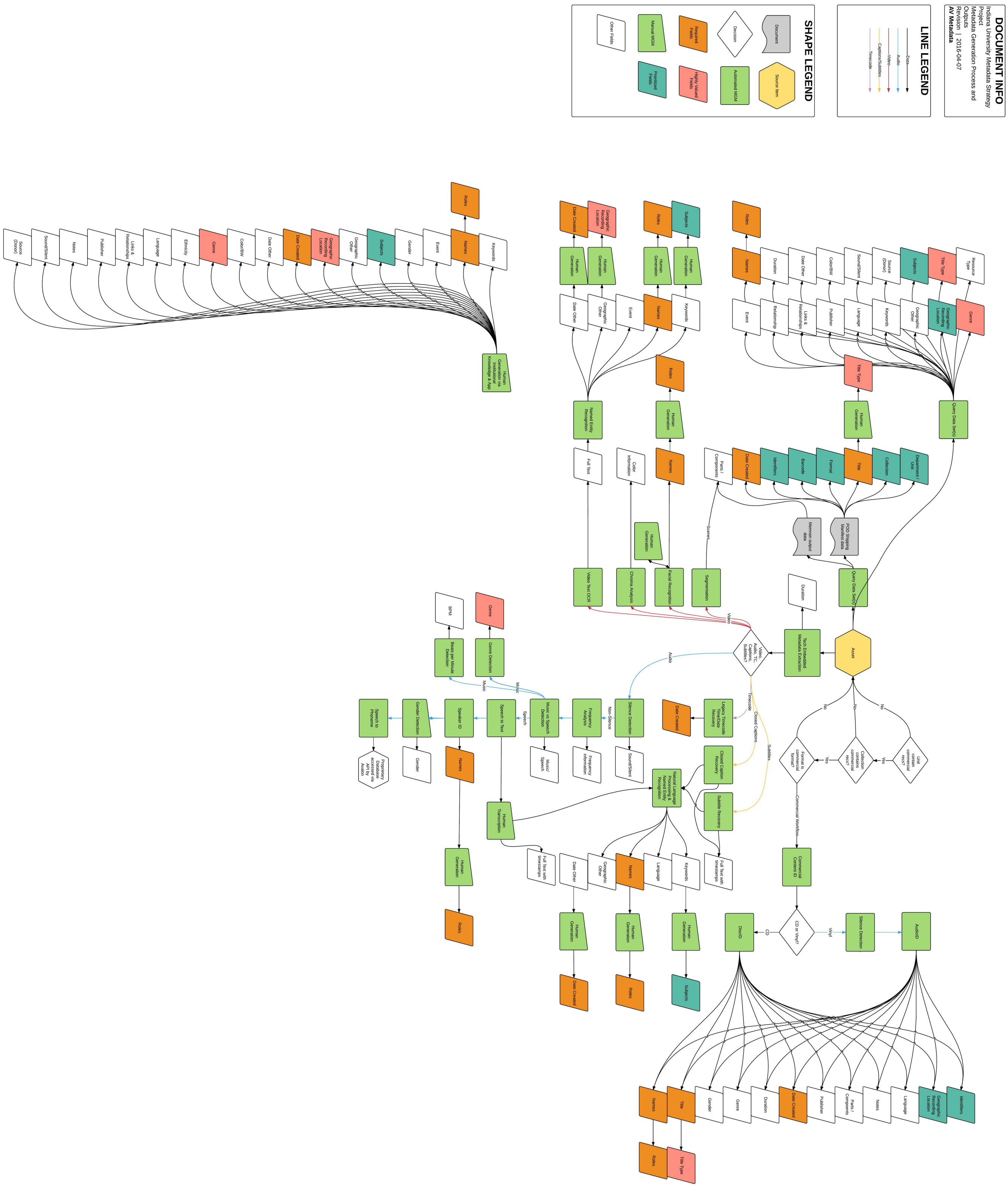
Google search, departmental collections

TECHNICAL PROFICIENCY

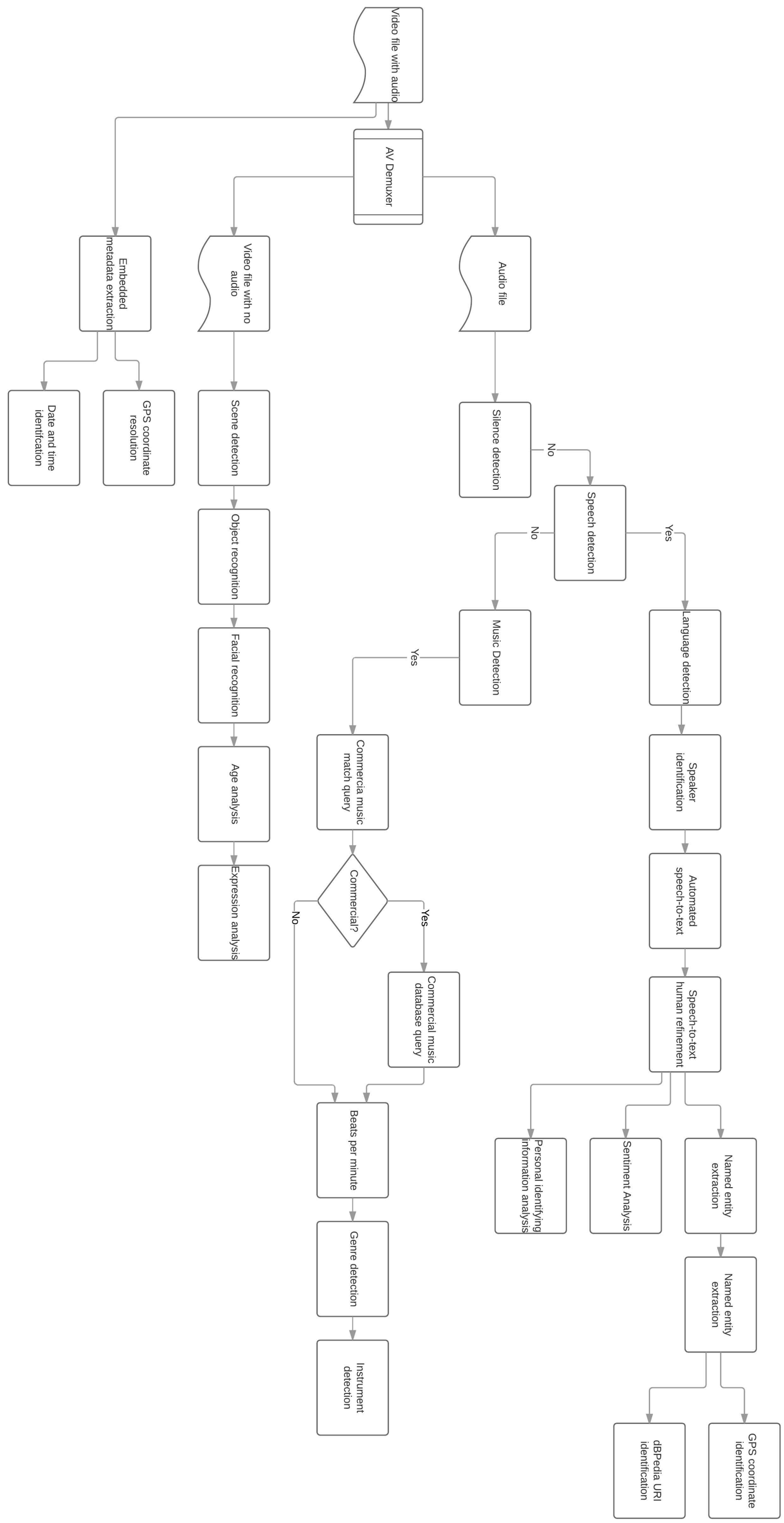
Moderate

Draft August 3, 2017

Appendix B. MGM flow (from prior AVP work with IU)



Appendix C. Example workflow



Appendix D. Functional requirements

ID	Category	Requirement (The system shall. . .)
FR-01	Workflow	Support confidence ratings for the output of MGMs, or flows of MGMs, to allow optimization of human refinement and interpretation for search and browse
FR-02	APIs	Provide API for content and metadata submission
FR-03	APIs	Provide API to retrieve metadata for use in target systems
FR-04	APIs	Support APIs for all UI functions
FR-05	Database	Perform backups of databases on a regular and frequent basis
FR-06	File management	Allow for multiple audio streams per media container when processing audio
FR-07	File management	Pass references to files in storage instead of passing files through the flow whenever possible
FR-08	File management	Allow user configuration of where files get processed—cloud, local, or HPC environment
FR-09	File management	Copy digital media file to MGM for processing OR MGM connects to a central digital media files store for processing
FR-10	Ingest	Support batch ingest of more than one file and/or file format at a time
FR-11	Ingest	Support ingest via multiple inputs including, e.g., DropBox, networked servers, APIs, upload from local computer
FR-12	Metadata	Support a flexible and extensible data model.
FR-13	Metadata	Produce event-based metadata to track provenantial history of workflows, batches, files, and metadata records
FR-14	Metadata	Support creation of unique and persistent IDs for both media and metadata
FR-15	Metadata	Data is transformed to and validated against MGM data model immediately before it is processed by the MGM
FR-16	Metadata	Data is transformed to and validated against AMP data model immediately after it is processed by an MGM
FR-17	MGMs	Support the creation, editing, configuration, registration, and addition of MGMs to AMP
FR-18	Organization	Support different persistent ID systems (e.g., DOI)
FR-19	Queueing	Enable job queueing so a job can run multiple times until it succeeds
FR-20	Queueing	Provide the ability for workflows to be scheduled
FR-21	Queueing	Provide the ability for workflows to be prioritized

FR-22	Queueing	Edit, cancel, delete workflows
FR-23	Reporting	Send an alert (email or another indicator) of any failure during any step in the flow.
FR-24	Reporting	Support configuration of push announcements about workflow activity
FR-25	Reporting	Enable user-defined report parameters for export to CSV, Excel, PDF
FR-26	Reporting	Provide statistics and analysis on processing success and failure, system errors, resource allocation and availability, performance, how long processes are taking, how many processors, backlog status
FR-27	Reporting	Track and generate reports on user access and usage (e.g., frequency of batch flows by a single user within a specific date range)
FR-28	Resources	Able to operate in cloud (e.g., AWS) or local hosting environment
FR-29	Resources	Automatically adjust when resource allocation is not optimized.
FR-30	Scaling	Process multi-TB audio and video files.
FR-31	Search	Support a Google-like metadata search mechanism for before, during, and after metadata is processed by MGMs.
FR-32	Tenancy	Provide the ability for each tenant to create, edit, and assign workflows related to that tenant's assets
FR-33	Tenancy	Provide the ability for multiple tenants to work in parallel without seeing another tenant's workflows or outputs.
FR-34	User Administration	Support standard access authentication mechanisms (e.g., Active Directory or Shibboleth)
FR-35	User Administration	Support user roles
FR-36	User Interface	Support interface to specify which fields to map from AMP to Target System.
FR-37	User Interface	Include or exclude an MGM in the workflow by selecting it from a list in the UI.
FR-38	User Interface	Support a GUI that is accessible by Content Owners.
FR-39	User Interface	Support a UI that allows AMP Administrators to configure AMP for local needs.
FR-40	Versioning	Support versioning of MGMs
FR-41	Workflow	Provide the ability to run workflow jobs in the background
FR-42	Workflow	Provide the ability for MGMs to work in parallel as defined by the workflow manager.
FR-43	Workflow	Embargo a file for review when an MGM process fails more than [x] times. [TBD]
FR-44	Workflow	Support configurable and interactive workflows so that user input at any point can restart a failed process or add or remove

		an MGM
FR-45	Workflow	Write workflow jobs so that they will not be broken by multiple executions.
FR-46	Workflow	For manual MGMs and other processes that last more than [X] period of time (TBD), poll system to see if job is complete (as opposed to running a continuous job).
FR-47	Workflow	Support the workflow manager working with more than one file at a time
FR-48	Workflow	Support the workflow manager working with more than one job at a time
FR-49	Workflow	Enable inclusion or exclusion of MGMs based on user need (with checkboxes, toggles, etc.)

Appendix E. Candidate MGMs

This list of potential metadata generation mechanisms (MGMs) came out of a workshop activity to capture both tool categories and specific tools that may be useful for AMP. The list below represents unmediated results as captured during the activity.

Genre Detection (Audio)	VAMP Suite
	Look for semantic tagging in Types
Legacy Closed Caption Recovery	Internal IU script leveraging FFmpeg
	CCExtractor
Speech to Text (STT)	PopUp Archive
	Speechmatics
	YouTube's transcription
	Google Cloud STT
	Baidu STT
	IBM Watson STT
	MS Azure Video Indexer
	PocketSphinx
	Kaldi
	Spoken data
OCR from paper	Abbyy
	Tesseract
	Google Cloud OCR
Phoneme Analysis	Nexidia
	VU Digital
Legacy Date/Time + Timecode Capture	Line 12/13 extraction script leveraging FFmpeg
Linked Data	MusicBrainz
	DBpedia
	id.loc.gov
	Open Metadata Registry
	Fuseki

	Virtuoso
	Blazegraph
Subtitle Recovery	Video OCR
	Extract from embedded stream when available
Tech Embedded Metadata	FFmpeg
	MediaInfo
	GPAC MP4Box
	ExifTool
	FITS
Queried/Imported Data Sets	Gracenote
	Agave API
	CKAN
	SRU
	IMDB
	OAI PMH
	Z39.50
	SPARQL Endpoints
	MINT
Segmentation	FFmpeg
	Shotdetect
	Fraunhofer AV-Analyzer
	LibROSA
	Google Cloud Video Intelligence
	MS Azure Video Indexer
Silence Detection	FFmpeg + Script
	LibROSA
	VAMP Suite
	Fraunhofer AV-Analyzer
Speaker Identification	VuDigital
	MS Azure Speaker Recognition API
	ARLO
	Google Cloud Speech API
	IBM Watson STT

	PopUp Archive
	Human generation
Content Matching	TinEye
	Shazam
	Muffin
	AcousticBrainz
Facial Recognition	Fraunhofer AV-Analyzer
	RAMP
	MS Azure Video Indexer
	MS Azure Computer Vision
	IBM Bluemix
	Google Cloud Vision API
Audio Frequency Analysis	AcousticBrainz
	LibROSA
	VAMP Suite
	FFmpeg
	SoX
	MediaInfo
Beats per Minute Detection	VAMP Suite
	LibROSA
Chroma Analysis	LibROSA
	QCTools
	FFmpeg ffprobe
Commercial Content Identification	MusicBrainz
	Gracenote
	Shazam
Human Generation	Oxygen XML Editor
	MarcEdit
	ArchivesSpace
	Open Video Annotation
	Amazon Mechanical Turk
	Sonic Visualizer
	Refer.cx
	Zooniverse

Human Transcription	Text Editor
	oTranscribe
	reCaptcha
	Amazon Mechanical Turk
	NY Public Library's Transcript Editor
Music/Speech Detection	Fraunhofer AV-Analyzer
	Frequency range analysis with script
	Semantic tagging (see Other Types)
OCR from Video	ConTEXTract Video OCR
	VuDigital
	Abbyy
	Tesseract
	YouTube
	MS Azure Video Indexer
Natural Language Processing (NLP)	Open Calais
	spaCy
	Natural Language Toolkit (NLTK)
	IBM Watson NLP
	Stanford NLP
	Term frequency–inverse document frequency (Tf/idf)
	Google Cloud Natural Language API
Text+Audio	AENEAS (forced alignment)
	HyperAudio (forced alignment)
	3PlayMedia (forced alignment)
Other Specific	Semantic Annotation Tools (Knight)
	Similarity Hashing (Simhash)
	Topic Clustering (Gensim, spaCy word2vec)
	Sentiment Analysis (Rosette)
	Audio Annotation (Sonic Visualizer)
	Audio Event Detection (SoundNet, Google AudioSet/VGGish)
	Source Identification (e.g. instrumental)

	Caption Format Converter (3Play Media)
	Haven OnDemand
Other Types	Date Estimation
	Geolocation
	Topic Segmentation
	Semantic Tagging
	Waveform Generator
	Language Identification
	Keyframe Extraction
	Aspect Ratio Detection