

Healing the Gap: Digital Humanities Methods for the Virtual Reunification of Split Media and Paper Collections

Stephanie Sapienza <sapienza_at_umd_dot_edu>, Maryland Institute for Technology in the Humanities
Eric Hoyt <ehoyt_at_wisc_dot_edu>, University of Wisconsin-Madison
Matt St. John <mdstjohn_at_wisconsin_dot_edu>, University of Wisconsin-Madison
Ed Summers <edsu_at_umd_dot_edu>, Maryland Institute for Technology in the Humanities
JJ Bersch <jbersch_at_wisc_dot_edu>, University of Wisconsin-Madison

Abstract

This paper introduces and unpacks several challenges faced by stewards who work with audiovisual resources, departing from the premise that audiovisual resources are undervalued and underutilized as primary source materials for scholarship and therefore receive less attention in the sphere of digital humanities. It will then present original research from the Maryland Institute for Technology in the Humanities (MITH), in conjunction with the University of Wisconsin-Madison and the Wisconsin Historical Society, on a project entitled *Unlocking the Airwaves: Revitalizing an Early Public Radio Collection*. As a case study, *Unlocking the Airwaves* successfully meets these challenges by employing strategies such as virtual reunification, linked data, minimal computing, and synced transcripts, to provide integrated access to the collections of the National Association of Educational Broadcasters (NAEB), which are currently split between the University of Maryland (audio files) and the Wisconsin Historical Society (paper collections). The project demonstrates innovative approaches towards increasing the discoverability of audiovisual collections in ways that allow for better contextual description, and offers a flexible framework for connecting audiovisual collections to related archival collections.

Introduction

It is difficult to talk about audiovisual archives as a field of practice, since audiovisual collections are never typical, existing in institutional repositories so varied as to make it impossible to talk about them as a monolithic whole. Audiovisual media collections are found within museums, libraries, historical societies, private collections; within media production units; and within traditional archives (only a small percentage of which are specifically dedicated to audiovisual collections). This wide array of communities and institutions has developed unique principles and standards, each individual community or institution borrowing the professional descriptive approaches needed for their particular circumstance, with no unifying methodology.

1

These splintered approaches have led to a particular set of challenges with a/v collections, which are woefully underdescribed and often segregated from contextual resources. This separation from contextual resources prevents understanding of the conditions that produced a/v materials, often limiting the types of research questions that can be answered by the a/v media alone. An additional challenge is what scholar Sarah Florini has called a "temporal commitment barrier," meaning that users of a/v collections can easily become overwhelmed by the sheer listening hours needed to research time-based collections [Vinson 2019] [Florini 2017]. The unfortunate result of the above three challenges is that a/v resources are undervalued and underutilized as primary source materials for scholarship, which in turn has meant that they also receive less attention in the sphere of digital humanities. The field of digital humanities is poised to innovate and meet these challenges by employing methods such as linked data and virtual reunification to virtually connect audiovisual collections to related contextual resources. DH can also exploit its ability to experiment and push boundaries by encouraging the use of new standards such as WebVTT and new approaches like minimal

2

computing, to enable distant reading and increase searchability via synced transcripts.

This paper will describe original research at the Maryland Institute for Technology in the Humanities (MITH), in conjunction with the University of Wisconsin-Madison and the Wisconsin Historical Society, on a project entitled *Unlocking the Airwaves: Revitalizing an Early Public Radio Collection*. The project aims to provide a model for innovation within digital humanities by virtually reunifying two geographically segregated collections of the National Association of Educational Broadcasters (NAEB), which are currently split between the University of Maryland (audio files) and the Wisconsin Historical Society (paper collections). The NAEB collections provide an in-depth look at the messages being broadcast to the general public through the rubric of 'educational radio,' which predated (and heavily informed) what we now know as public radio, between 1950-1970. By coordinating the expertise of archivists, humanities researchers, and digital humanists, the creation of this new online resource for humanities research will deliver enhanced access to important, mostly hidden, archival audiovisual materials.

3

Research Questions

At the heart of this project lies a question that we have identified as being central to many ongoing conversations surrounding the role of the audiovisual in digital humanities: *how can we utilize digital humanities methods to encourage the use of audiovisual collections as primary historical records, both in and outside of the academy?* To move towards an answer to that question, it was necessary to unpack why A/V collections are underutilized in the first place. We identified three core reasons: A/V collections are a) underdescribed as a whole, trending towards skeletal descriptions at the item level, b) are often segregated from related contextual resources, and c) present an additional challenge in the form of a temporal commitment barrier. The below sections of this paper will walk through each of these three reasons more in depth, before returning to our central research question.

4

Why are audiovisual collections underdescribed?

The term underdescribed is used here as an overarching descriptor for the whole of audiovisual materials across the globe, which as a general rule, do not enjoy a level of descriptive attention anywhere close to their textual counterparts. A/V collections exist within and have borrowed strategies from libraries and museums [Becker 2001]. However, this paper will limit discussions to scenarios when A/V collections exist in an archival context — a situation that can be particularly complex due to the ways that audiovisual materials can challenge archival conceptions of uniqueness and authorship. The provenance of an archival audiovisual collection can fall into one of these categories:

5

1. It is part of a traditional, primary textual or paper-based archival collection;
 - Ex: Mixed paper/audiovisual collections at the Smithsonian Archives of American Art; the Black Film Center/Archive at Indiana University, Bloomington;
2. It has been constructed entirely separately from a textual archive, but has a shared provenance, or a substantial number of shared characteristics or authorities;
 - Ex: Several distinct collections related to the Jam Handy Organization (JHO), a producer of sponsored films, exist at the Detroit Public Library (paper), Stanford Archive of Recorded Sound (audio), the University of Michigan (paper), and the Prelinger Archives (paper and film)
3. It was once part of a whole collection which was subsequently split up;
 - Ex: the National Association of Educational Broadcasters (NAEB) collections at University of Maryland Libraries (audio) and Wisconsin Historical Society (paper).

Some collections do inevitably get described at the collection level as per traditional archival principles, particularly in category 1, where they are often (but not always) folded into the container list of a finding aid as a separate series. But many stewards of audiovisual collections tend to focus on item level description for various reasons. They sometimes do this to maintain bibliographic control, in circumstances like where the collection comprises primarily completed works, or in instances when 'the work' is clear and has available bibliographic information. This approach is employed for the majority of a/v holdings at the UCLA Film & Television Archive, which includes records for their materials in the

6

MARC-based UCLA Library Catalog [Leigh 2006]. More often, however, item-level inventories are created to prepare for a preservation effort, frequently in the form of inventories concentrating on condition assessment. Many of the materials in these collections are rapidly decaying, or their physical format is in a state of obsolescence such that they cannot even be played back without some level of repair or preservation. Thus, for many such collections, there is less description even possible without reformatting or digitization. This is the case with the mass surveys undertaken by Indiana University-Bloomington and the Smithsonian, both of which undertook initial inventories, simple counts that took years just to get to the point of funding a preservation effort, which needed to happen before the inventory records could become part of any collection- or item-level descriptive process [Indiana University Bloomington 2011] [Forsberg 2017]. In these cases, the requirements for preservation often necessitate descriptive choices which cannot always be reconciled with collection-level processing.

Michael Heaney has previously described a typology for collection description, which now exists as the basis for Dublin Core's Collection Description Type (CDType) Vocabulary [Heaney 2000]. Using this vocabulary, the collection level approach most used by traditional archivists would be classified as a unitary finding aid (contains information solely or primarily about the collection as a whole), and the item-level approach would be classified as an analytic finding aid (contains solely information about individual items and their content, much like a library catalog). Andrea Leigh's 2006 piece "Context! Context! Context! Describing Moving Images at the Collection Level" in *The Moving Image* goes into great detail describing how these different approaches apply to A/V collections, using the UCLA Film & TV Archive's various collection categories as a case study [Leigh 2006]. UCLA began shifting its approach to accommodate collection-based description when it began to receive large collections of home movies, commercials, promotional or educational materials which were sometimes (but not always) related to other collections, or were too vast to describe with the amount of bibliographic control given to the core film collections [Leigh 2006].

When thinking about the very salient points that archivists make to defend the use of each of these different approaches, a useful exercise is to review an online discussion in 2013-14 that occurred between Megan McShea of the Smithsonian Archives of American Art and AVP's Josh Ranger. Ranger originally published a blog post entitled 'Does the Creation of EAD Finding Aids Inhibit Archival Activities?' which posited that using Encoded Archival Description (EAD) in the most traditional sense is problematic for A/V collections, since EAD often has issues with discoverability through Internet search engines, and partially because the lack of item-level information does nothing to prepare/plan for preservation efforts [Ranger 2013]. McShea replied to this post with a longer retort on the values of EAD for certain situations, addressing a significant number of Ranger's points as valid, but arguing that in many cases, finding aids become the easiest way for archivists with mixed collections to deal with institutional realities. And that, if proper measures are taken (as with the practices she'd established at the Smithsonian), archivists can effectively combine item-level and collection-level processing [McShea 2013] [McShea 2015]. This approach is in line with category 1 above. Ranger replied that every one of McShea's points was also valid, but then delivered a set of final points about EAD. He posited that "item level processing is really the only way to tell what's what and find the right pieces to preserve or transfer for access, and that EAD does not achieve this level of need. In my view the reliance on EAD has resulted in it becoming an endpoint or cul de sac, not a pivot point" [Ranger 2014].

In many ways, these two outlooks are equally valid, as both integrate concerns over utilizing More Product, Less Process (MPLP) techniques to maximize descriptive efficiency, and also reflect the nature of each person's position and institutional mandate. Ranger frequently worked on consulting projects with archives who often needed a complete item-level inventory to prepare for a preservation project with as much metadata as possible, whereas McShea works at a research institution with a substantial amount of both textual and audiovisual collections, whose goal is to make materials available for research. In either case, institutional needs determine the level of description, both of which lead to collections being 'underdescribed' in different ways. With an item-level approach, descriptive data about the collection can remain inaccessible to researchers for longer; whereas with a collection-level approach, the audiovisual resources can either become subsumed by their textual counterparts, or segregated from them entirely.

Why are audiovisual collections segregated from related contextual resources?

Over the past two decades, the archival field has taken great strides in the area of capturing contextual associations as

part of the description and arrangement continuum. Heany's typology of a "unitary finding aid" was written in 2000 [Heany 2000]. Joseph Deodato wrote an evocative piece in 2006 calling for archivists to employ postmodern strategies in description and arrangement as part of a call for archivists to be 'responsible mediators' of history, recalling and building upon the earlier work of Terry Cook, Wendy Duff, and Verne Harris in evoking the role of the archivist in constructing meaning [Deodato 2006] [Cook 2001] [Duff and Harris 2006]. What Deodato called a "'creation continuum" is "but one aspect of a complex provenance that also includes the context in which the records were created, the functions they were intended to document, and the record-keeping systems used to maintain and provide access to them" [Deodato 2006]. Since that time, standards have emerged such as Describing Archives, a Content Standard (DACS), which is "consciously designed to be used in conjunction with other content standards to meet local institutional needs for describing collection materials," [DACS 2019] and includes a newly revised Part II section devoted to the creation of context through archival authority records, or records about the people, corporate bodies, and families associated with archival collections [DACS 2019]. The emerging standard Encoded Archival Context — Corporate bodies, Persons, and Families (EAC-CPF) was designed to standardize the encoding of such archival authorities [Leigh 2006].

What Heaney invoked in his discussion of a unitary finding aid was similar to other movements in the field to address ideas of context and distributed collections, going as far back as documentation strategy in the late 1980s [Samuels 1986], Carole Palmer's notion of 'Thematic Research Collections' in 2004 [Palmer 2004], and virtual reunification, which has appeared in the literature as early as 2004 and evolved into its current form over the past five years or so due to the work of Ricardo Punzalan. As one of the preeminent scholars continuing to define and contextualize the forms and trajectory of distributed collections and the 'Archival Diaspora' that results from dispersed materials, Punzalan has continuously championed and focused on virtual reunification as an emerging and flexible strategy which exhibits some of the more useful qualities of several other methodologies and theories. In both his 2013 dissertation [Punzalan 2013] and subsequent pieces in *Library Quarterly* and *American Archivist* [Punzalan 2014a] [Punzalan 2014b], he tracks and analyzes different projects which have employed the strategy, while drawing parallels and distinctions between it and other approaches. At the heart of his analysis is an acknowledgement that all virtual reunification projects require an 'unprecedented' amount of inter-institutional collaboration a broad swath of technical expertise, and doing more than simply reuniting geographically separated collections, but producing an entirely different digital entity.

11

Acknowledging that A/V collections are also part of the 'Archival Diaspora', it follows that they are often constructed in an entirely different sphere from the paper materials of their various creators (producers, artists, studios, etc.). Compounding this are the issues described in the above section, where long-standing descriptive standards for processing paper collections simply diverge from many of the needs of an audiovisual collection. The facts surrounding this phenomenon can widely vary, but they lead to a number of different scenarios where you end up with A/V collections which have little to no contextual information available to their stewards, although that context exists elsewhere in the information sphere. With the exception of certain categories of orphaned works (home movies, educational, sponsored, experimental, and amateur films), most of the time there is some extant documentation of a media object's creation and dissemination — transcripts, production and field recording notes, press kits, photos, correspondence, provenance and copyright materials. That documentation can exist in different divisions of the same institution, either at the same or a different geographic location, or at an entirely different institution altogether.

12

Within that context, archival collections are often "mixed," meaning that at some point in their chain of custody (at or before the point of accession) they include both media materials as well as the related textual or photographic materials that are part of a shared provenance. It is a very common scenario when these mixed collections are accepted into archival repositories, accessioned, and then broken apart and processed using very different and separate techniques, guidelines, and description schemas. The Academy of Motion Picture Arts & Sciences (AMPAS) contains several divisions comprising one of the world's largest repositories devoted to film history, including the Academy Film Archive, the Margaret Herrick Library, the Academy Oral History Projects, and the Science & Technology Council. AMPAS has been receiving donations and gifts in the form of these "mixed" collections for years, from producers, artists, private collectors, and studios. Depending on the division receiving the accession, it was standard practice for years to split them, a practice that they've actively been shifting away from over the past 5-7 years by integrating an enterprise

13

collections management software product called Adlib, which grants them the capability to unite and ontologically link film and paper collections sharing provenance or shared authorities across different divisions. These shifts, and the earlier pivots taken by UCLA, demonstrate that these two repositories, as two of the world's largest collections of our audiovisual history, have had to adapt to changing realities in the field and within their institutions. Both have the resources for some form of solution when they receive and handle 'mixed' collections comprising both A/V and paper resources. It also stands to reason that the types of collections that each of these repositories receive most often (with notable exceptions) are more cleanly identifiable by a shared provenance, shared authorities, and some level of bibliographic control.

But quite often, either in situations where the institution doesn't have the kind of resources needed to direct attention to this kind of work, or when the collections fall under the category of so-called 'orphaned' works, the two collections never get near each other again – physically or ontologically. It is these circumstances which lead to culturally rich collections being siloed, underdescribed, under-contextualized, and thus largely ignored by the larger sphere of possible users. For the category of orphaned works, the value of the materials for researchers frequently remains unrealized, as the archival practices outlined above tend to separate them from the contextual associations that clarify their historical and social importance. Scholars such as Devin Orgeron, Marsha Gordon, and Dan Streible have addressed this problem, noting in *Learning with the Lights Off: Education Film in the United States* that nontheatrical films have experienced limited availability in archives, which has in turn caused a lost sense of their historical significance [Orgeron et al. 2012]. In recent years, sustained archival and scholarly interest has led to a number of projects attempting to reinstate this historical value by providing context for orphaned media objects. For example, the Canadian Educational, Sponsored, and Industrial Film (CESIF) project, led by Charles Acland and Louis Pelletier, in part contextualizes orphaned films through a set of circulating institutions, showing the range of nontheatrical materials that were produced in Canada [Acland 2016, 138]. Other projects have invited user contributions in these efforts. Mark Williams's Media Ecology Project at Dartmouth College allows users to add information, including metadata, to create contextual connections across archival collections, intended to "facilitate a dynamic context of research that develops in relation to its use over time by a wide range of users" [Williams 2016, 336]. This approach has been applied to orphaned collections in the Media Ecology Project's Historical News Media Study.

14

For split collections, contextual information that can highlight the significance of the audiovisual materials already exists, so the obstacle becomes connecting the media objects and their contextual materials. This challenge provides an opportunity for enterprising archivists who are able to approach a more postmodern approach to description, or for digital humanists who have the ability to help archivists innovate in these areas, to think about virtually unifying the two collections. The collections of the National Association of Educational Broadcasters (NAEB), which are at the heart of the case study below, are a prime example of this lost potential for discoverability.

15

What is a 'temporal commitment barrier,' and how does it pose challenges for the use of audiovisual collections as primary resources?

In a 2017 talk at MITH, scholar Sarah Florini discussed how black social media fandom employed strategies to mitigate the affordances of particular platforms, including podcasts, invoking the term "temporal commitment barrier." [Florini 2017] Lifting that out and employing it here, it becomes a useful term to invoke the notion of how users of A/V collections can easily become overwhelmed by the sheer listening hours needed to research time-based collections. Obviously, this barrier increases in size exponentially with a large corpus of materials and many thousands of hours of content. A solution to overcome this barrier, the use of time-stamped, synchronized transcripts, has been increasingly adopted by the audiovisual archiving field. An early adopter was the WGBH Media Library and Archives, who employed it for their project *Vietnam: A Television History* as early as 2006. Use of this approach has been fueled in part by access to automated speech-to-text tools, which has lowered the barrier of entry and made it so that human labor was not needed to transcribe entire time-based works. Oral histories are most often embracing the technology, as evidenced by the popularity of the open source software OHMS (Oral History Metadata Synchronizer) out of the University of Kentucky [Breaden et al. 2016]. OHMS employs both transcripts and indexing functionality to gain access to subparts of a time-based asset. The consulting company AVP (formerly AVPreserve) integrates OHMS with its popular platform Aviary, which AVP licenses as a robust front-end interface. Aviary can be integrated with a variety of platforms, providing

16

a functional, usable display next to descriptive context. The AMPAS Oral History Projects department uses OHMS to index and display synced transcripts for its collection of filmmaker interviews.

Scholar Tanya Clement has discussed the use of transcripts as a means to link a single audio or video event together with transcripts as a sort of scholarly primitive, invoking John Unsworth's use of the term [Clement 2015], while also acknowledging the limitations and limited vision of such approaches, stating "It is also our inability to conceive of and to express what we want to do with sound — what Jerome McGann (2001) calls 'imagining what you don't know' — that precludes us from leveraging existing computational resources and profoundly inhibits DH technical and theoretical development in sound studies" [Clement 2015]. The use of transcripts to 'read the text' of time-based media objects may indeed be a form of scholarly primitive, but it's still only just gaining traction in the field, with many adopters still struggling to implement synced transcripts in a usable fashion. It was only this year when the OHMS software was even able to introduce the ability to import pre-existing timestamps for transcripts — previously the tool required its users to manually create the sync points through a very clunky process of hearing beeps and marking the text you hear. The Pop Up Archive, a nonprofit which for several years received copious financial support from various government and private funders, provided speech to text services as well as integrations and features to correct, reuse, and embed those transcripts in other applications. The service fulfilled a need which was acknowledged and lauded by the funding community, before it closed down services in 2017 after being bought out by Apple. Since then, former users of the service have splintered off and utilized different methodologies to achieve this same need. Some of them go towards open source solutions, some develop their own approaches. One of the transcript file formats that Pop Up Archive provided included WebVTT. Since 2012 WebVTT has been supported in all major Web browsers. The standardization of WebVTT was not without controversy, due to its decision to forgo the use of the XML based transcription standard TTML. However, browser support for WebVTT and the Web Audio API [Adenot and Toy 2018] make it a logical choice for audio transcription projects because it greatly simplifies the synchronization of audio and transcripts in the browser.

17

Increasing the use of audiovisual collections for scholarly study

Scholars have addressed the enormous challenges of arrangement and description for the vast amount of audiovisual material now produced every day, such as Virginia Kuhn's discussion of this problem in her argument for a mixed methods approach to analyzing filmic media [Kuhn 2018, 301]. But the issues outlined above have contributed to a situation in which archival A/V collections similarly face problems of underdescription and temporal barriers, leading them to become neglected materials despite their availability to researchers. Many collections lack the detailed metadata and clear contextual explanation that would make them more accessible, so a/v materials often remain underexplored by scholars and educators. Digital humanists have an opportunity to promote the use of a/v materials by restoring the relationships that were severed by split collections, highlighting their value as primary historical records alongside contextual paper materials. As one example of a project with this goal, *Unlocking the Airwaves* is presented below. The section starts off with some background into both the NAEB collections and the project itself, and then lays out the DH techniques that were deployed through a description of the application's design, object model, descriptive standards, and workflows.

18

Case Study: The Unlocking the Airwaves project

Unlocking the Airwaves was in development and fundraising between 2013-2018 prior to being funded by the National Endowment for the Humanities in April 2018. As of June 2020, the project team has launched a beta version of the website, and has made substantial progress towards a final launch in Spring 2021. Although we still have five months to complete the final stages of the project, including the publication of exhibits and teaching tools, we have leveraged a range of digital strategies to bring together the NAEB's split paper and audio collections. Throughout this project, we have successfully combined existing tools and frameworks, such as a minimal computing application design, linked data, and synced transcripts, to make the collections accessible to users while minimizing cost and long-term maintenance. Although every split collection poses different problems for potential reunification efforts, *Unlocking the Airwaves* provides one method for digital humanists to reunite collections, creating new access points to A/V materials and positioning them within their historical context.^[1]

19

The NAEB Collections

Before the ubiquity of National Public Radio (NPR) and the Public Broadcasting Service (PBS), the National Association of Educational Broadcasters (NAEB) was the primary institution responsible for promoting and distributing public broadcasting content in the United States. The NAEB was initially established as the Association of College and University Broadcasting Stations in 1925. Member stations attempted to share programming resources in various ways until the organization created a distribution network in 1949, which was run from the NAEB's first national headquarters at the University of Illinois in Champaign-Urbana beginning in 1951. The organization grew steadily during the 1950s and 1960s, achieving a key moment in its history when the NAEB's director of radio distribution Jerrold Sandler successfully lobbied to insert language supporting public and educational radio into the Public Broadcasting Act of 1967. While this led to the creation of NPR, it also caused the demise of the NAEB in 1981 and preceded the separation of its historical records. Thousands of audio recordings were transferred to NPR in Washington DC and later added to the University of Maryland Libraries' National Public Broadcasting Archives, whereas the paper records were mostly archived in the Wisconsin State Historical Society (now Wisconsin Historical Society, WHS), deposited by longtime NAEB President William G. Harley. This separation of the NAEB's audio and paper collections complicates a full understanding of broadcasting's history in the United States.

20

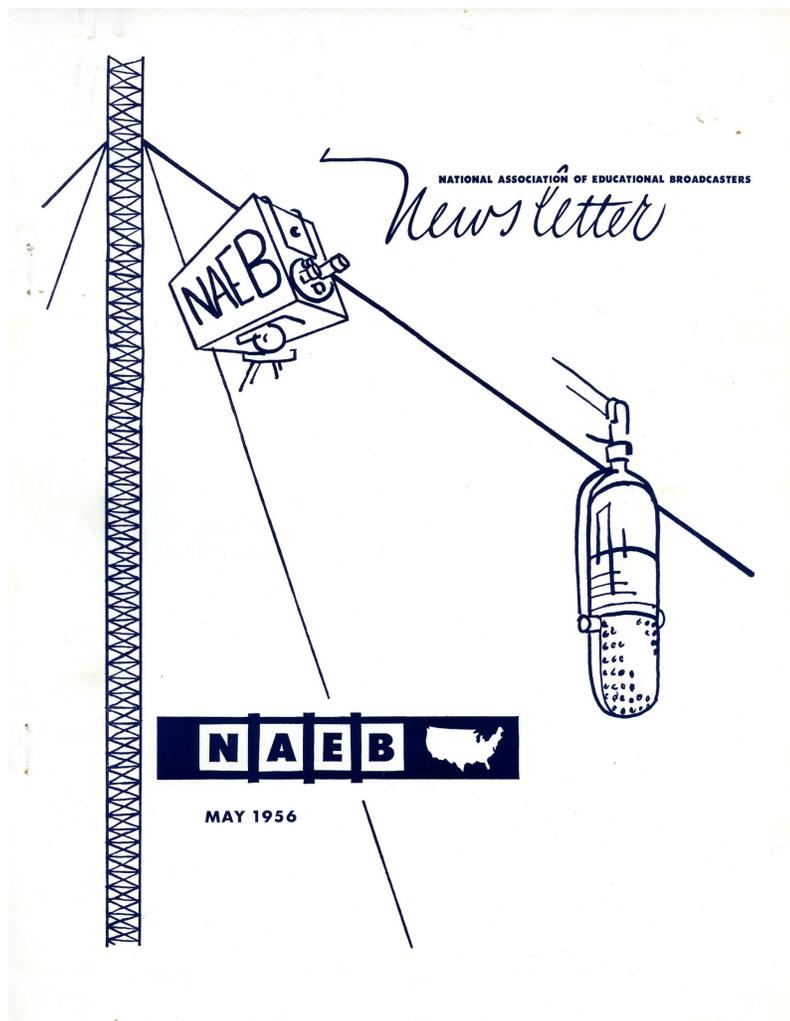


Figure 1. NAEB Newsletter, archived in the collections of the Wisconsin Historical Society.

Unfortunately, although researchers at the WHS could see these aspects of the NAEB, they could not *hear* them. The split between the paper and audio materials in the collection prevented the discovery of connections between the everyday practices of the NAEB and the media it presented to the public. Considering the two sides of the archival collection together can contextualize the philosophy, objectives, and practices of the organization that broadcast the

21

radio programs. In the daily correspondence of the NAEB's staff and key members and publications like the *NAEB Newsletter* (see Figure 1), the paper material covers everything from aesthetic norms to political aims, and illustrates how the values and goals of the organization influenced the types of programs it broadcasted and the eventual development of the public radio that we know today. The importance of the NAEB cannot be fully understood without restoring the connections between the audio and the paper materials, and the NAEB collection is just one of many collections that could benefit from this reunification of split materials.

The availability of A/V materials, like the NAEB radio programs, can allow scholars and educators to answer significant research questions that might not be possible with only a collection's paper materials. In the case of the NAEB, the content of thousands of episodes reveals distinct perspectives on key social and political topics from the period. For example, *People Under Communism* focuses on various aspects of life in Soviet Russia, including music and literature, and *Seeds of Discontent* examines the disadvantages facing various "discontented forces" in the United States through interviews with groups ranging from black artists in the entertainment industry to public school teachers. Other programs like *American Adventure* and *The Jeffersonian Heritage* seek both to educate and entertain with dramatized figures and moments from American history. Even when the paper materials might include summaries or outlines of a program's individual episodes, accessing the audio itself affords the study of aesthetic questions beyond content alone, like music, performance, audio effects, and timing. The topics generated by the audio programs can fruitfully intersect with the paper materials, like the correspondence and reports from the NAEB that reveal the program committee's aesthetic priorities and tastes. The committee's discussions cover everything from the quality of acting to the adequacy of episode introductions in program submissions, providing context for the types of radio that the NAEB preferred, as well as the responses to programs that did not fit the members' interests.

22

Background & Project Goals

At the onset of the project, the radio programs were already digitized and transcribed through the University of Maryland Libraries' prior collaborations with the American Archive of Public Broadcasting (AAPB) and Pop Up Archive. The paper collections had been described in an online finding aid, but not yet digitized. As of June 2020, we've achieved four of the project's five major milestones and are quickly approaching the completion of the fifth:

23

1. Digitize an identified subsection of the NAEB paper collection in-house at the Wisconsin Historical Society;
2. Create new metadata about the digitized material, including a set of archival authority records about early educational and public broadcasting;
3. Design the backend structure of the application utilizing a combination of select linked data and minimal computing methodologies;
4. Design and test a user interface informed by gathered user stories;
5. Integrate exhibits and teaching guides in the application, as well as curated access points both in and outside of the application.

System Design and Object Model

Unlocking the Airwaves employs a minimal computing application design, which combines static site generation, client-side web framework and indexing technologies to minimize server side dependencies, and greatly reduce the cost of deployment and long-term maintenance. As the Internet has become more central to the practice of humanities research and cultural heritage, the tools to publish websites have become increasingly complex. This is especially true for websites meant to serve as digital collections or archives. While many libraries, archives, and museums now recognize the value of having their collections accessible online — and while many humanities scholars build personal or thematic digital research collections as part of their scholarship — the tools and infrastructure to achieve this can strain the capabilities and resources of even privileged institutions. One of the primary goals of minimal computing is to reduce the dependence on costly web and Internet service infrastructures.

24

"Static site generators" are pieces of software for generating static content that can then be copied to and served up by web servers. This web content is called "static" because of its representation as simple files on disk (HTML, CSS,

25

JavaScript, images, video) that do not require content management software (CMS) to access, and can simply be viewed as-is in a web browser. Static websites are useful for sustainability since they require very little in terms of maintenance and monitoring. But this sustainability is achieved by pushing some of the complexity of a dynamic website into the static website's build process. Fortunately, this build process happens just once when the site is deployed, rather than every time a resource is fetched by a user. Although tools such as Columbia University's Wax and University of Idaho's Collection Builder have used static site generators to provide a generalized framework for building digital exhibitions and collections in an easily deployable, portable and low-maintenance web application, none of them have attempted to do so while simultaneously reunifying distributed collections. Combining minimal computing and static site generation with the aforementioned methodologies of virtual reunification, linked data, and synced transcripts allows us to examine facets of digital collection building and maintenance that have been explored with "standard" online platforms and software, but which have not been investigated in these more complex frameworks.

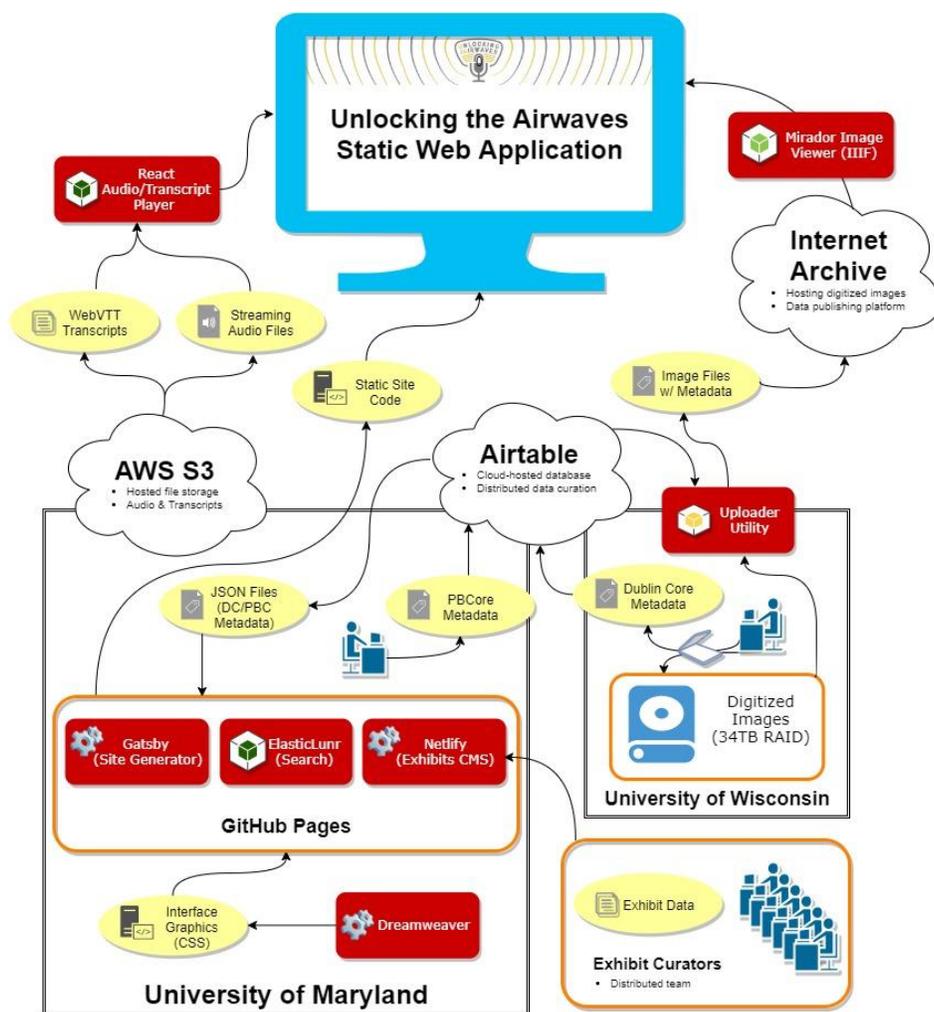


Figure 2. *Unlocking the Airwaves* System and Process Flow Diagram.

The above system diagram displays all the various components of the *Unlocking the Airwaves* application (Figure 2). The application website is being generated as a static site using Gatsby which is a static website generator written in the NodeJS development environment. Even though the website is deployed as a set of static files to MITH's webserver, it relies on several other services during its *build* and *runtime*. The *build* is a process that happens once when the site is deployed, and the *runtime* is the process run by a browser when a user accesses the site. These services include:

- GitHub: platform where the source code for the website is stored and versioned (*build*);
- Airtable: a cloud-based database containing audio and document metadata (*build*);
- Internet Archive: hosts the digitized document scans which are then made available through their IIIF

Service [<https://iiif.archivelab.org/iiif/documentation>] (*build + runtime*);

- Amazon S3: storage of MP3 audio files and WebVTT audio transcripts (*runtime*);
- Netlify: content publishing service for exhibit data (*build*);
- ElasticLunr: client-side search powers navigation and discoverability of the site's content. (*runtime*)

The audio files are redundantly stored on the servers of the American Archive of Public Broadcasting (AAPB) and on hard drives at UMD Libraries, with streaming copies hosted in an Amazon Web Services S3 bucket. Machine-generated audio transcripts are currently being stored as WebVTT files, also in an S3 bucket, which feeds into the application and merges the audio and transcripts for display in a WebVTT player designed by MITH Developer Ed Summers. This display creates one way to work around the 'temporal commitment barrier' that often accompanies audiovisual collections, as users will be able to search for terms that appear in the episode transcripts or scroll through and navigate to certain parts of episodes based on their interests, without having to listen to them in their entirety.

27

The digitized images of the paper collections are being redundantly stored on a 34TB RAID at the University of Wisconsin-Madison's CommArts Instructional Media Center, as well as on Internet Archive (IA) servers. Co-PI Eric Hoyt has been utilizing the Internet Archive for materials he and his team have produced for the Media History Digital Library, which creates automatic access derivatives including OCR files. Since the Internet Archive offers integration with the International Image Interoperability Framework (IIIF), the project team made the decision to use the IA as part of our digitization workflow, not only for redundant storage/access points and creating derivatives for all scanned materials, but also so we could utilize the Mirador image viewer. Mirador supports the IIIF standard, and the new version is written in the JavaScript library React, which dovetails nicely with Gatsby. This approach has saved us significant time and effort, as we can capitalize on linked data technology to display the documents, as opposed to hosting them on our own servers.

28

To track and consolidate metadata for the paper and audio materials, we opted to utilize the cloud-based relational database software tool Airtable for distributed data curation. This Airtable database has the added benefit of functioning as a project management tool for the digitization process, and as a means for publishing the digitized materials to the Internet Archive. Summers also developed a custom uploader program that utilizes Airtable's Application Programming Interface (API) to automatically take a folder of individual digitized images from a given folder, bundle them into one package, upload them to the NAEB collection on the Internet Archive, and map the metadata from Airtable into the Internet Archive's metadata schema. More about our choices for descriptive workflows and standards is included in the section below.

29

The exhibits and teaching tools are created by a distributed team of curators, who enter text, links, images, and captions into a graphical user interface using Netlify CMS. Netlify watches the airwaves GitHub repository, automatically builds a distinct staging site used for development and testing, then generates the exhibit pages in the application. As a plugin, Netlify is also responsive to our content model, which means that curators are able to directly attach links to people, organizations, or programs directly from its user interface.

30

The CSS files comprising the front end design graphics are being generated by MITH Designer Kirsten Keister using Dreamweaver, then published to the Airwaves application in GitHub. Tweaks and modifications to text on static pages can be modified directly in GitHub, but design changes must be separately committed by Keister. Select screen captures from the beta version of the application website can be viewed below (see Figure 2 Figure 3 Figure 4 Figure 5) and also through a screencast with an example of a recording [<https://vimeo.com/437614686>].

31



HOME ABOUT EXPLORE THE ARCHIVE EXHIBITS TEACHING TOOLS SEARCH



Unlocking the Airwaves

Unlocking the Airwaves: Revitalizing an Early Public and Educational Radio Collection is a comprehensive online collection of early educational public radio content from the National Association of Educational Broadcasters (NAEB). The forerunner of CPB and its arms, NPR and PBS, the NAEB developed and distributed educational radio programs and accompanying print materials to schools and communities across the United States. What's more, the NAEB lobbied extensively to unlock the airwaves—to access precious frequency space—in order to bring the voices of poet Robert Frost, architect Frank Lloyd Wright, anthropologist Margaret Mead, and conservationist "Ranger Mac," among many other individuals, into American homes and classrooms.

The NAEB's history is the dramatic story of idealists who believed in the utopian possibilities of technology for education and social uplift and who faced considerable challenges in pursuit of those goals, including economic depression, world war, and the scarcity of the electromagnetic spectrum. It's a story that has much to tell us about 20th century American culture, as well as the 21st century's environment of online educational technology and podcasting that we live in today.

Despite its historic importance and contemporary relevance, most of the NAEB members' programs were never heard again after their initial brief moments on the air. The archives for the radio programs and their related paper documentation have been split for over 25 years between two institutions: the University of Maryland and the Wisconsin Historical Society. Unlocking the Airwaves reunites the split collections, finally realizing the potential of the collections of the NAEB for exploration and the broader public.

[READ MORE ABOUT THE PROJECT](#)

[EXPLORE THE ARCHIVE](#)

Featured Exhibits



RADIO DURING THE COLD WAR



AMERICAN JAZZ IN THE 60S

[MORE EXHIBITS](#)



Unlocking the Airwaves has been made possible in part by a major grant from the National Endowment for the Humanities' Humanities Collections and Reference Resources (HCRR) program.

Figure 3. Unlocking the Airwaves Home Page.



Mississippi manuscript

Series	Document: Deep South
Air Date	1954-01-01
Duration	00:29:17
Description	The Importance of waterways. Including the Mississippi River. Colonel A.F. Clark, Jr., U.S. Army Corps of Engineers in Memphis, analyzes the river's Jekyll-and-Hyde behavior.
Subject(s)	Southern States--Economic conditions Southern States--Commerce Mississippi River Radio programs--United States Water resources development Rivers
Creator(s)	Bannerman, Leroy (Producer) University of Alabama (Producer) Bannerman, Leroy (Writer)
Contributors	Whitaker, Walter (Narrator) Jones, Walter B. (Walter Bryan), 1895-1977 (Advisor) Clark, A.F., Jr. (Speaker) Fund for Adult Education (U.S.) (Funder)
Genre(s)	Documentary News
Geographic Region(s)	regions
Time Period	1951-1960

▶ 0:00 / 29:23 🔊 ⋮

[00:05 - 00:07] Document Deep South

[00:07 - 00:16] Pole fields of cotton forests

[00:16 - 00:19] and factories in the heart of Dixie.

[00:19 - 00:24] A revealing story of progress documented with

[00:24 - 00:29] on the spot recordings and produced by the

[00:29 - 00:33] radio broadcasting service extension

[00:33 - 00:39] division of the University of Alabama. For the

[00:39 - 00:43] next 30 minutes

[00:43 - 00:47] you will make a transcribed trip through the

[00:47 - 00:52] side. You will see the significance of

[00:52 - 01:05] a new industrial solve a changing South.

[01:05 - 01:11] You will see how determined people are

[01:11 - 01:17] nature's plenty into prosperity.

[01:17 - 01:19] That more than ever is making itself felt in our

[01:19 - 01:32] nation's economy.

[01:32 - 01:37] A in A.

[01:37 - 01:42] List a sippy manuscript.

[01:42 - 01:54] A witness.

[01:54 - 01:59] in a.

[01:59 - 02:12] in a.

[02:12 - 02:17] Here on the Mississippi River on the deck of an

[02:17 - 02:19] old times derved we look.

[02:19 - 02:42] At the proud negro deckhand also has a

[02:42 - 02:54] weighted line into the water and you listen to

[02:54 - 01:59] his car.

[01:59 - 02:12] It's an important song. The words and the way

[02:12 - 02:17] he sings them tell the riverboat captain

[02:17 - 02:19] whether or not he's in safe water.

[02:19 - 02:42] This Genting the depth soundings is quite an

[02:42 - 02:54] art. Joe Gold one heard here is one of the

[02:54 - 01:59] few remaining deckhands who can do it.

[01:59 - 02:12] And then at the critical stage nine feet or less

Figure 4. *Unlocking the Airwaves* program landing page, with embedded transcript viewer/audio player. PBCore metadata is displayed on the left. Clicking on individual subjects, genres, or people leads users to a canned search on the Search page.



air pollution]

REFINE RESULTS 3027

[SORTING STUFF HERE]

DOCUMENT
Utilizing Instructional Television: Teacher's Manual For Demonstration Kits (NAEB, 1964) p. 169
Kit Number 6 ADDITIONAL INFORMATION ABOUT THE LESSON EXCERPTS FIRST EXCERPT Series PLACES IN THE NEWS is a series of weekly telecasts, * each approximately 20 minutes in length, produced by WNBT-TV in New York, with Mr. Jerry Si Iverstein as commentator. The series is made available through the Great Plains Instructional Television Library, in Lincoln, Nebraska. Mr. Si Iverstein presents background material for understanding crucial events and issues of national or international concern. Each tel

MEDIA
Pollution in your backyard
Hold your breath
This program seeks to personalize the impact of air pollution by focusing on a woman who has been fighting air pollution in Chicago for twenty years.

MEDIA
Agricultural problems caused by air pollution
Hold your breath
This program focuses on the financial costs of air pollution, particularly as it relates to agriculture.

MEDIA
Air pollution and cancer
Hold your breath
This program focuses on air pollution role in causing cancer.

MEDIA
Air pollution control in the community, part one
Hold your breath
This program focuses on air pollution control from a community perspective.

MEDIA
Air pollution and the industries, part two
Hold your breath
This program continues to focus on how industries deal with air pollution.

MEDIA
The private citizen in the fight for clean air
Hold your breath
This program focuses on the impact that private citizens can have when they battle air pollution.

MEDIA
Air pollution control in the community, part two
Hold your breath
This program focuses on air pollution control from a community perspective.

MEDIA
Air pollution and the industries, part three
Hold your breath
This program focuses on how coal and chemical industries deal with air pollution.

MEDIA
Day to day health effects

FILTER BY TYPE
Document 2998
Episode 29

FILTER BY SUBJECT
Programming 248
National Association of Educational ... 193
Clear Channel Hearings 169
Educational television 163
United States. Federal Communicati... 121
Television 100
NAEB Board of Directors 92
NAEB staff 85
Institute for Education by Radio 83
Public broadcasting--Law and legisla... 73

FILTER BY CREATOR
National Association of Educational ... 2998
WKAR (Radio/television station : Eas... 13
Michigan State University 13
Ford, Patrick 13
Westinghouse Broadcasting Company 10
Group W Productions 10
Daly, John Charles, 1914-1991 10
Wiser, Norman 2
KPCC-FM (Radio station : Pasadena, ... 2
California Institute of Technology 2

FILTER BY CONTRIBUTOR
Schooley, Frank E. 1044
Skornia, Harry J. 964
Hill, Harold E. 741
Harley, William G., 1911-1998 686
Paulu, Burton 598

Figure 5. Unlocking the Airwaves Search page, displaying a canned search for subject=Air Pollution. Facets can then be used to filter results.

Descriptive Standards and Workflows

The challenge inherent with possible data models for this application is that we want to enhance discoverability between these two major collections, while retaining the semantic and descriptive properties of each. Utilizing virtual reunification and linked data, our approach allows us to meet the challenges presented by the Archival Diaspora, while respecting institutional mandates and priorities. Below (see Figure 6) is the object model, which recalls key elements and components of the project data such as radio series, radio episodes, archival series, archival folders, and authorities (Person, Organization, Topic).

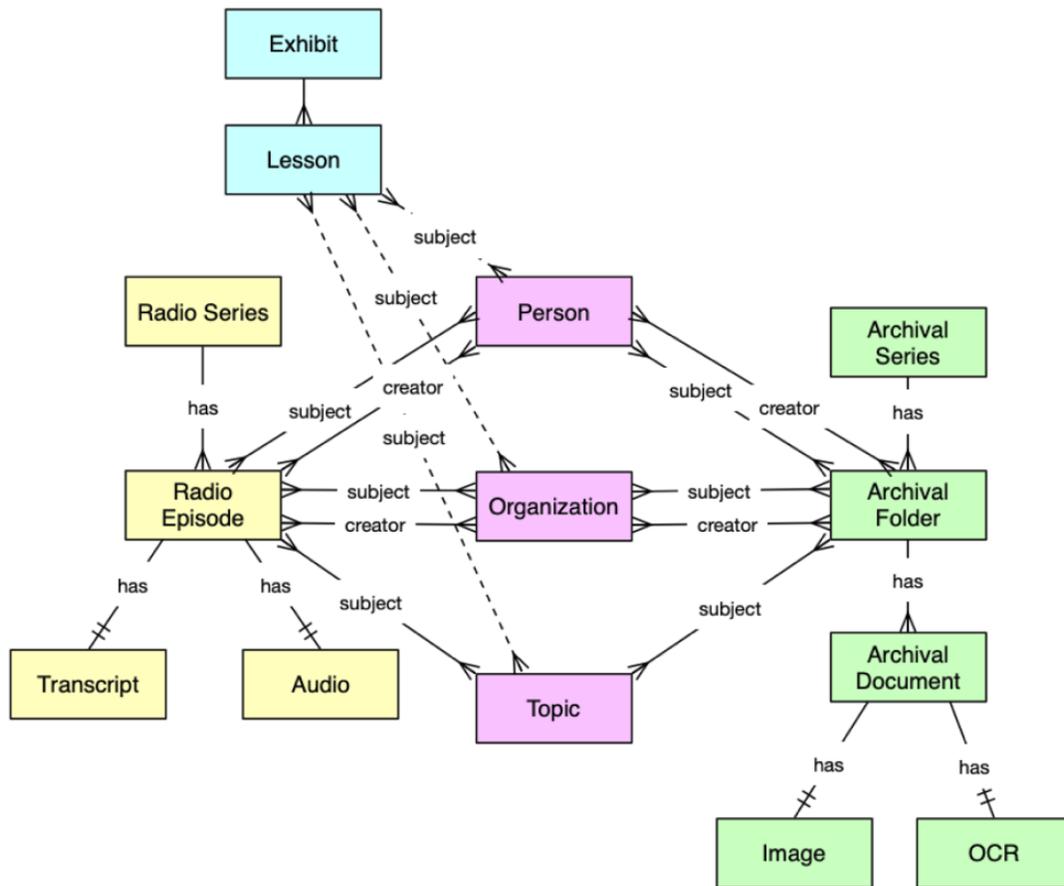


Figure 6. *Unlocking the Airwaves* Object Model.

Both the NAEB paper materials (at WHS) and the NAEB audio materials (at UMD) had already been cataloged and described according to established professional standards. The finding aid for the paper materials at WHS were encoded with the Encoded Archival Description (EAD) standard (see Figure 7). The audio materials had been described at the item level using the PBCore standard, which is based on the Dublin Core metadata schema with a number of added elements useful for media, making it possible to extend basic descriptive records by specifying sources, taxonomies, and parts of a media object.



National Association of Educational Broadcasters Records, 1925-1977 (bulk 1950-1970)

[outline view](#) | [full-text/printable view](#) | [Add this to my bookbag](#)

Search within this finding aid:

- About this Finding Aid
- Summary Information
- Biography/History
- Scope and Content Note
- Administrative/Restriction Information
- Contents List
 - Central Correspondence Files
 - William G. Harley Speeches
 - Subject File
 - National Educational Radio Files
 - Office of Research and Development Files
 - Publications and Newsletter Files

Summary Information

Title: National Association of Educational Broadcasters Records
Inclusive Dates: 1925-1977 (bulk 1950-1970)

Creator:
 National Association of Educational Broadcasters

Call Number: U.S. Mss 76AF; PH 3962; Audio 1244A

Quantity: 44.2 c.f. (112 archives boxes), 3 tape recordings, and 21 photographs

Repository:
 Wisconsin Historical Society, Division of Library, Archives, and Museum Collections

Archival Locations:
 See the catalog entry for information on possible additional materials and shelf locations.
 Wisconsin Historical Society (Map)

Abstract:
 Records, mainly 1950-1970, of the National Association of Educational Broadcasters (NAEB), an association of educational broadcasters and public broadcasting stations. The NAEB traced its origins to the formation of the Association of College and University Broadcasting Stations in 1925. The largest part of the collection is a subject file which documents the NAEB's board of directors, committees, conventions, conferences, seminars, and workshops; the development of the Educational Television Facilities Act of 1962; projects funded through Title VII of the National Defense Education Act of 1958; a curriculum development project directed by Marshall McLuhan; programming material on such offerings as the series "Jeffersonian Heritage"; and files on the National Educational Television and Radio Center, the Fund for Adult Education, the Joint Committee on Educational Television, and other organizations and foundations interested in educational broadcasting. The NAEB disbanded in 1981.

Language: English

URL to cite for this finding aid: <http://digital.library.wisc.edu/1711.dl/wiarchives.uw-whs-us0076af>
 † Bookmark this †

Figure 7. EAD-encoded finding aid for the National Association of Educational Broadcasters Records, 1925-1977. Available online at <http://digital.library.wisc.edu/1711.dl/wiarchives.uw-whs-us0076af>.

Recognizing the aforementioned challenges with different institutional imperatives driving archival description choices, we wanted to identify and track descriptive data about digitized materials in a manner that didn't interrupt or fundamentally change each collection's canonical metadata. The choice of Airtable as an interlocutor of sorts helped us enable our virtual reunification strategy, serving as a centralized location where we could create these links within the database by harnessing Airtable's various features, while also exploiting its API to connect this connective tissue out to various endpoints at various stages in the workflow. 34

Folder level details already present in the EAD were imported into an initial table in the Airtable base, which was then used to a) identify and prioritize boxes and folders to digitize, b) monitor their digitization and description progress, and c) use the Dublin Core metadata schema to describe the paper materials at the folder level. 35

PBCore records for the audio materials were also imported into the same Airtable base as a separate table. This allowed us to connect the paper to the audio by linking key descriptive elements where they align across the Dublin Core and PBCore schemas: by Creator, Contributor, Subjects, Temporal and Spatial Coverage, and Dates. These linkages were achieved through separate tables for People, Corporate Bodies, Subjects, and Geographic Locations. Figures 8-12 below illustrate how all of these various tables were structured within the Airtable base. 36

ID	Title	Series	Box	Folder	Digitize?	Priority	Progress	Date	IA URL
49	Committees Programming Practices, 1968-1969	Subject File	034	01	Yes	2	Completed	1968-1969	https://archive.org/details/na
50	Committees Publications, 1954-1964	Subject File	034	02	Yes	2	Completed	1954-1964	https://archive.org/details/na
51	Committees Public Relations, 1954-1957	Subject File	034	03	Yes	2	Completed	1954-1957	https://archive.org/details/na
52	Committees Public Relations, 1958-1960	Subject File	034	04	Yes	2	Completed	1958-1960	https://archive.org/details/na
53	Committees Public Relations, 1961	Subject File	034	05	Yes	2	Completed	1961	https://archive.org/details/na
54	Committees Public Relations, 1962	Subject File	034	06	Yes	2	Completed	1962	https://archive.org/details/na
55	Committees Public Relations, 1963-1965	Subject File	034	07	Yes	2	Completed	1963-1965	https://archive.org/details/na
56	Committees Radio Network Program, 1961-1962	Subject File	034	08	Yes	2	Completed	1961-1962	https://archive.org/details/na
57	Committees Radio Network School, 1953-1956	Subject File	034	09	Yes	2	Completed	1953-1956	https://archive.org/details/na
58	Committees Radio Network School, 1960-1964	Subject File	034	10	Yes	2	Completed	1960-1964	https://archive.org/details/na

Figure 8. Airtable base tracking extant data (including box number, folder number and title, and archival series) pulled from the EAD finding aid for the NAEB paper collections at Wisconsin Historical Society.

ID	Type(s)	Format	Relation	Series Relation	Description
49	Text Correspondence	Meeting minutes and reports	image/tiff		1968 and 1969 correspondence, survey rep
50	Text Correspondence	Meeting minutes and reports	image/tiff		1954 to 1964 correspondence regarding the
51	Text Correspondence	Meeting minutes and reports	image/tiff		1954 to 1957 correspondence regarding the
52	Text Correspondence	Meeting minutes and reports	image/tiff		1958 to 1960 correspondence regarding the
53	Text Correspondence	Meeting minutes and reports	image/tiff		1961 correspondence regarding the NAEB p
54	Text Correspondence	Meeting minutes and reports	image/tiff		1962 correspondence regarding the NAEB r
55	Text Correspondence	Rules and By-laws	image/tiff		1963 to 1965 correspondence regarding the
56	Text Correspondence		image/tiff	References Special of the V	1961 to 1962 correspondence regarding the
57	Text Correspondence		image/tiff	References Tales from Four	1953 to 1956 correspondence regarding the
58	Text Correspondence		image/tiff	References Nous Parlons Fr	1960 to 1964 correspondence regarding the

Figure 9. Table tracking descriptive Dublin Core metadata at the folder level for the NAEB paper collections being digitized at the Wisconsin Historical Society. Dublin Core Elements, such as "Type," are described in the table.

Unlocking the Airwaves-Metadata & Digitiz...

Authorities (People & Entities) Authorities (Subjects) Authorities (Geographic/Locations) Series IDs Series Relation PBCore Metadata (Audio) Series Genre(s)

Grouped by Series 45 hidden fields 1 filter Grouped by 1 field Sort Color

	Title 1 (Series)	Title 2 (Episode)	Description 1 (Series)	Description 2 (Episode)
TITLE 1 (SERIES) Count 13				
▼ Portrait of the American				
2438	cpb-aacip-500-804xms5g	Portrait of the American	Who are we as Americans?	Series that examines assessments of the American using the themes of in... This program seeks an answer t
2439	cpb-aacip-500-b355q6b	Portrait of the American	Moralists and misfits	Series that examines assessments of the American using the themes of in... This program looks at various n
2440	cpb-aacip-500-9290dh32	Portrait of the American	The American Adam and evil	Series that examines assessments of the American using the themes of in... Innocence unawares: The Ameri
2441	cpb-aacip-500-hh6c0h8r	Portrait of the American	Immigrants: Again and again	Series that examines assessments of the American using the themes of in... Innocence and moralism as the
2442	cpb-aacip-500-xp6v2q6h	Portrait of the American	The affluent American	Series that examines assessments of the American using the themes of in... Second theme: American affluen
2443	cpb-aacip-500-h41jnc63	Portrait of the American	The burdens of plenty	Series that examines assessments of the American using the themes of in... Affluence continued; critique of
2444	cpb-aacip-500-f3m1h1n	Portrait of the American	Rugged and not-so-rugge...	Series that examines assessments of the American using the themes of in... Third theme: the American succ
2445	cpb-aacip-500-vh5chd7p	Portrait of the American	The advent of Babbitts	Series that examines assessments of the American using the themes of in... The big change in economic his
2446	cpb-aacip-500-9316p51	Portrait of the American	The happy company men	Series that examines assessments of the American using the themes of in... The force of conformity in busin
2447	cpb-aacip-500-0v89m50x	Portrait of the American	How to know you've made it	Series that examines assessments of the American using the themes of in... What success means today and
2448	cpb-aacip-500-v97zqv7j	Portrait of the American	Quest for self	Series that examines assessments of the American using the themes of in... Fourth theme: the American sel
2449	cpb-aacip-500-319s5f68	Portrait of the American	The aggressive American	Series that examines assessments of the American using the themes of in... Self-reliance of Emerson and fir
2450	cpb-aacip-500-rx93d057	Portrait of the American	Patterns and paradox	Series that examines assessments of the American using the themes of in... Summary of series showing whi
+				
TITLE 1 (SERIES) Count 13				
▶ Prepare for survival				
TITLE 1 (SERIES) Count 7				
▶ Prison document				
TITLE 1 (SERIES) Count 12				
▶ Prospect of a union				

Figure 10. Airtable base tracking item-level PBCore metadata about individual radio programs, grouped in this view by Series.

Unlocking the Airwaves-Metadata & Digitiz...

Corporate Bodies Authorities (Geographic/Locations) Themes Progress Tracking Series IDs Series Relation Series Genre(s) Exhibits SHARE BLOCKS

VIEWS Grid view 1 hidden field Filter Grouped by 3 fields Sort Color

	ID	Box	Folder Name	Assigned to	Progress	Uploaded to IA?	Internet Archive URL
ASSIGNED TO J.J. Count 46							
PROGRESS Completed 30							
PROGRESS In Progress (Metadata) 13							
BOX 103 4							
182	naeb-b103-f01	3-f01	103 Publications, 1949	J.J.	In Progress (...)		http://archive.org/details/na
183	naeb-b103-f02	3-f02	103 Publications, 1949	J.J.	In Progress (...)		http://archive.org/details/na
184	naeb-b103-f03	3-f03	103 Publications, 1949	J.J.	In Progress (...)		http://archive.org/details/na
185	naeb-b103-f04	3-f04	103 Publications, 1949	J.J.	In Progress (...)		http://archive.org/details/na
BOX 104 4							
BOX 105 5							
PROGRESS In the Queue 3							

Figure 11. Table tracking digitization progress by production assistants at the University of Wisconsin-Madison.

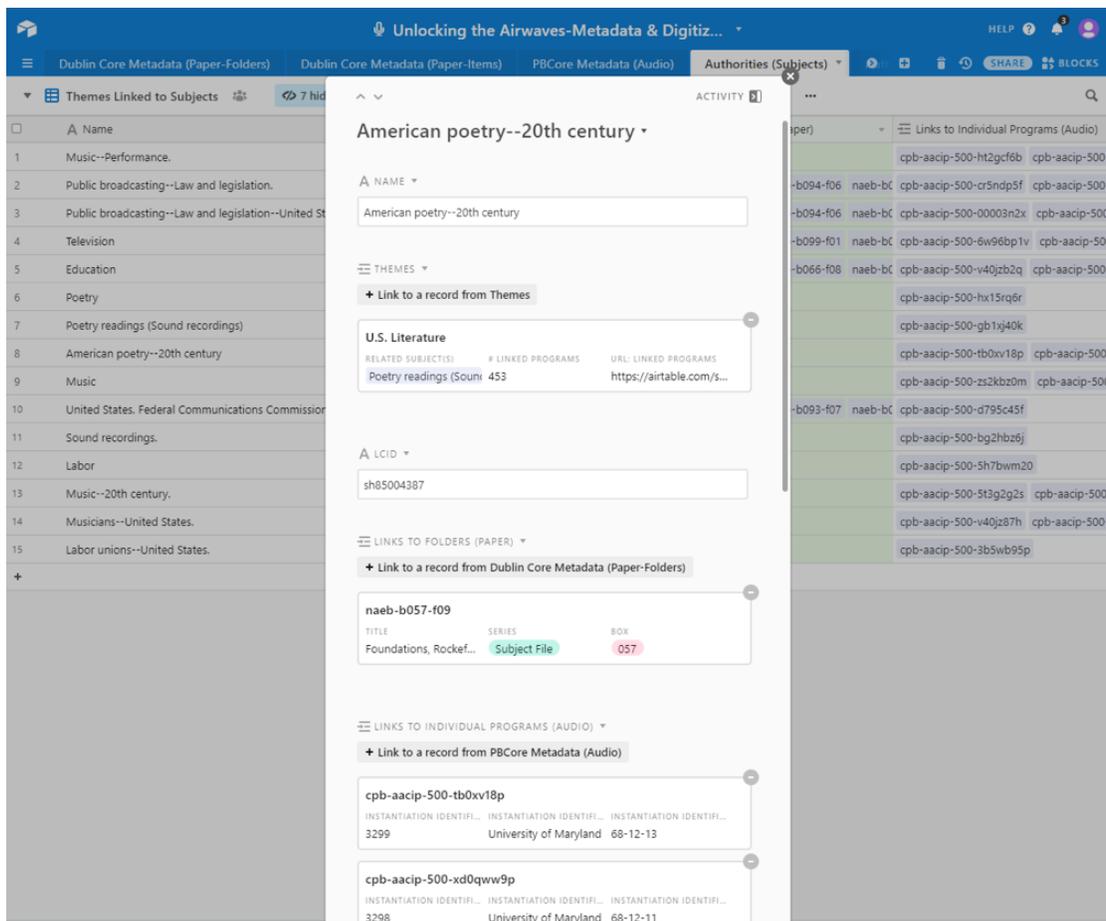


Figure 12. Airtable base tracking subject authorities as a linked table to item-level PBCore metadata. These authorities are also linked to the records in the same base tracking descriptive Dublin Core metadata about individual folders.

In the final phases of the project, we are establishing workflows and documentation for porting the newly created metadata back out to select endpoints. This involves utilizing Airtable's API to generate JSON files, which we can then be transformed into different formats:

37

1. valid PBCore XML records to deliver to University of Maryland Libraries and to the AAPB in order to respect the descriptive formats and choices of the institutional stewards of the collections;
2. EAC-CPF records to submit to the Social Networks and Archival Context (SNAC) project in order to encourage wider adoption of standardized authority records about early educational radio by other institutions with complementary holdings (thus bridging the Archival Diaspora); and
3. mapping archival authorities to Wikidata, in order to further extend the use of the newly created authority records to a wider audience.

For the latter two, we are identifying a key subset of archival authorities (roughly 200) for People and Corporate Bodies which are either central/crucial to the NAEB's development, initiatives, or programs, or not currently represented in the broader world of published authorities, i.e. in the Library of Congress, SNAC, VIAF, or Wikidata. All of the above choices reflect a calculated assessment of established and emerging descriptive standards and practices in the archival field, while also harnessing digital humanities tools, methodologies, and workflows to deploy them in innovative ways.

38

Conclusions and Possible Futures

Digital humanities projects like *Unlocking the Airwaves* not only make A/V materials more accessible to scholars, but they also position A/V within their specific contexts of production and distribution. Split collections make A/V materials less accessible and obscure the connections that can emerge from contextual resources, so virtual reunification projects

39

offer the possibility of elevating A/V materials as research sources. In the case of the NAEB, the paper records tell the story of an organization as it developed an infrastructure for national educational radio, and the A/V materials include hundreds of programs covering a wide array of topics and approaches. The NAEB's programs, produced between 1952 and 1970, represent a period of astonishing growth, turmoil and social change in the United States. Bringing the NAEB's papers and audio together can reinstate the centrality of the radio programs in the story of the NAEB, while presenting them in the context of their circulation. Beyond its obvious relevance to scholars interested in the history of educational and public media there is also an important pedagogical added value to such an approach. In its present form, the online collection can be integrated into curricula in American History, Politics, Social Science, African American History and Culture, Art and Music. This approach thus not only enriches research possibilities, but also restores the status of A/V collections as primary sources worthy of attention and analysis.

Although we are still in the project's final stages, including the development of all exhibits and teaching tools, the team is already thinking through ways we can improve our interoperability and linked data approaches to exploit the possibilities of the semantic web. We have been thinking through transformation of project data into JSON-LD, developing workflows with OpenRefine, using Wikidata to display knowledge graphs within the application, and integrating our archival authorities with other related collections within SNAC. For these future phases, we welcome inquiries and ideas from potential collaborators working in similar ways with similar resources.

A major benefit of this work will be demonstrating new and innovative digital humanities approaches towards increasing the discoverability of audiovisual collections in ways that allow for better contextual description, and a flexible framework for connecting audiovisual collections to related archival collections. With the publication of this resource, we will have documented a tangible solution for this set of challenges.

Notes

[1] To view a demonstration of the interface and functionality of the Unlocking the Airwaves beta web application, please reference the Vimeo video for "The NAEB Radio Collection: Document Deep South and Mississippi Waterways", by Stephanie Sapienza (<https://vimeo.com/419551316#t=462s>).

Works Cited

- Acland 2016** Acland, C. "Low-Tech Digital." In C. R. Acland and E. Hoyt (eds), *The Arclight Guidebook to Media History and the Digital Humanities*. Reframe Books, Sussex (2016), pp. 132-144.
- Adenot and Toy 2018** Adenot, P. and Toy, R. "Web Audio API". World Wide Web Consortium. (2018) Retrieved from <https://www.w3.org/TR/webaudio/>.
- Becker 2001** Becker, S. "Family in a Can: Presenting and preserving home movies in museums" *The Moving Image*, (1)2 (2001): 54-62.
- Breaden et al. 2016** Breaden, C., Holmes, C. and Kroh, A. "Beyond Oral History: Using the Oral History Metadata Synchronizer to Enhance Access to Audiovisual Collections" *Journal of Digital Media Management*, vol. 5, no. 2 (2016):133–150.
- Clement 2015** Clement, T. (2015). "When Texts of Study are Audio Files." In S. Schreibman, R. Siemens and J. Unsworth, *A New Companion to Digital Humanities*, Wiley-Blackwell, Hoboken (2015), pp. 348-357.
- Cook 2001** Cook, T. "Fashionable Nonsense or Professional Rebirth: Postmodernism and the Practice of Archives" *Archivaria*, 51 (Spring 2001): 14-35.
- DACS 2019** DACS, "Describing Archives: A Content Standard" DACS (2019): <https://saa-ts-dacs.github.io>.
- Deodato 2006** Deodato, J. "Becoming Responsible Mediators: The Application of Postmodern Perspectives to Archival Arrangement & Description" *Progressive Librarian*, 27 (Summer 2006): 56.
- Duff and Harris 2006** Duff, W. M. and Harris, V. "Stories and Names: Archival Description as Narrating Records and Constructing Meanings" *Archival Science*, 2.3 (2002): 263-285.
- Florini 2017** Florini, S. "Oscillating Networked Publics: Contingent Uses of Black Digital Networks" Digital Dialogues lecture series at MITH (2017): <https://mith.umd.edu/dialogues/dd-fall-2017-sarah-florini/>.

- Forsberg 2017** Forsberg, W. "Yes, We Scan: Building Media Conservation and Digitization at the National Museum of African American History and Culture" Maryland Institute for Technology in the Humanities Digital Dialogue (November 7, 2017): <https://mith.umd.edu/dialogues/dd-fall-2017-walter-forsberg/>.
- Heany 2000** Heany, M. "An Analytical Model of Collections and their Catalogues" *Collection Description*, 3:1 (2000): <http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf>.
- Howard 2017** Howard, J. "Pop Up Archive Filled a Need for Audio Archiving, and Apple Noticed" *HUMANITIES*, Volume 38, Number 4 (Fall 2017): <https://www.neh.gov/humanities/2017/fall/feature/pop-archive-filled-need-audio-archiving-and-apple-noticed>.
- Indiana University Bloomington 2011** Indiana University Bloomington Media Preservation Initiative Task Force. "Meeting the Challenge of Media Preservation: Strategies and Solutions" Indiana University (2011): <http://hdl.handle.net/2022/14135>.
- Kuhn 2018** Kuhn, V. "Images on the Move: Analytics for a Mixed Methods Approach." In J. Sayers (ed.), *The Routledge Companion to Media Studies and Digital Humanities*, Routledge, New York (2018), pp. 300-309.
- Leigh 2006** Leigh, A. "Context! Context! Context! Describing Moving Images at the Collection Level" *The Moving Image*, 6:1 (2006): 33-65.
- McShea 2013** McShea, M. "Megan McShea responds to "Does the creation of EAD finding aids inhibit archival activities?" [Blog Post]. 18 December 2013. Retrieved from <http://archivesnext.com/?p=3617>.
- McShea 2015** McShea, M. "Putting Archival Audiovisual Media into Context: An Archival Approach to Processing Mixed-Media Manuscript Collections" Council on Library and Information Resources (2015): www.clir.org/wp-content/uploads/sites/6/McShea.pdf.
- Orgeron et al. 2012** Orgeron, D., Orgeron, M. and Streible, D. *Learning with the Lights Off: Education Film in the United States*, Oxford University Press, New York (2012).
- Palmer 2004** Palmer, C. L. "Thematic Research Collections." In S. Schreibman, R. Siemens and J. Unsworth (eds), *A Companion to Digital Humanities*, Blackwell Publishing, Hoboken (2004), pp. 348-382
- Pfeiffer 2019** Pfeiffer, S. "WebVTT: The Web Video Text Tracks Format" World Wide Web Consortium (2019): Retrieved from <https://www.w3.org/TR/webvtt1/>.
- Prelinger 2009** Prelinger, R. "Points of Origin: Discovering Ourselves through Access" *The Moving Image*, 9:2 (2010):164-175.
- Punzalan 2013** Punzalan, R. "Virtual Reunification: Bits and Pieces Gathered Together to Represent the Whole" Thesis, University of Michigan, Ann Arbor (2013): <http://hdl.handle.net/2027.42/97878>
- Punzalan 2014a** Punzalan, R. "Understanding Virtual Reunification" *Library Quarterly: Information, Community, Policy*, 84:3 (2014): 294-323.
- Punzalan 2014b** Punzalan, R. "Archival Diasporas: A Framework for Understanding the Complexities and Challenges of Dispersed Photographic Collections" *The American Archivist*, 77 (2014): 326-349.
- Ranger 2013** Ranger, J. "Does the creation of EAD finding aids inhibit archival activities?" [Blog Post]. 22 May 2013. Retrieved from: <https://www.avpreserve.com/blog/does-the-creation-of-ead-finding-aids-inhibit-archival-activities/>.
- Ranger 2014** Ranger, J. "Data is a simple machine" [Blog Post]. 10 January 2014. Retrieved from <https://www.avpreserve.com/blog/data-is-a-simple-machine/>.
- Samuels 1986** Samuels, H. "Who Controls the Past?" *The American Archivist*, 49, no. 2 (1986): 109-124.
- Tai 2013** Tai, A. "WebVTT versus TTML: XML considered harmful for web captions?" Presented at Balisage: The Markup Conference 2013, Montréal, Canada, August 6 - 9, 2013. In Proceedings of Balisage: The Markup Conference 2013. Balisage Series on Markup Technologies, vol. 10 (2013) Retrieved from <http://www.balisage.net/Proceedings/vol10/html/Tai01/BalisageVol10-Tai01.html>.
- Vinson 2019** Vinson, E. "Reassessing A/V in the Archives: A Case Study in Two Parts." *The American Archivist*, 82:2 (2019): 421-439.
- Williams 2016** Williams, M. "Networking Moving Image History: Archives, Scholars, and the Media Ecology Project" In C. R. Acland and E. Hoyt (eds), *The Arlight Guidebook to Media History and the Digital Humanities*, Reframe Books, Sussex (2016) pp. 335-345.

Wisser 2011 Wisser, K. M. "Describing Entities and Identities: The Development and Structure of Encoded Archival Context — Corporate Bodies, Persons, and Families" *Journal of Library Metadata*, 11:3-4 (2011):166-175.