



# Broadcasting A/V Data

Using linked data  
and archival authority records  
to enhance discoverability  
for broadcasting collections.

NEH Grant #  
HAA-277247-21

Submitted on  
September 30, 2023

## WHITE PAPER

### Broadcasting Audiovisual Data

*Using linked data and local authority  
aggregators to enhance discoverability for  
broadcasting collections*

---

#### Principal Investigators

**Stephanie Sapienza**

Digital Humanities Archivist  
University of Maryland (MITH)

**Eric Hoyt**

Professor of Media & Cultural Studies  
University of Wisconsin-Madison

# Introduction

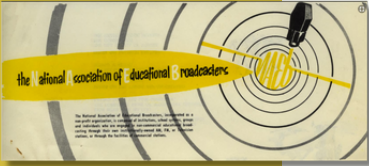
*Broadcasting Audiovisual Data* (BA/VD) is an initiative to enhance discoverability of archival radio collections using a linked open data framework to encourage the use of audiovisual collections inside and outside the academy. A partnership between MITH and the University of Wisconsin-Madison, the project is an expansion of the previous NEH-funded project *Unlocking the Airwaves*. While *Airwaves* was centered on virtually reuniting one set of geographically separated collections, the National Association of Educational Broadcasters' (NAEB) paper and media collections, *Broadcasting A/V Data* connects the linked NAEB collections to three additional complementary collections of educational radio, community radio, and public radio history. These include [The National Federation of Community Broadcasters](#) (NFCB) collections at UMD Libraries; the [Wisconsin Public Radio](#) (WHA) collections at University of Wisconsin-Madison Libraries; and the [WLB/KUOM collections](#) at University of Minnesota Libraries.

Educational and public broadcasting collections are a window into the history of the American experience. These collections are not just about unique content, they're also about unique people and organizations. BA/VD uses people and organizations as connective tissue between siloed collections of historic educational radio to promote new discoveries not just about the history of broadcasting, but about the history of how Americans shared their stories with each other during some of our nation's most culturally tumultuous decades.

This project represents a substantial shift from thinking about collections through a content-centric lens to a network-centric lens. Now, instead of viewing the collections themselves as snapshots of a particular historical moment or trend, *we can view them as products of networked knowledge flows governed by institutional structures and individual whimsy*. The project created new access points to these four collections across [Wikidata](#), the [Social Networks and Archival Context Cooperative](#) (SNAC) website, and Wikipedia, which help draw attention to and contextualize these collections for scholars, educators, journalists, and the general public. The project website includes a series of new exhibits curated by leading scholars in the field of broadcasting history, as well as a network visualization and a suite of simple tutorial videos and clear documentation for other stewards of cultural history who wish to implement our methods and workflows.

# Prologue: Unlocking the Airwaves

This project builds on the work of a previous NEH-funded project, *Unlocking the Airwaves: Revitalizing an Early Public Radio Collection* (PW-259067-18). The graphic below provides a brief visual overview of the Airwaves project, which can be researched more thoroughly by reading the [white paper](#).



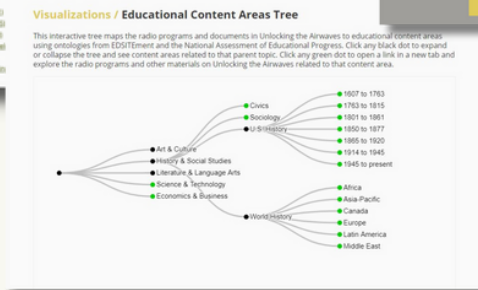
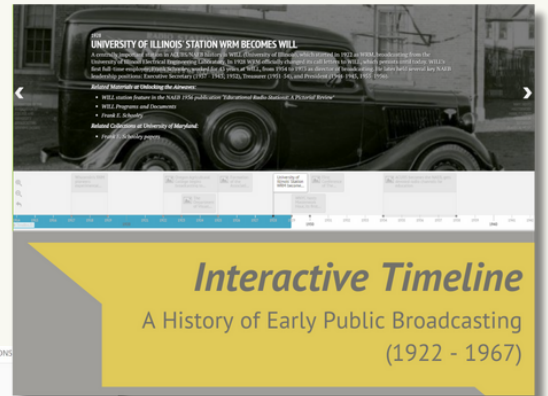
**NAEB  
Papers  
(Wisconsin)**

+

**NAEB  
Audio  
(Maryland)**

**VIRTUAL REUNIFICATION**

**Airwaves Website**



**Collections linked by:**

- **Subjects/Themes**
- **Entities (People/Organizations)**
- **Radio Series**
- **Full Indexed Search (OCR/Transcripts)**

**Also features:**

- **Exhibits**
- **Visualizations (Timeline, Map, Educational Content Tree)**

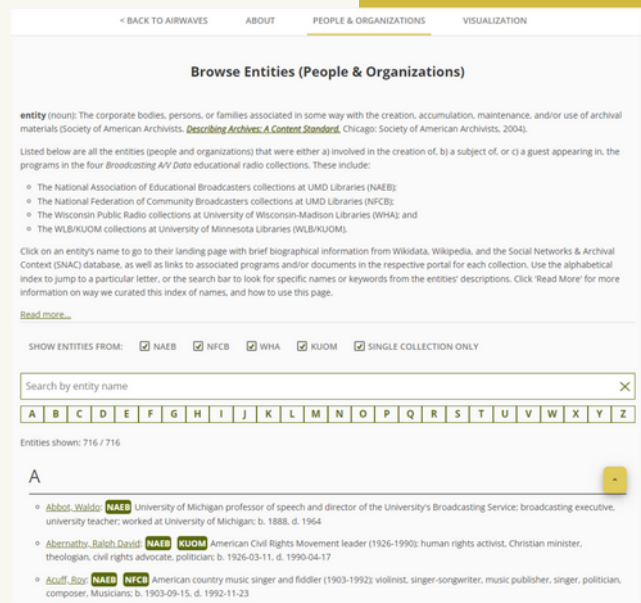
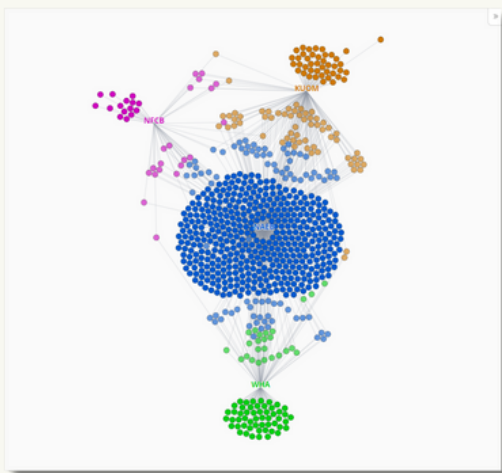
# Now: Broadcasting Audiovisual Data

With similar brevity, below is a graphic visualization of the BA/VD project, demonstrating how it fundamentally differs from Airwaves in execution:



## VIRTUAL REUNIFICATION LINKED DATA

BA/VD Website



### Collections linked by:

- **Entities (People/Organizations)**
- **Browse/Limited Search (Wikipedia blurbs)**

### Also features:

- **Exhibits**
- **Linked Data Tutorials/Documentation**
- **Network Visualization**

# Networks, Linked Data, & Research Questions

Below is a more in-depth exploration of our central research questions, as well as more contextual background information about networks and the usages and affordances of linked data.

## **QUESTION #1: HOW CAN THE DIGITAL HUMANITIES ENCOURAGE THE USE OF AUDIOVISUAL COLLECTIONS AS PRIMARY HISTORICAL RECORDS, BOTH IN AND OUTSIDE OF HIGHER EDUCATION?**

University libraries, historical societies, local and national archives, and personal collections are rich with audiovisual broadcasting collections – recordings dating from the 1930s that could enrich our understanding of American history and life. These collections do not merely reflect the history of broadcasting, but are a window into the history of the human experience.

In many cases, archival audiovisual materials have been collected and maintained separately from other kinds of (primarily textual) archival sources. At present, researchers must often discover and manually reunite audiovisual collections and their related materials if they want to understand a broadcast not just as an audiovisual object but as a medium that relays information within a set of historical contexts (time, place, related events, etc.).

Although a/v collections exist within and have borrowed strategies from libraries and museums, they do not enjoy a level of descriptive attention anywhere close to their textual counterparts. As a result of being woefully underdescribed and segregated from related, contextual resources, audiovisual 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.

As online collections are being asked to do so much more in a time of increased distance education and research, 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 materials.

# Networks, Linked Data, & Research Questions (continued)

## QUESTION #1 (CONTINUED)

We've identified three core reasons why a/v collections are underutilized in the first place: a) A/V collections are underdescribed as a whole, tending towards skeletal descriptions at the item level, b) A/V collections are often segregated from related contextual resources, and c) A/V collections present an additional challenge in the form of a temporal commitment barrier, where large time-based collections require an unfeasible time commitment for researchers. *Unlocking the Airwaves* successfully demonstrated methods for addressing some of these challenges, much of which is documented in a peer-reviewed piece written by *Airwaves* team members for the 'AV in DH' edition of *Digital Humanities Quarterly* in 2021 ("[Healing the Gap: Digital Humanities Methods for the Virtual Reunification of Split Media and Paper Collections](#)").

For BA/VD, we extended our reach to incorporate new approaches with linked data and authority aggregators, to move a/v collections towards having a more national platform alongside other major archival and digital humanities initiatives.

There are many possibilities that are afforded by the use of linked data, which connects information across the web together through concepts (subject terms), places (geographic terms), or people and organizations (entities). Linked data has the advantage of encouraging exploration, illuminating interesting connections between disparate collections which aren't immediately apparent to researchers who are looking at a collection in a silo, fostering surprise and spontaneity.



Above: Linked data graphic, courtesy of Stanford Linked Data Workshop

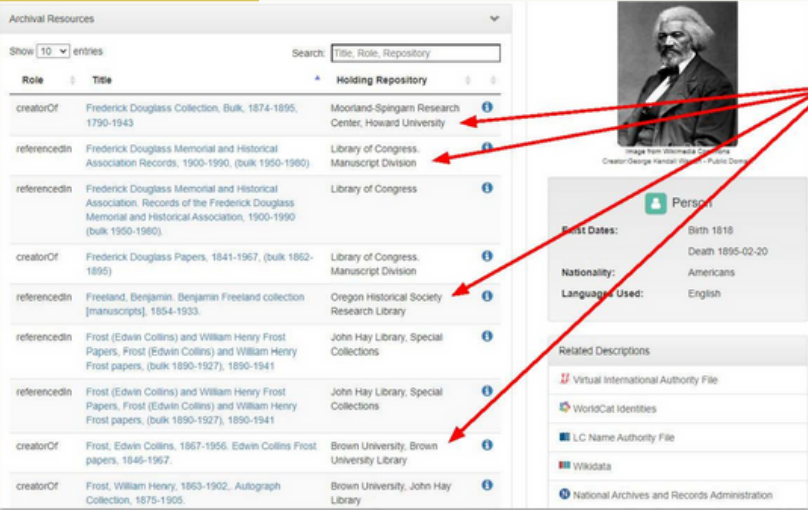
Below: Wikidata data model, demonstrating how a centralized archival authority record works with linked data to connect users to resources about an entity. (Source: <http://w.wiki/32q>)

label	<b>Marie Curie</b> (Q7186)	item identifier
description	Polish-French physicist and chemist Maria Skłodowska-Curie   Maria Salomea Skłodowska   Madame Curie ▶ In more languages	aliases
property	award received	value
rank		qualifiers
statement group		opened references
	<b>Nobel Prize in Physics</b> point in time: 1903 together with: Henry Becquerel, Pierre Curie prize money: 36,339 Swedish krona	
	▼ 2 references reference URL: <a href="http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html">http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html</a> retrieved: 3 August 2015 publisher: Nobel Foundation language of work or name: English title: The Nobel Prize in Physics 1903 (English) reference URL: <a href="https://www.nobelprize.org/nobel_prizes/about/amounts/">https://www.nobelprize.org/nobel_prizes/about/amounts/</a> + add reference	
	Willard Gibbs Award point in time: 1921 ▶ 1 reference + add value	collapsed reference

# Networks, Linked Data, & Research Questions

## QUESTION #1 (CONTINUED)

The same can be said about the usefulness of archival authority records, or records which standardize access points by uniquely identifying an entity, its associated/variant names, and other contextual information. The idea behind standardizing and publishing archival authorities online in a centralized aggregator website, is to a) encourage reuse by and interoperability with other systems, and b) provide a centralized hub for collections which are connected TO the entities – people and organizations – attached to them.



The screenshot displays the SNAC interface. On the left, a table lists archival resources with columns for Role, Title, and Holding Repository. On the right, a detailed view of a person's record is shown, including a portrait, birth and death dates, nationality, and language used. Red arrows point from the text on the right to specific entries in the table and the person's record.

Role	Title	Holding Repository
creatorOf	Frederick Douglass Collection, Bulk, 1874-1895, 1790-1943	Moorland-Spingarn Research Center, Howard University
referencedIn	Frederick Douglass Memorial and Historical Association Records, 1900-1990, (bulk 1950-1980)	Library of Congress, Manuscript Division
referencedIn	Frederick Douglass Memorial and Historical Association, Records of the Frederick Douglass Memorial and Historical Association, 1900-1990 (bulk 1950-1980)	Library of Congress
creatorOf	Frederick Douglass Papers, 1841-1967, (bulk 1862-1895)	Library of Congress, Manuscript Division
referencedIn	Freetand, Benjamin, Benjamin Freetand collection [manuscripts], 1854-1933.	Oregon Historical Society Research Library
referencedIn	Frost (Edwin Collins) and William Henry Frost Papers, Frost (Edwin Collins) and William Henry Frost papers, (bulk 1890-1927), 1890-1941	John Hay Library, Special Collections
referencedIn	Frost (Edwin Collins) and William Henry Frost Papers, Frost (Edwin Collins) and William Henry Frost papers, (bulk 1890-1927), 1890-1941	John Hay Library, Special Collections
creatorOf	Frost, Edwin Collins, 1867-1956: Edwin Collins Frost papers, 1845-1967.	Brown University, Brown University Library
creatorOf	Frost, William Henry, 1863-1902. Autograph Collection, 1875-1905	Brown University, John Hay Library

From Frederick Douglass's SNAC record, we can easily discover links to his archival collections, no matter the custodial institution.

A key objective of SNAC is to provide researchers with convenient, integrated access to historical collections held by multiple private and public archives and libraries around the world. This effort also sets the stage for a cooperative program for maintaining information about the people documented in the collections.

Description in SNAC helps resolve the challenge of *archival diaspora* by offering search and discovery through one central system of description.

snac Social Networks & Archival Context

## QUESTION #2: HOW DOES IT CHANGE THE USER RESEARCH EXPERIENCE TO ACCESS RADIO COLLECTIONS THROUGH A NETWORK-CENTRIC LENS?

The *BA/VD* resource we created is substantially different in terms of its user experience, in that it is fundamentally focused through a network-centric lens, as opposed to a content-centric lens.

# Networks, Linked Data, & Research Questions (continued)

## **QUESTION #2 (CONTINUED)**

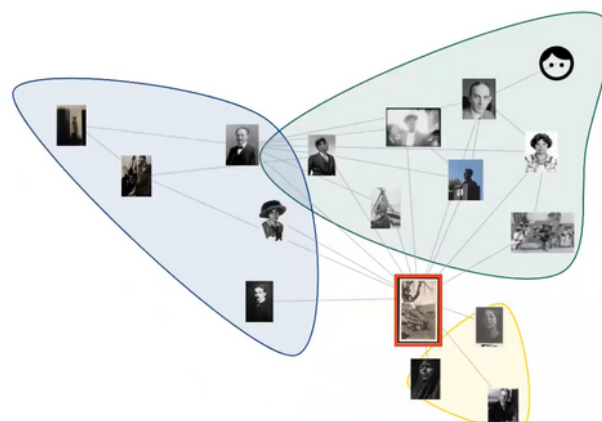
Project team member Christina Gibson helped us in the early phases of the project by investigating and writing about what it means for researchers to make this cognitive switch. She wrote: “I think about Karl Marx’s observation about the importance of social relationships/status and then Pierre Bourdieu’s extrapolation about the locus of knowledge being in hidden ways of being, which he calls ‘habitus.’ Librarians and archivists naturally think in terms of networks. Everything we do is creating and maintaining networks, both human and technological.

Historically this work has been undervalued because the owner of that intellectual process is not the person who connects one person to a resource, but the researcher-author who writes up the findings and constructs the narrative.

[These projects] present expansive, mutually informative networked data from multiple repositories to academic, K-12, and journalist audiences (among more). Those audiences have been largely taught to accrue knowledge in ways that erase the networking piece of the research process. Academic training from very young to very old is oriented toward a single actor/ author and their journey rather than the messiness of collaboration and interaction and the changeability that entails.”

To test this research question, we employed a team of exhibit curators to use the newly-developed resource to create seven new digital exhibits for the BA/VD website. We created a Curator Guide to guide them in this process, and employed user testing feedback cycles along the way to ensure we improved the resource, and were able to write about the results of our testing in the final project deliverables. The exhibits and the Curator Guide are described and linked to in sections further down in this report, along with more details on user feedback and research conclusions.

Mabel Dodge Luhan Network Visualization



*Above: Early prototype network visualization for a project on Mabel Dodge Luhan, created by Christina Gibson and Raffaele Viglianti. The BAVD team looked to this research when thinking through how users navigate network-centric research.*

# Networks, Linked Data, & Research Questions (continued)

## QUESTION #3: HOW DOES LINKED DATA HELP BETTER CONTEXTUALIZE RADIO COLLECTIONS BY CONNECTING THEM TO THE VAST ARRAY OF RESOURCES ON THE WORLD WIDE WEB?

Employing the use of contextual information to connect disparate collections is a more widespread strategy towards solving a problem that scholars like Ricky Punzalan have called the Archival Diaspora.

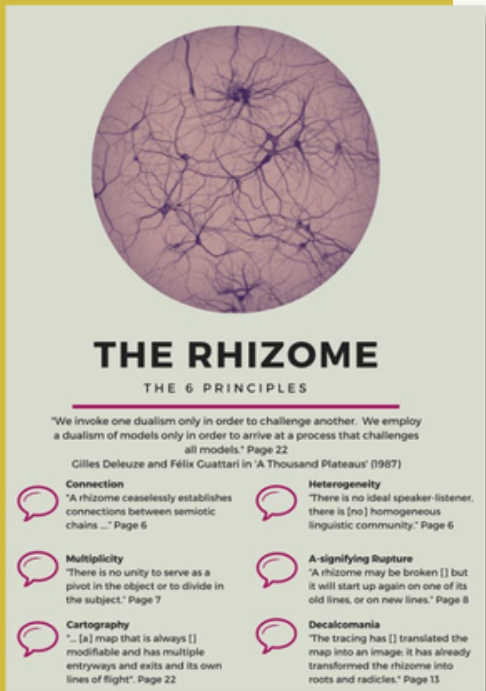
Acknowledging that a/v collections already face their own exponential description challenges, it follows that institutions with a/v collections are even further behind in these conversations about context and linked data.

While our second research question circles around the fundamental shifts in thinking that network-centric research represents for researchers, there is a broader implication for what it might then mean to connect those networks (and related collections) to a much broader contextual universe. To explore that question, we chose to frame it through the lens of how linked data fosters a *rhizomatic* ethos, by allowing users to branch out and explore related, contextual information about these people, organizations, and related collections on the web.

The concept of the rhizome is often invoked as a theoretical metaphor for linked data. A rhizome is a mass of interconnected things, with multiple pathways through and between, which can start in multiple places and lead to unexpected, serendipitous discovery experiences. In some ways the web itself is a rhizome. Wikipedia can also be seen as a rhizome.

Our project started by looking at four complementary collections of early educational radio, determined where there was overlap between them in terms of entities, and then mapped those entities to two key 'hubs:' Wikidata and SNAC (Social Networks and Archival Context). Linking to Wikidata also allowed us to connect entities to Wikipedia, the Library of Congress, the National Archives, and WorldCat.

Many of the resources referenced or linked to in the below sections go into great detail about the processes we employed to create this new framework, and to document our workflows so that they can be replicated by other archival professionals.



Above: Principles of the rhizome from Deleuze and Guattari (1980/2013), graphic by Jenny Mackness.



# Project Goals & Deliverables (continued)

## THE BA/VD WEBSITE RESOURCE (CONTINUED)



Images courtesy of UW-Madison Libraries and Wisconsin Public Radio

### WHA (Wisconsin Public Radio)

- Collecting Institution: University of Wisconsin-Madison
- Scope: ~5,400 radio programs (332 digitized)
- Time Period: 1920 - 1950
- Major Themes: educational content for farmers, home economics, 4-H programs, and Wisconsin folklore
- Examples (Program Titles): 4-H Songs; Homemakers Program; Farm Program; Abe Lincoln Slept Here

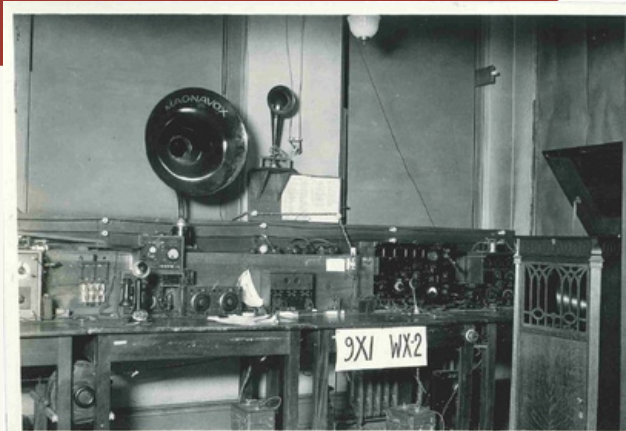
Wisconsin Public Radio—Where Broadcast Radio Began



This mural in the lobby of "Old Radio Hall" on the University of Wisconsin campus depicts early developments in the history of 9XM/WHA-AM. It has been operating continuously since 1917 making it the "oldest station in the nation" with an unbroken history of providing public radio programming services. Mural painted by John Stella

### Station WLB/KUOM

- Collecting Institution: University of Minnesota
- Scope: ~4,600 radio programs
- Time Period: 1920 - 1950
- Major Themes: educational information, political conversations, public affairs topics, and social issues
- Examples (Program Titles): Twin Cities School Problems; Masters of the Mesabi; Winifred Witch and the Ungrateful Grace; Songs of the Revolution; Secrets of the Water World



Images courtesy of University of Minnesota Libraries

# Project Goals & Deliverables (continued)

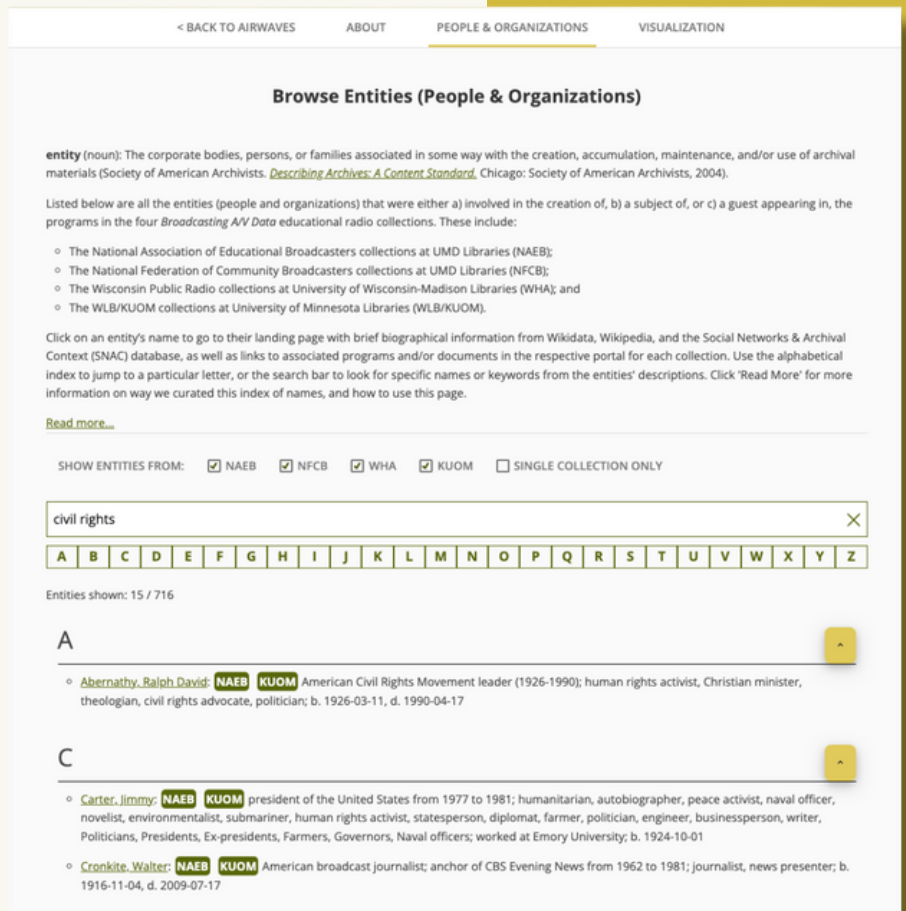
## THE BA/VD WEBSITE RESOURCE (CONTINUED)

Within the website, an About page contains basic background data about the project, a guide on how to use the website, links to documentation and training resources, links to newly-created digital exhibits, and a list of contributors.

Links to radio series and programs in each of the collections can be accessed through individual website pages that each represent a specific entity. To discover the entities, users can browse by clicking on the “People & Organizations” link in the top navigation bar. This leads to an alphabetized index of all entities which meet our selection criteria:

- Entities that overlap more than one of the four collections (nodes in a network, in other words);
- Entities which are represented frequently enough in any one collection that they are considered important to the history of educational radio, regardless of any overlap across the four collections.

The browsable entities index also contains short text blurbs that include cursory basic contextual information about each entity. These were either extracted from its label and description data on Wikidata or Wikipedia, or created from scratch for this project. That text is indexed as part of the site’s static index, and can be searched by keyword by using the search bar above the alphabetical index links.



(Above) The People & Organizations page, which allows for browsing and search of entities. Here you see the results for a search for ‘Civil Rights,’ which filters the list of entities by select metadata.

# Project Goals & Deliverables (continued)

## THE BA/VD WEBSITE RESOURCE (CONTINUED)

Once users navigate to an individual entity landing page (in addition to links to the represented collections), users will also see the following information:

- Links to content across all four radio collections that are related to that entity;

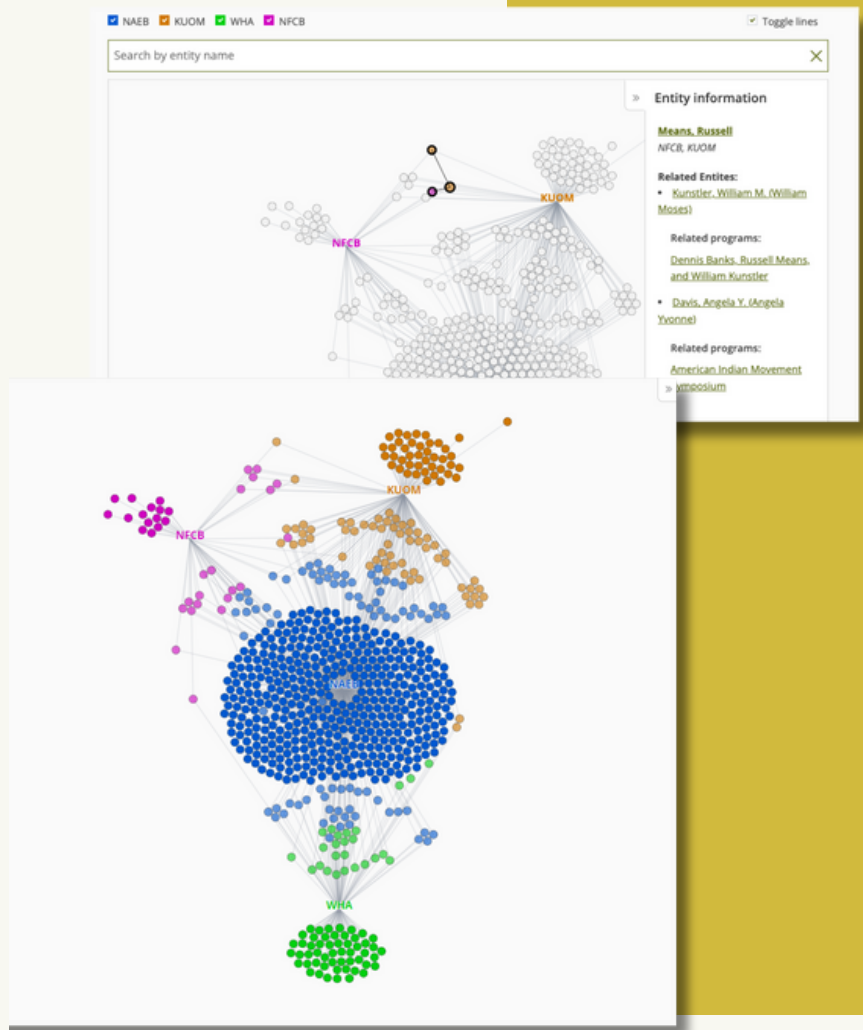
- Links to associated linked data for each entity across the web, either harvested or newly-created for this project and pushed back out to those sources. These include Wikidata, Wikipedia, SNAC, Library of Congress, the Virtual International Authority File (VIAF), the National Archives, and WorldCat Identities;
- A variety of different contextual information about each entity, including Wikimedia photos, short bios, dates of birth/death, alternate names, occupations, fields of work, and associated places. These were either extracted (and sometimes enhanced) from Wikidata and/or SNAC, or they were newly-created for this project.

# Project Goals & Deliverables (continued)

## THE BA/VD WEBSITE RESOURCE (CONTINUED)

Due to the project's emphasis on networks, the website also features a [visualization](#) that helps show where and how often overlaps occur between the entities represented in these four collections. The visualization is a customized force-directed graph representing the network of entities. Entities are in turn represented by color-coded circles, which are positioned on the graph according to the collections to which they fall under. When entities belong to more than one collection, they fall in between the two 'nodes' and closer to the graph's center.

The circles in the different collection nodes can be browsed, or searched using the search bar to search by name. Users can also filter by collection, which limits the visualization display to just those collections. Hover over a circle on the visualization to bring up the entity's name and connected collections, and click on the circle to view more information about the entity's connections to other entities on the visualization, as well as a display of sidebar information listing common programs that those two entities collaborated on. For example, users interested in the connections between one person and one radio station can use the visualization to browse the radio programs that they had in common.



# Project Goals & Deliverables (continued)

## NEWLY-CREATED SCHOLARSHIP ON EARLY RADIO

A team of exhibit curators selected specifically for this project were commissioned to test our research questions. Using a Curator Guide (linked to below in Appendices) along with a beta version of this website, curators were asked to explore the collections thinking about these networked knowledge flows, as opposed to exploring them thinking about pre-determined, content-centric subjects or themes. The seven new digital exhibits include:

### Exhibits

**Dance on the Radio: An Exploration of Dance Programs in the BA/VD Archives**

Bill Kirkpatrick



**From "Good Morning Homemakers!" To Singing the "Custom Made Women's Blues": Talking to Midwestern Farm Women over the Radio**

Jennifer Wang



**Burton Paulu and Musical Programming**

Christina Gibson



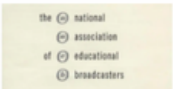
**Inventing a Service: How Educational Radio Became Public**

Alexander Russo



**The NAEB and the NETRC**

Allison Perlman



**Pioneers and Protectors of PBS Journalism**

Glenn Clatworthy



**The Global Politics of American Culture and the English Language: A Voice of America Survey**

Marit MacArthur



**See All Exhibits at Unlocking the Airwaves**



- [Dance on the Radio: An Exploration of Dance Programs in the BA/VD Archives](#) (Bill Kirkpatrick)
- [From "Good Morning Homemakers!" To Singing the "Custom Made Women's Blues": Talking to Midwestern Farm Women over the Radio](#) (Jennifer Wang)
- [Burton Paulu and Musical Programming](#) (Christina Gibson)
- [Inventing a Service: How Educational Radio Became Public](#) (Alexander Russo)
- [The NAEB and the NETRC](#) (Allison Perlman)
- [Pioneers and Protectors of PBS Journalism](#) (Glenn Clatworthy)
- [The Global Politics of American Culture and the English Language: A Voice of America Survey](#) (Marit MacArthur)

All seven BA/VD exhibits can be accessed on either the BA/VD website [About](#) page, or the *Unlocking the Airwaves* Exhibits page, alongside the original set of exhibits created for that project.

# Project Goals & Deliverables (continued)

## NEW OR ENHANCED ARCHIVAL AUTHORITY RECORDS

---

After an exhaustively daunting process of getting over 10,000 total unique entities represented in the collections down to a subset of 721 entities represented on the *BA/VD* website based on team-determined criteria (detailed more thoroughly in the Making Curatorial Decisions section under Discoveries and Outcomes), the project team was able to either enhance existing authority records already on the web, or create brand new archival authority records. We also created a small number of brand-new Wikipedia articles, and enhanced select existing articles. Below is a brief list containing statistics about those results, including links to summarized lists of the data in each category:

### **Wikidata**

Entities successfully reconciled against existing Wikidata items: 501

New Wikidata items created: 215

Existing Wikidata items edited: 99

### **Social Networks and Archival Context**

Entities successfully reconciled against existing records: 479

New SNAC CPF constellations created: 237

New SNAC resource records created: 5,073

### **Wikipedia**

Wikipedia articles edited: 7

Wikipedia articles submitted to create: 7

Wikipedia articles pending review: 4

Wikipedia articles accepted: 3

By doing this incredibly difficult but rewarding work, we were able to not only enhance and bolster the functionality and usefulness of the project website, but we were able to substantially contribute new and enhanced knowledge about these entities back out to the web. By 'planting trees' in this way, we increased the discovery potential of the networks represented in these collections in the broader sphere of semantic knowledge.

# Project Goals & Deliverables (continued)


## TRAINING RESOURCES AND DOCUMENTATION

A major goal of *BA/VD* is to increase understanding of these concepts in practice, and to address the many challenges involved with institutions' ability to even create (much less publish) archival authorities, which are much more labor-intensive than traditional bibliographic authorities created for library catalogs. Similarly, we aim to bridge the gap present in cultural stewards' ability to actually understand (much less implement) linked data, which so often runs up against legacy practices and stalls when faced with vastly different institutional priorities and resources.


The Appendices section contains links to the various types of documentation and training we created for the project, including the project's site code; data models; workflow diagrams; scripts for transforming various types of data; curatorial guidelines for selecting and prioritizing entities; metadata field mappings; guides to extracting, creating and formatting various types of records (Wikidata, Wikipedia, and SNAC records), and slide decks/videos of presentations and recorded workshops with demonstrations of our workflows.

### Documentation and Training


Another goal of *BA/VD* is to make these processes transparent to and replicable by other stewards of cultural heritage materials who want to use linked data processes to make their collections more discoverable across the semantic web. Below you'll see links to various types of documentation and training materials we've created to help guide others towards this goal.



GitHub Repository



Video Tutorials



Workflow Documentation

## Linked Data Tutorials


Share



Enhancing Geographic Metadata Using ...

MITH in MD


This video demonstrates how UMD ISchool graduate students Erica Gresham and Christine Anne Scully were able to use batch operatio...



Linked Data at University of Maryland (P...

MITH in MD

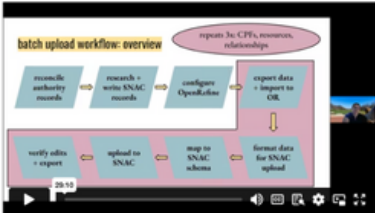
Thursday, May 25, 2023, 1-5pm Maryland Institute for Technology in the Humanities (MITH) Moderator: Stephanie Sapienza (MITH)...



Linked Data at University of Maryland (W...

MITH in MD

Thursday, May 25, 2023, 1-5pm Maryland Institute for Technology in the Humanities (MITH) Created and led by Emily Frazier (MITH) This...



Using SNAC and OpenRefine with the B...

MITH in MD

# Discoveries & Outcomes

## COMPLEXITY OF CURATORIAL WORKFLOWS AND CRITERIA

The level of complexity of the work required to first extract the entities from the metadata about the four collections, then to combine, disambiguate, and connect them together, was a Herculean challenge in and of itself.

To do this, we had to devise ways to first isolate the entity data from the series and program metadata, then ensure that this data was maintained in separate (but linked) Airtable databases. Then we worked to compare and disambiguate the entities that overlapped more than one collection and merge together duplicates. This combined data was then stored in a newly-created Airtable database that we called The Projector.

From there we had to work to determine how to get an original 10,529 total unique entities down to a manageable number. To do this, we had to start off by performing an initial reconciliation of all 10K entities using OpenRefine to match them to existing records on the web. This gave us results in four categories:

	BAVD name	BAVD CPF Landing Page	Wikidata match status	SNAC match status	Enhancement group
ENHANCEMENT GROUP					
>	Enhancement group 1				Count 19
ENHANCEMENT GROUP					
▼	Enhancement group 2				Count 11
20	Rarig, Frank M.	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
21	Phillips, Glen	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
22	Kastendieck, Miles	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
23	Gelb, Phillip	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
24	Wilhelm, Ross	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
25	WNAD (Radio station : Norman, Okla.)	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
26	Hagen, Carlos	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
27	McNeel, Wakelin	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
28	Bartell, Gerald	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
29	West, Robert	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
30	Agard, Walter R.	<a href="https://bavd.unlockingthe...">https://bavd.unlockingthe...</a>	Matched	Matched	Enhancement group 2
ENHANCEMENT GROUP					
>	Enhancement group 3				Count 32

### WD-match/SNAC-fail

- Has a matching item in Wikidata, but not SNAC

### WD-fail/SNAC-match

- Has a matching item in SNAC, but not Wikidata

### WD-match/SNAC-match

- Has a matching item in both Wikidata and SNAC

### WD-fail/SNAC-fail

- Does not have a matching item in Wikidata or SNAC

# Discoveries & Outcomes (continued)

## COMPLEXITY OF CURATORIAL WORKFLOWS AND CRITERIA (CONTINUED)

From here, we created a set of criteria (linked to in the Appendices section, "Enhancement priorities and work plan") based on two levels of Criteria:

**Criteria 1:** Based solely on a threshold for the number of radio programs linked to a given entity, and/or the number of collections it overlaps.

**Criteria 2:** Based on criteria not quantified by the number of associated programs, this criteria is crucial to grant considerations based on diversity, equity, and inclusion (DEI). Because women and people of color were historically not afforded the same opportunities as white men, they frequently have fewer production credits as a result. Basing enhancement criteria solely on the number of production credits, therefore, risks repeating and perpetuating these same patterns of exclusion. An intentional curatorial approach, with an emphasis on DEI, is necessary in order to break this cycle and bring to light the under-appreciated contributions of women and people of color. As such, application of Criteria 2 supersedes Criteria 1 for any authorities it affects.

Reconciliation Categories	WD-match/SNAC-fail			WD-fail/SNAC-match			WD-match/SNAC-match			WD-fail/SNAC-fail		
	P	CB	Total	P	CB	Total	P	CB	Total	P	CB	Total
<b>SUBTOTALS</b>	0	0	84	0	0	62	0	0	417	0	0	158
Criteria 1, Group 1									177			27
Criteria 1, Group 2									83			20
Criteria 1, Group 3									157			111
Criteria 2, Group 1												
Criteria 2, Group 2												
Criteria 2, Group 3												

Enhancement Actions	
<b>STAGE 1</b>	Enhancement group 1: - BAVD Total Count greater than or equal to 200 –OR– - Overlaps 2+ collections <a href="#">Show less</a>
<b>STAGE 2</b>	Enhancement group 2: - 220 = New Wikidata Entries - 242 = New SNAC Entries 1 Priority Specs; <b>Create</b> SNAC record 41 1 Priority Specs; <b>Update</b> SNAC record 196 2 Priority Specs; <b>Create</b> SNAC record 31 2 Priority Specs; <b>Enhance</b> SNAC record 94 3 Priority Specs; <b>Create</b> SNAC record 170 3 Priority Specs; <b>Enhance</b> SNAC record 189

# Discoveries & Outcomes (continued)

## COMPLEXITY OF CURATORIAL WORKFLOWS AND CRITERIA (CONTINUED)

Each of those criteria was then further subdivided into groups, which determined the level of enhancement we would grant each entity. This means, how many and which pieces of metadata to add to an entity record.

Field	Source	Group 1	Group 2	Group 3
		= 237	= 125	= 359
label	WD	✓	✓	✓
description	WD	✓	✓	✓
instanceOf_p31	WD	✓	✓	✓
altLabels/altNames	WD	✓	✓	✓
Minimum One:	WD	✓	✓	✓
snacARK_p3430	WD	—	—	—
loc_p244/locURL	WD	—	—	—
viaf_p214/viafURL	WD	—	—	—
worldCat_p7859/worldCatURL	WD	—	—	—
nara_p1225/naraURL	WD	—	—	—
birth_date_p569	WD	✓	✓	—
deathDate_p570	WD	✓	✓	—
inceptionDate_p571	WD	✓	✓	—
occupation_p106	SNAC	✓	✓	✓
fieldOfWork_p101	SNAC	✓	✓	✓
employer_p108	SNAC	✓	✓	✓
entityType	SNAC	✓	✓	✓
biogHist	SNAC	✓	✓	✓
wikidataURL	SNAC	✓	✓	✓
subjects	SNAC	✓	✓	✓
places	SNAC	✓	✓	✓
associatedWith	SNAC	✓	✓	✓
activities	SNAC	✓	✓	✓

- ✓ = Required
- ✓ = Required but can be minimal
- ✓ = Preferred/Can do without
- ✓ = Optional/cursory research

Past all of this, the work of creating new archival authorities for those entities in the categories which met our criteria, but did not have an existing authority on the web, was exhaustively challenging and time-consuming. In these cases, the entities often had very little information available with which to create the records. Thus, creating them involved a careful examination of the programs or documents in the collections where they were found, or using other forms of primary and secondary sources in archival repositories or the web.

To do this kind of work requires a kind of rare skill set that is a combination of creativity, incredible attention to detail, and a gift for problem-solving. The project was lucky enough to make use of the talents of Emily Frazier for this monumental task. But it did not escape our attention that this level of difficulty was not for the faint of heart. Future stewards who want to approach any similar linked data project that combines entity data from multiple collection sources would be advised to keep this in mind when planning for the amount of labor required, and the types of skill sets and technical knowledge needed to meet these challenges.

# Discoveries & Outcomes (continued)

## CHALLENGES AND BENEFITS OF WIKIDATA AND WIKIPEDIA

The section above describes a fraction of the sheer complexity of these workflows. At the center of this complexity was the choice to utilize two interrelated Wikimedia projects, Wikidata and Wikipedia. Both of these projects afforded us the ability to capitalize the immense power of the semantic web, but each of them also presented significant challenges getting up to speed with the intricacies of harnessing that power.

Wikidata is a platform that leverages crowdsourcing to enrich metadata and access linked data content from various data projects, such as the Library of Congress Linked Data Service, Getty vocabularies, the Social Network of Archival Content (SNAC), and others. Editorial decisions are made by the Wikidata community.

*BA/VD* relied on Wikidata IDs as our central URIs, because they centralized all the other linked data services in one centralized place. But to even get to the point where we could first harvest, then add back TO Wikidata, we had to do a substantial amount of maneuvering of this editorial process, including the creation of new Wikidata 'stub' entries for those entities without an existing entry, so that we could have an ID to start with, prior to any enhancements or additional linking. As an extension of the first outcome in this section, this requires a significant and unrelenting attention to detail, paying particular attention to dependencies and careful ordering of operations in our workflows.

Once we did have Wikidata IDs for the entities we selected for prioritization, OpenRefine was the tool we selected to allow us to first extend (pull in existing data) and contribute back to the service, using its built-in Wikidata plugin. Going through the careful first steps of creating stub entries and eliminating bottlenecks ended up saving us hundreds of person-hours to vastly expand our metadata about these entities, and to connect to the entity records in other services such as LoC and SNAC.

Enhancement Actions			
STAGE 1	#	= No Stage 1 Action, Move to Stage 2	
	#	= Needs Wikidata Entry	
	#	= No Stage 1 Action, Move to Stage 2	
	#	= Needs Wikidata Entry	
STAGE 2	#	= Match Wikidata Entry to Group 1 Priority Specs; <b>Create</b> SNAC record	41
	#	= Match Wikidata Entry to Group 1 Priority Specs; <b>Update</b> SNAC record	196
	#	= Match Wikidata Entry to Group 2 Priority Specs; <b>Create</b> SNAC record	31
	#	= Match Wikidata Entry to Group 2 Priority Specs; <b>Enhance</b> SNAC record	94
	#	= Match Wikidata Entry to Group 3 Priority Specs; <b>Create</b> SNAC record	170
	#	= Match Wikidata Entry to Group 3 Priority Specs; <b>Enhance</b> SNAC record	189

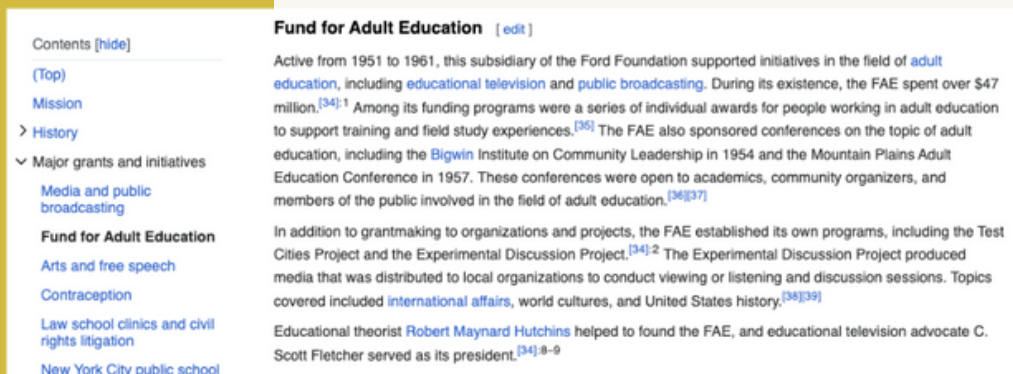
220	=	New Wikidata Entries
242	=	New SNAC Entries

# Discoveries & Outcomes (continued)

## CHALLENGES & BENEFITS OF WIKIDATA AND WIKIPEDIA (CONTINUED)

Wikipedia is the gold standard for community-monitored democratization of knowledge on the web, and has proven time and again to be one of the most effective tools to point users to resources connected to a given article about an entity. Once we used OpenRefine to connect entities to their Wikidata ID, we were also able to extract valuable information like Wikipedia 'Lead Sections' (brief biographical parenthetical information at the top of most articles), many of which you'll see on the *BA/VD* entity landing pages.

We also aimed to get an undetermined number of new Wikipedia articles published, in order to drive users towards the entity's connected resources. But there continue to be significant hurdles to getting a new article published, specifically in terms of getting it past the point of what Wikipedia calls 'notability requirements.' Wikipedia articles are subject to constant edits and potential deletion if not adequately maintained. So when we ventured to create new articles, we had to quickly get up to speed with those requirements, which are strict in nature.



The screenshot shows a Wikipedia article titled "Fund for Adult Education" with an "edit" link. On the left is a navigation menu with items like "Contents [hide]", "(Top)", "Mission", "> History", "Major grants and initiatives", "Media and public broadcasting", "Fund for Adult Education", "Arts and free speech", "Contraception", "Law school clinics and civil rights litigation", and "New York City public school". The main text of the article begins with "Active from 1951 to 1961, this subsidiary of the Ford Foundation supported initiatives in the field of adult education, including educational television and public broadcasting. During its existence, the FAE spent over \$47 million.<sup>[34]:1</sup> Among its funding programs were a series of individual awards for people working in adult education to support training and field study experiences.<sup>[36]</sup> The FAE also sponsored conferences on the topic of adult education, including the Bigwin Institute on Community Leadership in 1954 and the Mountain Plains Adult Education Conference in 1957. These conferences were open to academics, community organizers, and members of the public involved in the field of adult education.<sup>[36][37]</sup>"

*(Left) Newly-created entry for the Fund for Adult Education (a subsection of the Ford Foundation article) on Wikipedia..*

Even after putting in the effort to create an article, Wikipedia presents significant sustainability challenges, in that users must commit to the ongoing task of monitoring and updating it to ensure its relevance and accuracy. Furthermore, when creating articles about underrepresented or marginalized people and groups, cultural stewards often encounter significant bias and systemic discrimination, which in turn perpetuates the underrepresentation of these subjects on the platform. The *BA/VD* team met with a team of UMD Librarians during the project, who shared their workflows and stories about surmounting these hurdles.

# Discoveries & Outcomes (continued)

## CHALLENGES & BENEFITS OF WIKIDATA AND WIKIPEDIA (CONT'D)

---

One such story included a new article about Elaine J. Coates, the first African American woman to graduate from Maryland College, being flagged for deletion less than 10 hours after it was published. Luckily the team had many other users come to their defense, and the page was maintained.

These hurdles were the main reason that the *BA/VD* team only ended up publishing roughly seven new articles as part of our project, and enhancing/updating another seven. The sheer complexity of the project overall was such that continuously keeping on top of Wikipedia's shifting and subjective permanence was an additional labor challenge that we did not have the time or person-hours to overcome. As such, cultural practitioners who wish to replicate these practices must consider accounting for that additional labor when planning a project involving Wikipedia.

## SNAC AS AN EVOLVING RESOURCE

---

The Social Networks and Archival Context (SNAC) Cooperative is at the heart of how we devised our original project plan. An international cooperative of archives, libraries, and museums, SNAC's most visible project is the SNAC Portal, which allows users to discover biographical and historical information about entities that created or are documented in collections held at cultural heritage institutions. The *BA/VD* project was designed to link into this portal, and for the resources and 'constellations' (archival authority records) we created and uploaded to also link users to other collections connected to those same entities.

To do this, we originally strived to align our authority records to the standard used by SNAC, called Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF). EAC-CPF records are meant to be encoded so that they can link directly to Encoded Archival Description (EAD) finding aids, the gold standard for collection-level archival description.

# Discoveries & Outcomes (continued)

## SNAC AS AN EVOLVING RESOURCE (CONTINUED)

While the library community has a long history of cooperatively describing library holdings and sharing the descriptive data, the archival community does not. This is arguably based on a fundamental difference in the nature of the holdings of the two professional communities: many and perhaps most of library holdings are not unique, while most of the archival holdings are. [...] They are nevertheless inextricably interconnected in ways that provide opportunities for cooperating that will significantly benefit both the archival community and the researchers, scholars, and members of the public that use their holdings.

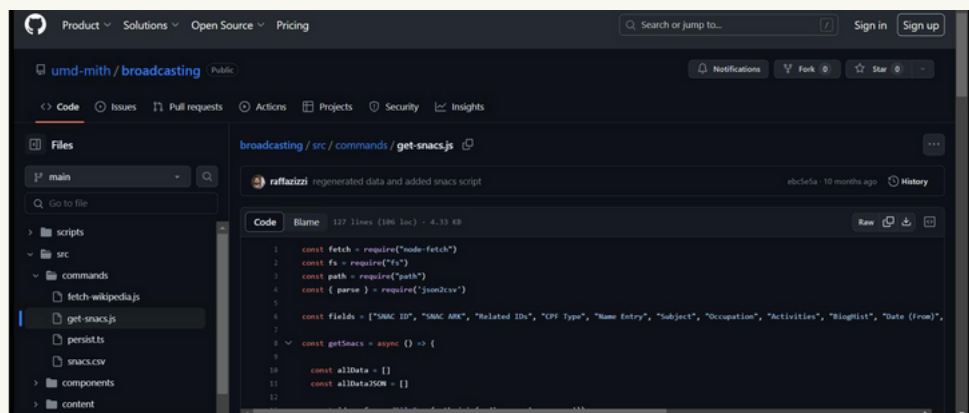
-- Daniel Pitti, "Rationale for an Archival Description Cooperative"

SNAC founder Daniel Pitti is a tireless advocate for developing and refining archival standards for description, with the goal that archival professionals begin working cooperatively to align those standards. SNAC is the largest and most well-known institutional initiative to bravely attempt bridging this gap in cooperative archival description, seeking economies to share labor and expense, and increase the discoverability of collections across the web. The SNAC Cooperative has received several impressive rounds of funding from Mellon and other sponsors, and is now in a state of rapid change as they seek to enhance their sustainability and ensure that their resources

persist into the future. To do this, they are moving to a membership model, and have also expanded and turned over many of their key staff in just the last two years.

All of this is positive in terms of catalyzing the archival field. But the timing of these changes often presented a challenge for the BA/VD team as we sought to utilize and align with SNAC's evolving priorities, technologies and technical standards. SNAC records were contributed en masse by hundreds of cultural stewards, there was much duplication in terms of authority records, some of which were being addressed as we plugged along (if slowly).

(Right) SNAC data harvesting script. There were technical challenges with parsing SNAC's EAC-CPF records which required us to create custom scripts to extract authority data from SNAC records.



The screenshot shows a GitHub repository for 'umd-mith/broadcasting'. The file 'src/commands/get-snacs.js' is open, displaying a JavaScript script. The script uses 'node-fetch' to fetch data from SNAC records and parse it into a JSON array. The code includes comments and variable declarations for fields like 'snac ID', 'SNAC ARI', 'Related ID', 'CPF type', 'Name Entry', 'Subject', 'Occupation', 'Activities', 'Bibliography', and 'Date (From)'. The script also includes a function 'getSnacs' and an 'allData' object.

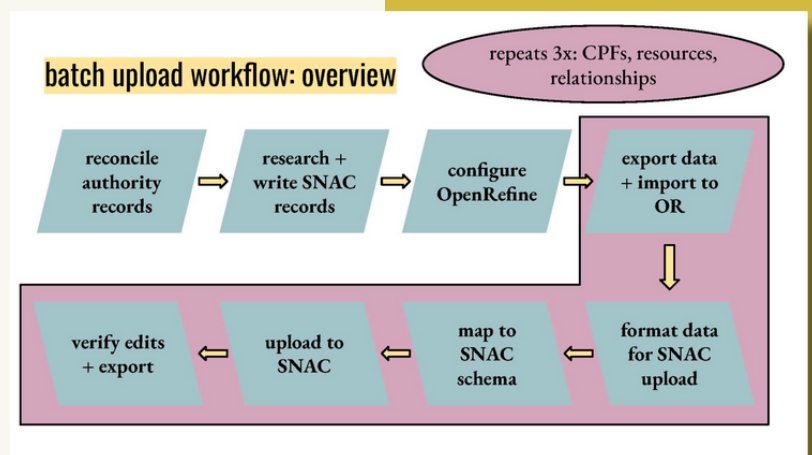
# Discoveries & Outcomes (continued)

## SNAC AS AN EVOLVING RESOURCE (CONTINUED)

There were technical challenges with parsing SNAC's EAC-CPF records which required us to create custom scripts to extract authority data from SNAC records. The SNAC OpenRefine API also faced many technical challenges, making the kind of automation that Wikidata allowed with ease, to be slow/delayed. We were able to extract SNAC records using OpenRefine, and use our custom scripts (available in the Appendices section) to extract and use the data from them, but there were significant challenges with pushing new or enhanced authority data back into SNAC.

When uploading a set of resource records, the BA/VD team had to make tough decisions about the hierarchical level we selected. For example, one producer or radio station might have had thousands of individual radio programs associated with it, which would have made browsing the resources related to that station bloated and overwhelming. Our solution involved linking directly to radio series, as opposed to radio programs, for just the NAEB records. This relates directly to the challenges described in the above section, in that not all of our content partners even had series metadata to work with.

On a positive note, many of these hurdles have started to ease right at the end of the project, as the SNAC team has worked tirelessly to address many of these challenges. BA/VD project team member Emily Frazier presented at the 2023 SNAC Annual Meeting, using our project as a case study to demonstrate how we were able to maneuver some of these challenges. Her presentation and slides are linked to in the Appendices below.



*(Above) Slide from Emily Frazier's presentation at the SNAC 2023 Annual Conference, demonstrating our batch upload workflows.*

# Discoveries & Outcomes (continued)

## FACING INSTITUTIONAL REALITIES

---

Another crucial area where the team we ran into the need for creative problem-solving revolved around the different types of metadata we received from the four partner institutions/content partners, as well how differing institutional realities and priorities pushed up against or complicated our ability to be cohesive and consistent with our workflows and the dissemination of new archival authority data back to those institutions.

Postmodern discussions of descriptive priorities have shifted in recent years towards discussions of how contextual information can supplement or enhance the work of data stewards, and fill in for gaps and shortages in labor resources needed to fully describe collections. Those very discussions are at the heart of our research questions. If an institution makes the decision to describe materials at the collection level as part of a More Process, Less Product (MPLP) mindset, there is nothing inherently wrong with that approach. It allows stewards to maneuver overwhelming backlogs with more speed, and get more material available to the public instead of letting them languish in the accession stage, hidden mostly or fully from view.

But it is very much worth noting that for the collections we selected for this project, and to craft the workflows that we created, required that a substantial amount of item-level description was already performed on the collections before we started work. This was in fact part of the rationale for selecting these four collections, since each of the newly-added collections were recipients of a previous CLIR Recordings at Risk grant.

But even with that said, the collections came to us in various forms of completion. The WHA collection in particular was partially digitized and those recordings were available through their Digital Library Services (DLS) portal with fully vetted item-level metadata. But there were an additional ~5,400 programs still in the queue for digitization, for which there was very good metadata, but that was not yet vetted for full accuracy. The UW team admirably championed an access-drives-preservation ethos for this project, meaning that they wanted records for the undigitized programs to be included in *BA/VD* regardless of their digitization status. They committed to digitizing a program if a researcher found it through the *BA/VD* website and

# Discoveries & Outcomes (continued)

## FACING INSTITUTIONAL REALITIES (CONTINUED)

requested it. But to get a permanent URI record for those undigitized records meant that the DLS team had to make exceptions to their policy for ingest and inclusion, and so they created 5,400 new skeletal records for us to use.

Similar challenges occurred with the NFCB and KUOM collections, wherein they were either in process of digitizing part of the collections, or added new records after we began work on the project. We had to account for the fact that many of the radio programs had title metadata at the program level, but didn't include metadata about a corresponding radio series (either because the program wasn't part of a series, or because that metadata was never originally captured). This required us to be careful in terms of choices around metadata consistency, and about how to display the programs on our website. The NAEB collections were always described hierarchically on the *Unlocking the Airwaves* website (Series --> Included Programs), so we had to adjust our data model to accommodate for differences between the collections.

*(Below) Screenshot from a Zoom meeting with the UW-Madison Libraries team, demonstrating the challenge of displaying radio series vs. programs in the WHA collections. 510 WHA programs did not have an assigned series, requiring us to find creative solutions to adjust our display parameters..*

The screenshot shows a Zoom meeting window. On the left, a web browser displays the 'WHA (UW-Madison) Metadata - BAVD' interface. The interface has a table with columns for 'Identifier (no prefix)', 'Title', and 'Series'. The first row is circled in red and shows '(Empty)' under 'Identifier', 'Count: 510' under 'Title', and 'Series' under 'Series'. Other rows include 'Merle Curti Lectures' (Count: 40), 'WSA-This our Democracy' (Count: 26), 'Wisconsin Communities' (Count: 11), 'University of Wisconsin Band Programs' (Count: 609), and 'Freshman Forum' (Count: 116). On the right, a 'Chat Messages' window shows a conversation. Emily Frazier asks for a URI for WHA, and Scott Prater provides a link. Eric Hoyt provides another link. Stephanie Lynn Saplenza responds to Emily's question.

Identifier (no prefix)	Title	Series
(Empty)	Count: 510	Series
Merle Curti Lectures	Count: 40	
WSA-This our Democracy	Count: 26	
Wisconsin Communities	Count: 11	
University of Wisconsin Band Programs	Count: 609	
Freshman Forum	Count: 116	

**Chat Messages**

Emily Frazier 19:00  
looks like we have this one for WHA:  
<https://digital.library.wisc.edu/1711.d/WHARadi>

Scott Prater 19:35  
<https://digital.library.wisc.edu/1711.d/WHARadi>

Eric Hoyt (he/him) 19:39  
<https://search.library.wisc.edu/digital/AJEOX5YPFQSVWZ83>

Eric Hoyt (he/him) 19:54  
[https://search.library.wisc.edu/search/digital?filter%5Bfacets%5D%5Bnames\\_facet-Bartel%2C+Gerald%5D=yes](https://search.library.wisc.edu/search/digital?filter%5Bfacets%5D%5Bnames_facet-Bartel%2C+Gerald%5D=yes)

Emily Frazier 42:08  
when we direct users to the records from the BAVD/Airwaves site, we can also have some way of noting which links will take them to live streaming audio and which are cite only

Stephanie Lynn Saplenza 42:41  
Yes thank you Emily that's what was (poorly) trying to say, but that is a much clearer way of stating it.

# Discoveries & Outcomes (continued)

## FACING INSTITUTIONAL REALITIES (CONTINUED)

---

Lastly, it's important to note that the same institutional limitations described above mean that many repositories don't currently have the systems or resources to maintain archival authority control over their collections. Thus, after we created or enhanced these authority records, we provided those data sets back to each content partner with the full knowledge that none of them currently have the ability to ingest or utilize that data in their respective systems. This places an additional sustainability burden on the BA/VD team to be the 'source of truth' for this centralized new authority repository, for as long as we choose to maintain the project.

## CHALLENGES CREATING TRAINING RESOURCES

---

The same challenges described above resulting from different institutional realities and priorities is connected to the related challenge of creating sufficiently replicable training and workflows. that are broadly construed enough to be widely applicable in the field yet detailed enough to bridge the afore-mentioned knowledge gap in terms of implementing complex metadata/descriptive workflows.

Cultural heritage practitioners often faced with a tension between seeing a viable solution for their daily challenges, and their institutions' willingness to let them expand their facility in tackling those challenges, become justifiably frustrated. Realities such as overwhelming backlogs of unprocessed materials, combined with a lack of labor or technological resources, can lead to feelings of overwhelm, resign, and adjusting to a less than optimal set of available solutions.

The goals of BA/VD were to test a series of solutions to some of these challenges, and then create tools and workflows for cultural stewards to learn how to tackle these challenges on their own. The reality that our team continuously faced was that there was only so much we could make those solutions viable and replicable, because the solutions were sufficiently technologically complex and situationally-specific. Additionally, some of these solutions were based on open-source software which were evolving and in flux (see the above section on SNAC as an evolving resource).

# Discoveries & Outcomes (continued)

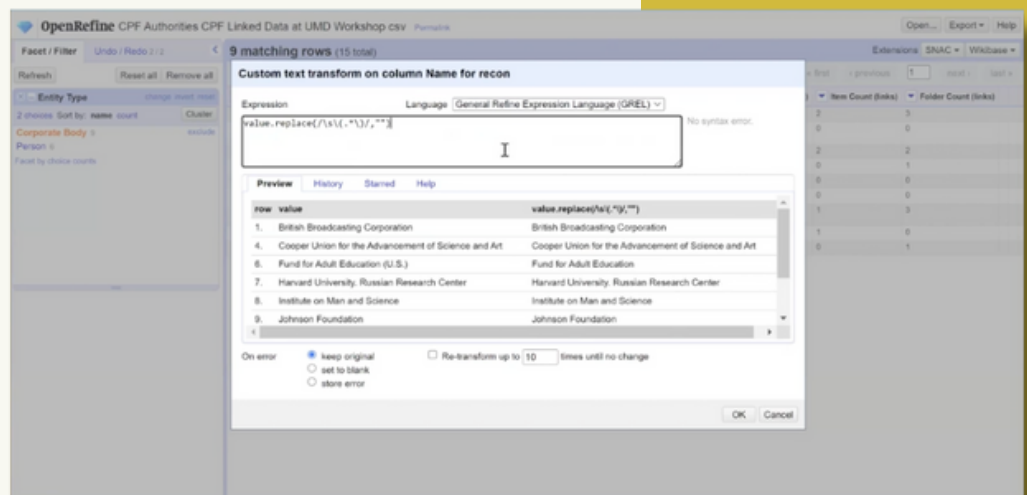
## CHALLENGES CREATING TRAINING RESOURCES (CONTINUED)

Even cultural stewards with a modicum of technological prowess and some knowledge of coding can find the implementation of OpenRefine's capabilities intimidating. Even to perform basic functions such as normalization, deduplication, and toggling between rows and records (an essential part of understanding the tool's facility) to accommodate for one-to-many or many-to-many relationships, or multiple values in a single descriptive field, required an understanding of OR's General Refine Expression Language (GREL).

In the May 2023 UMD Linked Data workshop hosted by MITH (with the UMD iSchool and Libraries), Emily Frazier provided a set of 'cheat codes' for users to perform some of these basic functions. In the workshop video linked to below in Appendices, she demonstrates a careful

and measured walkthrough of each GREL function and how it operates, so that users can both a) cut and paste the GREL expression to move through the workshop, but also b) understand how they might modify the expression based on their own institutional needs. But even with this, workshop participants continuously expressed frustration with not understanding how to parse or modify these expressions for their own workflows. Thus, our team learned that we could only do so much in terms of making this work replicable and usable for a broader community of users. At some point, stewards have to bridge that knowledge gap on their own. Which is precisely where they can become stalled by the exact aforementioned institutional realities.

*(Below) Screenshot from Emily Frazier's May 2023 workshop at MITH, in which she carefully walked participants through how to perform transformations on their data in OpenRefine, a step required before reconciliation steps can occur.*



# Discoveries & Outcomes (continued)

## CURATORS SHIFTING TO NETWORK-CENTRIC RESEARCH

The first users of the *BA/VD* website experienced a productive challenge: how to take a network-oriented approach to historical research?

It's common to begin a research project into radio history with a question that foregrounds subject matter and program content (for example, how did educational radio programs in the 1960s address the environment and pollution?). It's far less common to foreground the relationships across individuals and organizations responsible for the production of programming. And yet that is what we asked the seven broadcasting historians who we recruited to serve as exhibit curators to do. In this section, we want to expand upon the opportunities and challenges of network-centric research, both as they pertain to our project and larger fields of Humanities research.

By the Fall of 2022, the *BA/VD* website was sufficiently functional and developed for the seven exhibit curators to begin using it. The goals at this stage were two-fold: first, to provide feedback about their experiences as users that could help us further enhance the functionality of the website; and second, to begin formulating an idea for an exhibit that they could research further and then write about for the *BA/VD* website.

Through a series of virtual meetings and email exchanges, we obtained valuable feedback from the exhibit curators. Some were excited by the site's functionality—especially the ability to scroll through entries, click on an individual entity's name, then quickly pull up a page containing a biographical description and the programs in the collection linked to that entity. For example, clicking on "Roy Acuff" opens an entity webpage that contains recordings that the country music artist and publisher participated in within both the NAEB and KUOM collections.

### Broadcasting AV Data Draft Curator Guide (August 20, 2022) By Christina Gibson, Eric Hoyt, and Stephanie Saplenza

Thank you for agreeing to develop one or more online exhibits for the [Broadcasting AV Data project](#)! Although the pandemic has created some big challenges for our project (as well as so many other things), it has also renewed our commitment to showcasing the long history of educational media / remote learning and expanding digital access to archival sources. We've compiled a set of instructions below to guide you through the creation of your exhibits.

#### [1. What is Broadcasting AV Data \(BA/VD\)?](#)

#### [2. Big Picture Goals](#)

#### [3. Generating Exhibit Ideas](#)

##### [Linked Collections](#)

[NAEB \(National Association of Educational Broadcasters\)](#)  
[NFCB \(The National Federation of Community Broadcasters\)](#)  
[WHA \(Wisconsin Public Radio\)](#)  
[WLB/KUOM](#)

##### [Themes](#)

[Local, National, and Cosmopolitan](#)  
[National Identity](#)  
[Gender](#)  
[The Structure of Public Radio](#)

#### [4. Audience](#)

#### [5. Exhibit Models](#)

#### [6. Browsing BA/VD Entities and Linked Resources](#)

#### [7. Timeline for Exhibit Curators](#)

#### [8. Format for Text Write-up and Delivery](#)

#### [9. Linked Resources](#)

[A\) Citations to primary sources within the linked data collections](#)  
[B\) Visual images for the right sidebar](#)  
[C\) Audio or Text examples](#)  
[D\) \(optional\) Citations to secondary sources](#)

[Appendix: Sample Exhibit \(Topic: Burton Paulu\) by Christina Gibson](#)

# Discoveries & Outcomes (continued)

## CURATORS SHIFTING TO NETWORK-CENTRIC RESEARCH (CONT'D)

---

Other exhibit curators had substantial trouble adapting to the network-centric model. They could wrap their head around the idea, but expressed to our team that the website and network approach were difficult to use in the way they were accustomed to pursuing research. Some of the specific suggestions that our curators offered were:

- Creating data visualizations that would allow for additional ways to explore the collection;
- Adding full text searchability across all of the entities;
- Changing the first name, last name listings of the entity list into the more familiar last name, first name listings (e.g., changing “Roy Acuff” to “Acuff, Roy”);
- Adding more metadata about identity—especially race and gender—to the entity landing pages and the programs.

In the next section, we share how we chose which suggestions to implement, and the products that ultimately resulted.

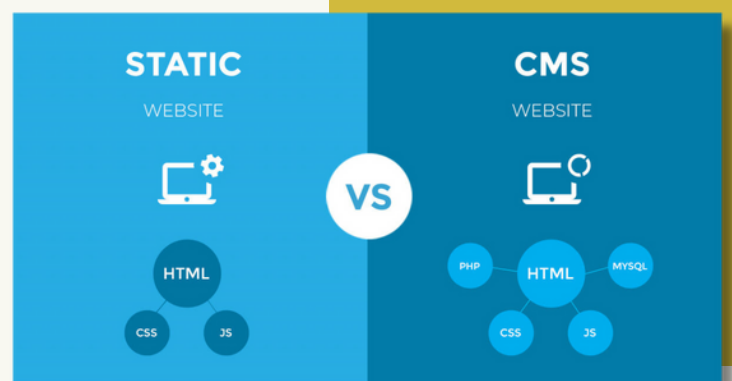
## EXPLORATION AND THE SEARCH CONUNDRUM

---

We chose to implement all of the suggestions from curators that we could that were A) feasible and B) enhanced the network-oriented research and exploration approach at the core of our project. However, we wanted to avoid making changes that we felt would undermine the network-oriented approach in favor of a more conventional, searchable catalog of entries.

Moreover, we needed to maintain the integrity of our website’s underlying data architecture, which was developed using a static site model (HTML, CSS, JavaScript, and JSON compiled through the metadata of an Airtable database, which itself aggregates information from our partner institutions).

*(Below) Graphic courtesy of Webbit's resource, "[Difference Between Static Website and CMS Website.](#)"*



# Discoveries & Outcomes (continued)

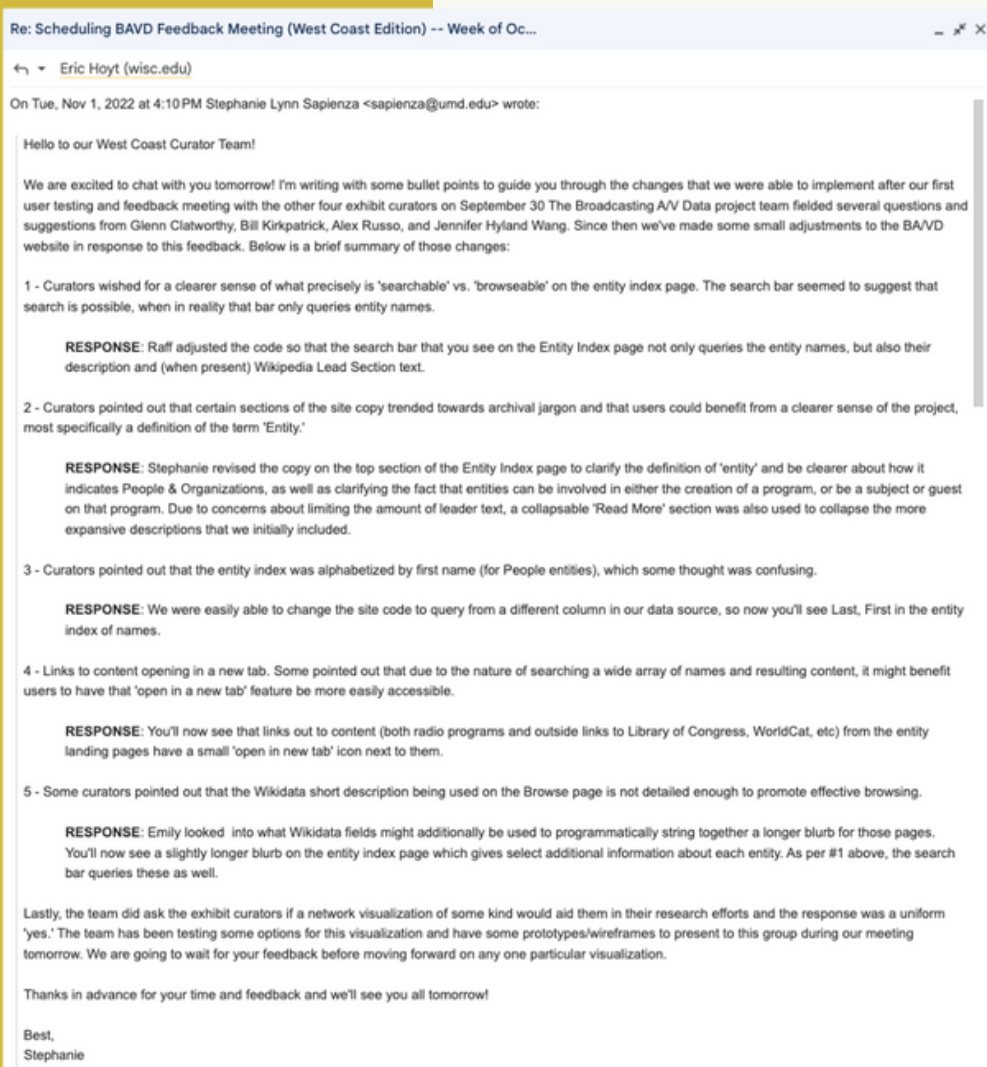
## EXPLORATION AND THE SEARCH CONUNDRUM (CONTINUED)

The static site approach makes the *BA/VD* website more durable, easier to preserve and sustain, and less vulnerable to online attacks that target open-source database frameworks, such as MySQL, and search indexes like Apache Solr. However, this same data architecture limits the possibilities for the site's full text searchability and adaptability to add metadata fields on the fly.

Ultimately, we opted for an in-between solution by implementing some very rudimentary search capacity on names, labels, and Wikipedia descriptions (when available), while suppressing the descriptions from the index page for readability. This approach encouraged users to browse serendipitously, while allowing some cursory level of search. We also designed data visualizations (described in more detail above) that applied a force directed graph design to the data, further emphasizing the network-oriented research approach while also allowing another entry point for exploration.

Some of our advisors and exhibit curators continued to find the exploratory methods to be challenging, since it differed from their past research methods.

By the end, though, the network-oriented approach yielded interesting findings and insights, many of which are thoughtfully described within the seven exhibits now on display on the final website.



# Appendices

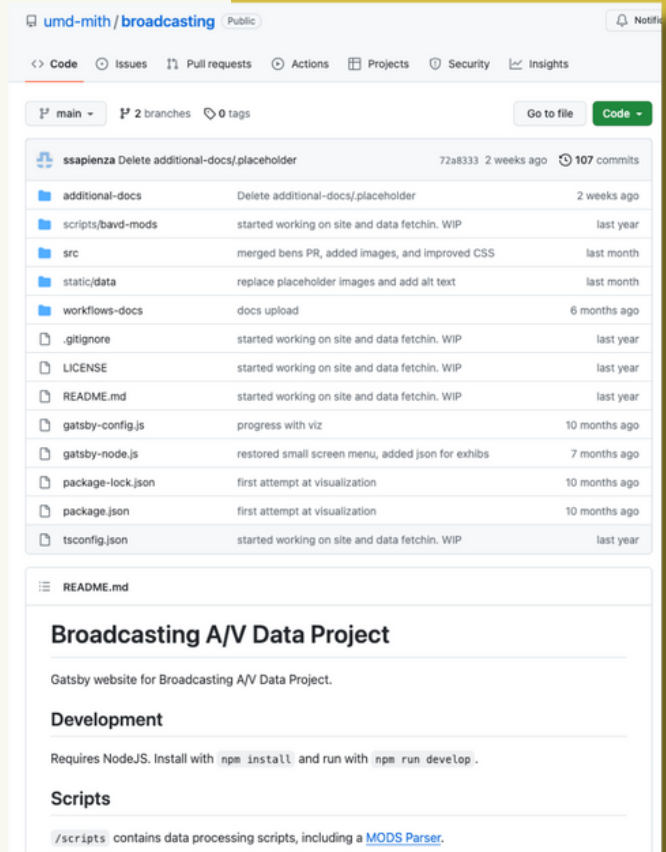
## [GitHub Repository](#)

The GitHub repository for BA/VD contains publicly-accessible site code.

## [BA/VD Workflows & Documentation](#)

The below list provides an overview of all the documentation shown in the Broadcasting A/V Data GitHub repository. Each entry represents one document found in the documentation folder, along with a brief description explaining what workflows are described by the document.

1. Broadcasting Audiovisual Data documentation index
2. Data model: Visual representation of the original project data model
3. Projector design schematic: Overview of design for central Airtable database, known as the Projector
4. MODS XML to CSV: Detailed specification for transforming WHA MODS XML records into CSV that describes location of elements, formatting requirements, etc.
5. JSON to CSV with OpenRefine: Instructions for transforming relatively simple JSON structure into CSV using OpenRefine
6. Enhancement priorities and work plan: Categories, criteria, and definitions used to enhance archival authority record for entities
7. Projector field mappings: Provides mappings/documentation for Wikidata, SNAC, and CPF landing pages fields in projector base
8. Wikidata reconciliation: Process of reconciling datasets against Wikidata using OpenRefine



# Appendices (continued)

## [BA/VD Workflows & Documentation](#) (continued)

9. Transforming SNAC records: Workflows for transforming SNAC constellations as exported into CSV for ingest into projector base
10. Creating new Wikidata items: Instructions for bulk Wikidata item creation in OpenRefine
11. Editing existing Wikidata items: Instructions for bulk Wikidata item edits in OpenRefine
12. Creating new SNAC records: Workflow for bulk SNAC constellation creation in OpenRefine
13. SNAC schema and formatting alignment guide: Table for mapping projector base fields to SNAC schema, with OpenRefine transformation instructions
14. Creating and editing Wikipedia articles: Workflow for identifying Wikipedia create/edit candidates, editing articles, and writing new drafts

## [Video Tutorials on Vimeo](#)

There are four videos in this Vimeo Showcase, each demonstrating a different aspect of how we applied linked data concepts in practice. The first video, “Enhancing Geographic Metadata Using Airtable, OpenRefine, & Wikidata,” was technically created for *Unlocking the Airwaves*. However, these workflows are the very processes we built on and expanded for *BA/VD*. The video concentrates heavily on how Airtable can be used as part of an overall workflow for getting geographic metadata prepped for use in OpenRefine, so that it can be used to find matches and geocoordinates with Wikidata. It then goes into detail about how to take the resulting geographic data and pull it back into Airtable. It features some demonstration on how to use particular Airtable extensions (now called ‘Apps’) like CSV Import, Batch Update, Deduplicate, to clean up and prep your enhanced geographic metadata for use in other applications such as map visualizations.

# Appendices (continued)

---

## [Video Tutorials on Vimeo](#) (continued)

The second and third videos are recordings from our May 2023 event at University of Maryland, Linked Data at University of Maryland. One of these is a panel with Diana Marsh (UMD iSchool) and Sarah Hovde (UMD Libraries), and Emily Frazier from the BA/VD team, talking about different ways we've harnessed linked data to enhance our various initiatives. The workshop video goes into great detail (with a demonstration using provided data sets and OpenRefine formulas) on how to use OpenRefine to extend metadata. We demonstrated this using three different data sets on Entities (People/Organizations), Subjects, and Geographic Terms. The fourth video is a Zoom recording of Emily Frazier's presentation at the 2023 SNAC Annual Meeting, using our project as a case study to demonstrate how we used SNAC in the project to extract and publish archival authorities (CPFs or 'constellations') from and to SNAC using OpenRefine.

## [Additional Documentation](#)

*NOTE: Some of the below events are also referenced in the above section, but below contains more links to slide decks and datasets which are complementary to the videos.*

## [Curator Guide](#)

"Linking Radio Networks: The Broadcasting A/V Data Project," Presentation at the Association for Recorded Sound Collections Annual Conference, May 2023. [Presentation Video](#) | [Presentation Slide Deck](#)

"Linked Data at UMD: A Panel and Workshop," Panel and Workshop at UMD, May 2023 [Folder with Slide Decks and Workshop Datasets](#)

"Broadcasting A/V Data to the SNAC Universe (and beyond!)," Presentation by Emily Frazier at the 2023 SNAC Annual Meeting, September 19, 2023 [Presentation Video](#) | [Presentation Slides](#)

# Project Team & Contacts

## PROJECT DIRECTORS

---

Stephanie Sapienza (Principal Investigator, University of Maryland)  
sapienza at umd dot edu

Eric Hoyt (Principal Investigator, University of Wisconsin-Madison)  
ehoyt at wisc dot edu

## PROJECT TEAM

---

Emily Frazier (Graduate Research Assistant, MITH)  
Trevor Muñoz (Director, MITH)  
Raffaele Viglianti (Senior Research Software Developer, MITH)  
Christina Gibson (Graduate Student, MITH)

Ben Pettis (Graduate Assistant, UW-Madison)  
Cat Phan (Archivist, UW-Madison)

Erik Moore (Archivist, University of Minnesota)  
Rebecca Toov (Archivist, University of Minnesota)

## EXHIBIT CURATORS

---

Christina Gibson, Towson University  
Glenn Clatworthy, Independent Researcher  
Jennifer Hyland Wang, University of Wisconsin-Madison  
Bill Kirkpatrick, University of Winnipeg  
Marit MacArthur, University of California, Davis  
Allison Perlman, University of California, Irvine  
Alexander Russo, The Catholic University of America