# Changing the tires while driving the car

A pragmatic approach to implementing linked data

David Seubert, American Discography Project, University of California, Santa Barbara, USA Shawn Averkamp, AVP, USA Michael Lashutka, ProperlySorted Database Solutions, Beacon, NY, USA





# Discography of American Historical Recordings (DAHR) - Background

- Discography = Bibliography of sound recordings
- Index of 326,000+ recordings by US record companies 1894 through 1950s
- Project began in 1964 by two US collectors to document the Victor Talking Machine Company
- Based at the University of California Santa Barbara since 2005
- Expanded to cover 8 additional recorded labels 2009-present
- Digitized audio added in 2011 in partnership with US Library of Congress -45,000 digital files now online





Browse

Search

Resources

News

vs About

#### **Featured Artist**



Harry Tally (composer)

On this Date... 11/4/1930

Blue moon blues (Victor Matrix BVE-62934, take 1 was recorded by Tiny Parham and his Musicians in Chicago, Illinois 952 North Michigan Ave.)

# **Discography of American Historical Recordings**

Welcome to the Discography of American Historical Recordings (DAHR), a database of master recordings made by American record companies during the 78rpm era. It is part of the American Discography Project (ADP)—an initiative of the University of California, Santa Barbara and the Packard Humanities Institute that is edited by a team of researchers based at the UCSB Library.

DAHR is an expansion of the Encyclopedic Discography of Victor Recordings (EDVR) incorporating previously published discographic data licensed from various publishers, folding them into the search and display framework of DAHR. Documentation of the activities of the Victor Talking Machine Company and RCA Victor derives primarily from consultation of publications of the companies and original documents held in the archives of Sony Music Entertainment in New York City. DAHR entries describing recordings made by Columbia Records, the Berliner Gramophone Co., OKeh Records, Zonophone Records, Leeds & Catlin Records, and Decca are based on authoritative published discographies, supplemented with new research, examination of recordings in the collections of the University of California Santa Barbara Library, and submissions from record collectors and archives.

To begin your research, enter names, titles, places, matrix or catalog numbers below, or click on the advanced browse or search buttons above.

#### News

#### Discos E y C y T de Columbia desaparecidos

September 23rd, 2020 By: David Seubert

Entre 1908 y 1923, Columbia Records emitió miles de grabaciones para mercados "étnicos" y extranjeros en su serie nacional de doble cara E, la serie de doble cara C española. ...

Read More »

#### Help Us Reconstruct Missing Columbia Data

August 11th, 2020 By: David Seubert



# Linked Data Project Overview

- With the addition of more digital audio, project leaders felt the database needed to be friendlier, provide more context to data, and move in the direction of an "audio encyclopedia" rather than a discography
- Every name or "talent" in database had been researched by DAHR editors and over 20,000+ unique names in database already had Library of Congress Name Authority File (LCNAF) record numbers in the database
- Each name was role specific, so individual records existed for a person as composer vs a person as performer, for example
- As LCNAF records have become prominent semantic links to interconnecting data sets online, we felt we could hook into these other data sets by leveraging our existing LCNAF through automated harvesting
- Helps establish DAHR as its own authority





# Initial Project Scope

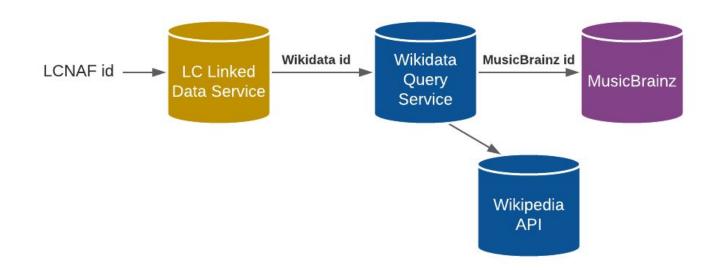
- Extract LCNAF record number from DAHR for AVP to use for harvest
- Use LCNAF to match and harvest identifiers from other crawlable or open datasets (VIAF, MusicBrainz, WikiData, Wikipedia, AllMusic, etc.)
- Merge our multiple role-based "talent" records under a new "Master Talent" record with a new URI and harvested data (more on this later)
- Harvest exact names matches from LC authority file to hopefully find more links between our names and other databases





# Data Harvesting: Strategy 1 Query LC Linked Data Service by LCNAF id

47,595 DAHR "talent" records with LCNAF records (21,535 unique artists)







# Data Harvesting: LCNAF

- 1) Query id.loc.gov LCNAF endpoint by record number, retrieve:
  - Name
  - Name type (personal, corporate, conference)
  - Birthdate
  - Deathdate
  - Notes
  - Sources

- VIAF id
- Wikidata id
- Getty id
- FAST id
- Related names
- Variant names

example: http://id.loc.gov/authorities/names/n83071529





### Data Harvesting: Wikidata

2) If Wikidata id is in LCNAF record, query Wikidata Query Service by Q id and retrieve:

- Name
- Birthdate
- Deathdate
- Place of birth
- Place of death
- Occupation
- Work period start
- Work period end
- Instrument

- MusicBrainz id
- LCNAF id
- Discogs id
- Spotify id
- iTunes id
- AllMusic id
- o ISNI
- VIAF id
- Wikipedia page title (en)

example: https://query.wikidata.org/bigdata/namespace/wdq/sparql with SPARQL query













# Data Harvesting: Wikipedia

3) If the Wikidata item has an English-language Wikipedia page, query the Wikipedia API by page title and retrieve:

- Images
- Description (exintro)

example:

https://en.wikipedia.org/w/api.php?action=query&prop=extracts&exintro=&format=json&titles=Bessie%20Smith





### Data Harvesting: MusicBrainz

4) If MusicBrainz id in Wikidata record, query MusicBrainz artist by id and retrieve:

- Name
- Sort name
- Type (group, person)
- Birthdate
- Deathdate
- Gender
- Area (country)

- LCNAF id
- VIAF id
- Wikidata
- o ISNI
- Related artists

example: http://musicbrainz.org/ws/2/artist/ffa28768-ecda-42c6-ac49-6ce5c7d33043



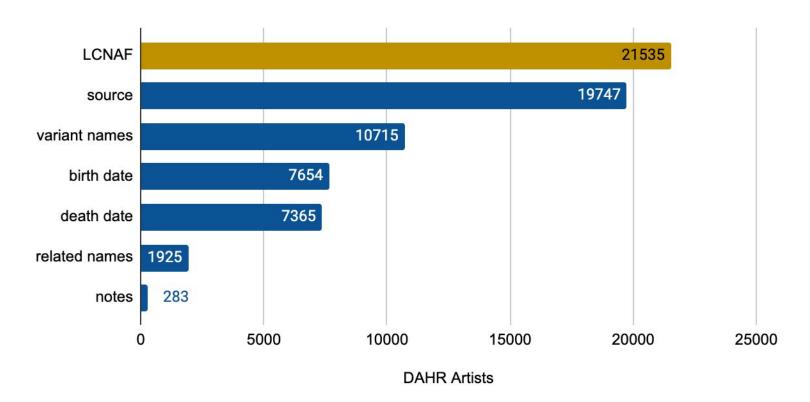






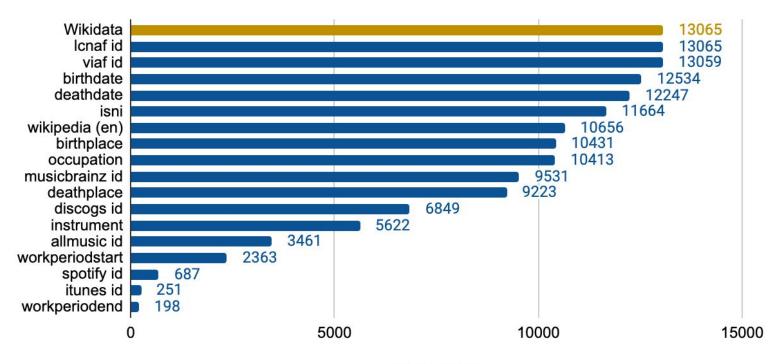
# Number of LCNAF properties populated

for 21,535 DAHR artists with LCNAF records



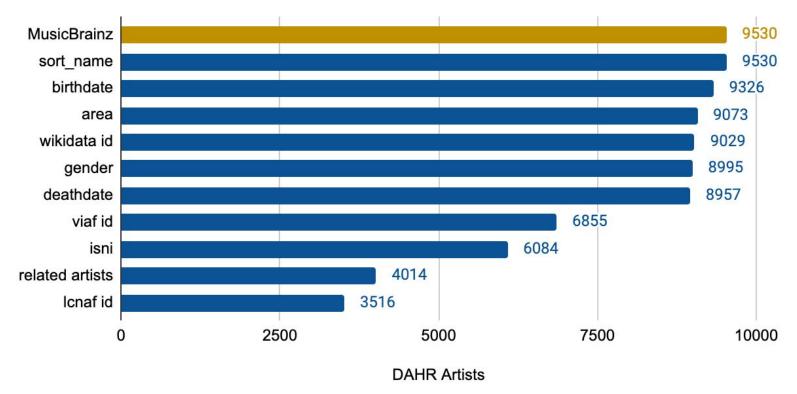
# Number of Wikidata properties populated

for 13,065 DAHR artists with Wikidata records

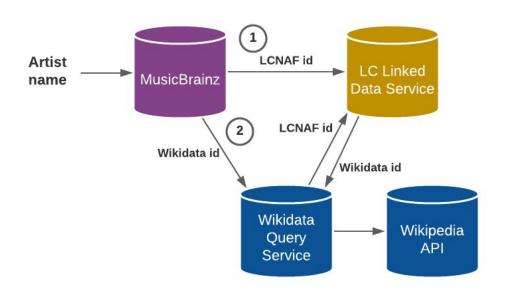


# Number of MusicBrainz properties populated

for 9,530 DAHR artists with MusicBrainz records



# Data Harvesting: Strategy 2 Query MusicBrainz by name



If "talent" record has no LCNAF id, query MusicBrainz artist endpoint by name

- If LCNAF id is in MusicBrainz reocrd, query LC Linked Data Service by id, then Wikidata (if Wikidata id is in LCNAF)
- 2) If Wikidata id is in MusicBrainz record, query Wikidata, then LC Linked Data Service by LCNAF (if LCNAF id is in Wikidata)

example: https://musicbrainz.org/ws/2/artist?query="Bessie%20Smith"







#### About Claris FileMaker



FileMaker is an extremely versatile, user/developer friendly, and comprehensive platform. Although FileMaker was originally a simple database, over the past three decades, it has become a full development platform.

- 24 Million Copies Sold in 15 Languages
- #1 Rapid Development Tool (via www.g2crowd.com)
- Connects to other databases including mySQL, Microsoft SQL Server, Oracle, DB2, PostgreSQL
- 50,000 Developers worldwide
- Wholly owned subsidiary of Apple





#### DAHR Technical overview

- Users at UCSB use a suite of custom tools in FileMaker app to curate data on over a million recordings, objects (like a 78 rpm record), and musical artists.
   FileMaker Server is used to host the app for Windows and macOS users.
- After an extensive "Pre-Flight" check, selected data is available to the public at http://adp.library.ucsb.edu





# FileMaker is "Low Code"

Low code is a development environment used to create application software through graphical user interfaces and configuration instead of traditional hand-coded computer programming.

# FileMaker is "Low Code"

Low code is a development environment used to create application software through graphical user interfaces and configuration instead of traditional hand-coded computer programming.

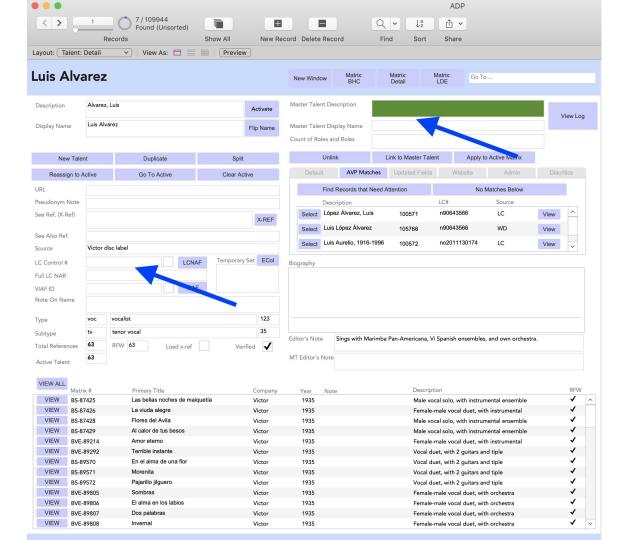


### Human QC and Data Ingest

- 5000+ of possible matches to new LCNAF records harvested from MusicBrainz needed manual verification
- New tool to allow for quick verification and adding LCNAF record number
- Creation of merged "Master Talent" table
- Combination of automated merges and manual merges. Talent roles with matching LCNAF numbers merged automatically, other close matches (e.g. identical names with role of both lyricist and composer merged automatically), others merged manually with new merge tool
- Second harvest on final merged talent after merging and verification of data

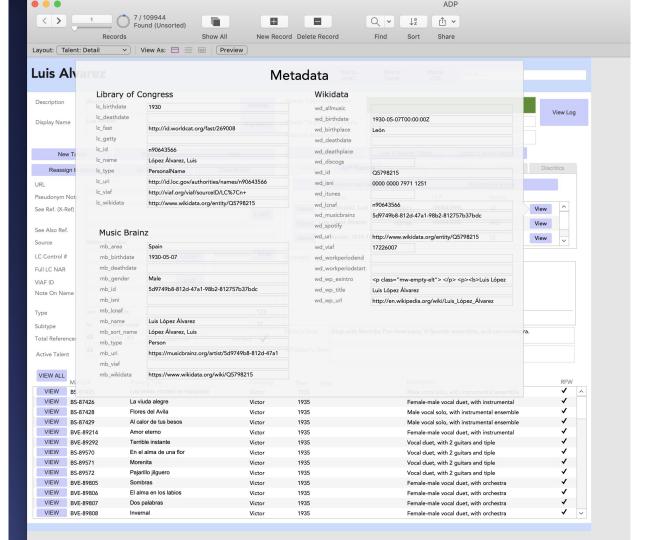


(video version)



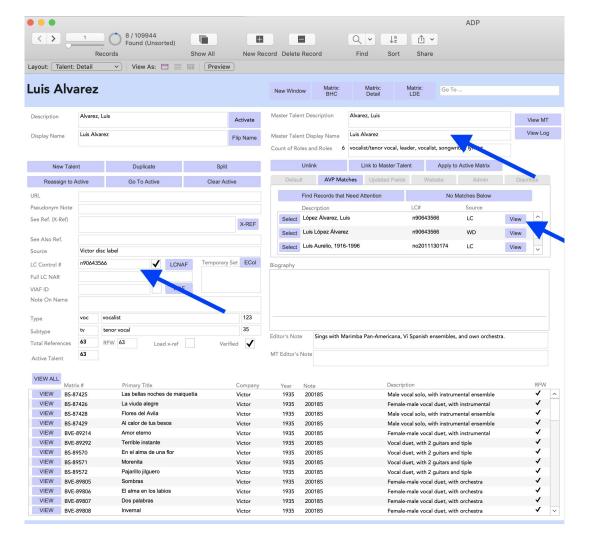
(slide version)

Reviewing AVP Metadata

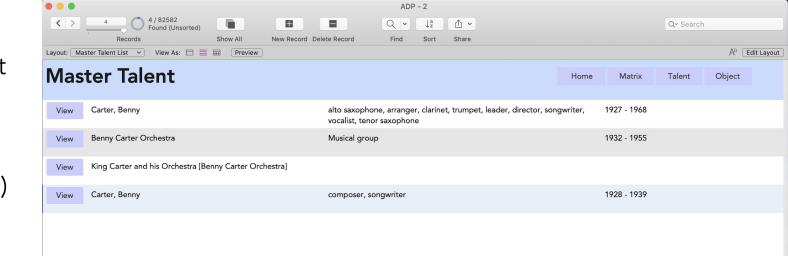


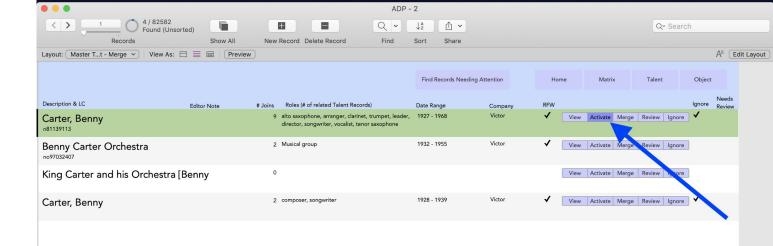
(slide version)

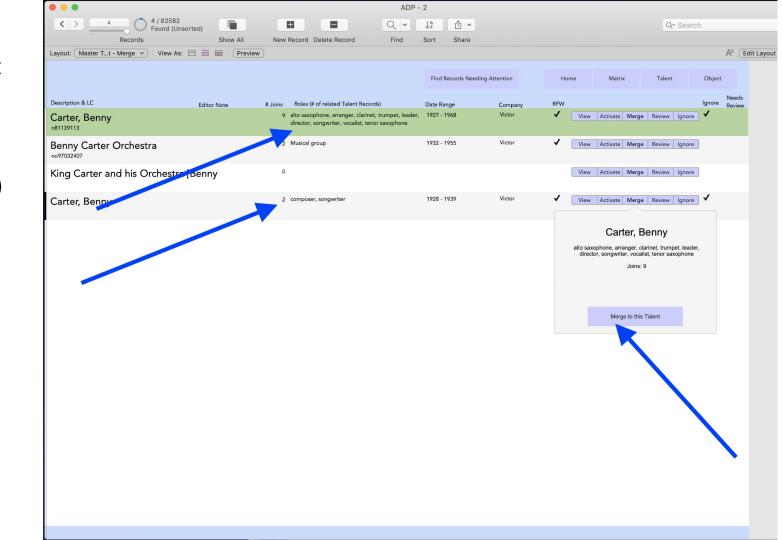
Selecting AVP record



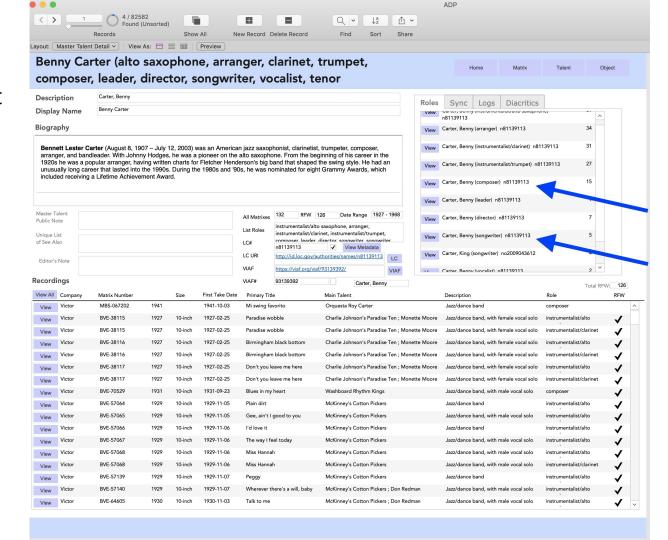
(video version)







#### Merged Master Talent Record In FileMaker



Push to website via Filemaker to MySQL Sync

(video version)

**Facets** 

refine search. Company

Show All

Show All Audio

Show All Issued Show All

Role

Use facets below to limit or

#### Discography Of **American Historical Recordings**

Search Resources **Browse** 

News

About

Push to website via Filemaker to MySQL Sync

(slide version)



Image Source: Wikipedia

#### **Benny Carter**

Bennett Lester Carter (August 8, 1907 - July 12, 2003) was an American jazz saxophonist, clarinetist, trumpeter, composer, arranger, and bandleader. With Johnny Hodges, he was a pioneer on the alto saxophone. From the beginning of his career in the 1920s he was a popular arranger, having written charts for Fletcher Henderson's big band that shaped the swing style. He had an unusually long career that lasted into the 1990s. During the 1980s and '90s, he was nominated for eight Grammy Awards, which included receiving a Lifetime Achievement Award.

Continue reading at Wikipedia...

Wikipedia content provided under the terms of the Creative Commons BY-SA license

Birth and Death Data: Born August 8th, 1907 (Harlem), Died July 12th, 2003 (Los Angeles)

Date Range of DAHR Recordings: 1927 - 1968

Roles Represented in DAHR: alto saxophone, arranger, clarinet, trumpet, composer, leader, director, songwriter, vocalist, tenor saxophone

#### Recordings (Results 1-25 of 126 records)

Company	Matrix No.	Size	First Recording Date	Title	Primary Performer	Description	Role	Audio
Victor	<u>BVE-</u> <u>38115</u>	10-in.	2/25/1927	Paradise wobble	Charlie Johnson's Paradise Ten ; Monette Moore	Jazz/dance band, with female vocal solo	instrumentalist, alto saxophone, clarinet	0
Victor	<u>BVE-</u> <u>38116</u>	10-in.	2/25/1927	Birmingham black bottom	Charlie Johnson's Paradise Ten ; Monette Moore	Jazz/dance band, with female vocal solo	instrumentalist, alto saxophone, clarinet	0
Victor	<u>BVE-</u> 38117	10-in.	2/25/1927	Don't you leave me here	Charlie Johnson's Paradise Ten ; Monette Moore	Jazz/dance band, with female vocal solo	instrumentalist, alto saxophone, clarinet	0
Victor	BVE- 57064	10-in.	11/5/1929	Plain dirt	McKinney's Cotton Pickers	Jazz/dance band	instrumentalist, alto saxophone	0

Push to website via Filemaker to MySQL Sync

(slide version) (Results 1-25 of 126 records)



#### Citation

Chicago Style >

Discography of American Historical Recordings, s.v. "Carter, Benny," accessed November 24, 2020, https://adp.library.ucsb.edu/names/103374.

#### **DAHR Persistent Identifier**

URI: https://adp.library.ucsb.edu/names/103374





#### **External Sources**

W Wikipedia: Benny Carter



Allmusic: Carter, Benny

#### **Linked Open Data Sources**

LCNAR: Carter, Benny - <a href="http://id.loc.gov/authorities/names/n81139113">http://id.loc.gov/authorities/names/n81139113</a>

**Wikidata:** Benny Carter - http://www.wikidata.org/entity/Q356715

VIAF: https://viaf.org/viaf/93139392/

MusicBrainz: Carter, Benny - https://musicbrainz.org/artist/25cf5731-bed9-4604-97a4-8b9655de1a30

#### **Feedback**

Send the Editors a message about this record.

#### UC SANTA BARBARA

#### Library

© 2008-2020 Regents of the University of California, All Rights Reserved. Home | Contact Us | Terms of Use | Accessibility





# Push to website via Filemaker to MySQL Sync

(slide version)

Composer information source: Disc label.

Loren L. Watson (session supervisor)

Personnel Notes McKinney's Cotton Pickers (Musical group) Benny Carter (instrumentalist : alto saxophone) Leonard Davis (instrumentalist: trumpet) Sidney De Paris (instrumentalist: trumpet) Coleman Hawkins (instrumentalist: tenor saxophone) Coleman Hawkins (instrumentalist: clarinet) Claude Jones (instrumentalist: trombone) Kaiser Marshall (instrumentalist : drums) Theodore McCord (instrumentalist: clarinet) Theodore McCord (instrumentalist: tenor saxophone) Don Redman (director) Don Redman (instrumentalist : clarinet) Don Redman (instrumentalist: alto saxophone) Joe Smith (instrumentalist: trumpet) Fats Waller (instrumentalist: piano) Fats Waller (instrumentalist : celeste)

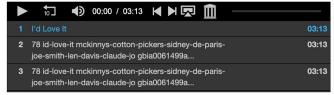
to a reissue compilation CD. Personnel listings here are not intended to imply that every instrument listed is played on every master recorded at the session.

Victor ledgers: "Mr. Watson, present."

Source(s): Frog CD notes (personnel); Rigler-Deutsch Index; Victor blue history card; trade catalog; Victor ledgers.

Take Date and Place	Take	Status	Label Name/Number	Format	Note
11/6/1929 New York, New York	1	Hold	RCA Victor LPV-520	LP	
11/6/1929 New York, New York	2	Master	Victor V-38133	10-in.	
11/6/1929 New York, New York	2	Master	Bluebird B-10706	10-in.	
11/6/1929 New York, New York	2	Master	Gramophone B-4967	10-in.	
11/6/1929 New York, New York	2	Master	Frog DGF 25	Compact disc	
11/6/1929 New York, New York	2	Master	JSP 928	Compact disc	

#### Audio of this recording from the Internet Archive





# Challenges

- Work was done while database was live and continued to be updated by editors
- Initial harvest was in Feb 2020. After COVID lockdown in March, all staff began working remotely. Fortunately, tools were already in place for remote work
- Staff re-assigned from other duties to QC and merging, providing variety and additional work for staff who ordinarily worked with the public or with physical collections





#### **DAHR Talent Statistics**

- 62,571 new merged DAHR Master Talent (personal and corporate) online
- 20,086 have matching LCNAF records
- 20,178 have matching VIAF IDs
- 8,635 have biographical information from Wikipedia
- 8,203 have matching Musicbrainz ID
- 5,881 have matching Discogs ID
- 3,046 have matching Allmusic ID
- 9,836 Names in Wikidata populated with DAHR Artist ID





#### Lacunae

- No outward facing DAHR API
- Real-time query of Wikipedia/Wikidata for bios and images
- Real-time harvesting of Linked Data as new names added to DAHR by editors (300/month)





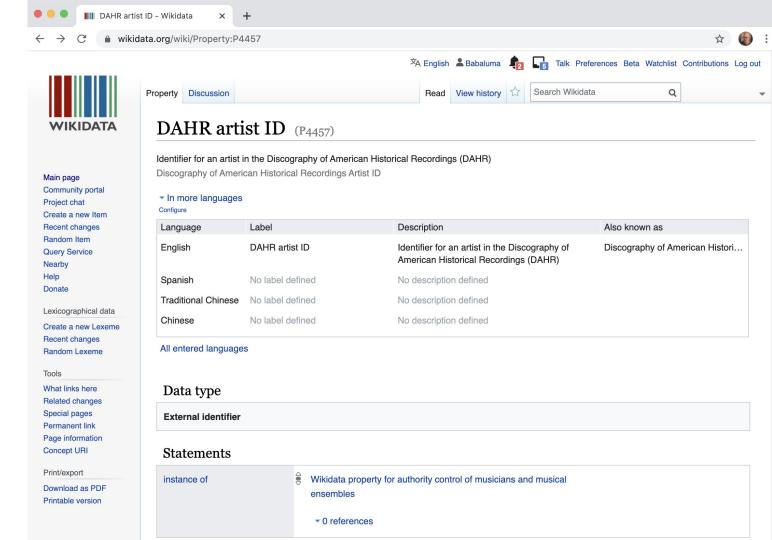
### Next Steps

Added DAHR Artist ID (URI) to Wikidata pages (with QuickStatements v2)





#### Wikidata DAHR Artist ID



### Next Steps (continued)

- Added DAHR URI to Wikidata pages (with QuickStatements v2)
- Do similar harvest from VIAF for names not in LANAF
- Add matching DAHR URI to MusicBrainz, Discogs and other databases
- Add DAHR artists to Wikidata

Suggestions?





#### Questions? Comments?

- David Seubert, DAHR Project Director, UCSB <u>seubert@ucsb.edu</u>
- Shawn Averkamp, Senior Consultant, AVP <a href="mailto:shawn@weareavp.com">shawn@weareavp.com</a>
- Michael Lashutka, Principle, ProperlySorted LLC michael@properlysorted.com



