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
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.
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
- ISNI
- VIAF id
- Wikipedia page title (en)

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

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

- Images
- Description (exintro)

Example:
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
- 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

- LCNAF: 21,535
- Source: 19,747
- Variant names: 10,715
- Birth date: 7,654
- Death date: 7,365
- Related names: 1,925
- Notes: 283
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

1) If LCNAF id is in MusicBrainz record, 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.

(slide version)
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
Link AVP data to Talent Records in FileMaker
(video version)
Link AVP data to Talent Records in FileMaker

(slide version)
Link AVP data to Talent Records in FileMaker

Reviewing AVP Metadata
Link AVP data to Talent Records in FileMaker

Selecting AVP record
Master Talent
Merge In
FileMaker

(video version)
<table>
<thead>
<tr>
<th>View</th>
<th>Name</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carter, Benny</td>
<td></td>
<td>alto saxophone, arranger, clarinet, trumpet, leader,</td>
<td>1927 - 1968</td>
</tr>
<tr>
<td></td>
<td></td>
<td>director, songwriter</td>
<td></td>
</tr>
<tr>
<td>Benny Carter Orchestra</td>
<td></td>
<td>vocalist, tenor saxophone</td>
<td>1932 - 1955</td>
</tr>
<tr>
<td>King Carter and his Orchestra</td>
<td>Benny Carter Orchestra</td>
<td>Musical group</td>
<td>1928 - 1939</td>
</tr>
<tr>
<td>Carter, Benny</td>
<td></td>
<td>composer, songwriter</td>
<td></td>
</tr>
</tbody>
</table>
### Master Talent Merge In FileMaker (slide version)

<table>
<thead>
<tr>
<th>Description &amp; I.C.</th>
<th>Editor Note</th>
<th># Joins</th>
<th>Roles of related Talent Records</th>
<th>Data Range</th>
<th>Company</th>
<th>RW</th>
<th>Needs Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carter, Benny</td>
<td></td>
<td>9</td>
<td>alto saxophone, arranger, clarinet, trumpet, leader, director, songwriter, vocalist, tenor saxophone</td>
<td>1927 - 1968</td>
<td>Victor</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Benny Carter Orchestra</td>
<td></td>
<td>2</td>
<td>Musical group</td>
<td>1932 - 1955</td>
<td>Victor</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>King Carter and his Orchestra [Benny]</td>
<td></td>
<td>0</td>
<td></td>
<td>1928 - 1939</td>
<td>Victor</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Carter, Benny</td>
<td></td>
<td>2</td>
<td>composer, songwriter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description &amp; LC</td>
<td>Editor Note</td>
<td># Joins</td>
<td>Roles (of related Talent Records)</td>
<td>Data Range</td>
<td>Company</td>
<td>RFW</td>
<td>Needs Review</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>---------</td>
<td>-----</td>
<td>-------------</td>
</tr>
<tr>
<td>Carter, Benny</td>
<td></td>
<td>9</td>
<td>alto saxophone, arranger, clarinet, trumpet, leader, director, songwriter, vocalist, tenor saxophone</td>
<td>1927 - 1968</td>
<td>Victor</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Benny Carter Orchestra</td>
<td>no9703407</td>
<td>2</td>
<td>Musical group</td>
<td>1932 - 1955</td>
<td>Victor</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>King Carter and his Orchestra, Benny</td>
<td></td>
<td></td>
<td>composer, songwriter</td>
<td>1928 - 1929</td>
<td>Victor</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Carter, Benny
- alto saxophone, arranger, clarinet, trumpet, leader, director, songwriter, vocalist, tenor saxophone
- Join: 9

Merge to this Talent
Benny Carter (alto saxophone, arranger, clarinet, trumpet, composer, leader, director, songwriter, vocalist, tenor)

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 1930s he was a popular arranger, having written charts for Fletcher Henderson's big band that set the style. He had an unusually long career that lasted into the 1990s. During the 1980s and 90's, he was nominated for eight Grammy Awards, which included receiving a Lifetime Achievement Award.
Push to website via Filemaker to MySQL Sync

(video version)
Push to website via Filemaker to MySQL Sync (slide version)
Push to website via Filemaker to MySQL Sync (slide version)
Push to website via Filemaker to MySQL Sync

<table>
<thead>
<tr>
<th>Personal</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>McKinney's Cotton Pickers (Musical group)</td>
<td></td>
</tr>
<tr>
<td>Benny Carter (instrumentalist: alto saxophone)</td>
<td></td>
</tr>
<tr>
<td>Leonard Davis (instrumentalist: trumpet)</td>
<td></td>
</tr>
<tr>
<td>Solrey De Paris (instrumentalist: trumpet)</td>
<td></td>
</tr>
<tr>
<td>Coleman Hawkins (instrumentalist: tenor saxophone)</td>
<td></td>
</tr>
<tr>
<td>Coleman Hawkins (instrumentalist: clarinet)</td>
<td></td>
</tr>
<tr>
<td>Claude Jones (instrumentalist: trombone)</td>
<td></td>
</tr>
<tr>
<td>Kaiser Marshall (instrumentalist: drums)</td>
<td></td>
</tr>
<tr>
<td>Theodore McCord (instrumentalist: clarinet)</td>
<td></td>
</tr>
<tr>
<td>Theodore McCord (instrumentalist: tenor saxophone)</td>
<td></td>
</tr>
<tr>
<td>Don Redman (director)</td>
<td></td>
</tr>
<tr>
<td>Don Redman (instrumentalist: clarinet)</td>
<td></td>
</tr>
<tr>
<td>Don Redman (instrumentalist: alto saxophone)</td>
<td></td>
</tr>
<tr>
<td>Joe Smith (instrumentalist: trumpeter)</td>
<td></td>
</tr>
<tr>
<td>Fats Waller (instrumentalist: piano)</td>
<td></td>
</tr>
<tr>
<td>Fats Waller (instrumentalist: celeste)</td>
<td></td>
</tr>
<tr>
<td>Loren L. Watson (session supervisor)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Take Date and Place</th>
<th>Take</th>
<th>Status</th>
<th>Label Name/Number</th>
<th>Format</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/9/1929 New York, New York</td>
<td>1</td>
<td>Hold</td>
<td>RCA Victor LPV-520</td>
<td>LP</td>
<td></td>
</tr>
<tr>
<td>11/9/1929 New York, New York</td>
<td>2</td>
<td>Master</td>
<td>Frog DG-25</td>
<td>Compact disc</td>
<td></td>
</tr>
<tr>
<td>11/9/1929 New York, New York</td>
<td>2</td>
<td>Master</td>
<td>JBP 928</td>
<td>Compact disc</td>
<td></td>
</tr>
</tbody>
</table>

Audio of this recording from the Internet Archive

1. I'd Love It
   03:13
2. 78 id-love-it mckinny's-cotton-pickers-sidney-de-paris-joe-smith-len-davis-claude-je gbia0061498a...
   02:13
3. 78 id-love-it mckinny's-cotton-pickers-sidney-de-paris-joe-smith-len-davis-claude-je gbia0061498a...
   03:13
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)
DAHR artist ID (P4457)

Identifier for an artist in the Discography of American Historical Recordings (DAHR)

Discography of American Historical Recordings Artist ID

Data type

External identifier

Statements

Instance of

Wikidata property for authority control of musicians and musical ensembles

0 references
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 seubert@ucsb.edu
- Shawn Averkamp, Senior Consultant, AVP shawn@weareavp.com
- Michael Lashutka, Principle, ProperlySorted LLC michael@properlysorted.com