
This paper describes the development of a database that helps users locate archival holdings of early country music 78 rpm records. The database includes recordings held at the Southern Folklife Collection, the Library of Congress, and the Country Music Hall of Fame. The database is a prototype for a more extensive project to follow. The recordings it includes are listed in Guthrie Meade’s Country Music Sources: A Biblio-Discography of Commercially Recorded Traditional Music.

Headings:

- Audio-visual archives – United States.
- Collectors and collecting.
- Popular music.
- Sound archives.
- Sound archives – United States.
- Sound recording libraries.
- Sound recordings.
THE CREATION OF A DATABASE FOR LOCATING ARCHIVAL HOLDINGS OF EARLY COUNTRY MUSIC 78 RPM RECORDS.

by

Carrie Boone

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina

July 2010

Approved by
I. Background and Research Problem

In 2002, the Southern Folklife Collection and The John Edwards Memorial Forum published a new kind of discography, entitled *Country Music Sources: A Bibliodiscography of Commercially Recorded Traditional Music*. In this critically lauded work, Dick Spottswood and Douglas S. Meade complete the life’s work of Guthrie T. Meade: “to identify, organize, and annotate the traditional music recorded by white country musicians for commercial consumption, in the period from 1921 to 1942 (Meade).”

Traditionally, discographies list recordings by song title or artist name, and provide information such as the record label\(^1\), matrix number\(^2\), and recording date\(^3\) associated with each recording. Meade’s work goes beyond this, and categorizes discreet recordings into broader categories, each of which contains a number of recordings grouped together by a kind of uniform title Meade terms “song”. Meade’s conceptualization of “song” as compared to “recording” is very similar to the recently popular FRBR concept of “work” as compared to “expression” (Taylor, *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools*). The level of musical analysis and breadth of musical

---

1. A record label is similar to a publisher; it is the company that produces and issues a record. It is not to be confused with the physical label which is affixed to the center of the record, although this physical label usually bears the name of the issuing record label. Examples of these include Columbia, Bluebird, OKe\(h\), etc.

2. At a recording session, each song which was recorded was assigned a unique number; this number is the matrix number. Multiple issues of the recording may have different issue numbers or be released on different record labels, but the matrix number serves to identify each of these as manifestations of a single, original recording.

3. The date the recording was made of a live performance. A recording might not mass produced and released for sale for weeks, months, or even years after it was made.
knowledge required to do this is what sets *Country Music Sources* apart from other
discographies, and makes it so valuable as a research tool.

While *Country Music Sources* is comprehensive in its attempt to categorize and
organize early country music recordings, actually locating extant, publicly available
copies of those recordings is beyond its scope. Once a researcher has used the
text to identify the specific recording he or she needs, no guidance is offered as to how to
find that recording. In his introduction to the discography, Meade notes that he drew on
the resources of private collectors and public archives; from the latter, he used materials
from the Library of Congress (both the Music Division and the Archive of Folk Song and
Culture), the University of Virginia, and the Harvard Library (Meade 2002). However,
many of these materials consisted of paper-based musical materials such as sheet music\(^4\)
and journals\(^5\). Much of the information about the actual recordings (record label, matrix
number, etc.) came from lists compiled by record companies, private collectors, or other
researchers. It is also worth noting that Meade died in 1991; the discography was
completed by others using the notes he left. A great many recordings have been acquired
and processed in archives – both the archives Meade used and in others - since 1991.
And while many archives maintain online or paper-based finding aids to their holdings,
there is not currently a single portal through which a researcher can search the holdings
of multiple archives for a particular title. From my experience working at one music
archive, I have discovered that one of the most frequently asked questions (either via

\(^4\) A commercial publication, usually less than 10 pages long, featuring musical notation and lyrics for
popular songs.

\(^5\) Specialty print journals sometimes included the musical notation and lyrics for these songs, in either
partial or complete versions.
phone or email) of the archivist is “Do you have recording X?” If the answer is “no”, then the inevitable follow-up question is “Do you know who does?” Sadly, the answer to this is usually “no” as well.

In order to address these issues, I have constructed a database that will allow researchers to quickly locate archival holdings of recordings listed in Meade’s discography. Because of the immense scope of the book, I have limited my focus for the purposes of this project to one section of the book: “American Ballads”. For each recording listed in this section, the database indicates which archives (if any) hold copies of the recording. Additionally, the database contains contact and usage information pertinent to each archive it includes. For its first iteration, this project is limited to a basic database with a single authorized person (me) who may update its entries. However, later iterations of this project should ideally allow an authorized person at each archive it includes to update its entries as more recordings are acquired and processed, thereby creating a living, responsive, and – above all – useful tool for country music researchers.
II. Currently Available Tools

Union Catalogs

The tool this database is most similar to in function is a union catalog. Libraries have long used union catalogs as a way to cooperatively list holdings beyond those available at any given institution. However, several differences between typical library holdings and archival 78 rpm record holdings have prevented the creation of a true union catalog of these holdings. First, the role of the Library Congress as a single, authoritative voice in cataloging in the United States greatly assisted with creating the level of bibliographic standardization needed for union cataloging. Even during the era of paper-based catalog cards, copy cataloging made relatively efficient the work of identifying and sorting out duplicates from large numbers of materials held by different libraries. However archives lack a central, unifying authority such as the Library of Congress, and it was much later, historically, that concerns were raised over standardizing the way their holdings are described and cataloged.

Second, the nature of items held by libraries and archives differ so greatly that the basic approach to describing and cataloging each type of institution takes varies greatly as well. Libraries’ approach to bibliographic description developed largely to handle printed materials individually. Archives needed a way to describe highly heterogeneous items – not individually – but as entities grouped by provenance. As a result, archival finding aids tend to be both broader and more idiosyncratic than their library catalog
counterparts, further complicating the development of any kind of a union catalog or union finding aid.

WorldCat

As libraries transitioned into electronic formats such as MARC, the use of description methods prioritizing printed materials continued. While MARC records are very well suited to describing books, they are not at all suited to describing 78 rpm records (some interesting developments at NUCMC notwithstanding). For example, the “main entry” field of a MARC record is relatively straightforward for a book: this can be an author, authors, or editor. But who would be the main entry for a 78 rpm record? The recording artist or artists, the composer, or the arranger? Some of these could go in the “added entry” field, but who should get the primary credit for the recording? What about the “title” field? A typical 78 rpm record would have at least two titles, one for each side. Which one is the title proper? The “edition” field is even more problematic. 78 rpm recordings were usually assigned a matrix number; this number, theoretically, is a unique identifier for that particular recording. But different record labels could (and did) issue that recording multiple times, assigning a different issue number in each case. Using the matrix number in the “edition” field would not provide enough information to distinguish one recording from another, but using the issue number could lead to unnecessary confusion and a high level of redundancy in the catalog. And, of course, all of these issues are compounded by the fact that 78 rpm recordings are two-sided, resulting in a single physical item that is really multiple intellectual items.
The largest and most widely-used collaborative bibliographic tool currently in use, WorldCat, is based on MARC records. This is not to say that WorldCat does not include 78 rpm recordings; on the contrary, WorldCat can be useful in locating some of these items. However, the number of 78 rpm records which can be located in WorldCat is extremely limited compared to the number of records which actually exist in archives. A representative example of this can be seen by searching WorldCat for the keywords “Bradley Kincaid” (a prolific recording artist during the 78 rpm era) and limiting the results to the years 1920-1940 (the heyday of 78 rpm record production). This search returns six results, every one of which is held by a single library, MTSU’s Center for Popular Music. A screen shot of one of these results shows the creative ways in which the cataloger tailored the available MARC fields to better suit 78 rpm records:

These limited results are typical for a WorldCat search for 78 rpm records. After conducting several searches by song title and artist name, even the most successful search – for Jimmie Rodgers’ “Blue Yodel” -- only returned holdings in six libraries worldwide.
Obviously, WorldCat can be helpful in locating some 78 rpm records, but the limited number of MARC records for these items, due to both MARC’s inappropriateness for these items and to archives’ traditional lack of use of MARC records, severely hinders its usefulness.

Archival Finding Aids

As more and more archives begin to convert their finding aids into electronic formats, searching for items using online finding aids is becoming a feasible research strategy. However, there are several limitations to this strategy. First, it is both tedious and time-consuming to search every archive’s finding aids individually. Second, the lack of standardization among archives, while improving, is still a hindrance to the researcher who must learn to navigate a different system for each archive. Third, many archives still do not have online finding aids for some or all of their holdings. And finally, even those archives which do maintain clear, online finding aids often have only a small percentage of their holdings cataloged and entered into a searchable online interface.

Other Tools

A few other tools exist that may be of some usefulness in locating 78 rpm records. One of these, OAIster, functions as a union catalog of digital items from archives. While searching for items in OAIster can be somewhat confusing, the potential it offers is unique among other tools: it can lead the researcher to digitally-recorded streaming audio of the 78 rpm records included in its database. For example, an author search for “Vernon Dalhart” leads to this page:
which is actually part of another useful tool for finding digitized copies of 78 rpm recordings: the Internet Archive. The Internet Archive (www.archive.org) contains a significant collection of digitized 78 rpm and cylinder recordings which may be browsed by title, artist, or subject/keywords (http://www.archive.org/details/78 rpm).

Both OAIster and the Internet Archive can be invaluable for locating the digitized audio content of 78 rpm records; however, neither can satisfy a researcher’s need for locating the original physical object from which the sound is taken.
III. Audience

In building this database, it was useful to focus on its intended audience. This database should serve the needs of two groups: researchers and public service librarians/archivists. In order to better assess the needs of these user groups, I interviewed Jamie Vermillion, who acts as the public service librarian for the Southern Folklife Collection. A partial transcript of the interview appears below:

Q: Do you receive inquiries (by phone or email) from patrons about whether or not you have a particular recording? If so, how frequently? Or are most patrons able to determine this from your web-based finding aids?

A: Yes, mostly email, and relatively frequently. These come mainly from two types of patrons: those who have been calling/emailing around to various archives looking for a particular recording (often made by a relative) and have been given our contact info by another institution, and frequent users of the SFC who know we have a large collection of uncataloged recordings. We probably get two or three of these types of requests ("do you have this recording?") a month, and some of our more familiar researchers will often send a list of recordings they are searching for.

Q: If you do not have a particular recording, do patrons ask you (via phone or email) if you know what other libraries/archives do have it? If so, how frequently?

A: If they don't ask I will usually try to be helpful and volunteer the information. I always refer them to the LOC (though usually with the assumption that they have already asked there), as well as to more specialized institutions depending on what they are looking for (Country Music Hall of Fame, Center for Popular Music at MTSU, Jazz Archives at Tulane, etc).

Q: Are you generally able to answer this question (the "where else can I find it" question)?

A: Almost never. Most often I can only suggest where to look. If I can find it in a WorldCat or Google search, I let them know where it is, but
odds are they already tried that (or the first library they contacted tried it) before they contacted us.

**Q:** What tools do you currently use to answer questions about the location/availability of particular recordings?

**A:** I can't spend a lot of time looking for recordings once I have established we don't have it here. A quick Google search and a WorldCat search (just in case they didn't already try that) and if nothing comes up I let the patron know that they should try one of the above institutions. The few times I have tried contacting record companies have gotten me nowhere fast, so I don't even suggest it anymore.

**Q:** How satisfied are you with these tools?

**A:** Well, WorldCat + Google should *theoretically* get you most cataloged material, but trolling through the Google search results trying to find something that looks like a finding aid or library database entry is incredibly inefficient (especially for material that has been published multiple times, or has a common title) and WorldCat is no good for 78 or 45 rpm records (most singles aren't MARC cataloged and aren't in there).

**Q:** Is there a specific information need (or needs) that, if met, would enable you to more efficiently and accurately field these sorts of inquiries?

**A:** If there was something like a Google search that limited for library material, that would help. If everybody cataloged everything in MARC and uploaded it to WorldCat, that would be awesome. Neither of these things are going to happen.

**Q:** In your experience, would patrons who are looking for a particular recording be satisfied with being able to hear a dub or a digitized version of that recording?

**A:** It depends on the needs of the patron: if it's the only recording that can be located of their grandaddy's cajun band, a dub is fine. If a record company is looking to put out a CD of 78 reissues, they are usually looking for the copy in the best available condition or even a specific matrix number. In those cases they are also interested in scans of the 78 label as well, so a dub wouldn't suffice.
From this interview, along with similar reports from various archival user studies published in the past 5-10 years, I determined that the database should fill the following information needs:

- It should be accessible to both researchers and librarian/archivists
- It should provide holdings information for multiple archives, especially those archives which are most likely to have early country music recordings
- It should provide contact information about the archives whose holdings it describes
- It should indicate whether the item is an original recording or a dub \(^6\)
- It should include the names of all personnel who are represented in the recording (for the benefit of people searching for recordings by family members, many of whom may not be the primary artist listed on the recording)
- It should be built in such a way as to allow for significant growth as uncataloged holdings are added to each archive’s collection of cataloged holdings

The main benefit for researchers of such a database would be the ease and convenience of searching a single tool, rather than finding and learning to use multiple tools scattered across a variety of locations. The main benefits to librarians/archivists would be the amount of time saved in fielding individual patron requests, as well as the improved level of service that would result from being able to confidently refer patrons to other institutions that can meet their needs.

\(^6\) A duplicated copy, usually made by someone other than the original artist or recording engineer. Usually in a different format than the original recording.
Some added benefits would include being able to search for recordings by entry points beyond artist and song title. For example, a researcher could find all items recorded in a certain year, or in a particular location, or all items released by a particular combination of record labels. The inclusion of Meade’s uniform titles for songs will also allow a researcher to locate all variations of a particular song, whether or not their actual titles are similar.
IV. Methodology

The Plan

The basic plan for this project was very simple. First, conduct a needs assessment to determine if this database would be a useful contribution to archives and their users. Second, determine what information the database should contain. Third, select a database management system to build the database. Fourth, populate the database with data. And finally, select a way to share the database with its intended user groups.

I based my needs assessment on five sources: informal interviews with researchers, archivists, librarians, and students; the more formal interview with an archivist referenced above; a survey of archival user studies from Dr. Helen Tibbo’s “Archival Access” class (Spring 2009); a query posted on the ARSC listserv, and my own experiences as a student worker at the Southern Folklife Collection. The general consensus of all of these was that there is a need for a simple-to-use, unified tool for locating audio recordings, particularly early and hard-to-find recordings such as 78s.

The choice to include or exclude information from the database was based on information from these interviews, combined with the information available in Meade’s book. Two types of data are included: descriptive and holdings. For the descriptive data, I chose to include any information that I thought might be a potential search/entry point for users, including artist(s) names, song titles, matrix numbers, record label and issue
numbers, recording date and location, and recording release date. To these I added two categories unique to Meade’s classification system: uniform title and broad category. I included these not because they are currently widely used by researchers, but because of their innate usefulness in parsing out what will certainly grow to be massive numbers of recordings. For the holdings data, I included the name, address, and contact information for each archive included in the database, as well as whether the archive is public or private. Additionally, I included the local call number for each item held by each institution, in the hopes that this would facilitate users’ ability to locate and inquire about these items.

The process of selecting a database management system (DBMS) was influenced by my desire to make the database as accessible as possible for as many users as possible. While Microsoft Access is certainly not the best available DBMS, it is one of the most ubiquitous, thus ensuring that most users will be able to use the database without needing to install extra programs on their computers. Additionally, Microsoft Access has a great deal of supporting literature – much of it free, enabling novice and intermediate users to interact with the data with a minimum of instruction. Finally, the fact that Microsoft Access is so widely used may help offset some concerns about the preservation of the database in the future; it is more likely to be available (albeit in an updated version) in the future than a less-widely used program, thereby enabling the content to be viewed by future users.
In order to add content to the database, I planned to use Meade’s discography in
conjunction with the online catalogs of the relevant archives. If possible, I planned to
take one or more short trips to the archives to make use of any paper-based cataloging
systems. This plan was subsequently modified.

As for sharing the database with its intended users, it is currently located on the
servers of the School of Information and Library Science at the University of North
Carolina at Chapel Hill. At a later date, I hope to find a host such as ibiblio in order to
improve access to the database while maintaining its integrity.

The Process

As I began work on this project, I quickly realized that I would need to modify
some aspects of my plan. I originally intended to include four archives in the database:
the Southern Folklife Collection, the Library of Congress, the Country Music Hall of
Fame, and the Center for Popular Music at Middle Tennessee State University. These
would serve as a sort of test group for the database; the small number would be
manageable while still providing enough recordings to have a significant amount of data
to organize and sort.

The Southern Folklife Collection and the Library of Congress each have online
databases of their cataloged 78 rpm records. However, both archives only have a small

---

7 A digital library and archive project, publicly available and open source. Currently curated and hosted by
the School of Information and Library Science and the School of Journalism at the University of North
Carolina at Chapel Hill.
percentage of their actual holdings cataloged; the rest are waiting in a significant backlog. I hoped to use the internal databases of each archive to augment the number of recordings I could enter into the database; however, this turned out not to be feasible.

The Center for Popular Music at MTSU has entered many of its recordings into WorldCat, but does not have an online catalog of its own recordings. Their staff will conduct a search for researchers, but this service is provided in exchange for a hefty fee. Rather than pay the fee or search for these holdings through WorldCat, each of which would have been prohibitively expensive in either money or time, I elected to delay adding this archive’s holdings until a later time.

The Country Music Hall of Fame has neither an online catalog nor any presence in WorldCat. In fact, there are only two ways to find out what recordings they have: through a trip to the archive or through the extremely helpful auspices of Becky Miley, the Associate Librarian at this repository. Ms. Miley’s assistance was invaluable to this project; through a series of emails, I have been able to track down about two hundred recordings held at the CHMF so far.

Once I narrowed down the scope of the project, I began to set up the database itself. After creating all of the tables and their relationships, the process of entering information into the database followed a fairly regular protocol:

1. Enter all pertinent descriptive information about a recording from Meade’s book. This includes: matrix number, record label, issue number, song title, main artist,
additional artists, place and date of recording, issue date, Meade’s assigned category, and Meade’s assigned uniform title.

2. Double-check the label and issue information against an outside source. This step was time-consuming, but necessary to ensure that the database was as free from errors as possible. The two sources that I used most frequently were Tony Russell’s *Country Music Records: A Discography, 1921-1942* and the Online Discographical Project (www.78discography.com). This step ensured that I found all available holdings within the scope of the project; before I began doing this, I was missing some holdings due to typos or other errors in Meade’s book.

3. Search for archival holdings of each recording. For the SFC and the LOC, this mostly entailed searching their online databases. For the CMHF, this involved creating an Excel spreadsheet of each song title, along with its label and issue information, and sending it to Ms. Miley. Ms. Miley would, in turn, send the spreadsheet back after filling in whether or not the CMHF had a copy of each record, along with its call number (if any).

Once the database was populated with a workable number of recordings, I added some queries that may be useful to users. Using these queries, a researcher can quickly find the following information: a list of all recordings (listed in the database) held by each respective archive, the location (if any) of a particular recording by title and artist, or the location (if any) of a particular recording by record label and issue number.

Finally, as a way to back up the database and as a way to share it with others, I uploaded a copy of it to a free online hosting site, www.wiggio.com. From here, any user
may download the database onto any computer, but the database will remain protected from any changes made to it.
V. Recommendations for Maintenance and Future Improvements

This project has always been intended as a pilot project for a much more comprehensive database. As such, it is currently limited in its scope and function. As it continues to grow, I hope to incorporate the following improvements:

- Adding more archives. A partial list of these, based on the significance of their 78 rpm record holdings, includes the Center for Popular Music at MTSU, the Performing Arts Collection at UCLA-Santa Barbara, the Archive of Recorded Sound at Stanford University, and any other archive wishing to be included.

- Moving to an open-source format. In order for this database to remain as current and comprehensive as possible, it must become the work of more than one person. In an open-source format, hosted online through an entity such as ibiblio, authorized individuals could update the database on a regular basis. Ideally, these individuals would be staff members at the included archives, who would update the database to reflect their newly acquired or newly cataloged holdings.

- Linking out to digitized versions, where applicable. Although this is partially accomplished by tools such as OAIster and the Internet Archive, the ability to link to a digitized recording of a particular 78 rpm record would be a convenient feature for those researchers who simply wish to hear a particular recording. This would be an added feature of the database rather than its central purpose. However, it would fit in with its goal of making locating these recordings as convenient as possible.
Bibliography

Abrams, Steven. The Online Discographical Project.


Vermillion, Jamie N. Southern Folklife Collection Assistant Carrie Boone. 7 April 2010.