Recorded Sound and Discovery Interfaces

Nara L. Newcomer,
Head, Music/Media Library, University of Missouri-Kansas City
ARSC 2013 Pre-Conference Workshop:
“Discovery Access Methods for Sound Recording Collections”
May 15, 2013 | newcomern@umkc.edu
First, the roadmap. Our first section is an introduction to the topic, and exploration of some of the basics behind recorded sound: what our users need, where we are, and how we got there.

The second half is the meat of the presentation. We’ll examine, one-by-one, attributes important to the discovery of recorded sound: what users need, challenges to meeting their needs, and possible solutions.

Also, a bit about myself here:

Prior to joining UMKC in January 2013, I was the Assistant Music Librarian – and music cataloger - at East Carolina University for 7 ½ years. So I come at this as a cataloger and a public services librarian. I also have a bias towards commercial recordings, Western Art music, and AACR2/MARC data, simply because I know these areas best. I’m really interested to hear from those here with expertise in other areas, and how your expectations and needs differ.

I “got into” discovery by being involved with selection, implementation, and/or configuration of 3 different discovery interfaces at ECU, specifically: SirsiDynix Symphony e-library; OCLC WorldCat Local, and Serials Solutions Summon. Throughout, I saw that music, including recorded sound, was poorly served. Soon, I found myself leading a the Music Library Association-endorsed group which created the Music Discovery Requirements.

in ARSC Pre-Conference Workshop: Discovery Access Methods for Sound Recording Collections
In this section, we’ll look at cataloging’s historical bias towards the book, start thinking about what attributes are important to recorded sound, examine the “puzzle pieces” for successful discovery, and learn about the history, purpose, and organization of the Music Discovery Requirements. We’ll also have a quick review of important points of FRBR as it relates to recorded sound.
In this section, we’ll look at cataloging’s historical bias towards the book, start thinking about what attributes are important to recorded sound, examine the “puzzle pieces” for successful discovery, learn about the history, purpose, and organization of the Music Discovery Requirements, and briefly review the important points of FRBR group I entities as they related to recorded sound.
To better understand the current discovery environment for recorded sound, let’s take a brief sojourn into discovery & cataloging history.

Library cataloging has long been biased towards the book, reflecting the preeminence of written and spoken language in our society.

Indeed, we can go all the way back to Cutter’s 3 access points (1876): author, title, subject.
Let’s consider these three access points, using a well-known musical work I’ll call for now “Beethoven’s 9th Symphony.”

The author, Beethoven, is clear, and, indeed, personal name is frequently used as an access point for music. However, due to prolific composers and the multiple expressions and manifestations libraries regularly hold of each musical work, personal name alone is an insufficiently precise access point.

There are many options for the title, thus the development of uniform titles which in actuality frequently combine many attributes like form, medium, work numbers, etc.

Subject is a problematic concept for music, because many musical works are arguably not “about” anything. Indeed, music “subject headings” often reflect attributes other than subject. The 9th Symphony probably is about something, but that’s an interesting topic for musicological debate, is not a good access point.

So, we see that author, title, and subject alone, though they may work fairly well for books, are insufficient access points for music.
So, we've seen that author, title, and subject aren’t often enough when finding musical works. A skilled reference librarian will draw the identifying attributes out of a patron through a reference interview. My colleague Wendy Sistrunk and I will enact two reference interviews for you. As you’re watching, jot down what attributes – features and characteristics – the reference interview draws out of the patron. We’ll talk about them after the skit.
On this slide, type in the attributes identified by attendees.
What we’ve observed anecdotally is also backed up by research. In 2007, David King summarized and synthesized the research to that date regarding user search strategies for printed music and recordings, concluding that music searching differs in specific ways from general searching, particularly in that known-item searching is more common for music materials and personal name is used more frequently as an access point. Similarly, a study of music reference service found that, in comparison to general reference, music questions "focus much more heavily on locating specific, known items" and were answered via different means, with music reference employees relying on "their local catalogs, the circulating collection and their own knowledge much more frequently than librarians in general reference situations."


Jin Ha Lee’s study is particularly interesting because it goes beyond Western Art Music. Lee harvested queries from Google Answers’ music category in 2005. Google answers was a service, now discontinued, where people could post their queries and then a “Google researcher” tried to supply an answer. These questions tended to be about “popular” music. Example: “I heard this song by a female singer in an ARBY’s. I believe it is from the 70s or early 80s. The main chorus of the song says, “over and over again.” Kind of a sad, slow, easy listening love song.” Lee then broke these apart into the attributes desired, similar to how we just did in our “reference interview” exercise. She found that, though “known item” queries were common, users often didn’t “know” a lot about the item sought and especially lacked commonly recorded bibliographic information.
For a successful discovery interface, three major pieces of the puzzle must work together: formulating data, encoding the data thus formulated, and indexing and displaying that formulated, encoded data. A problem or breakdown in any piece can cause the system to fail, creating difficulties for users in finding materials and information.
When a puzzle piece breaks down, users won’t know which one it is. They’ll just know it doesn’t work. But, as librarians, it’s our job to understand the pieces, pinpoint the problem, and fix it.

No score for the *Surprise Symphony*? And the only recordings we have are on LP? What’s wrong with this library?!?
To help understand the puzzle pieces, I’ve listed some of the standards we currently use for each.

AACR2, and now RDA, provide the rules for formulating data—determining what bits of information to record down and how to express them.

Once that data is formulated we encode it in “metadata format” like MARC. The metadata formats listed here are the ones for which we provided mappings in the Music Discovery Requirements.

Finally, we present the data in a user interface. Since the card catalog died, this has been largely a free-for-all without a recognized standard. Furthermore, interface designers regularly operate from a book-centric mindset, leaving music poorly served.

The Music Discovery Requirements’ purpose is to provide the final piece of this puzzle.

A bit of history on the MDR...
Due to today’s rapidly changing bibliographic landscape, and the urgent need for a document to aid in discovery implementations, the Music Discovery Requirements is not a standard but a set of “recommendations and possible best practices.” While the work was in many ways precipitated by the onslaught of faceted and web-scale discovery tools, and strongly informed by AACR2, RDA, and MARC, the MDR can be used with any recorded and encoded data.

The MDR was created under the auspices of MLA’s Emerging Technologies and Services Committee and bears the official endorsement of MLA’s Board of Directors.
It is available on MLA's web site and published in the March 2013 issue of MLA’s journal, *Notes.*
The MDR’s target audience moves beyond music specialists to include those people who often design and control the systems we all use: vendors, programmers, system administrators, etc.
The MDR focuses on musical works, and their expressions and manifestations in scores and recordings, rather than secondary literature about music, because musical works present the most unique discovery needs.
The MDR consists of a main document plus 3 appendixes. The appendixes are designed to be practical: three spreadsheets containing mapping guidelines and content/carrier indexing which came out of the committee members’ own experiences and frustrations creating similar documents for our institutions’ discovery implementations. In fact, if you don’t think your vendor/system administrator/developers will read the whole document and appendixes – point them to the appendixes.
The MDR, like RDA, is based on FRBR, with section II focusing on attributes & relationships for musical works, and section III on attributes and relationships for expressions and manifestations.
Much of what I’ll cover today is based on FRBR, especially FRBR’s Group I entities, so here’s a quick review. There are four group 1 entities, with a basically hierarchical relationship. From the top down: work, expression, manifestation, item.
A concrete example is helpful in understanding WEMI. Thankfully, recorded sound, especially Western art music, provides one of the clearest examples of the work-expression-manifestation-item structure.

[Talk through WEMI tree on slide]
So, we've explored broadly the attributes of recorded sound. We have established that recorded sound has many attributes which are important for discovery, but current systems do a poor job handling these important attributes. How can we make the situation better? Here's where it gets nitty-gritty. We will look at each of 11 attributes, one-by-one.
For each attribute, I’ll reference the related section(s) in the Music Discovery Requirements, so you can refer there for more details. This presentation is not a spoken version of the MDR. Rather, I am going to highlight and expound on selected important issues. I am not covering these in the same order as the MDR, and I am not covering every attribute, particularly as a few are specific to printed music. I’m also combining some subsections. If you are actually configuring a discovery system, I urge you to read and consult the MDR for complete, detailed recommendations, which are much better suited to written documentation than to a spoken presentation.
Titles (MDR II.B)
Book titles are pretty straightforward. They are a pretty good way to identify works.
For sound recordings, this isn’t the case.
1. Here’s one formulation of this title, for a musical work
2. This is an album title [you could argue this isn’t a bibliographically significant title]
3. This is an album title that actually isn’t an album title… (actual 245 is “The Beatles”) [but this one is an album title also, and you could argue that it is bibliographically significant]
4. Or, supplied titles in oral history #76821375
Now we’re going to take a trip into music uniform titles, especially generic uniform titles. This is less applicable to distinctively titled works (Appalachian Spring), a category which includes most “popular” and jazz titles. These attributes are not the title of the work. Nonetheless, people often use these attributes to refer to musical works in writing and in speech, especially formal writing and speech. Librarians took these elements and controlled the vocabulary and the order of elements, creating uniform titles.
Just as in regular speech, the composer is crucial (people talk about “Beethoven’s 5th Symphony, not “the” 5th Symphony), so it is in uniform titles. In these three examples, the title portion of the uniform title is the same. Only the author serves to distinguish.

In MARC-speak, the 100 + 240 and the entire 700 must display and function together.
This example, from Stanford’s Blacklight instance, shows the creator and title functioning together very effectively.
Identifying Numbers (MDR II.C)
Identifying numbers as a work-level attribute are most common in Western Art Music – that is, opus, thematic index, and serial numbers like BWV 565 or “no. 4” or “op. 25”.

Identifying numbers as expression-level attributes are very important to the study of recorded sound, especially jazz and popular music, i.e. matrix numbers. We’ll talk about these later.

Many discovery systems allow for searching these numbers as part of general keyword indexes, but not in their own separate indexes. This can result in many false hits, especially when the numbers are low, as the keyword search will pick up timings, publication years, etc.
The Digital Mozart Edition, the free online version of the Neue Mozart Ausgabe, nicely includes searching by Kochel numbers.

It even provides cross-referencing between the different versions of the Kochel catalog.

Obviously, this is much easier in an interface designed solely to handle the works of one particular composer.

[Note for expression-level numbers-a good example might be the encyclopedic victor discography or another online recording site, maybe from American Memory, or even a print discography]
Numbers can also be useful at the expression and manifestation levels. For example, the Encyclopedic Discography of Victor Recordings allows searching by catalog number...
...and browsing by matrix number. Similar searching capability is common in printed discographies. These numbers are invaluable to recorded sound researchers. They are less useful to people who care more about the work than about the particular expression or manifestation. OCLC’s interfaces – I’m thinking here of WorldCat as well as Connexion – do a nice compromise: they display the info, and they make searchable in advanced and command line searches. This means that it’s available, but not front and center. Actually, this is overall a good approach for attributes which many people don’t care much about, but a few people desperately need.
Medium of Performance (MDR II.D)
Medium of performance is an attribute not only unique, but very important to music. It’s also a really problematic access method. Thinking back to the puzzle pieces -- we have a data recording problem. AACR2 did not recognize medium of performance except in the service of creating unique text strings -- uniform titles -- to identify musical works. We did have subject headings, and some libraries coded MARC 048 fields for medium of performance. For scores, classification provides pretty useful access, but few libraries classify recordings.

There is hope for the future, because of two closely related developments. First, RDA. RDA recognizes medium of performance as a distinct data piece, and so facilitates machine manipulation of these elements for powerful searching. Second, the in-progress Library of Congress Medium of Performance Thesaurus, which will replace LCSH and provide vocabulary for recording RDA’s “medium of performance” attribute.
Let’s look at some examples. The University of Virginia’s Blacklight instance uses MARC 048 data to allow faceting by instrument.
Ball State has a number of helpful tools, separate from their catalog, for finding various kinds of media.
The one for music highlights a problem: if 048 fields are used, the search is most effective if it only considers each 048 as a unit, not all the 048’s in the record. This is because each 048 is for an individual piece. If the whole record is searched, and you are seeking, say, flute and clarinet duets, you could end up with a recording that contained a piece for flute and piano and a separate piece for clarinet and piano.

Also, the Ball State Chamber Music Finder uses IAML medium of performance codes, which are more specific than the MARC medium of performance codes, but not widely used.
JW Pepper is a commercial vendor of sheet music, not recordings. They provide a whole slew of facets for medium of performance. The “department” “voicing” and “instrument” facets all address medium. Obviously, medium is a very important attribute for Pepper’s customers.
Musical key has a variety of uses: as an identifying element for generically titled works (“Sonata in C”), to differentiate different versions (A version, E version, etc. – For vocal music, it might be called “high voice/low voice/medium voice.”) Some instruments, such as handbells or carillon have non-standard ranges. This element is not regularly coded in a way suitable for machine processing, different instruments, and vocalists, need different things here, and there are searching problems with “a” (as in “A major”) frequently being a stopword, or in searching the musical sharp and flat signs.

My personal opinion: in most situations, this musical key/range is both problematic and less crucial, so practical considerations may often mean we don’t do much with it. Nonetheless, it’s worth thinking about and considering your particular situation and needs regarding this attribute.
The single most important aspect of dates related to sound recordings is that there are A LOT of them! Dates are important, but the easiest date to grab onto is not always the best one to use. So you need to think and plan, not just grab onto manifestation date, or whatever.
You have a handout with this exercise on it. Here is a CD from the UMKC library. Follow the instructions and think of as many dates associated with this item as you can. Focus first on listing as many dates as you can think of, then work on the other questions if you have time.

If it is a small group/time, ask for students to name dates they thought of. If it works, could talk about each one as they mention it, skipping over slide 36 somewhat.
Here are the dates I came up with. Did anyone come up with others?

As I said before, some of these are not included here, or in anywhere in our library catalog, in fact. I had to do some other research. I’ve arranged these in rough WEMI order. Let’s talk about each one, its usefulness to patrons, and how easy (or not) it is to exploit in a discovery system.
On this slide, I’ve organized the dates from the previous slide in FRBR WEMI style. The point here is not so much FRBR as an organized way to think about the various dates.

The first is the creation date of the work – when Mozart composed the concertos. This is a neglected access point in bibliographic data, and in fact, is not present anywhere in UMKC’s bib record for this CD. Patrons do look for specific years or ranges, including broad eras of music history, often indicative of style. The information is sometimes contained in bibliographic records, in MARC 045, in subject headings, in free text in notes, and, more promising, as an element in RDA authority records.

Year the program notes were written – actually, in most cases I’d say it’s safe to ignore this one. We have to pick and choose.

Year of recording – this is prominent on this particular CD, and coded in the MARC data. This is not always the case, though. In FRBR terms, recording date is an expression-level attribute. The publication date of the CD, inferred from the 1990 sound recording copyright date, is a manifestation level attribute. Publication date is one of the most easily and regularly coded dates, and the most frequently exploited...
for all kinds of date limits. Yet, for recorded sound, it’s one of the least useful. Composition date or recording date – work and expression level dates – are much more useful. In many discovery systems, the best we can do may be to simply clearly label the types of dates exploited, so users aren’t confused.

Dates of other releases are also relevant, especially the first release. Not so relevant here since it was originally a radio broadcast, so recording and release date are the same.

Item level dates – ironically, many library systems can easily sort by the date an item was acquired or cataloged, but this has little use except “new titles” lists. If a patron has an item checked out, the date it’s due is of great interest. This is something usually handled in a special “my account” type area.

K488 - Recorded Mar. 3, 1946, 3:00-4:30 p.m., from broadcast over CBS network, of original performance at Carnegie Hall, New York, per #610238785. Releases in 1950 (#29216895).. and others.
Persons and corporate bodies are similar to dates in that numerous persons and corporate bodies are associated with recorded sound, and it is important to distinguish them. In some ways they are a bit easier to deal with because they do tend to be recorded, and, in MARC, their relationship to the material is often even recorded.
Persons/Corporate Bodies

- Composer/author
- Performer (all kinds)
- Interviewer, Interviewee
- Librettist, lyricist
- Arranger
- Producer
- Author of program notes
Let's take this opportunity to talk about standardized vocabulary. Persons and corporate bodies aren't the only situations where libraries employ standardized vocabulary – uniform titles and subject headings are two other important examples. As we know, it's important to distinguish between people with the same or similar names, and libraries have taken great pains to make this possible through authority control.

Libraries spend a lot of effort not just distinguishing between similarly named people, but also creating standardized text strings to represent those people – their authorized form of name. So, for example, in the U.S., the Library of Congress/NACO authority file headings for these 3 men are all different.

Adams, John, 1735-1826
Adams, John Quincy, 1767-1848
Adams, John, 1947-
In our current environment, one common problem is discovery systems failing to take advantage of this standardized vocabulary.

Here, in worldcat.org, we’ve found a recording of Nixon in China by John Adams (the composer born in 1947). But if you click on his name to “search for more by this author”, ....
...it just conducts an author keyword search on the words “John” and “Adams” and thus you are flooded with results from the two presidents.

Clicking on a link within a record for a work composed by John Adams the composer (Adams, John, 1947-) should locate only materials related to this person, not items related to either president or to other persons named John Adams. In many current systems, this will mean utilizing the bound text string for the full authorized form of name as the link. Alternately, a unique identifier could function behind the scenes, regardless of interface display.
Notice here, the full authorized form of name being displayed, and the keyword search which is conducted, on the full authorized form of name. Be aware that these are two separate decisions. Merely displaying the full form of name doesn’t necessarily mean that the entire string will be searched. In my opinion, the searching is the more important part.
Let’s look to the future a bit.

First, despite what some may say, it’s very necessary to make these distinctions. Even wikipedia has discovered this – have you seen their “disambiguation” pages?
To be honest, Wikipedia’s page is probably easier to navigate than the LC/NAF, where you’re presented with a long list organized by birth and death date. You know the 2nd president, 6th president, and composer are different people, but do you have their birth and death dates memorized? Other attributes can be more useful.
One specific helpful development is the advent of RDA. RDA allows for recording many attributes for persons & corporate bodies, and the MDR addresses potential benefits.

“In addition to seeking specific known persons and corporate bodies, users may seek works associated with persons/corporate bodies possessing particular attributes, such as date, nationality, language, or field of activity.”

“If these fields were regularly encoded, this data could be exploited to answer questions like “What music do you have by French women composers born before 1950?” through use of facets or limiters.”

However, adopting RDA will not cause all this data to magically appear. For one thing, we have the large body of AACR2/MARC legacy data which lacks this information. Second, RDA instructs to add these elements when “applicable and readily ascertainable,” i.e. not all the time. However, it might be possible to exploit other sources of this data and link in the information.
I’d be remiss here not to mention the Virtual International Authority File, a joint project of national (and several transnational) libraries hosted by OCLC, which matches and links their authority files. Here, we not only see the headings used in different countries, we get help in selecting the correct John Adams.
The WorldCat Identities project exploits authority data from VIAF, subject headings, and WorldCat data to create summary pages for every name in WorldCat.
I’d argue that true topical subjects don’t need to be handled much differently for recorded sound than for anything else. An interview about WWII is “about” WWII in pretty much the same way that a book about WWII is “about” WWII. Does this mean recorded sound subject headings are a breeze? No! The problem – and I’m talking most specifically about LCSH here – is that subject headings for recorded sound often reflect aspects other than true “aboutness.”
Music “Subject Headings” Express

- **music of national, religious, and ethnic groups** often by adding the form subdivision “Music” to an otherwise topical heading (II. J, III.K)
- **source of text set**, by employing the title of a liturgical text or adding the subdivision “Musical settings”
- **temporal coverage**, often via date subdivisions (II.F, III.I)
- **medium of performance** (II.D, III.G)
- **genre/form** (II.I).

Talking specifically regarding music. Much music, especially purely instrumental music, is not objectively “about” anything. Therefore, music subject headings often reflect other attributes, listed here with corresponding sections of the MDR.
It's also important to remember that LCSH was not conceived as a faceted vocabulary, and many subject headings get their meaning from their location within the entire string, making it important for the entire string to be able to function and display as a unit.

There are many changes underway in subject vocabulary right now, including

- Library of Congress Medium of Performance Thesaurus for Music (LCMPT)
- Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT)

So the important thing in discovery systems right now is flexibility to accommodate changes.
Let’s talk about genre/form a bit.
FRBR provides a definition for form, but not for genre.

"The form of work is the class to which the work belongs (e.g., novel, play, poem, essay, biography, symphony, concerto, sonata, map, drawing, painting, photograph, etc.)."

FRBR 4.2.2
In MARC, the 655 field is defined for genre, form, and/or physical characteristics of the materials being described.
Here is a further definition of these terms. You see that “genre” is an intellectual concept, while both “form” and physical characteristics refer to physical character. So, “sonata” and “poem” are form terms, while “diary” or “directory” are form terms. Seem fuzzy? That’s probably why we put them all in the same field!
Here are some current genre/form headings. The thing to keep in mind is that genre/form headings represent what something “is” rather than what it is “about.”

From a discovery perspective, it is important to be aware that many changes are going on. Look carefully at your data. Are genre/form terms regularly coded in 655, or are they mixed between 650 and 655? It won’t do any good to try to make the distinction in your interface if your data won’t support it. So you might want to just mix them together. But if you DO have data in various fields, make sure you include all the fields where the data might live! Also consider your users – what do they need to do?
We wrote in the MDR that geographic area is a “problematic and undercoded aspect.” There are several reasons for this. First, geographic area tends to be more important for music other than Western art music, but Western art music has often received priority. Second, many geographic areas can be associated with works, expressions, and manifestations, so there is the intellectual task of sorting out and prioritizing all these. Because of these complications, data is often not present to support geographic area features in discovery interfaces.
First – we’re not covering here the geographic area something is “about” because that is a subject attribute. Do note that Library of Congress Subject Headings have been applied beyond geographic “subjects.”

Out of these geographic areas, note the stickiness of identifying these places—there are many cases where a person’s ethnic/national background is not straightforward, folk music defies being tied to a precise “region,” and music may be popular in multiple regions.

The most definite attribute is the place of publication, so it’s frequently recorded and coded, even though it’s probably also the least valuable attribute.
With today’s mapping technology, though, you can really do a lot of cool things with geographic area, including browsing on a map. This is perhaps most useful for special and digital collections, both because users will be interested in the geographic area and because the data will likely be there to support it. This is a non-music example. If you click on a county, it will take you to materials associate with that county. Now, it would probably be possible to do something like this for a collection based on the place of publication, but it wouldn’t be very useful!
Here's an example where geographic access is very valuable – Vanderbilt University’s Global Music Archive. A great deal of geographic access is provided – region, district, and (a sort of geographic element) ethnic group.
Content and carrier type are frequently crucial aspects of our users’ sound recording discovery needs. Broadly speaking, this is where we distinguish that we’re dealing with sound recordings, as opposed to books, scores, maps, etc., and where we can distinguish particular carriers, i.e. LP, CD, cassette tape, 78, cylinder, etc.
So, under AACR2, we had the GMD, that “sound recording” that showed up in titles. RDA does not use the GMD, instead assigning 3 attributes: content type, media type, and carrier type. The vocabulary may be new, but the concept is something sound recording discovery has needed for a long time.
So, here are the RDA definitions, along with examples for sound recordings. You’ll notice that they still don’t quite get at what our users seek. For example, LPs, 78s, and 45s are all examples of the carrier type “audio disc.” In the current environment, MARC coding is probably a better way to accomplish the needed distinctions. That’s what most systems that allow these distinctions have been using already.
On the discovery front, my experience with multiple systems has shown that it is crucial to pay attention to this discovery need. Most of us work with discovery systems that are not only for recorded sound. This means that our users must be able to distinguish all sound recordings – from the other content types in the system, like video, audio, texts, etc. Furthermore, they must be able to distinguish specific carriers: a CD vs. an LP vs. a 78. Especially outside of archival settings, where materials can be checked out, our users don’t want just any carrier, because they don’t have playback equipment for every carrier out there. Make sure that your discovery system allows both of these, not just one of them.
Here are a few examples of facets and limits which allow limits by both content and specific carriers. They all do it a bit differently. #1 uses post-search facets in a hierarchy layout. #2 uses a similar hierarchy, but in a pre-search limit. #3 uses two completely separate attributes.
In addition to limiting to specific content or carriers, users need to identify the content/carrier in results lists. Icons are frequently useful for this task, as shown here.
Compilations are the norm for most audio collections.
Multiple expressions are frequently issued together in a single manifestation. This is called a “compilation” or, sometimes an “aggregate.” If you’re interested in problems of defining and discussing aggregates per FRBR and RDA, see the report cited on this slide.

To put it all in concrete terms: Normally, LPs and CDs contain more than one piece of music. (Why would you waste a whole LP/CD for just one piece?) A quick note that before the LP this was not such a problem, and now, with digital downloads, it may be changing again.
Frequently, our users are searching for just one work on the compilation. Our existing bibliographic data, based at the manifestation level, has not done a good job connecting all the bits of information about the various works/expressions together in a way that a machine can process.

So here, you have a screenshot of some of the bibliographic data for a Yo-Yo Ma CD. We list all the different works, and the other performers, and the subject headings, but there’s no machine way to tie together which works have which performers and which instrumentation.

This is a problem, and the current solution is usually to favor recall over precision – that is, if you search for a piece for cello, by Bach, you might well pull up a CD which has a piece for cello by Brahms and a piece for piano by Mozart.