

Opportunity Knocks: Updating an Online Music Catalogue

Alastair Boyd (University of Toronto)

A number of academic and large public libraries have started drifting away from the web catalogues supplied by the vendors of their integrated library systems. Frustrated catalogue users were showing signs of bypassing these ILS catalogues in favour of search engines they considered to be faster and simpler, like Google or amazon.com; and frustrated cataloguers were seeing much of their carefully structured data languishing unused, its potential unexploited. The efforts of competing ILS vendors to create useful web-based catalogues have been cautious and uncoordinated. The resulting clumsy hybrids have lost much of the “analogue” functionality of the old card catalogues, while managing to gain little of the speed and simplicity of the “digital” Age of Amazon. So an ever-growing number of librarians have had a radical thought: what would happen if we married the flexibility of popular online search sites with the mountains of rigorously controlled bibliographic data stored in library databases?

ILS vendors might justly protest that until we could all agree on what we wanted, they could hardly justify spending money on experiments. However, it now seems that a pivotal moment has arrived. A consensus is forming, inspired partly by IFLA’s *Functional Requirements for Bibliographic Records* (FRBR), and in part by 10 years’ experience in coping with the shortcomings of off-the-rack web interfaces. The tools of online commerce are available for our “not-for-profit” world of scholarship, and a growing list of libraries has started to use them.

We’ve Only Just Begun

On that list is the University of Toronto Library. It has been almost two years since a Selection Committee chose the Endeca platform, with which to build a new “discovery layer” on top of the current Sirsi database. U of T was not the first library to choose Endeca, so we were able to profit from the experience of predecessors like North Carolina State and (closer to home) McMaster University. I use the word “build” deliberately. Endeca is best regarded as a toolkit, with which you can create whatever interfaces and indexes you like, limited only by the quality of your existing data (more on this below), and your supply of time and money for programming. After about eight months of work, the first version of the new catalogue was offered as an option in September 2008, and it became the default in January 2009.

However, we’re not finished yet—or so I hope. Much of the effort so far has been devoted to “under the hood” work refining the indexes, fixing problems with exotic diacritics, and adjusting the timing of data loads from Sirsi into Endeca. And also, as I hinted a moment ago, we have spent a great deal of time cleaning up errors in some of the bibliographic records to make the system work better. A made-to-measure catalogue provides the opportunity to make effective use of all sorts of data that has been lying fallow in bibliographic MARC records almost since we began to use MARC. But a side-effect of building precise filters and indexes on previously ignored MARC data is that long-dormant errors are suddenly bathed in a glaring spotlight. A shared database like U of T’s, with some records more than 25 years old, and with more than 30 contributing libraries, is bound to contain inconsistencies.

For example, our new catalogue makes extensive use of certain MARC control field codes in order to allow precise filtering of search results according to format: e.g. books, scores, or recordings (the last further divided into LPs, cassette tapes, or CDs). This is terrific, or would be

if all our records for scores were coded as such, and all cataloguing for sound recordings had the necessary precise format codes. Unfortunately, our database contains thousands of ancient records dating from the dawn of U of T's MARC database. Because these were transcribed from catalogue cards using a single book template for all formats, the oldest records for music scores showed up as books in the new catalogue. Consequently when catalogue users tried to whittle down hitlists to scores only, a great many useful results were left behind. Happily there are ways to identify and correct many such deficiencies through automated batch processes; thanks to the shrewd oversight of our Metadata Librarian, U of T has made great inroads in this kind of "data cleansing" (to use Endeca's own clinical terminology).

Evolution or Revolution?

We now have an opportunity to demonstrate that our existing bibliographic records contain great potential for improved online discovery of music resources. Quite a lot is at stake here, because a chorus of critics has been suggesting that the problem with web catalogues is old-fashioned cataloguing rules (AACR2) and data formats (MARC 21). Cast off these shackles, it is said, and all will be well. Indeed, for some of these critics even RDA, the proposed successor to AACR2, is insufficiently radical because of its intended compatibility with existing records. It is certain that if we consider cataloguing principles and data structures only in the light of online keyword searching we arrive at something different from RDA, and from the principles and structures currently used. But an uncompromising insistence on such a position, however logical and ideologically pure, reminds me of Bernard Shaw's crusade to reform English spelling, or the Music Notation Project's proposals to abolish accidentals in favour of a brand new chromatic musical staff. It is unrealistic to ignore the past, whether we're talking about several centuries' worth of book and music publishing, or 145 million existing MARC records. Updated methods that build on and supplement the status quo will be less elegant and efficient, but can still improve resource discovery while incorporating, rather than discarding or rebuilding, our existing databases.

For pragmatists who see tools like Endeca as a way out of this bind, a crucial moment has arrived. Now is our chance to show how these new "discovery layers" that sit on top of existing databases can do the work of reinterpreting old data structures to work in new ways. A perfect example of this is the old-fashioned uniform title, as prescribed in AACR2. In a music card catalogue, uniform titles gathered entries for various editions, arrangements, excerpts, and translations of works in an alphabetical sub-arrangement under the heading for the composer, by using a standardized title for any given work. In an online catalogue incorporating the principles of FRBR, uniform titles as coded in MARC 21 records provide a ready-made method for identifying *Works* and *Expressions*, and in establishing FRBR *whole/part*, *equivalent*, and *derivative* relationships—that is, as long as we bear in mind that uniform titles are only one half of a name-title heading. Such headings contain all the necessary information and tagging, thanks to subfield codes that identify *Work* elements (composer plus a standardized title), *Expression* elements (e.g. format terms such as "vocal score", or the language(s) of translations), and *whole/part* elements such as the titles of component parts of a work (for example, arias within an opera, or single movements from a suite).

Manipulation of these elements as they are fed into Endeca (or "ingested", to use the company jargon) allows us to create flexible ways of displaying and sorting search results which contain multiple instances of the same *Work* and *Expression*, but different *Manifestations* (editions). In the illustration below, eight hits from a search for the title *Moonlight sonata* have been combined in the initial FRBR-based display of results, which represent records for different editions of this

1	Beethoven, Ludwig van, 1770-1827. Sonatas, piano, no. 14, op. 27, no. 2, C# minor (8 records)			
2	Beethoven, Ludwig van, 1770-1827. Piano music. Selections			
	Beethoven masterpieces for solo piano : 25 works / Ludwig van Beethoven. Mineola, NY : Dover Publications, 2004.			Score 
	Music	Bindery	M22 .B4 D63 2004	N/A
3	Beethoven, Ludwig van, 1770-1827. Sonatas, piano. Selections			
	Five great piano sonatas / Ludwig van Beethoven ; edited by Heinrich Schenker ; introduction by Carl Schachter ; performance notes by Anton Kuerti. Mineola, NY : Dover, 1999.			Score 
	Music	Due Jun 2, 2009	M23 .B414 S22 1999	N/A

single work. Clicking on this “rolled up” hitlist entry will create an expanded list showing brief entries for all eight of these different editions.

It Must Be There Somewhere...

There is room for further improvement here, however. The display has been based on the composer plus uniform title, which is the heading currently available to represent the Work. But does this cumbersome heading have to be what we display in the hitlist? This is where the concept of “authority records” residing in an “authority file” can be put to new uses.

A quick digression on authority records: these provide an efficient mechanism for establishing an agreed-upon standard form of a heading, such as a composer’s name, or a name plus uniform title, combined in a single record with variant forms of the name and/or title. When such records in MARC format are linked to MARC bibliographic records, cross-references don’t need to be inserted into every catalogue record that employs a given heading; they can be stored, and updated, in a single place. If these authority records are then fed into a system such as Endeca, the resulting index can assist catalogue users by directly associating the variants in name or title headings with bibliographic records containing the authorized heading. In the card catalogue, or in previous online catalogues incorporating an authority file, anyone searching “Skriabin”, for example, would be directed to search instead for “Scriabin”; now, in our new catalogue, the single search “Skriabin” will immediately return the same results as “Scriabin”, since the former spelling is a variant in the authority record.

In the same way, using the same mechanism, anyone searching the terms “Beethoven” and “Moonlight sonata” can now retrieve the same results as if they had searched by the uniform title displayed in the sample hitlist shown above. Better yet, given the relational database structure of bibliographic and authority record databases, it would be perfectly feasible to create a more elegant heading to represent this particular work. AACR2 prescribes the uniform title “*Sonatas, piano, no. 14, op. 27, no. 2, C# minor*” because it was designed to create a browseable alphabetical subarrangement in the card catalogue under the main entry for Beethoven. Having inherited headings like these, automated validation in many MARC databases also currently depends on literal matching of these cumbersome alphanumeric strings between bibliographic and authority records. But there is no need to duplicate the actual text of the heading in both places; the bibliographic record need only carry a link (inserted at the point of cataloguing) to the name-title authority record. And taking this a step further, if the authority record contains enough

information for cataloguers to correctly identify the work, then the work heading as displayed in the catalogue could be something more user-friendly than the long string shown in the previous example. Perhaps “*Piano sonata no. 14, op. 27, no. 2, C # minor, by Ludwig van Beethoven (1770-1827)*”. This kind of permutation of the existing heading could be achieved programmatically, thanks to its subfield coding and punctuation. Ultimately, the best solution would be to update the cataloguing rules to recognize the post-card-catalogue reality that the work heading itself doesn’t necessarily have to contain all the data, as long as sufficient information can be found in the authority record. For catalogue users’ benefit, I think we should display headings that are as short as possible without becoming ambiguous. The current version of RDA threatens to go in the opposite direction, however. Where AACR2 prescribes a uniform title like “*Variations, piano, woodwinds, op. 6*” RDA proposes “*Variations, piano, bassoon, clarinet, flute, oboe, op. 6*”. Simply by adding a preferred work display title to the authority record (“*Variations, op. 6*” might be enough), we could let our systems show something more concise.

Another benefit of controlling the display of headings from the authority records involves the perennial problem of titles translated from other languages. For books or music with text, it is difficult to see any real alternative to AACR2’s directive to use the original title, rather than the most commonly used title in the language of the catalogue. Otherwise, when creating headings for translations, one winds up with paradoxical headings like “*Magic flute*” to represent scores or recordings in German, and “*Magic flute. English*” to represent a translation. But for works without words, we could choose to be more flexible. The current use of “*Vesna sviashchennaia*” instead of “*Rite of spring*” or “*Sacre du printemps*” as the established uniform title tends to puzzle all catalogue users except those who know the Library of Congress romanization of Весна священная. Again, by marking a title in the authority record as the one preferred for display, we can show whatever we like in the catalogue without having to update all the bibliographic records. (And we can instantly flip the display to something else, if we have second thoughts.)

What Else Do They Have?

What I have described so far has mostly to do with so-called “known item” searches. There are also intriguing possibilities for helping users find music scores when they don’t have a title in mind. For example, the Library of Congress Classification for music provides a systematic way to discover what is available based on the medium of performance. Catalogue users could start

Chamber music	Large ensembles	Music for children
<u>One solo instrument</u> – M6-176	<u>Orchestra</u> – M1000-1075	<u>Instrumental music</u> – M1375-1420
<u>Two or more instruments</u> – M177-990	<u>String orchestra</u> – M1100-1160	<u>Secular vocal music</u> – M1990-1998
	<u>Band</u> – M1200-1270	<u>Sacred vocal music</u> – M2190-2196
	<u>Other ensembles</u> – M1350-1366	
	<u>Aleatory & Electronic music. Mixed media</u> – M1470-1480	

Secular vocal music	Sacred vocal music	Collections, collected editions
<u>Dramatic music</u> – M1500-1527.8	<u>Dramatic music</u> – M2000-2007	<u>Musical sources</u> – M2-2.3
<u>Choruses</u> – M1530-1610	<u>Choral services, etc.</u> – M2010-2017.6	<u>Composers' collected works</u> – M3-3.3
<u>Solo voice</u> – M1530-1610	<u>Choruses</u> – M2010-2017.6	
<u>Folk, ethnic, popular music</u> – M1627-1853	<u>Solo voice</u> – M2102-2114.8	
	<u>Hymnals, Hymn collections</u> – M2115-2146	
	<u>Liturgy and ritual</u> – M2147-2188	
	<u>Popular religious music</u> – M2198-2199	

from an overview; then by clicking any of these top-level links, they could expand subsets of numbers and quickly navigate to a call number browse of holdings for a specific medium.

Chamber music	Large ensembles
One solo instrument — M6-176	Orchestra — M1000-1075
Two or more instruments — M177-990	String orchestra — M1100-1160

Duets M180 - 298.5 (20458) Trios M300-386 (6382) Quartets M400-486 (8092) Quintets M500-586 (3929) Sextets M600-686 (1293) Septets M700-786 (600) Octets M800-886 (777) Nonets + larger M900-986 (2339)
--

The ability to build our own indexes and search interface could also enable searching by instrumentation, based on the precise catalogue record instrument codes rather than subject headings. The example shown here (based on the Canadian Music Centre’s catalogue search interface) illustrates a search for precise instrumentation. In most catalogues, that type of search is impossible, since LC subject headings for vocal music do not include the names of accompanying ensemble instruments, and the classification scheme lumps together at a single number all songs for voice accompanied by two or more instruments.

Category	Instrument	Number
Woodwinds	Flute	1
Strings, plucked	Harp	1
Voices	Soprano	1

Reality Check

It is not difficult to imagine designs for web catalogues that utilize existing bibliographic and authority data more effectively, as well as incorporating new ways of linking to other online objects and services. (Audio clips? Thematic catalogues?) The real constraints, unsurprisingly, are time and money. It can be difficult to persuade programmers to devote time to building catalogue features of special interest to music libraries, given the competing demands on IT resources, and the prevailing concept of a standard catalogue interface. The obvious question arises: can a really useful music catalogue interface cohabit happily with a general catalogue designed for use by all and sundry? This is not a problem for specialized institutions like the Canadian Music Centre, which can tailor the interface and features for their particular clientele. But in a general academic library catalogue, it might be strategically sensible to argue for a customized interface for use within the Music Library. Then there could be no objections to including search and display options of interest only to musicians. That still leaves the problem of getting enough of the programmers’ attention. Perhaps some of us should take up Java programming in our spare time!

(This article is loosely based on a presentation by Alastair Boyd and Suzanne Meyers Sawa at the 2009 CAML Conference in Ottawa).