The Weakest Link:

Subject Access to Music in Online Catalogues

by Alastair Boyd, Cataloguing Department,
Robarts Library, University of Toronto

When we talk about "Subject" access to music, what do we actually mean? For books and textual material, subject cataloguing means adding terms to represent the topic of an item. Library users looking for material on a topic can then use these assigned subject terms to find what is available. Since music (printed music or sound recordings) is not *about* anything in the way a book is about something, subject access to music has nothing to do with topics. Instead, the assumption is made that catalogue users will want to find musical works in a certain *form*, for a certain *medium*, and perhaps fulfilling a certain *function*. For instance, a Library of Congress heading such as *Sonatas (Violin and piano)* describes both form and medium; *Waltzes* refers to the form only; *Piano trios* refers to the medium; *Wedding music* refers to the function. Music cataloguers assign form and genre terms so that people can find music based on what it *is*, not what it is *about*.

Confusion can arise when these form and genre headings co-exist with topical headings under the collective label of "subject headings", as they do in the *Library of Congress Subject Headings*. LCSH is a general list, created by LC's literary warrant over many years, and originally designed to produce convenient alphabetical arrangements in card catalogues. This warrant—creation of headings specifically for items catalogued by LC—together with the physical constraints of searching in cabinets full of cards, has given rise to fluidly constructed, "precoordinated"

headings. In other words, LCSH headings often combine a number of distinct terms denoting different concepts, to produce single phrases that describe topics: *Communism and music*, *Women concentration camp guards*, *Composers in literature*, and so on. The choice and order of terms in such phrases is often arbitrary, so filing additional cards with cross-references helps steer library patrons in the right direction: e.g. from *Music and communism* to *Communism and music*. Where it has been considered useful to create a hierarchical, consecutive sequence of related headings, these headings are inverted to bring them into alphabetical compliance: *Choruses, Sacred (Mixed voices, 4 parts), Unaccompanied* is a particularly twisted example familiar to all music cataloguers. In this case, the broadest heading *Choruses* can be followed immediately in the subject catalogue by the narrower heading *Choruses, Sacred*, which can be further sub-arranged according to vocal medium and accompaniment. The more natural form of the heading—*Unaccompanied sacred choruses*—would push the material so described to the far end of the alphabet, and thus into a cabinet conceivably several yards away from *Choruses* in general.

As computer catalogues (OPACs) have increasingly replaced cards, problems inherent in the structure and principles of LCSH have become more and more apparent. Examinations and criticisms of LCSH have been proliferating since the late 1980s, both in

general terms and from the standpoint of music cataloguing. One such document even emanates from the Library of Congress itself—a proposal from the LC Music Subject Group on improving music subject access. In the preamble there is no mincing of words: the "current system of music subject headings, developed for a card catalog, is antiquated and unsuitable [for] the online catalog.... Headings are a combination of form and medium. Application of headings involves subtle practices which are difficult for catalogers to master and confusing to users."[1]

The fundamental problem is the inconsistency with which LCSH includes terminology for medium and form. Sometimes these terms are stated specifically, sometimes only generally, and sometimes they are omitted altogether. Also, there is a different treatment for music featuring solo instruments or voices. Some examples:

1. *Concertos (Violin)* versus *Concertos (Violin with string orchestra)*. The first implies a full orchestral accompaniment, the second explicitly indicates the string orchestra.

2. *Violin with instrumental ensemble.* Here there is a solo instrument with an accompaniment only broadly indicated. To include the names of accompanying instruments requires instead a heading such as:

3. *Octets (Piano, bassoon, flute, percussion, violin, viola, violoncello).* Now all the instruments are named, but the relationship of solo violin to ensemble accompaniment is lost. Also, few users can see the logic behind the order of elements and browse accordingly. The instruments don't even appear to add up to

eight, because "percussion" is used collectively here.

4. *Piano quintets.* Here the heading alone stands for piano, two violins, viola, cello. Compare this with String quintets (Violins (2), viola, violoncellos (2)), where the heading indicates broad medium and ensemble size, with instrumental details given in score order.

5. *Instrumental ensembles.* This is the only valid chamber music heading as soon as more than nine instruments are involved. Suddenly no instrument names at all can be specified; we know only that there are ten or more instruments, involving two or more families (e.g., winds and strings).

6. *Sonatas (Bass clarinet and piano).* Because this is a duet, the specific member of the clarinet family is named here, and the piano, regarded as accompaniment, comes second. If a cello is added to this ensemble, then we must use *Trios (Piano, clarinet, violoncello)* instead. The clarinet loses its "bass" designation, and even if the composer calls the piece a "sonata," the heading no longer describes it as such. Note that the piano is now named first.

Once vocal works enter the picture, things become still more inconsistent. Consider the heading:

7. *Songs (High voice) with instrumental ensemble.* "Instrumental ensemble" in this context (accompanying a vocal piece) now stands for only two or more instruments, which must remain unnamed.

8. *Choruses, Sacred (Mixed voices, 4 parts), Unaccompanied.* Does this imply SATB chorus? Probably—but headings for choral works specify only the number of choral parts, not the specific voices involved. Moreover, the number is indicated only when there are eight or fewer parts, and only for a cappella works,

[1]Library of Congress Music Subject Group, "Improving Subject Access for Music Materials: A Proposal." April 28, 1993. Published in *Music Cataloging Bulletin* 24 no. 7 (1993): 3-6.

or those with keyboard accompaniment. In addition, all solo voices are entirely ignored.

And so on. The point is these inconsistencies stem from the card catalogue, when such headings were a pragmatic, even creative, method of providing subject access while constrained by an alphabetical, card-based list. In our brave new online world, however, keywords are the instinctive search strategy, especially when the terminology and heading structure are unfamiliar. In a music library these days, most OPAC users treat these carefully formed LCSH headings as no more than a quarry to be mined for keywords, disregarding the structure and logic behind them. This is not only sad for the cataloguers who have patiently learned how to build and apply them, but, far more to the point, the strategy does not work well for the catalogue user. As we see in the examples above, LCSH headings cannot be relied on for keyword searching by form and medium since they do not consistently contain the necessary terms.

Imagine that a user is looking for works like Schubert's "Der Hirt auf dem Felsen," by searching with the keywords soprano, piano and clarinet. Because the LCSH heading for this work is simply *Songs (High voice) with instrumental ensemble*, there is no reliable way to search an LCSH catalogue for this work or others using the same instruments. Nor will the LC classification scheme provide much help, beyond enabling users to restrict the search to songs with piano and one other instrument. Anyone attempting such apparently simple searches will have to throw in title- and note-field keywords as well, and hope for the best. This sacrifice of precision, and the uncertainty of outcome, means that the catalogue fails in one of its basic functions: to inform users reliably of all library holdings that meet their search criteria.

"Clearly what is needed is a new and simplified system for listing forms and media, stripping away the intricate practices developed over past years," says the previously-quoted LC Music Subject Group 1993 manifesto. In fact, one such system has been around for nearly thirty years—but not the one the Group has in mind. Another solution for filling in the gaps left by LCSH was inspired by the advent of the MARC music format and the prospect of computerized catalogues: the MARC control fields for form and medium. First of all, the control fields 008 and 047 in music records have been defined to carry information about musical forms to supplement LCSH. Cataloguers can choose from a list of sixty-five two-character codes representing common musical forms or genres, such as "fg" for fugues, "sy" for symphonies, "tc" for toccatas, etc.

Second, and even more useful, the precise medium of performance can be coded in field 048, again using a two-character code. Each code may be followed by a two-digit number (01-99) to indicate the number of parts or performers, if applicable. Subfield codes distinguish between soloists and other performers or ensembles. For example, a piano trio is represented by the 048 string "$a ka01 $a sa01 $a sc01"; a song such as Schubert's "Der Hirt auf dem Felsen" by "$b va01 $a wc01 $a ka01".

These codes are conspicuously unhelpful to catalogue users in this opaque form. However, once indexed they need only be linked to drop-down lists of genres/forms and instruments (voices, ensembles) in the OPAC. Searchers could then select from these natural language lists (this would help cataloguers in applying the codes too). Of course, 048 coding has to be consistently applied to all music records of the same level within a catalogue, if

15

this form of access is to be useful and reliable. Regrettably, in the early 1990s the Library of Congress discontinued this form-and-medium coding, and many libraries have followed suit. This decision was made on the grounds that although "the 048 field of the USMARC format was created specifically to provide access via instrumentation ... this field is not used by many libraries because 1) it requires time-consuming coding instead of natural language, 2) the codes are not comprehensive (for example, no code exists for contrabassoon), and 3) many library catalogs do not index this field."[2] These arguments are rather circular, and, ten years too late, these objections seem rather easy to answer. After all, to abandon such coding is to render useless all the labour ever spent in providing it, since its value as an access point depends on consistency. However, history has moved in another direction, based on keyword indexing of natural language fields. Libraries that have discontinued 048 field coding will probably be reluctant to resume the practice.

This is where the Music Thesaurus Project could come to the rescue. For ten years Harriette Hemassi of Rutgers University has been leading the work in developing a dedicated thesaurus for the discipline of music. Inspired by the success and methodology of the *Art & Architecture Thesaurus,* and using terminology primarily from LCSH, it will permit faceted access to music, by means of a standard vocabulary in a systematic hierarchy. Music Library Association members were able to read all about the underlying principles in MLA's *Notes* (March 1994), with updates

presented at the annual MLA Conferences.[3] A comprehensive bibliography relating to the project is available online through the MLA Web site, along with online versions of proposals and working documents, so I will give only a very brief summary.[4]

Hemassi and the MLA Working Group on Faceted Access to Music have been sorting and converting around 14,000 LCSH music headings into seven facets or provisional categories: Agents, Events, Forms/ Genres, Geocultural Attributes, Sound Devices, Texts, and Other. The idea is to break apart the multiple-concept, "pre-coordinated" style of LCSH subject strings into thesaurus-style single concept terms. For example, the LCSH heading *Suites (Recorder and harpsichord)* combines a form-genre term (Suites) with terms for sound devices; *Folk songs, Russian* combines geo-cultural and form-genre terms. We can use the online *Art & Architecture Thesaurus* browser to see the possibilities for contextual searching provided by a properly structured thesaurus.[5] It demonstrates the features we could expect in the Music Thesaurus: a comprehensive vocabulary, arranged in a hierarchy that reflects the discipline rather than the alphabet, with a complete cross-reference structure to link broader, narrower, and related terms.

Such a thesaurus would be a widely useful addition to the reference resources for the field

---

[2]Working Group on Faceted Access to Music, "Discussion Paper, Faceted Access to Music: Possibilities and Ramifications", MLA Web site 22 Oct. 2002: <http://www.musiclibraryassoc.org/BCC/BCC-Historical/BCC94/94WGFAM1.html>

[3] Harriette Hemassi, "The Music Thesaurus: Function and Foundation," *Notes* 50, no. 3 (March 1994): 875-882.

[4]Music Thesaurus Project Advisory Task Force, "Music Thesaurus Project Bibliography," MLA Web site 22 Oct. 2002: <http://www.musiclibraryassoc.org/BCC/MTP/mtp-bib.html>

[5]Getty Research Institute, Art & Architecture Thesaurus Browser, 22 Oct. 2002: <http://www.getty.edu/research/tools/vocabulary/aat/>

of music. Of particular interest to music cataloguers is its potential as a catalogue retrieval mechanism. Back in 1993, the LC Music Subject Group made some specific proposals: "1. List musical forms and media ... in a single MARC Field 654.... Terms would come from the Music Thesaurus.... There would be no prescribed order for instruments/ voices.... 2. List music descriptors (headings for genre, style, ethnicity, seasonal use, etc.) in MARC Field 655.... Terms would come from the Music Thesaurus."[6] There follow some examples illustrating how thesaurus terms would be carried in a catalogue record. Soon after, MLA's Working Group on Faceted Access to Music also published a discussion paper in which Harriette Hemmasi provided her own examples.[7] Although differing in certain details, taken together these sources suggest how cataloguers might apply thesaurus terminology in a music MARC record, and how systems might index it.

For example, compare the following LCSH headings with their possible faceted equivalents.

Item: a Russian folk song for voice and accordion.
- Current LCSH: 1. *Folk songs, Russian*. 2. *Songs with accordion.*
- Faceted Approach. Form/genre: Folk songs; Geo-cultural: Russian; Sound devices: solo voice, accordion.
- MARC coding: 654 $c f $a folk songs $c g $b Russian $c s $b voice $c s $b accordion $t 2

(Subfields: $c = facet, $a = primary term, $b = secondary term, $n = number, $t = total number)

---

[6] Library of Congress Music Subject Group, *op. cit.*, p. 4.

[7] Working Group on Faceted Access to Music, *op. cit.*

Item: a suite for string sextet (pairs of violins, violas and cellos.)
- Current LCSH: 1. *Suites (Violins (2), violas (2), violoncellos (2))*
- Faceted Approach. Form/genre: Suites; Sound devices: violin (2), viola (2), violoncello (2) (total number=6)
- Coding: 654 $c f $a suites $c s $b violin $n 2 $c s $b viola $n 2 $c s $b violoncello $n 2 $t6

The order of elements follows the score or, for sound recordings, the order listed in the notes. Since the field is going to be indexed as a source of keywords, rather than to generate a browseable list of headings, the order of terms is immaterial. Therefore, no intricate rules will regulate the order or number of elements, which should save cataloguing time and reduce errors. These strings of terms would also be consistently complete and thus more reliable for keyword searching than current LCSH headings.

Of course there will be new rules to learn. Even if the terms themselves are natural language rather than the arbitrary codes of the 048 field, at first glance the subfield coding makes the sample strings above look just as complex as those "time-consuming" 048 tags. Here again, improved cataloguing software could help, by presenting a labelled input screen which would supply the coding automatically. An online thesaurus resembling the *Art & Architecture* browser would help cataloguers choose the correct terminology.

Assuming that Music Thesaurus terms could be input easily and accurately, how would they be indexed, and how would they be used in searching the catalogue? The AAT online search is fine for discovering terms, defining them, and displaying their context. But library users may be unenthusiastic if such searches stand alone, yielding results that they must

then enter themselves in a subsequent search of the catalogue. It is important to make a useful connection between the indexed information and a search interface.

Without precise indexing, conflicting sources of subject terminology can create problems. (Even now, if terms from other sources such as the MESH or AC lists find their way into an LCSH subject index, users get confusing and contradictory feedback about preferred terms and cross-references). MARC coding allows us to create a separate index for faceted Music Thesaurus terms—in theory, even separate indexes for specific facets, and/or primary versus secondary terms. Therefore, it seems sensible to keep Music Thesaurus terms well away from a general keyword index, and design instead an additional OPAC search specifically geared to music scores and recordings, based on a 654/655 field index. It is improbable that users would, in a single search, want to find both books about music and the music itself. The default OPAC subject search would remain one for books and textual materials, but users opting for a "music" search would be offered a screen in which to input keywords for instrument or ensemble names, numbers of instruments, forms and genres. This would fulfil one of the Automation Requirements for Music Materials defined by the Music Library Association: "The system should allow construction of a separate index for form and genre terminology (655 field and 650 $v)." Such a feature, says MLA, would "considerably enhance the effectiveness of the system for handling music information."[8] Already, the latest Web catalogue for the

University of Toronto offers separate searches for "Subject" (i.e., LCSH terminology in one index), "Medical Subject" (i.e., MESH terminology in a separate index), and "Genre" (i.e., terminology from 655 fields in yet another index). All that is lacking, in this last case, is the actual terminology....

A helpful interface for performing keyword searches will already be familiar to librarians who have used Folio Views and LC's Cataloger's Desktop. The searcher inputs a string of keywords into the appropriate box; then, as terms are added, a tree is generated showing occurrences of the terms, singly and then in combination. In this way, anyone searching can see at a glance which terms are found together and which are not. This avoids the frustration of getting a "0 records found" result, with no indication which term or terms may have rendered the search null and void—as happens in many Web catalogues that model themselves on Web search engines.

Ideally the search window would supply (via authority records) the thesaurus reference structure to guide users in choosing terms, just as happens now with LCSH-based subject searches. The crucial point is that modern OPACs should leave it to the system to broaden search results upon request through the "post-coordination" of separate headings, rather than by lumping everything into a single index at the outset. You cannot easily perform a precise search on a broad index, but you can easily get broad results by combining those from several precise indexes.

Will the Music Thesaurus approach succeed? It has been more than ten years now since proposals for reform began. Creation of the Music Thesaurus, which underlies the whole operation, is taking longer than expected, as the initial job of reorganising thousands of pre-coordinated LC subject

---

[8] Music Library Association Subcommittee on Automation, "Automation Requirements for Music Materials", MLA Web site 22 Oct. 2002: <http://www.musiclibraryassoc.org/committee/co_adm_autoreq.htm>

strings has revealed further problems and inconsistencies. This re-mapping of LCSH terminology is now complete, but the problems are by no means over. Even if the first version of the Music Thesaurus were to be published tomorrow, the problem of converting existing bibliographic databases remains to be dealt with. If faceted subject access based on the Music Thesaurus is to have any viability then LC and OCLC will have to support the idea. Even purely local implementations at large music libraries would be futile if their existing bibliographic records could not be updated in some way. And yet, how could this be done? Perhaps libraries could share a conversion table of some sort to automate the translation of certain LCSH headings into faceted equivalents. This could work for headings like my before-and-after examples earlier on (e.g. the LCSH heading *Suites (Violins (2), violas (2), violoncellos (2)* translated into the faceted heading $c f $a suites $c s $b violin $n 2 $c s $b viola $n 2 $c s $b violoncello $n 2 $t6). But it could not be done for those headings whose very lack of instrumental detail is the reason they need to be converted in the first place. Of course, so long as a record with the vague heading *Songs (High voice) with instrumental ensemble* also has a precise 048 field coded "$b va01 $a wc01 $a ka01", then an automated solution is still possible, thanks to the detailed medium information carried in the 048. But then, what of all the 048-less cataloguing produced in the last ten years? It *is* going to be difficult.

Before administrators will consider paying for this conversion, or even for the introduction of faceted access, they will have to be convinced that there is a demand for precise, reliable subject searches. And perhaps any such demand is fading, in an era when library catalogues are being remodelled as Web portals, and the distinction between local catalogue searches and Google Web searches is being increasingly blurred.[9] I think we must try and make the case. I think there is a case to be made, even in the face of anecdotal evidence about Web-weaned undergraduates who hazard some key words and are grateful for any results at all; as they sift patiently through hit lists, they are unaware of having bypassed the inherent precision of the catalogue's controlled access points and cross-reference structure.

Music cataloguers are not alone in finding LCSH to be somewhat problematic. The recent Faceted Application of Subject Terminology (FAST) proposal from OCLC also begins by breaking apart LCSH strings into components. The goal is to make it easier for staff who lack training in indexing and classification to apply controlled topical terms to Web resources—a less ambitious project than the Music Thesaurus. Perhaps the Music Thesaurus Project has been too ambitious and long-delayed. Let us hope not, or we will have missed an opportunity to simplify the cataloguing of music while improving access to it. While we wait and see, and make do with our current makeshifts and workarounds, I have a suggestion: why don't we all start coding field 048 again? It would be a good start.

This paper is based on a presentation entitled "The Role of a Music Thesaurus in Tomorrow's OPAC," given at the CAML/CUMS conference at the Edward Johnson Building, University of Toronto, on May 28, 2002.

---

[9] For a recent discussion of what the differences are and why they matter, see Bernhard Eversberg, "On the Theory of Library Catalogs and Search Engines," 22 Oct. 2002 <http://www.biblio.tu-bs.de/allegro/formate/tlcse.htm>