

## **Deriving Faceted Terms from Library of Congress Subject Headings for Music: Challenges and Possibilities**

**MLA Annual Meeting 2018, Portland, OR**

**February 2, 2018**

*Summary written by Felicia Piscitelli, Texas A & M University*

*Session presenters: Casey A. Mullin, Western Washington University; Gary Strawn, Northwestern University*

In this session, Casey Mullin articulated a need to convert Library of Congress Subject Headings (LCSH) in existing catalog records into form, genre and media of performance terms better suited to music, and Gary Strawn discussed a tool he has developed to accomplish this task. The necessity for this conversion and the difficulties it would entail were emphasized rather than the specifics of assigning faceted terms.

Casey Mullin is the outgoing chair of the MLA Vocabularies Subcommittee. He began his presentation by pointing out the strengths and weaknesses of LCSH. The alphabetic arrangement, use of subdivisions to clarify a topic, terminology that is kept up to date (more or less) and widespread implementation are among its advantages. However, it is not so effective in distinguishing between “aboutness” (that is, what an item is about) versus “is-ness”, or what the thing is, which is particularly problematic for music. For example, the term “Symphony” refers to works *about* the symphony as a musical form, whereas its plural, “Symphonies”, is applied to actual pieces in that form. Another problem is a lack of granularity. A piece like “Dover Beach” by Samuel Barber, for baritone voice and string quartet, would receive the subject heading: Songs (Medium voice) with string ensemble. But, a “medium voice” could be a mezzo-soprano, and a “string ensemble” can mean any combination of stringed instruments, a string quartet being only one of them. Some attributes are not obvious. For instance, “American” in “Songs, American” denotes the country of origin, but no comparable heading exists for “Songs, Mexican”; for the latter, one needs to employ the heading “Songs, Spanish \$z Mexico” (wherein “Spanish” refers to the language). Mullin asks, “If LCSH for music is this difficult for *librarians* to use and understand, what hope do our end users have?”

The development of “faceted access with purpose-built vocabularies”, along with corresponding MARC fields, provides terminologies that are easier to use for music than LCSH. These vocabularies include the Library of Congress Medium of Performance Thesaurus for Music (LCMPT) and MARC field 382; Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), coded in field 655; and the Library of Congress Demographic Group Terms (LCDGT), with MARC fields 385 and 386. Additional MARC fields include 370 for geographic area (formulated according to LCSH and the LC/NACO authority file), and 046 and 388 for chronology. Language is already encoded in fields 008, 041, and 546; ideally, online systems should be able to parse these fields into usable data. There are many benefits to using these thesauri: they are constructed to NISO standards; they are post-coordinated, like FAST headings, so there’s no need for subdivisions, and they are easier to navigate; each facet has specific metadata elements; they’re freely available on the Web, and compatible with linked data.

While these new vocabularies and MARC fields have been implemented in current cataloging, catalogers are still applying LCSH as well, thus doing double work. The value of faceted terms for helping our users find music materials will not be realized until a critical mass of our records contain them. However, as of

January 1, 2017, only 8% of the records for scores and sound recordings in WorldCat contain the 382 field. How can a greater number of records be enhanced with these new terms and fields in an efficient way?

Between 2014 and 2017, the MLA Vocabularies Subcommittee has partnered with Gary Strawn to develop algorithms to convert LCSH to LCMPT and LCGFT/ LCDGT automatically. Each LCSH should “beget” at least one LCMPT and/or one LCGFT term. Selective MARC composition and format of music codes (from the fixed field and from field 047) can fill in gaps in LCSH terminology. The algorithm should work equally well on records for both scores and sound recordings; duplicate fields should be removed, and the machine-generated fields should conform to best practices.

If a record for one of J.S. Bach’s Preludes arranged for cello is correctly encoded (“pr” in Form of Composition, “l” in Format of music) and has 650 (LCSH) of Cello music, Arranged, the algorithm will generate three LCGFT terms: Preludes (Music), Arrangements (Music), and Scores. This is an example straightforward, “lossless” conversion.

Of course, not all transformation will be so easy. Human language in LCSH adds complexity to the automatic process. For example, can the algorithm be programmed to “know” that the LCSH term “Symphonies” implies a medium of performance? How can it account for differences in the vocabularies, i.e., LCSH “mixed voices” versus LCMPT “mixed chorus”, or “Canons, fugues, etc.” vs. “Canons” and “Fugues”? Will the presence of subdivisions like “History and criticism” enable it to determine when the topical LCSH is legitimate? Could the algorithm be applied to records for videorecordings? The lack of granularity in LCSH may result in generic LCMPT output. Also, there is as yet no good way to handle religious demographic terms, as in “Buddhist hymns”, even with LCDGT; a new MARC field may be needed to code for a demographic group that is neither a subject, intended audience, or creator/contributor. Because of these and other issues, the LCSH headings are retained for the present; the new fields are simply being added.

Gary Strawn described how he programmed a tool for OCLC Connexion that will perform these transformations. The tool is a macro that can be assigned to a button or keystroke that will activate a DLL. He created a spreadsheet containing lists of values for each field and how it should be mapped. For example, List 7 includes Format of Music codes from the fixed fields or 047 and maps them to 655 LCGFT terms.

The tool is ready for testing, and the presenters encourages everyone to do so. Those wishing to install this program should e-mail Gary at [mrsmith@northwestern.edu](mailto:mrsmith@northwestern.edu) or Casey at [casey@mullingroup.com](mailto:casey@mullingroup.com) The documentation for this program is available at: <http://files.library.northwestern.edu/public/Music382/documentation/>

