

8-14-2019

## Beyond the Subject: Non-Topical Facets for Exploration and Discovery

Kelley McGrath

### Recommended Citation

McGrath, K. (2019). Beyond the Subject: Non-Topical Facets for Exploration and Discovery. *OLA Quarterly*, 25(1), 9-16. <https://doi.org/10.7710/1093-7374.1970>

© 2019 by the author(s).

*OLA Quarterly* is an official publication of the Oregon Library Association | ISSN 1093-7374

# Beyond the Subject:

## Non-Topical Facets for Exploration and Discovery

**by Kelley McGrath**  
Metadata Management Librarian,  
University of Oregon  
kelley@uoregon.edu



Kelley McGrath is Metadata Management Librarian at the University of Oregon. She is an experienced media cataloger and an active member of Online Audiovisual Catalogers (OLAC). She has a long-standing interest in the potential of faceted navigation in library catalogs. She is particularly interested in ways to make library metadata more useful for humans and machines and ways to design discovery interfaces to make better use of library metadata.

*New developments in the cataloging world can help libraries better answer questions like: What music do you have for string quartets? What young adult fiction do you have by African American male authors? Do you have any diaries written by pioneer women in Oregon in the late nineteenth century? Do you have any recent movies from China?*

### **Introducing New Vocabularies and MARC Fields**

Historically, the Library of Congress Subject Headings (LCSH) have included terms both for what something is about (topic) and for what something is (genre or form). Many users are looking for something either as a topic, or as a genre or form, and not for the two things mixed together. Sometimes LCSH makes a clear, albeit not intuitive, distinction. For example, “Symphony” is used for works about symphonies while “Symphonies” is used as a genre-form heading on records for scores and performances of symphonies. Sometimes subdivisions, such as “History and criticism” are appended to headings to indicate aboutness. So “Science fiction” is used for a collection of science fiction stories while “Science fiction—History and criticism” is used for works about science fiction. However, the guidelines for LCSH tell catalogers to omit genre and form information for individual works of literature. So, while an anthology of science fiction stories would get the heading “Science fiction,” a novel published on its own wouldn’t. All this creates an unpredictable and unwieldy landscape for users to navigate.

Starting in 2007, the Library of Congress (LC) began work on a new vocabulary, now known as Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), to use for genre and form instead of LCSH (Library of Congress, Policy and Standards Division, 2017). The initial genre-form terms for film and television were issued in September 2007. Since then, LC has coordinated projects to develop basic sets of terms for additional areas such as music, literature, and legal and religious materials. Terms for visual and artistic works, the last planned project area, were published in 2018. New terms



are added to these basic sets of terms on an ongoing basis. In October 2018, Online Audio-visual Catalogers (OLAC) released a genre-form vocabulary for video games to compensate for a gap in LCGFT that LC currently does not have the resources to address (2018).

In the process of separating genres and forms from topical LCSH, the Library of Congress realized that LCSH contains other types of information that are neither topical nor genre and form. These include intended audience (e.g., “Conversation and phrase books (*for medical personnel*)”), type of creator (e.g., “*Prisoners’ writings*”), and instrumentation (e.g., “*Sonatas (Flute and piano)*”). Since one of LC’s goals is to support faceted browsing (Young, 2017), which requires categories based on “clearly defined, mutually exclusive and collectively exhaustive” characteristics (Taylor, 2000, p. 274), they had to find somewhere else to put this information.

As a result, they created two additional new vocabularies: Library of Congress Medium of Performance Thesaurus for Music (LCMPT) for instruments and voices and Library of Congress Demographic Group Terms (LCDGT) for audiences and creators. An initial 800 medium of performance terms were published in 2014 and are now in widespread use for music cataloging (Library of Congress, Policy and Standards Division, 2019b). The first demographic group terms were published in 2015. However, the demographic group vocabulary is currently frozen following a pilot implementation, while LC reevaluates its underlying structure and principles (Library of Congress, Policy and Standards Division, 2019a). In 2012-2013, three new fields were created in the MARC 21 format as places to record this kind of information: 385 (Audience Characteristics), 386 (Creator/Contributor Characteristics) and 382 (Medium of Performance).

### **How Improving the Searchability of Non-Topical Characteristics of Resources Can Help Users**

The following examples demonstrate some possible uses for these fields, as well as a number of other MARC fields for non-topical aspects of library resources. The illustrations are taken from a sandbox view of the University of Oregon’s Primo discovery interface. If you wish to experiment with your own searches, you can try it out at [http://alliance-primo-sb.hosted.exlibrisgroup.com/primo-explore/search?vid=UO\\_NRWG](http://alliance-primo-sb.hosted.exlibrisgroup.com/primo-explore/search?vid=UO_NRWG).

#### **INSTRUMENTATION**

Music students often come to the library looking for scores for their instrument. LCSH for instrumentation are inconsistent in their formatting and may be combined with other types of information. Even if a user limits their search to scores to eliminate resources about, say, saxophone music, they may have trouble searching effectively. In a traditional subject browse list, saxophone music may appear under “S” for saxophone, but it also shows up under the names of other instruments, after genres, and following terms such as quartets and woodwind trios:

- Didjeridu and saxophone music
- Marches (Saxophone and piano)
- Quartets (Piano, saxophones (2), vibraphone)
- Saxophone and piano music
- Saxophone music
- Saxophone music (Saxophones (2))



In a keyword search, the user must know to look for both singular and plural forms. Untangling these different characteristics and recording them in a consistent and predictable way eases the burden on the user. For example, in a discovery interface that includes separate facets for non-topical information, a search for Beethoven can be easily limited by the search facets, “Resource Type” and “Instrument or Voice (Music)” to scores that contain a part or parts for the violin.

Active filters: violin x Scores x  
 Reset filters  
 Refine my results  
 Sort by Relevance v  
 Availability ^

PAGE 1 291 Results

- SCORE  
**Concert for piano, violin and cello, op. 56**  
 Ludwig van Beethoven, 1770-1827.; New York, N.Y. : E.F. Kalmus; 196-  
 Available at Knight Library Music (M1040 .B4 op.56 1960z) >
- SCORE  
**The string trios**  
 Ludwig van Beethoven, 1770-1827, composer.; New York : Lea Pocket Scores; 1959  
 Available at Knight Library Music Minature (M350.B4 L4 1959) >

These results can be further refined by genre or form, total number of performers, additional instruments or exact instrumentation.

**Genre/Form** ^

- Arrangements (Music) (49)
- Art music (11)
- Cadenzas (3)
- Chamber music (177)
- Concertos (54)

**# of Performers (Music)**

- 2 performers (113)
- 4 performers (71)
- 3 performers (62)
- ensemble(s) (33)
- 5 performers (13)

**Complete Instrumentation**

- violin (1); piano (1) (94)
- violin (2); viola (1); cello (1) (59)
- violin (1); orchestra (1) (27)
- violin (1); viola (1); cello (1) (19)
- violin (1); cello (1); piano (1) (19)

**Instrument or Voice**

- cello (154)
- piano (151)
- viola (114)
- orchestra (35)
- violin (solo) (34)

### CREATORS, AUDIENCES, AND DEMOGRAPHIC CHARACTERISTICS

In many cases, users are interested in the perspective from which something is written or created. This is particularly true for all kinds of autobiographical materials, including archival materials, and literature and the arts. Demographic characteristics of the creator, such as those found in LC Demographic Group Terms, can potentially be helpful. For example, users might wish to search for short stories by Mexican-American authors, poetry by Oregonians, or autobiographies by women written in the 17th century.

The screenshot shows a search results interface. On the left, under 'Active filters', there are two filters: 'Short stories' and 'Mexican Americans', each with an 'x' to remove it. Below these is a 'Reset filters' button. Under 'Refine my results', there is a 'Sort by' dropdown menu currently set to 'Relevance'. On the right, under '7 Results', two results are shown. Result 1 is a 'PRINT BOOK' titled 'Mexican-American anthology II : poetry, prose, essays, stories, songs, dichos, corridos, art', with a small book cover image. It is 'Available at Knight Library General Collection (PS508.M4 A5 1976)'. Result 2 is a 'PRINT BOOK' titled 'Voces : an anthology of Nuevo Mexicano writers', also with a small book cover image, and is 'Available at Knight Library General Collection (PS571.N6 V63 1987)'.

Demographic group terms can also be useful for bringing out the intended audience of a resource. Historically, this has most commonly been done for materials intended for children, and it is possible to leverage existing fixed field data in MARC records to bring this out. Existing data can also sometimes be used to distinguish among resources on technical, medical, or legal topics intended for experts, the general public, or children. The 385 field (Audience Characteristics) supports more fine-grained distinctions in audience and could identify such things as math textbooks for third graders. This new field expands the possibilities for expressing audience to things like Chinese language phrase books targeted at businesspeople versus tourists, or books on the job search process aimed at college students versus people over fifty.

### PLACES, DATES, AND ORIGINAL LANGUAGE

Places associated with works are often of interest. For example, a user doing a search for folk music might be interested in refining their search by place of origin. Users might also be interested in novels set in New York City or movies filmed in their hometown. Someone might also be interested in exploring performances of Beethoven's *Ninth Symphony* by date of recording or browsing 19th century Russian novels. Currently, this information is most likely to be found as part of LCSH strings or in note fields. All of these searches will be more successful when non-topical information is more widely recorded in separate fields as structured data.

The transition to structured data for recording these characteristics is well underway in moving image cataloging. The definition of MARC 257 (Country of Producing Entity) was expanded in 2009 to permit its use in non-archival cataloging. 041 subfield h (Language code of original) was redefined in 2011 to include the original language of a work even when the resource being cataloged does not contain a translation. OLAC has promoted the use of these subfields and 046 subfield k (Beginning or single date created) in its widely-used best practices for cataloging DVDs and streaming video (Online Audiovisual Catalogers, 2018).



These fields support targeted searches. Unfortunately, Primo, the University of Oregon’s discovery interface, does not support faceted browsing prior to entering a search term, but a keyword search for “films” limited to French under the original language facet provides a list that can then be narrowed by country of production or original date of release.

The screenshot shows a search interface with the following elements:

- Active filters:** French (with an 'x' to remove), 2000-2100 (with an 'x' to remove), and Algeria (with an 'x' to remove). There is a green 'Reset filters' button below the filters.
- Refine my results:** A link to refine the search results.
- 4 Results:** A list of search results.
  - Result 1:** DVD VIDEO, **Disruptive film. Vol. 1 : everyday resistance to power**. It is available at Knight Library Audio & Video Room (DVD 05722).
  - Result 2:** DVD VIDEO, **Hors la loi = Outside the law**. There is a 'Check holdings' link next to it.

Three other related MARC fields, 370 for associated place, 377 for associated language, and 388 for time period of creation were added to the MARC 21 format in 2014, but have been more commonly used in authority records to date.

### Roadmap to Improving Access to Non-Topical Characteristics of Library Resources

A lot of work has gone into the development of these new vocabularies and MARC fields for non-topical characteristics of library resources, and they have great potential to support exploratory search for many types of library resources. However, much work remains to be done before they can be optimally realized.

#### BEST PRACTICES, DOCUMENTATION, AND TRAINING

In order to accurately and consistently add these fields, catalogers need a shared understanding of how to use them and many decisions remain to be made about best practices. Documentation will then be needed to record these choices, and training will be needed to disseminate the decisions that were made and raise awareness in the broader cataloging community.

There are some existing training materials and documentation, but much is still in draft form. LC’s draft documentation for demographic group terms and genre-form terms, as well as their documentation for the medium of performance thesaurus is freely available online (Library of Congress, Policy and Standards Division, 2018a-c). The Music Library Association has created best practices for LCMPT and for LCGFT for music (2019). OLAC has developed best practices for LCGFT for moving images, although these have not been updated in almost eight years (Online Audiovisual Catalogers, 2011). Adam Schiff of the University of Washington has developed training materials for some of these fields (2019).

The American Library Association Subject Access Committee (2017) has released a white paper on implementation of these new faceted vocabularies and charged a subcommittee with the development of best practices and training materials (American Library Association Subject Access Committee, 2019). This work should result in more authoritative and comprehensive guidance.



**RETROSPECTIVE RECORD ENHANCEMENT AND RECALL**

Even if catalogers agree on a shared practice and begin routinely adding these fields to new records, there are many, many existing records that lack these data points. Since it is not practical to recatalog all of these materials, it will be necessary to use automated or semi-automated methods to enhance existing records. Music catalogers have made the most progress on this front. The Music Library Association, in collaboration with the programmer librarian Gary Strawn, has developed a music toolkit, which was released in April 2018 (Mullin, 2018a). The toolkit works on individual records in OCLC's Connexion client cataloging software. It is run as a macro and uses complex algorithms to analyze existing LCSH and some coded information in records for scores and musical recordings. It then adds new fields to the record, including genre and form, and medium of performance. The music toolkit is only as good as the existing metadata in the record. Therefore, it must be reviewed by a cataloger and may need to be corrected or expanded manually. However, assuming the existing metadata is sound, it is largely accurate in its inferences and saves significant typing time. Mullin (2018b) anticipates that the tool will be adapted for larger-scale automated enhancement projects with selective human review. It is likely to be more difficult to expand this approach to other materials, such as individual works of literature, where the desired information is often not explicitly recorded in the bibliographic record in any form.

Although tools are being developed to populate existing records with these fields, this will take time and will likely be less complete and less granular than is possible in new cataloging. Libraries may be reluctant to expose these new fields to the public due to concern about incomplete retrieval. Clearly some minimum level of recall is necessary. On the other hand, all large datasets have errors and omissions and 100% retrieval for all queries can only be an aspiration. Flaherty and Morgan (2019) argue that "one of Google's original competitive advantages was that it recognized that precision was more important than recall in the context of searching an almost infinite data source like the web—in other words, displaying only relevant results for a query is better than returning every result that could potentially be relevant (at the penalty of including many low-value hits)." Questions to consider moving forward include: *What is a minimally acceptable level of recall and how will we know when we have reached it? Will the relevancy ranking provided by library discovery systems be sufficient to provide a satisfactory user experience?*

**INCORPORATION INTO DISCOVERY INTERFACES**

Although getting the data into the records is an essential first step, it is of little use if public-facing discovery systems can't take advantage of it. In some cases, such as LC's demographic group terms in the audience and creator characteristics fields, the data is already in a form suitable for basic display, search, and use in facets. All that is needed is a system that provides local flexibility in deciding which fields to display, to include in search indexes and to use for creating facets. However, some of these fields require more manipulation in order to be useful. This, in turn, requires either a system that is hardcoded by the vendor or developer to perform such manipulation or which provides tools that a library can use to do this itself. An example of this type of data is a typical 382 (Medium of Performance) field (see table below). A raw, unmanipulated display of the MARC field is unlikely to be helpful to users (e.g., "viola 1 clarinet 1 piano 1"). McGrath and Lowery (2018a, 2018b) were able to use the data manipulation tools of Primo's Back Office to produce the more human-readable display on line 3 for the Orbis-Cascade Alliance's shared Primo discovery layer. However, it



was impossible to reproduce the Music OCLC User Group’s recommended display (Belford, 2015) due to technical constraints on the types of data modification Primo supports. Coded data, such as dates from the 046 field or original language information from 041, will also require transformation to be useful to end users.

Formatting Issues with MARC 382 (Medium of Performance) Field	
MARC field	382 01 \$a viola \$n 1 \$p clarinet \$n 1 \$a piano \$n 1 \$s 2 \$2 lcmpt
Unmanipulated display	viola 1 clarinet 1 piano 1
Orbis-Cascade Alliance display	viola (1); piano (1) (alternate instrumentation: clarinet)
Music OCLC Users Group (MOUG) recommended display	viola or clarinet (1); piano (1)

### Conclusion

LC’s new vocabularies and related MARC fields have great potential for helping library users find resources based on their non-topical characteristics, but much work remains to be done to turn these possibilities into reality. 🐼

### References

American Library Association, Subject Access Committee. (2019). SAC Subcommittee on Faceted Vocabularies. Retrieved March 31, 2019, from <http://www.ala.org/alcts/mgrps/camms/cmtes/ats-ccsacfv>

American Library Association, Subject Access Committee, Working Group on Full Implementation of Library of Congress Faceted Vocabularies. (2017). A brave new (faceted) world: Towards full implementation of Library of Congress faceted vocabularies. Retrieved from <https://alair.ala.org/handle/11213/8146>

Belford R. (2015). WorldCat Discovery display preferences for medium of performance. Retrieved from: [http://musicoclcusers.org/wp-content/uploads/WCD\\_Medium\\_Report\\_201504291.pdf](http://musicoclcusers.org/wp-content/uploads/WCD_Medium_Report_201504291.pdf)

Flaherty, K., & Morgan, K. (2019). The dangers of overpersonalization. Retrieved from <https://www.nngroup.com/articles/overpersonalization/>

Library of Congress, Policy and Standards Division. (2017). Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials. Retrieved from <https://www.loc.gov/aba/publications/FreeLCGFT/2017%20LCGFT%20intro.pdf>

Library of Congress, Policy and Standards Division. (2018a). Library of Congress Demographic Group Terms PDF files. Retrieved March 31, 2019, from <https://www.loc.gov/aba/publications/FreeLCDGT/freelcdgt.html>

Library of Congress, Policy and Standards Division. (2018b). Library of Congress Genre/Form Terms PDF files. Retrieved March 31, 2019, from <https://www.loc.gov/aba/publications/FreeLCGFT/freelcft.html>

Library of Congress, Policy and Standards Division. (2018c). Library of Congress Medium of Performance Thesaurus for Music PDF files. Retrieved March 31, 2019, from <https://www.loc.gov/aba/publications/FreeLCMPT/freelcmpt.html>

Library of Congress, Policy and Standards Division. (2019a). Introduction to Library of Congress Demographic Group Terms. Retrieved May 24, 2019, from <https://www.loc.gov/aba/publications/FreeLCDGT/2019%20LCDGT%20intro.pdf>

Library of Congress, Policy and Standards Division. (2019b). Introduction to Library of Congress Medium of Performance Thesaurus for Music. Retrieved June 2, 2019, from <http://loc.gov/aba/publications/FreeLCMPT/2019%20LCMPT%20intro.pdf>

McGrath, K. & Lowery, L. F. (2018a). Getting more out of MARC with Primo: Strategies for display, search and faceting. Presented at the ELUNA Annual Meeting, Spokane, WA. Retrieved from <http://documents.el-una.org/1746/>

McGrath, K., & Lowery, L. (2018b). Getting more out of MARC with Primo: Strategies for display, search and faceting. *The Code4Lib Journal* (41). Retrieved from <https://journal.code4lib.org/articles/13600>

Mullin, C. (2018a). New OCLC Music Toolkit for generating faceted music data [Blog post]. Retrieved from <https://tinyurl.com/y2sfhzpw>

Mullin, C. A. (2018b). An amicable divorce: Programmatic derivation of faceted data from Library of Congress Subject Headings for music. *Cataloging & Classification Quarterly*, 56(7), 607–627. <https://doi.org/10.1080/01639374.2018.1516709>

Music Library Association. (2019). MLA best practices. Retrieved from [https://www.musiclibraryassoc.org/mpage/cmc\\_mlabestpractices](https://www.musiclibraryassoc.org/mpage/cmc_mlabestpractices)

Online Audiovisual Catalogers. (2011). Library Of Congress Genre-Form Thesaurus (LCGFT) for moving images: Best practices. Retrieved from <https://tinyurl.com/y3fpokjv>

Online Audiovisual Catalogers. (2018). OLAC Video Game Vocabulary. Retrieved from <https://www.olacinc.org/olac-video-game-vocabulary>

Schiff, A. (2019). Using Library of Congress faceted vocabularies. Presented as an Oregon Library Association-Washington Library Association Preconference Workshop, Vancouver, WA. Retrieved from <https://faculty.washington.edu/aschiff/>

Taylor, A. (2000). Wynar's introduction to cataloging and classification. (9th ed.) Englewood, CO: Libraries Unlimited.

Young, J. L. (2017). Unlimited opportunities for enhanced access to resources: The Library of Congress' faceted vocabularies. Presented at Subject Access: Unlimited Opportunities, IFLA WLIC Satellite Conference, Columbus, OH. Retrieved from <http://library.ifla.org/2074/>

