

Retrospective Implementation of Faceted Vocabularies for Music

Efforts Led by the Music Library Association and Recommendations for Future Directions

A Technical Report

Prepared by Casey Mullin
Chair, MLA Vocabularies Subcommittee (2014-2018)
Head of Cataloging and Metadata Services, Western Washington University Libraries
April 19, 2018

Contents

Introduction.....	1
Background, Rationale and Parameters	2
Available Tools.....	3
Future Directions.....	4
Conclusion	6
Glossary	6

Introduction

This document describes the rationale and process for automatically generating faceted data for inclusion in descriptive metadata for music resources (e.g., MARC bibliographic and authority records). Over the past several years, MLA has collaborated with Gary Strawn of Northwestern University to develop specifications for generating faceted terms based on the presence of existing legacy metadata (mostly *Library of Congress Subject Headings (LCSH)*). Although implementation of newly-developed faceted vocabularies by music catalogers in current cataloging has reached a critical mass, the benefits of access to music resources offered by these new vocabularies will not be fully realized until a preponderance of music records in a given database carries these terms. The endeavor described here was prompted by that need.

Background, Rationale, and Parameters

In 2014, the years-long development of a new suite of LC faceted vocabularies began to come to fruition. These new vocabularies include the *Library of Congress Medium of Performance Thesaurus for Music* (LCMPT), the *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT), and the *Library of Congress Demographic Group Terms* (LCDGT). As these vocabularies developed, MARC elements have been specified to encode terms from these vocabularies and other faceted data. These include MARC fields 046 (Special Coded Dates), 370 (Associated Place), 380 (Form of Work), 382 (Medium of Performance), 385 (Audience Characteristics), 386 (Creator/Contributor Characteristics), and 388 (Time Period of Creation), defined in both the Bibliographic and Authority formats, as well as 655 (Index Term - Genre/Form), defined in the Bibliographic format.

Current implementation by music catalogers in the U.S. commenced almost immediately upon the release of the vocabularies, thanks in large part to training sessions at conferences and online, and to the best practices documents for LCMPT¹ and for music terms in LCGFT² promulgated and maintained on an ongoing basis by MLA. Some music catalogers have also begun to include other fields, such as the 046 and 370, in bibliographic records for musical resources and authority records for musical works and expressions.

Despite the rapid and enthusiastic uptake of these new vocabularies and other facet-friendly metadata elements by music catalogers in current cataloging (at least in the U.S.), full implementation of the faceted approach to indexing and discovery of music resources requires that all legacy metadata be enhanced with the same types of faceted data that catalogers are manually inputting now. In response to this imperative, in 2014 the Music Library Association's Subject Access Subcommittee (now the Vocabularies Subcommittee (MLA/VS)) began a multi-year project to analyze the content of LCSH music headings and MARC codes in bibliographic records for notated and performed music. The objective was to develop specifications for machine generation of faceted data that could be encoded in the aforementioned MARC fields.

Ideally, in such a retrospective process, each LCSH heading that describes what a music resource *is* (rather than what it is *about*) should beget at least one faceted data field. In most cases a heading will generate a medium of performance statement in a 382 field and/or one or more genre/form terms in 655 fields. In other cases, terms for audience and creator characteristics, coded dates, and geographic place terms can also be automatically generated.

¹ http://www.musiclibraryassoc.org/resource/resmgr/BCC_Resources/BPsForUsingLCMPT.pdf

² http://www.musiclibraryassoc.org/resource/resmgr/BCC_Resources/BPsForUsingLCGFT_Music.pdf

LCSH syntax for music headings is complex but systematic, fairly well documented³ and largely predictable. That said, the complexities of this syntax defy one-to-one crosswalking of terminology. Many LCSH headings are amenable to one-to-one conversion (e.g., LCSH **Old-time music** is equivalent to LCGFT **Old-time music**), but many headings contain multiple disparate components that must be “decoupled” in order to be repurposed as faceted data. For example:

650 _0 Sonatas (Viola and piano), Arranged \$v Scores and parts.

corresponds to the following faceted data

382 01 viola \$n 1 \$a piano \$n 1 \$s 2 \$2 lcmpt

655 _7 Sonatas. \$2 lcgft

655 _7 Chamber music. \$2 lcgft

655 _7 Arrangements (Music) \$2 lcgft

655 _7 Scores. \$2 lcgft

655 _7 Parts (Music) \$2 lcgft

Given the complexity of variables involved with LCSH pattern music headings, an enumerative table listing each possible LCSH permutation and its corresponding faceted data output would be impractical. Furthermore, an exhaustive description of these complexities is beyond the scope of this document. Suffice it to say, an effective process (or “Algorithm”) should be sufficiently detailed to account for all nuances built in to LCSH practice, but also succinct enough to be comprehensible by an implementer and actionable by a programmer.

Available Tools

In order to instantiate and prove in concept the viability of the intellectual endeavor of mapping LCSH headings to faceted data and their corresponding MARC fields, MLA engaged Gary Strawn of Northwestern University, whose prowess in developing tools to manipulate library descriptive metadata is well established.⁴ MLA/VS and Strawn proceeded to collaborate in developing machine-actionable specifications for retrospective generation of faceted data based on legacy metadata. MLA/VS developed the intellectual essence of the Algorithm, providing music expertise and deep knowledge of LCSH practice for music, and Strawn created a program (a Dynamic-link library, or “DLL”) that runs the Algorithm on MARC bibliographic records. Both the MLA Algorithm and Strawn DLL have been undergoing ongoing testing and refinement.

³ The LC Subject Headings Manual contains detailed instructions on formulating such “pattern” headings: https://www.loc.gov/aba/publications/FreeSHM/H1917_5.pdf

⁴ Strawn’s Authority Toolkit is one recent example: <http://files.library.northwestern.edu/public/oclc/documentation/>

Subsequently, at the urging of MLA/VS, in 2017 Strawn created an OCLC toolkit (the “Music Toolkit”), a macro that “calls” the DLL and writes the results of the DLL into a single bibliographic record within OCLC Connexion. The toolkit documentation is available online.⁵ The Algorithm, as instantiated in Strawn’s DLL is fully described in the document *Deriving 046, 370, 382, 385, 386, 388 and 655 fields in bibliographic records for notated music and musical sound recordings*⁶ and its accompanying spreadsheet.⁷ These documents are subject to ongoing revision by MLA/VS and Strawn as the Algorithm and DLL are refined over time.

The Algorithm documentation is freely available, and community feedback on it is encouraged. Additionally, music catalogers are encouraged to install and use the Music Toolkit in day-to-day cataloging,⁸ and to report unexpected behavior to MLA. Feedback on the Algorithm and Music Toolkit may be submitted using a Google form.⁹ Implementers wishing to test the DLL on entire bibliographic databases using batch processing should contact Strawn.¹⁰ Note that Strawn’s DLL source code is proprietary.

In addition to revising to its core program as needed, Strawn may amend his DLL with optional separate modules, which potential implementers should evaluate alongside the core program.

Future Directions

Although the MLA Algorithm and Strawn’s DLL and Music Toolkit have been refined and tested significantly already, MLA/VS recognizes their current limitations. To wit, these products are still in “beta” status. The following areas of study and development are in MLA/VS’s long-range plan:

- Analyzing multiple MARC fields in combination
- Linking faceted data fields that originate from a single source field

⁵ <http://files.library.northwestern.edu/public/Music382/documentation/>

⁶ Available at <http://files.library.northwestern.edu/public/Music382/Docs/>

⁷ The spreadsheet is included in the Music Toolkit installation package, and gets added to the operator’s local hard drive during installation, in the same folder as the DLL and the other configuration files. This folder will vary, depending on the Windows version, but the last element in the path will be \MusicDllWrapper. For example: C:\Program Files (x86)\MusicDllWrapper.

⁸ Note that the Music Toolkit strictly writes additional fields to existing MARC bibliographic records. It does not remove existing data, control headings, or replace records; these tasks are the responsibility of the cataloger. The cataloger is also responsible for reviewing MLA Best Practices and term scope notes in evaluating the results, and adjusting, deleting, or adding fields as necessary according to MLA Best Practices.

⁹ <https://goo.gl/forms/P0rbGQfxaGIMtrba2>

¹⁰ mrsmith@northwestern.edu

- Denoting the presence of machine-generated faceted data in a record by the use of a “marker” (such as the 883 field in MARC¹¹); additionally, devising a means to indicate that such data has subsequently been reviewed and remediated by a human operator
- Utilization of non-controlled terminology (e.g., 500 notes describing medium of performance)
- Utilization of authorized access points for musical works and expressions (e.g., medium of performance statements in subfield \$m, “arranged” statements in subfield \$o)
- Utilization of coded data in 045 and 048 fields
- Implementing the Algorithm on authority data for musical works and expressions
- Expanding the scope of the Algorithm to include moving image resources that include music
- Instantiating the Algorithm in non-MARC environments (e.g., MODS, BIBFRAME)
- Incorporating URIs for vocabulary terms into Algorithm output

The Algorithm and DLL will also need to be amended on an ongoing basis to incorporate new and revised terms in LCMPT, LCGFT and LCDGT.

It should be emphasized that while the Music Toolkit provides an excellent “laboratory” for testing the Algorithm and Strawn’s DLL, the ultimate goal of this endeavor is to perform retrospective implementation on entire databases. MLA recognizes that no instantiation of the Algorithm will ever be perfect, and that any full-scale implementation of the Algorithm will require a significant and thoughtful component of human review and remediation of Algorithm output. Strawn’s DLL and its associated documentation do account for this aspect, and implementers as well as other potential developers are advised to consider it as well.

Another long-term goal associated with retrospective implementation of LC faceted vocabularies in particular is the wholesale reassessment of LCSH practice for music. Many LCSH music form/genre/medium headings could be cancelled, now that equivalent methods exist in LC’s faceted vocabularies for describing the same attributes. Other headings may need adjustments in scope and granularity. MLA/VS will seek to collaborate with LC’s Policy and Standards Division to work towards making the appropriate changes to LCSH. Lastly, as LCSH practices for current cataloging are reduced and streamlined, certain LCSH headings can and should be removed from legacy metadata in order to ensure consistency across databases and mitigate retrieval problems (e.g., false drops) in discovery environments.

¹¹ <https://www.loc.gov/marc/bibliographic/bd883.html>

Conclusion

MLA endorses Strawn's DLL and Music Toolkit as the best means currently available for enabling retrospective implementation of music faceted vocabularies. MLA/VS will continue to refine the Algorithm, collaborate with Strawn and others to refine instantiations thereof, and facilitate the testing and implementation of the Algorithm on entire bibliographic databases (including OCLC WorldCat). Efforts to advocate for full-scale implementation of faceted vocabularies more broadly are described in the ALCTS white paper *A Brave New (Faceted) World: Towards Full Implementation of Library of Congress Faceted Vocabularies*.¹²

Glossary

Algorithm: The specifications developed by MLA's Vocabularies Subcommittee for automatically deriving faceted data from legacy music metadata, primarily LCSH headings for music but also select MARC codes.

DLL (Dynamic-link Library): The instantiation of the MLA Algorithm programmed and maintained by Gary Strawn.

Implementer: A cataloger, metadata creator/developer or database manager pursuing retrospective implementation of faceted data fields to metadata for music resources.

Music Toolkit: The OCLC macro, created by Gary Strawn in 2017, that runs the DLL on a single MARC bibliographic record in OCLC Connexion.

¹² Available here: <https://alair.ala.org/handle/11213/8146>