

A Brave New (Faceted) World:

Towards Full Implementation of Library of Congress Faceted Vocabularies

A white paper prepared by the Working Group on Full Implementation of Library of Congress Faceted Vocabularies, ALCTS/CaMMS Subject Analysis Committee, Subcommittee on Genre/Form Implementation

*Submitted to SAC on June 16, 2017
Re-submitted with corrections on July 13, 2017*

Working Group Members

| | |
|---------------------|---|
| Casey Mullin, chair | Western Washington University |
| Karen Anderson | Backstage Library Works |
| Lia Contursi | Columbia University Law Library |
| Beth Iseminger | Independent Music Cataloging Contractor |
| Mary Mastraccio | MARCIVE, Inc. |
| Robert Maxwell | Brigham Young University |
| George Prager | New York University Law Library |

The Working Group wishes to thank the following individuals for their input:

| | |
|--------------------|---|
| Liz Bodian | Colgate University |
| Jennifer Bromley | Library Consultant |
| Karen Coyle | Library Consultant |
| Rosemary Groenwald | Mount Prospect Public Library |
| Stephen Hearn | University of Minnesota |
| Daniel Joudrey | Simmons College |
| Adam Schiff | University of Washington |
| Matthew Wise | New York University |
| Janis Young | Library of Congress, Policy and Standards Division |

Table of Contents

| | |
|--|----|
| Executive Summary | 3 |
| Introduction | 4 |
| The Current Situation | 4 |
| A New Way Forward | 5 |
| Brief Histories of the New LC Vocabularies..... | 6 |
| LCGFT | 6 |
| LCMPT | 7 |
| LCDGT..... | 8 |
| Other Vocabularies Available for Non-Topical Attributes | 9 |
| Accomplishments in Training and Implementation | 10 |
| Resources Developed by the Library of Congress | 10 |
| Audiovisual..... | 11 |
| Music..... | 11 |
| Law | 12 |
| Cartographic Materials..... | 12 |
| Rare Materials | 13 |
| Revisions to Policy and Other Documentation Needed to Support Full-Scale Current Implementation... | 14 |
| LCGFT | 15 |
| LCMPT | 16 |
| LCDGT..... | 17 |
| Changes Needed to LCSH Practice | 19 |
| Challenges and Possibilities with Retrospective Implementation..... | 19 |
| Bibliographic Records--General Issues | 20 |
| Bibliographic Records--Music Resources | 22 |
| Bibliographic Records--Literature | 23 |
| Bibliographic Records--Law Resources..... | 25 |
| Authority Records--Use of LCMPT, LCGFT, and LCDGT | 27 |
| Display, indexing and searching of faceted non-topical attributes | 32 |
| Conclusion: The Argument for Full-Scale Implementation | 34 |
| Appendix: Working Group Charge | 37 |

Executive Summary

Over the past decade, the Library of Congress, in collaboration with the ALCTS/CaMMS Subject Access Committee and other constituencies, has developed a suite of new controlled vocabularies which collectively show potential for a new era in resource discovery. These vocabularies are designed to be used to describe various non-topical attributes of resources, attributes that have heretofore been described using LCSH headings and MARC control data fields, with mixed success.

This white paper summarizes work done thus far to develop and promote implementation of these new LC vocabularies, and suggests next steps for achieving full-scale current and retrospective implementation of faceted vocabulary terms in bibliographic and authority metadata. This new era represents a sea change, arguably on the same scale as RDA and Linked Data implementation; indeed, it will require significant buy-in and investment of time and resources by groups such as LC, the Program for Cooperative Cataloging (PCC), OCLC and other entities. Although these entities are the primary intended audience for this paper, further discussion and effort throughout the English-language cataloging community is desired.

The *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT), the *Library of Congress Medium of Performance Thesaurus* (LCMPT), and the *Library of Congress Demographic Group Terms* (LCDGT), have been developed by LC with significant assistance and collaboration from the ALCTS/CaMMS/SAC Subcommittee on Genre/Form Implementation and specialist cataloging communities. For the non-LC groups, this has included work on the content of the vocabularies themselves; updates to the MARC formats that enable the granular encoding of these attributes; best practices for cataloger implementation; training sessions, both online as webinars and face-to-face at conferences; and, research into programmatic retrospective assignment of faceted terms in legacy bibliographic metadata.

To bring these aforementioned efforts to fruition, an expansion in scope of effort is necessary. Though formidable talent exists within ALA and specialist communities, none of these groups alone has the resources to pursue full-scale current and retrospective implementation. Further, until a critical mass of bibliographic metadata includes these faceted attributes, the vision of optimal user discovery experience that these attributes enable remains out of reach.

Full-scale implementation requires a combination of broad and comprehensive training of catalogers (“current implementation”) and retrospective implementation through the development of nuanced and powerful machine algorithms. Work on both of these fronts is already underway, and now requires the buy-in and support of national and international entities like PCC, LC, OCLC and library systems vendors. Such support will enable the following:

- Comprehensive faceted vocabulary training for catalogers working in shared environments
- Routine creation of work-level (and in many cases expression-level) authority records for works (or expressions) which are embodied in or are likely to be embodied in multiple manifestations
- Retrospective implementation of faceted vocabulary terms using algorithms developed, vetted, and tested by expert communities

- Display and granular indexing of all faceted data, including (but not limited to) MARC bibliographic fields 046, 370, 382, 385, 386, 388 and 655 (or their equivalents in other encoding standards)
- Display and granular indexing of authority data, including attributes (including MARC fields 046, 370, 372, 374, 375, 380, 382, 385 and 386) and syndetic structures

Though the above components are expressed in terms of MARC metadata, the intellectual work will be transferrable to emerging metadata formats such as BIBFRAME, insofar as those formats are sufficiently granular and expressive.

Introduction

The Current Situation

In the days before computer-mediated access to library collections became ubiquitous, the card catalog consisted of three main components: an author file, a title file and a subject file. This tripartite mode of access was in keeping with Charles Ammi Cutter's first objective as described in his *Rules for a Dictionary Catalog* (first published 1876).¹ For the third of these card files, many libraries in the English-speaking world have deployed *Library of Congress Subject Headings* (LCSH), a robust list of terms developed and expanded over many decades since its establishment in 1898. Since LCSH is designed to be amenable for describing nearly any type of resource LC might collect (and, by extension, any resource most any library might collect), headings have been formulated to describe facts about resources that really aren't "subjects" at all. For example, LCSH headings exist to describe: literary genre and country of origin (**Short stories, American**); musical medium of performance (**Violin and piano music**); creator characteristics (**Holocaust survivors' writings**); intended audience (**Children's literature**) and time period of creation (**Rock music--1971-1980**), just to name a few. In the MARC bibliographic format, all of the above headings are designated with the same tag: 650; this further calcifies the conflation of these attributes from a machine processing perspective.

This conflation of distinct attributes may have been necessary in the staid card catalog environment, but with the advent of faceted browsing and limiting capabilities inherent in the current generation of library discovery tools, the problems caused by this conflation are amplified in this environment. A facet that is scoped to include headings found in MARC 650 subfield \$a will necessarily include many descriptors that do not fit within a (necessarily pithy) label applied to the facet such as "Topic." The presence of genre/form terms designated in MARC field 655 (coming from any number of possible thesauri, including LCSH) means that some discovery tools also have a separate facet for "Genre/Form"; however, this is misleading to the user as many genre/form headings are "trapped" in the "Topic" facet because of their MARC coding.² Some discovery tools feature additional facets that are populated with data from MARC 6XX subfields defined for geographic, chronological and form subdivisions. The FAST

¹ Freely available online in its several editions; the fourth edition can be found at <https://archive.org/details/rulesforadictio06cuttgoog> (viewed May 12, 2017).

² For this reason, some libraries have suppressed this facet altogether.

(Faceted Application of Subject Terminology) endeavor,³ developed by OCLC in collaboration with LC, seeks to facilitate this further by deconstructing subdivided LCSH headings into more user-friendly, independently-comprehensible faceted descriptors. This is a step in the right direction, but still leads to recall failures, since many LCSH headings are not subdivided when attributes that would be described by those subdivisions are inherent or implied by the root heading, and some qualifiers are not separately subfielded (e.g., **String quartets, Arranged; Anonymous writings, French**).

The left-anchored syntax of LCSH headings and subdivisions further hinders discovery, even in a “classic” browse environment, which is still the optimal method of engagement with LCSH. In the case of musical medium of performance, a user seeking music written for a flute and harp duet must be instructed to search under any heading that includes these two (and only these two) instruments: **Flute and harp music; Sonatas (Flute and harp); Suites (Flute and harp); Variations (Flute and harp)**, and so on. These headings occur some distance from one another in an alphabetical browse list, requiring multiple searches to amass a comprehensive pool of search results that satisfy the user’s query. Present-day users might forego this method altogether and simply execute a keyword search on “flute harp,” an approach that favors recall over precision. Such a Google-inspired solution surely satisfies the user in the moment, but the robust bibliographic metadata present in library catalogs can accommodate a much better experience, leading users to resources otherwise obscured from their view. Indeed, there is a third path, between the labor-intensive recourse to iterative browsing of an alphabetic index, and “quick and dirty” keyword searching with deceptively ineffective limiting mechanisms. This white paper lays out that path, building upon years of foundational work performed within the library metadata community.

A New Way Forward

Over the past decade, the Library of Congress, in collaboration with the ALCTS/CaMMS Subject Access Committee and other constituencies, has developed a suite of new controlled vocabularies that collectively show potential for a new era in resource discovery. Beginning with the *Genre/Form Terms for Library and Archival Materials* (LCGFT), an effort to better describe the nature of what resources *are*, in addition to or as opposed to what they are *about*, was underway. LC has sought to develop term lists that, unlike the longstanding LCSH, are true thesauri, conforming to the ANSI/NISO Z39.19 standard⁴. These include LCGFT as well as the *Medium of Performance Thesaurus* (LCMPT). The *Demographic Group Terms* (LCDGT) is not, strictly speaking, a thesaurus according to Z39.19, but it does display some of its features and, as described below, is purpose-built just like its peers. These vocabularies are designed to be used to describe various non-topical attributes of resources, attributes that have heretofore been described using LCSH headings and MARC control data fields, with mixed success.

This white paper summarizes work done thus far to develop and promote implementation of these new LC vocabularies. In particular, contributions by non-LC bodies are highlighted below. Most importantly, this paper suggests next steps for achieving full-scale current and retrospective implementation of faceted vocabulary terms in bibliographic and authority metadata. This new era of resource discovery

³ <http://www.oclc.org/research/themes/data-science/fast.html> (viewed May 12, 2017).

⁴ http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

represents a sea change, arguably on the same scale as RDA and Linked Data implementation; indeed, it will require significant buy-in and investment of time and resources by groups such as LC, the Program for Cooperative Cataloging (PCC), OCLC and other entities. Although these entities are the primary intended audience for this paper, further discussion and effort throughout the English-language cataloging community is desired. Although the LC vocabularies are not the only ones available for achieving these aims, they are arguably the most popular and amenable for widespread use in a shared metadata environment. Nonetheless, the suggestions below can be adapted for the purposes of implementing other faceted vocabularies as well.

The Working Group strongly advocates for full implementation of faceted vocabularies via the following five components, which will be addressed in more detail throughout the paper.

- Comprehensive faceted vocabulary training for catalogers working in shared environments
- Routine creation of work-level (and in many cases expression-level) authority records for works (or expressions) which are embodied in or are likely to be embodied in multiple manifestations
- Retrospective implementation of faceted vocabulary terms using algorithms developed, vetted, and tested by expert communities
- Display and granular indexing of all faceted data, including (but not limited to) MARC bibliographic fields 046, 370, 382, 385, 386, 388 and 655 (or their equivalents in other encoding standards)
- Display and granular indexing of authority data, including attributes (including MARC fields 046, 370, 372, 374, 375, 380, 382, 385 and 386) and syndetic structures

Brief Histories of the New LC Vocabularies

LCGFT

The Library of Congress' Policy and Standards Division (PSD) began developing a genre/form thesaurus with broad application in 2007. One purpose of this new thesaurus was to reduce the need to refer to multiple thesauri when supplying genre/form access. The new genre/form thesaurus, *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT), is distinct from *Library of Congress Subject Headings* (LCSH) because its terms describe what something *is* as opposed to what it is *about*. An additional distinction is that LCGFT terms are faceted, as defined by PSD. The PSD definition considers each facet a single concept (term or phrase).⁵ This means LCGFT terms are single words or phrases that convey one concept and are entirely contained within a single metadata element (e.g., in MARC Bibliographic Format, 380/655 subfield \$a). When complex concepts are needed, then multiple terms are used to bring out each aspect. For example, the film *Hobbit, the desolation of Smaug* would be assigned the genre/form terms **Action and adventure films**, **Fantasy films**, and **Feature films**, not **Action and adventure fantasy feature films**, **Fantasy films—Feature**, or something similar. In addition, LCGFT terms do not repeat data supplied elsewhere in the bibliographic record. For example, the language of the work can be coded in several places within the descriptive portion of the MARC record (e.g., in the

⁵ PSD Library of Congress Genre/Form Terms Manual (Online), p. GF-ii, viewed August 9, 2016. <http://www.loc.gov/aba/publications/FreeLCGFT/freeLCGFT.html>

008, 041, 130, 240, and 546 fields), so no provision is made for language within LCGFT.⁶ The same is true for creator and audience characteristics, which are instead handled with a separate thesaurus, LCDGT (described below). In addition, unlike LCSH headings, genre/form terms may not be subdivided in any way (topically, geographically, chronologically, or by form).

A phased discipline-by-discipline approach was undertaken to evaluate LCSH genre candidates and migrate terms to the genre vocabulary. In June 2010, after approximately 800 terms were established, it was determined that LCGFT should be formally separated from LCSH. A necessary step in the separation of the genre/form authority records was the revision of the LCCNs and MARC coding that indicated the term was part of LCSH. On May 24, 2011 the genre/form authority records, coded as LCSH, were deleted and reissued with new coding. The new coding used a prefix of “gf” for the LCCN and the 008/11 was set to “z”.

By early 2011, PSD had approved genre/form terms for moving images (films and television), non-musical sound recordings, cartographic resources, and law. Terms in three other disciplines (music, literature, and religion) and general terms were next to be developed. Throughout 2015, LC released large batches of approved authority records for general, music, literature, and religious genre terms. As of January 2016 there have been 1868 terms established. Additional music and literature terms are being reviewed and art is under development.⁷

LCMPT

Development on the *Library of Congress Medium of Performance Thesaurus* (LCMPT) began in 2009 as a collaboration between the LC’s Policy and Standards Division (PSD) and the Music Library Association’s Cataloging and Metadata Committee (then the Bibliographic Control Committee). Initially, the project addressed genre/form and medium of performance terms together, and was envisioned as a complete music thesaurus within the (still fairly new) LCGFT. Not long after work began, it was decided that medium of performance (i.e., the “instrumentation”, “scoring”, etc. of a musical work/expression) is a distinct bibliographic facet unto itself, and ought to be described by a separate thesaurus. Thus, work on LCMPT was launched in 2012 and released in full in February 2014. Though based roughly on medium of performance terminology residing in LCSH, LCMPT was built from the ground up as a true thesaurus (just like LCGFT), incorporating categorization concepts from LCSH (e.g., grouping terms into widely understood families of instruments such as wind, brass and string) as well as the taxonomy described by Sachs and Hornbostel in their seminal work published in the *Zeitschrift für Ethnologie* in 1914 (e.g., grouping instruments into aerophones, chordophones, idiophones, and so on). Thus, in the hybrid taxonomy of LCMPT, new terms can be placed in the hierarchy in the most precise and culturally neutral way possible.

Perhaps the most distinct feature of LCMPT is its granularity and the complex permutations of terms that can result from describing a musical ensemble in complete detail. Whereas LCGFT terms are employed as single concepts which, when combined, describe a resource kaleidoscopically, LCMPT

⁶ https://www.loc.gov/catdir/cpsso/genre_form_faq.pdf

Frequently Asked Questions about Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT) (Revised June 6, 2011).

⁷ PSD Library of Congress Genre/Form Terms Manual (Online), p. GF-v, viewed August 9, 2016.

<http://www.loc.gov/aba/publications/FreeLCGFT/freeLCGFT.html>

terms are designed to be combined into whole statements and be retrievable as such. That is, a work for viola and piano is assigned both LCMPT terms **viola** and **piano**, but must also be encoded to make clear that the work calls for one of each of these instruments, and two instruments in total. This permutation has been handled by LCSH syntactically (i.e., **Viola and piano music**). Conversely, LCMPT terms can be encoded in a purpose-built MARC field that has been refined over the past several years to be sufficiently expressive and machine actionable. To wit, what was encoded in a single subfield in LCSH (650\$a) is encoded thus in 382:

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

This allows retrieval on the individual terms, as well as on total number of performers. Unlike LCSH, whose syntax rules⁸ place an upper limit on individual medium components that may be described in a single textual string, LCMPT and MARC 382 allow for an indefinite number of terms and subfields. So, describing an ensemble of 15 different instruments results in a 382 with 15 occurrences each of \$a and \$n, and concluding with \$s 15. This increase in granularity poses a difficulty for retrospective implementation, as will be described below. Nevertheless, the intellectual limitations of describing musical resources fully in a controlled fashion, as enforced by LCSH, are not present in this new faceted environment. Other considerations, such as time limitations for the cataloger, can result in “shortcuts” that resemble LCSH headings (e.g., encoding only the single term **instrumental ensemble** instead of listing all 15 instruments). However, such truncations are dictated by practical need rather than imposed by the vocabulary itself.

LCDGT

During the development of LCGFT it was determined that there were other aspects of LCSH headings that would be out of scope for the faceted LCGFT vocabulary. LC’s policy with LCGFT (with a few exceptions) is to establish terms that represent only genre or form and not other aspects that may be included within LCSH form headings. For example, LCSH has always included headings that describe the audiences of resources and the creators and contributors of resources. Headings often include demographic groups (e.g., **Children’s films**; **Landscape painting, Australian**; **Nurses’ writings**), but subdivisions are also frequently employed to bring out the characteristics of the audience or the creators and contributors, as in the heading strings **English language—Conversation and phrase books (for businesspeople)** and **Emotions—Juvenile literature**, which denote audience, and **English fiction—Welsh authors**, which denotes the characteristics of the creators. Once LCGFT is fully implemented, these out-of-scope aspects will need to be recorded somewhere else in bibliographic and authority records. Two of these aspects are the category of persons who created or contributed to a work or expression and the intended audience for a resource.

MARC Discussion Papers Nos. 2012-DP04⁹ and 2012-DP05¹⁰ were formulated to provide input on how to handle audience characteristics and creator characteristics in MARC 21. These discussion papers led to

⁸ Described in the instruction sheet H 1917.5, “Music Form/Genre Headings: Medium of Performance”, in the Subject Headings Manual, available here: https://www.loc.gov/aba/publications/FreeSHM/H1917_5.pdf

⁹ <http://www.loc.gov/marc/marbi/2012/2012-dp04.html>

¹⁰ <http://www.loc.gov/marc/marbi/2012/2012-dp05.html>

acceptance of the MARC proposals to define the 385 and 386 fields in 2013. These proposals provided a limited set of examples for terms that could be used in these fields.

During the discussions on these fields and consultation with the ALA/CaMMS Subject Analysis Committee, staff at LC felt it would be beneficial to have a separate controlled vocabulary specific to demographic groups and creator characteristics that could be used in these fields. Since audiences, creators and contributors belong to demographic groups, terms in this vocabulary would be used to describe the characteristics of intended audiences, and creators and contributors of resources. This vocabulary is known as the *Library of Congress Demographic Group Terms* (LCGDT). The Subject Analysis Committee provided initial advice on the scope and structure of the LCGDT.

The LCGDT has been developed in three phases. The first two phases were undertaken by specialists within LC's PSD. The first phase began in May 2015 with LC staff putting out draft principles on the scope and structure of LCGDT under the title *Library of Congress Demographic Group Terms: Introduction and Guiding Principles of the Pilot*¹¹, and an initial 380 proposed terms for library community discussion. This phase was completed in June 2015 with the approval of 387 terms.

During Phase 2, from June 2015 to January 2016, PSD used feedback received from the community and further testing to revise existing terms, refine the guiding principles, and explore relationships. By January 2016, 440 new terms were created and a draft manual was published for community review- *Demographic Group Terms Manual*¹².

Phase 3 began in January 2016. This phase is intended to test the LCGDT principles and policies in a production environment. Proposals are being accepted from the library community for terms needed for new cataloging. These proposals should follow the guidelines and instructions contained in the draft *Demographic Group Terms Manual* and will be approved according to the same standard practices for LCSH proposals.

Other Vocabularies Available for Non-Topical Attributes

Although this white paper focuses on implementation of the LC faceted vocabularies, it should *not* be taken to imply that other vocabularies are obsolete or ineligible for use in 655, 382, 385, 386, or 656 fields of bibliographic records or related fields in authority records.

LCGFT appears in

- MARC Genre/Form Code and Term Source Codes, <http://www.loc.gov/standards/sourcelist/genre-form.html>.

This document enumerates thesauri and vocabularies that are eligible for use in bibliographic 655 fields. LCGFT appears in this list, but there are currently over 50 other authorized sources, in addition to LCGFT, any of which may legitimately appear in a bibliographic 655 field. Additionally, this document states that “Many general subject lists and thesauri, such as LCSH and the *Art and Architecture Thesaurus* (AAT),

¹¹ <https://www.loc.gov/catdir/cpsolcdgt-principles.pdf>

¹² <http://www.loc.gov/aba/publications/FreeLCGDT/freelcdgt.html>

also contain controlled vocabularies for specifying genre/forms. Thus those general sources may also be used in usage elements identified for genre/form terms, with the appropriate source code (see [Subject Heading and Term Source Codes](#)).” Vocabularies commonly used in bibliographic 655 fields, aside from LCGFT, include AAT, LCSH, GMGPC, and the rare book thesauri (e.g. rbgenr, rbtyp, etc.)

LCMPT appears in

- Music Instrumentation and Voice Code Source Codes, <https://www.loc.gov/standards/sourcelist/musical-instrumentation.html>.

Six other authorized vocabularies appear here which may be used in 382 fields. It seems likely that in PCC practice, however, only LCMPT will be used in this field.

LCDGT appears in three lists:

- Occupation Term Source Codes <https://www.loc.gov/standards/sourcelist/occupation.html>
- Gender Code and Term Source Codes <https://www.loc.gov/standards/sourcelist/gender.html>
- Subject Heading and Term Source Codes <https://www.loc.gov/standards/sourcelist/subject.html>.

Any of the occupation vocabularies, as well as appropriate terms from the general lists, may be used in bibliographic 385, 386, and 656 fields and authority 374 fields. Commonly used vocabularies, aside from LCDGT, include LCSH, DOT, and ILOT.

MARC authorizes any of the gender vocabularies to be used in authority 375 field; and appropriate terms from vocabularies in the general list may be used as well. PCC policy is being formulated at this time, but will likely recommend using terms from LCDGT in 375 in NACO name authority records.

Accomplishments in Training and Implementation

Resources Developed by the Library of Congress

The Library of Congress Policy and Standards Division published a draft of their *Genre/Form Terms Manual* in January 2016 to provide guidance for using LCGFT in cataloging.¹³ Additional information on LCGFT may be found on the PSD genre/form web page¹⁴. LC’s Cataloging Distribution Service provides a web page with free cataloging documentation and training resources, some of which relate to genre/form terms.¹⁵ An example of the conference and online presentations provided by LC is the

¹³ <http://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html>

¹⁴ <http://www.loc.gov/catdir/cpsd/genreformgeneral.html>

¹⁵ <https://www.loc.gov/cds/products/catManual.php>

webcast by Janis L. Young, *Enhancing Access to LC's Collections*, presented July 8, 2012.¹⁶ So far, much of the training has been basic coverage of the concept of using genre terms along with facets.

Audiovisual

Because the first set of terms published in LCGFT comprised motion picture related terms, the audiovisual community was among the earliest to begin implementation of LCGFT. LCGFT is now almost universally used for genre/form in audiovisual records. The organization Online Audiovisual Catalogers (OLAC) has published a best-practices guide for LCGFT.¹⁷

Music

In the music cataloging community, implementation of LCGFT and LCMPT in current cataloging has been rapid and enthusiastic. This is thanks in part to the release of best practices documents by the Music Library Association's Cataloging and Metadata Committee. The document *Provisional Best Practices for Using LCMPT* was released on April 7, 2014 (less than two months after the release of LCMPT) and re-released in a slightly corrected version on May 16, 2014. Significantly revised and expanded versions of this document, re-titled *Best Practices for Using LCMPT*, were released on February 22, 2016 and (most recently) February 14, 2017.¹⁸ The parallel document for LCGFT, *Best Practices for Using LCGFT for Music Resources, Version 1.0*, was released on June 8, 2015 (just four months after the initial release of LCGFT music terms), and re-released in a slightly corrected version on June 21, 2016.¹⁹

In addition to promulgating best practices, members of MLA's Cataloging and Metadata Committee have given numerous presentations and training sessions in the years since the endeavor began. Most noteworthy among these were a half-day workshop given to attendees of the Music OCLC Users Group (MOUG) 2015 Annual Meeting in Denver, Colorado²⁰, and an ALA/MLA-co-sponsored webinar in October 2015²¹. Response to these events has been positive, as has the overall level of engagement from members of the music cataloging community regarding the availability of the new terms and potential for their full exploitation, as evidenced by frequent and robust discussions on e-mail lists such as MLA-L and MOUG-L. With this and with the appearance of the new vocabulary terms in recently created and

¹⁶ https://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=5870

¹⁷ Freely available on line at www.olacinc.org/drupal/capc_files/LCGFTbestpractices.pdf

¹⁸ Freely available online at http://c.ymcdn.com/sites/www.musiclibraryassoc.org/resource/resmgr/BCC_Resources/BPsForUsingLCMPT_14022017.pdf

¹⁹ Freely available online at http://www.musiclibraryassoc.org/resource/resmgr/BCC_Genre_Form_Task_Force/BestPractices160621.pdf

²⁰ PowerPoint slideshow from this workshop is available here: <http://musicoclcusers.org/wp-content/uploads/LCMPT-LCGFTWorkshop1.pptx>

²¹ Archival slideshow available on SlideShare: <http://www.slideshare.net/ALATechSource/mla-workshop-introduction-to-lcs-music-medium-of-performance-and-genre-vocabularies>

upgraded bibliographic records in OCLC WorldCat , the authors conclude that a preponderance of music catalogers in the Anglo-American community have embarked upon current implementation.

Law

In the law cataloging community, implementation of LCGFT in field 655 has been uneven. LC approved the list of law genre/form terms in November 2010. As early as January 2011, catalogers on LC's Law Team started to use selective terms from the list. Other law libraries, including ones in the BIBCO program, are using LCGFT terms to the fullest extent possible. Some other law libraries are only accepting them in copy cataloging.

AALL's Classification and Subject Cataloging Advisory Working Group (CSCAG) is the group within AALL most involved with genre/form terms. It has developed brief guidelines intended to supplement the LCGFT manual, entitled: *Best Practices for the Application of Law Genre/Form Terms* (released June 2014; updated October 2015).²²

Questions on use of law genre/form (or other LCGFT terms) in field 655 are occasionally discussed on the CSCAG or AALL Cataloging email lists. CSCAG members also make SACO proposals for new and revised LCGFT terms as needed, or refer them to the SACO Law Funnel.

CSCAG has yet to discuss the use of law LCGFT terms in field 380 (form of work) and as qualifiers in authorized access points (AAPs). Catalogers of AALL need to decide whether they want to have community-specific guidelines, or if they prefer to rely upon (as yet unwritten) guidelines from PCC.

Among law libraries, LCDGT terms are mostly used in field 374 (Profession) and 375 (Gender) in personal name authority records, and as qualifiers in AAPs in bibliographic and authority records. Much less common is the use of LCDGT terms in fields 385 and 386 of bibliographic records or work/expression authority records, including series authority records. CSCAG has been considering whether to propose the use of LCDGT terms in preference to LCSH terms in fields 374 and 375 of authority records, and as qualifiers in AAPs. The use of these terms in fields 385 and 386 remains to be discussed.

AALL has not yet offered any training sessions on the use of LCGFT and LCDGT. It seems advisable to wait until some of the above policy questions are addressed.

Cartographic Materials

In mid-2009, LC PSD released the first set of LCGFT terms for cartographic materials. To the Working Group's knowledge, implementation among cartographic catalogers has proceeded successfully, facilitated by ALA's Map & Geospatial Information Round Table (MAGIRT).

²² Freely available online at https://docs.google.com/document/d/1wzSdAaAyyANfp0k-RTAd3cvGhu51D8-TQZplq_Yoal/edit?pli=1

Rare Materials

Implementation of LCGFT and the related vocabularies LCDGT and LCMPT in the rare materials cataloging community has been uneven and has not been particularly promoted. This is probably because the rare materials cataloging community has a well-established tradition of recording genre/form information using their own fully-developed and widely-implemented suite of vocabularies, the *RBMS Controlled Vocabularies* (<https://rbms.info/vocabularies/>), maintained by the ACRL/RBMS Controlled Vocabularies Editorial Group (<http://www.ala.org/acrl/rbms/acr-rbmcveg>). Unlike LCGFT, LCDGT, and LCMPT, these vocabularies are not intended to be completely faceted.

Rare materials catalogers were also among the earliest to embrace the bibliographic 655 field when it was first implemented in the mid-1980s: ever since rare materials cataloging has encoded genre/form in 655, where many other communities, including LC, continued for decades to use 650 to encode genre/form.

The rare materials cataloging community also has a longstanding practice of cooperating with other cataloging communities to ensure full integration of rare materials records into cataloging databases. For example, in addition to controlled genre/form vocabularies, RBMS continues to maintain its own relationship designator vocabulary (https://rbms.info/vocabularies/relators/alphabetical_list.htm). However, to ensure its integration into the wider sets of relationship designators the Controlled Vocabularies Editorial Group maintains a close relationship with the editors of the MARC Relator Codes/Terms vocabulary <http://www.loc.gov/marc/relators/> to ensure that all its relationship designators are included in the MARC list and are assigned a relator code.

Ways to link LCGFT (and where appropriate, LCDGT and LCMPT) vocabularies with the RBMS vocabularies should be explored, so that systems can understand when terms are equivalent. The Working Group recommends formal communication between the editors of the LC and RBMS vocabularies so that the vocabularies themselves (vs. the way they are implemented) can develop as much as possible along parallel paths. The rare materials cataloging community should promote the use of the new LC vocabularies in rare materials cataloging records, in tandem with the RBMS vocabularies.

Revisions to Policy and Other Documentation Needed to Support Full-Scale Current Implementation²³

The following recommendations are intended to illustrate the current state of cataloging documentation as it pertains to the use of faceted vocabularies. As such, they represent a starting point for discussion with the various stakeholders who are responsible for maintaining these resources. They should not be read as firm demands by the current Working Group.

²³ Policy documents referred to:

CONSER Cataloging Manual (CCM; revised modules) <https://www.loc.gov/aba/pcc/conser/more-documentation.html>; earlier edition is available in Cataloger's Desktop, a subscription product

CONSER Editing Guide (CEG); available in Cataloger's Desktop

Descriptive Cataloging Manual Z1 : Name and Series Authority Records (DCM Z1) (last updated Oct. 2016), freely available through at: <http://www.loc.gov/catdir/cpso/dcmz1.pdf> (and in Cataloger's Desktop)

Library of Congress Demographic Group Terms Manual (LCDGT Manual)"; draft, March 23, 2016), available at: <https://www.loc.gov/aba/publications/FreeLCDGT/freelcdgt.html>

Library of Congress Genre/Form Terms Manual (LCGFT Manual); draft, March 23, 2016), available at: <https://www.loc.gov/aba/publications/FreeLCGFT/freelcft.html>

Library of Congress-Program for Cooperative Cataloging Policy Statements (LC-PCC PS) available in RDA Toolkit, a subscription product; updated quarterly

MARC 21 Format for Authority Data (MARC 21 Authority) <http://www.loc.gov/marc/authority/>

MARC 21 Format for Bibliographic Data (MARC 21 Bibliographic) <http://www.loc.gov/marc/bibliographic/>

Music Library Association's document: Best Practices for Using LCMPT (revised Feb. 22, 2016) http://c.ymcdn.com/sites/www.musiclibraryassoc.org/resource/resmgr/BCC_Resources/BPsForUsingLCMPT_22022016.pdf

NACO Training (current Version: July 2013, with revisions through Nov. 15, 2016) <https://www.loc.gov/catworkshop/courses/naco-RDA/index.html>

NACO Series Training (Preliminary Edition: May 2014, with updates through May 2016)

PCC guidelines [for] 382 Medium of Performance (Feb. 2014?) <https://www.loc.gov/aba/pcc/rda/PCC%20RDA%20guidelines/Z01%20382%202014feb.pdf>

PCC RDA BIBCO Standard Record (BSR) Metadata Application Profile (June 13, 2016 revision) <http://www.loc.gov/aba/pcc/bibco/documents/PCC-RDA-BSR.pdf>

RDA CONSER Standard Record (CSR) Metadata Application Profile (June 15, 2016 revision) <https://www.loc.gov/aba/pcc/conser/documents/CONSER-RDA-CSR.pdf>

LCGFT

General recommendation: The addition of genre/form terms should be a core requirement for PCC BIBCO records, whenever they are appropriate. All PCC documentation should state this. (At some point in the future, genre/form terms would appear on authority records for works, and would not need to be manually added at the time of cataloging for every bibliographic record).

LCGFT Manual (*Still in draft form as of this writing*)

Recommendation: It needs to be finalized, and kept as up to date as possible. All other documentation that deals with genre/form should link to this manual.

BSR: In the section of the document about MARC data, there is this general statement about subject and genre/form access: “Use judgment in assessing each resource. As appropriate, assign a complement of access points that provide access to at least the primary/essential subject and/or form of the work at the appropriate level of specificity. Assign such access points from an established thesaurus, list, or subject heading system.” A hyperlink is provided to “Subject headings and term source codes” at <http://www.loc.gov/standards/sourcelist/subject.html>.

Recommendation: The language in this section should be strengthened to indicate that genre/form terms will often be applicable to a given resource, and in the case of certain domains like music and literature, at least one genre/form term will always be applicable. Currently, subject and form are given in a single sentence, which arguably conveys the idea that the two attributes are interchangeable, and that providing access to the primary/essential subject of the work alone satisfies the overall requirement.

More specific instructions are given for rare books: Rare books Adding genre/form terms from one of the rare books and manuscripts section (RBMS) Controlled Vocabularies is strongly recommended. Assign terms from other thesauri as appropriate. A hyperlink is provided to <https://rbms.info/vocabularies/index.shtml>.

Form of work (130, 240, 380 7XX). BSR policy is to record form of work (RDA 6.3) in these fields whenever it is needed to differentiate one work from another.

Recommendation: See general recommendation above.

CCM: There is no mention of genre/form headings in the revised module 15 on subject headings, nor, to the Working Group’s knowledge, does “form of work” appear to be covered elsewhere in the manual.

Recommendation: See general recommendation above.

CEG

Field 380: Not covered.

Field 655: Gives standard coding instructions, but no recommendations as to whether or not to use this field.

Field 755: Gives examples of the use of this field: Added entry—Physical characteristics, which can be coded from Lists of genre/form terms, with source given in \$2. This field has been obsolete since 1995.

Recommendation: Add a short section on field 380. Remove 755 documentation from CEG. See also general recommendation above.

CSR: Same general statement as in BSR about subject and genre/form access.

Form of work (130, 240, 380 7XX). CSR policy is the same as BSR

DCM Z1

380 Form of work instruction sheets state preference for taking terms from controlled vocabularies such as LCSH, LCGFT, or MeSH.

Recommendation: Prefer to use terms from genre/form vocabularies such as LCGFT, *Genre Terms: A Thesaurus for Use in Rare Book and Special Collections Cataloguing* (RBGENR), or AAT, rather than subject vocabularies such as LCSH, in fields where genre/form is expressed (380 in bibliographic or authority records, 655 in bibliographic records). If no suitable terms exist in LCGFT or another genre/form vocabulary, then supply a term from any other controlled vocabulary such as LCSH. If necessary, the cataloger may devise a suitable term.

LC PCC PS

6.3.13 Form of work. Offers guidelines on practice for series. Also states preference for taking terms from controlled vocabularies.

Recommendation: See above under DCM Z1 380.

MARC 21: Offers adequate coding guidelines for use of fields 380 in bibliographic and authority records, use of field 655 in bibliographic records, and use of x55 fields in authority records.

LCMPT

General recommendation: Strongly encourage the addition of 382 fields, using LCMPT terms, whenever appropriate in bibliographic and authority records.

Library of Congress Medium of Performance Terms Manual (in process)

Once published, this will be a definitive source for the use of these terms. The Music Library Association's *Best Practices for Using LCMPT* (referred to below) will continue to be a definitive (and likely significantly more detailed) source as well.

Recommendation: This manual should be completed as soon as possible, and kept regularly up to date, taking advantage to the greatest extent possible the assistance of the Music Library Association. All other documentation relating to LCMPT terms should link to this manual.

MARC 21 Bibliographic and Authority Format: The formats provide many examples of the proper coding for field 382 in bibliographic and authority records. The Authority format under x62, 162, 462, 562, and 762 offers basic coding guidelines for medium of performance terms from LCMPT and other controlled vocabularies.

DCM Z1. Detailed guidelines in 382 instruction sheet are offered for authority records. Most of these guidelines can be followed for bibliographic records as well. However, some discrepancies exist with

MLA *Best Practices*, which account for all recently defined 382 subfields (most notably \$e, \$r, and \$t), which are not authorized for use in NACO authority records as of this writing.

Recommendation: This instruction sheet should be updated and further harmonized with the MLA *Best Practices* and the forthcoming LC Medium of Performance Terms Manual.

PCC guidelines

382 Medium of Performance (undated, but from the file name, it appears that these guidelines are from February 2014).

Recommendation: Are these guidelines necessary? If so, they should be updated as needed, and a latest revision date should be included somewhere prominently in the document. If not, PCC could provide a link to the MLA *Best Practices*.

MLA's Best Practices for Using LCMPT

Self-described as a supplement to the forthcoming LC manual on LCMPT usage.

Recommendation: All other LCMPT documentation should link to this document, as well as to the LCMPT manual, when completed. Inasmuch as LC's forthcoming manual may only include broad guidance on term assignment, MLA's *Best Practices* should be seen as the canonical source for guidance on MLA's policies on LCMPT implementation.

BSR: It instructs to record medium of performance if needed to differentiate. (fields 130, 240, 382, 7XX).

Recommendation: See general recommendation.

CCM, CEG, & CSR: These are very general guidelines, and do not cover Medium of performance terms (which are unlikely to be used in serial records).

LCDGT

General Recommendation: The section below entitled "Authority Records--Use of LCMPT, LCGFT, and LCDGT" describes the Working Group's recommendations for the full deployment of LCDGT terms in authority records. PCC should consider these recommendations carefully and decide how and when these fields should be used in both national level bibliographic and authority records. Once PCC decides, then that policy should be reflected in all the policy documents below.

LCDGT Manual: The draft LCDGT manual offers detailed guidelines for the use of these terms in bibliographic and authority records. It focuses on Audience (field 385) and Creator/Contributor Characteristics (field 386). Use of these terms for gender is briefly covered in instruction sheet L 525, but only in the context of fields 385 or 386. These terms may also be used in name authority records for persons in field 375 (Gender). Similarly, use of these terms for occupation/field of activity are covered in instruction sheet 545, but also only in the context of field 385 or 386 (MARC bibliographic field 656 is not discussed, which may be used to describe the occupation and avocation reflected in the contents of materials. This field could use terms from LCDGT, but is not required and rarely appears in PCC records.)

Recommendation: Instruction Sheets 525 and 545 should briefly refer to the use of LCDGT terms for occupation and gender in personal name authority records, and field 656 for bibliographic records.

DCM Z1: 385 and 386 Instruction Sheets have only the general statement: LC/PCC catalogers may use this field but are encouraged to wait until best practice guidelines are developed by the PCC.”

Recommendation: DCM Z1 instruction sheets should refer the user to the *Library of Congress Demographic Group Terms Manual*, once this manual is no longer in the draft stage.

MARC 21: Offers adequate coding guidelines for use of fields 374-375, and 385-386. It also offers guidelines on use of bibliographic field 656.

BSR and CSR: They make no mention of using LCDGT terms in bibliographic records.

Recommendation: These documents need to be updated with some (brief) general information on LCDGT terms, and how they should be used, especially for fields 385 and 386. At a minimum, they should include a link to LC’s *Demographic Group Terms Manual*.

CCM: Module 4 is entitled “Persons, families, and corporate bodies associated with a work”, but it only deals with access points (fields 100-111 and 700-711). Use of audience or creator/contributor characteristics is not covered.

Recommendation: Once PCC policy has been determined, then a section should be added to Module 4 (or a new module) on the use of LCDGT terms, especially in fields 385 and 386.

CEG: No coverage of LCDGT or use of fields 385 and 386 in bibliographic or authority records.

Recommendation: Policy needed. Then brief information can be added under instruction sheets for fields 385 and 386 (they can refer the user to the LCDGT manual for more information).

LC-PCC PS

9.7 [Proposed]. RDA recommends that terms for gender be taken from a standard list, if available.

Recommendation: Add an LC-PCC PS, recommending a preference for terms from LCDGT (lacking a suitable term there, taking it from another controlled vocabulary).

9.16.1.3 *Recording Professions or Occupations* (100 field in bib. and auth.; 374 field in auth.). It recommends taking a term from a controlled vocabulary such as LCSH and MESH.

NACO [Names] Training and NACO Series Training: NACO training Module 2 (Describing persons) and Module 6 (Describing works and expressions) contain information about the use of LCDGT and recommends its use in authority records for persons and works/expressions (<https://www.loc.gov/catworkshop/courses/naco-RDA/index.html>). NACO series training materials are still under development and do not as yet include information about LCDGT.

Changes Needed to LCSH Practice

With full implementation of these faceted vocabularies, certain practices within LCSH will need to change. The role of form subdivisions, especially when they duplicate a genre/form term (in meaning if not in exact verbiage), is called into question. Even more significant are the entire areas of music and literature. For resources that *are* works of music or literature (not works *about* them), the preponderance of LCSH headings applied are not “subject” headings at all but rather headings that convey only form/genre, medium of performance, creator/contributor, audience, geographic, language and/or chronological characteristics. After these headings have been satisfactorily mapped to faceted terms and encoded in their proper designations, they ought to be removed entirely from bibliographic records. Moreover, in certain cases the corresponding LCSH authority records ought to be cancelled; these cases include music headings that are never appropriate for topical use, for example **Symphonies** (cf. **Symphony**, the corresponding topical heading) and **Quintets (Piano, clarinet, violins (3))** (arguably too specific and esoteric to be useful as a topical heading). This final phase of retrospective implementation should not be undertaken lightly, and semantic loss should be carefully avoided. Another factor to consider is the fact that many music form/genre/medium LCSH headings are actually pattern headings, not backed by authority records; thus, the cancellation of an authority record cannot always be relied upon to trigger removal of that particular string in bibliographic records. Instead, this will likely mean that the same criteria used to identify LCSH headings for faceted term generation should be used to identify such headings for removal from bibliographic records. Still, these are separate processes that needn’t (and arguably shouldn’t) occur simultaneously.

Decisions regarding changes in LCSH policy and consequent cancellation of authority records lie squarely within the purview of the Library of Congress Policy and Standards Division. Decisions regarding removal of LCSH headings from bibliographic records, on the other hand, are a matter for the administrators of bibliographic databases: OCLC in the case of WorldCat, and LC, consortia and other individual libraries in the case of their respective databases. While LC is a lynchpin in this process, all stakeholders should be involved in decisions at both the policy and execution levels. Given LC’s track record for collaboration through the SACO program and through the development of LCGFT, LCMPT and LCDGT, we are optimistic that similar collaboration will occur in this phase. That said, it cannot be taken for granted and so we acknowledge it as an imperative.

Challenges and Possibilities with Retrospective Implementation

Controlled retrospective implementation (i.e., the assignment of faceted vocabulary terms based on existing metadata such as LCSH headings) is critical to avoid non-retrieval issues caused by the genre/form information not being supplied in standard fields. To facilitate retrospective implementation, both a fully automated method and a semi-automated method, with manual intervention, must be employed. Consideration should be given to enhancing both existing bibliographic records and authority records.

At this time, LC/PSD has decided to retain LCSH form subdivisions to accommodate current systems.

Bibliographic Records--General Issues

Bibliographic records have considerable data that can be harvested to supply non-topical facets. Extensive algorithms must be written and exhaustive testing undertaken to ensure the highest percentage of accurate retrospective implementation.

There are limitations as to what can be safely done in a fully automated process. Even with manual intervention, there are some types of non-topical attributes that cannot be assumed for global insertion or remapping. The most problematic field to interpret is the 650 field (defined in MARC as “Subject Added Entry-Topical Term”), in which many LCSH headings are designated that are in fact genre/form headings. Accordingly, many 650 fields may be used to generate an appropriate genre/form term in 655 and possibly an audience term (385) or a creator/contributor term (386). However, 650 fields with content only in subfield \$a (i.e., those without subdivisions) cannot universally be interpreted as describing genre/form. Although many can (depending on the type of resource and discipline), there are certainly some 650 fields with content only in subfield \$a that are used to describe a topic rather than a genre/form. These cases require manual review and/or sophisticated logic. To wit, depending on the nature of the resource²⁴, certain fixed field codes can be utilized to determine whether a given 650 describes a topic or a genre/form.²⁵

While fixed fields have data that can be used to verify whether a 650 in the record describes a topic or a genre/form, sometimes the code refers to a portion of the content rather than the full item. Just because an item has the code for a form does not guarantee that the resource as a whole *embodies* that form; it may just refer to part of the item. For example, the codes for “Bibliography” (000/6-7=”am” + 008/24=”b”) can refer to the entire item or just a part of it.

Terms in LCSH that can describe either a language or an ethnic group can be quite difficult to identify and parse as separate facets. It may not be sufficient to build an algorithm based on how LCSH intends the qualifier to be used, since cataloger application may vary from LC’s intention (if that intention can be clearly ascertained from scope notes and the Subject Headings Manual; in many cases, it cannot). For example, in a shared cataloging environment one library may interpret “Hymns, English” to mean the hymns were written by English authors, another library might use it for any hymn originally written in the English language, while yet another might use it for any collection of hymns in the English language, even if translated from another language. An added layer of complexity is the difficulty in determining if the qualifier refers to the creator or the audience; in LCSH, the term “Children” is particularly problematic in this way.

The moving of medium of performance terms from the 650 field to 382 is more challenging but some can still be done in an automated process. A critical element in the algorithm is that the Leader/06 must be either “c”, “d”, or “j” before a 650 can be a candidate for automated mapping to the 382 Medium of Performance field. The parsing of 650 fields that include both genre/form and medium terms is discussed in further detail below.

²⁴ Music resources, for example, require special handling. See the section below.

²⁵ Arguably, through replacement with faceted vocabulary terms, certain fixed field data could be deprecated, and the longstanding practice of redundant input of the same information in coded form (i.e., in MARC fixed and control fields) and textual form (i.e., in MARC variable fields) could be discontinued. In this way, the redundancy in legacy data would not need to be perpetuated in a post-MARC environment.

Although it only contributes to a small portion of the genre/form and demographic terms that need to be generated, there is data in the bib record that can be easily mapped to the newly defined facets. An obvious place to start is with the subject fields, form subdivisions (6XX \$v). A few of the subdivisions do not have a direct mapping to an LCGFT term. Some of these have no equivalent and others have multiple possibilities that require manual review, or there may be only a specific sub-category with an established narrower term. However, a large number of the form subdivision terms have a direct match to a genre/form term (LCGFT) and possibly a demographic term (LCDGT).

A fully automated program can generate an appropriate 655 from all form subdivisions where a direct mapping is available. In some cases a 385 can also be generated from a Form subdivision. For example:

```
650_0 $a...$v Conversation and phrase books (for first responders)
385__ $n occ $a First responders $2lcdgt
655_7 $a Phrase books $2lcfgt
```

```
650_0 $a...$v Juvenile films
385__ $n age $a Children $2lcdgt
655_7 $a Motion pictures $2lcfgt
```

With the addition of faceted data, it may be possible to remove LCSH form subdivisions that have been successfully mapped to other fields. However, the context that the pre-coordinated LCSH string provides would be lost. Further study on this issue is needed.

In addition to the form subdivisions, there are a few fixed field codes that can facilitate the generation of non-topical facets. The list of usable codes is short but will supply some audience terms (385) and some genre/form terms (655). In addition, these codes can be used in an automated process to confirm whether a 650 genre/form term is about the genre/form or is the genre/form. For example:

```
If Leader/06-07=am + 008/22=c      insert 385__ $_n age $_a Preteens $_2lcdgt
If Leader/06-07=am + 008/33=p      insert 655_7 $a Poetry $2lcfgt

If 650_0 $a Detective and mystery stories
And Leader/6-7=am + 008/33=0      DO NOT match to a genre/form term

If 650_0 $a Jazz + Leader/06-07=am  DO NOT match to genre/form term
If 650_0 $a Jazz + Leader/06=j      generate      655_7 $a Jazz $2 lcfgt
```

Classification numbers are another useful bit of data to consult when determining if an item is a particular genre, rather than about that genre. For ease of programming and based on volume it is recommended to use Library of Congress Classification (LCC) found in the 050 (bib records) or 053 (authority records) fields. Given the size of the LCC, it would require a group effort to provide a table of genre terms (LCGFT), audience terms (LCDGT), and creator terms (LCDGT) uniquely associated with specific classification numbers. The program should start with terms that do not have a fixed field code, or other data, to identify the genre, but ideally, the table should provide genre terms for all unique LCC numbers used only for collections or works of a specific genre. One issue with classification numbers is that not all bib records, or authority records, have this data. Another issue is that even when the classification data is present it may not always be used consistently as defined by the LCC Schedules.

Ideally, a fully automated method will be used for these to produce the most significant return for the least expenditure of resources (staff and budget). Using a fully automated method for the easily programmable portions will allow time for manual review of more complex issues.

The ALA/ALCTS/SAC Genre/Form Implementation Subcommittee -LCGFT Working Group has compiled a 185/155 mapping spreadsheet for form subdivision terms as well as a fixed fields mapping spreadsheet to assist those wanting to generate 655 fields based on this data in the bib record.

It is important to note that not all libraries using field 655 have separately indexed this field in their OPACs or discovery systems. This severely limits accurate searching by genre/form, and remains a major barrier to widespread implementation of field 655. The reluctance to index this field is based on the likelihood that executing a search by a genre/form term will for the most part only retrieve materials cataloged within the past several years, therefore leaving users with a misleading impression about what the library has in a particular genre/form. Many library systems administrators will allow indexing of genre terms only when a critical mass of records has been retrospectively enriched with these terms.

Three disciplines where retrospective implementation presents particular challenges are described below.

Bibliographic Records--Music Resources

Retrospective implementation for music resources has numerous complexities that require careful study and detailed algorithms. Most notably, the preponderance of LCSH headings assigned to musical works/expressions are pattern headings, indicating medium of performance, with or without a genre/form term (e.g., **Quartets (Clarinet, flute, cello, double bass), Sonatas (Violin and piano)**). Other LCSH headings denote intended audience (e.g., **Piano music--Juvenile**), creator characteristics (e.g., **Jews--Music**), chronological coverage (e.g., **Jazz--1951-1960**), language (e.g., **Hymns, Latin**) and format of notated music (e.g., **Symphonies--Scores and parts**), in combination with form/genre and/or medium of performance. Some of these facets are separately subfielded in 650, but many are not, and so must be programmatically decoupled.

Three other difficulties complicate the automated mapping of LCSH music headings to faceted terms:

- 1) Certain genres/forms have an implied medium of performance. For example, **Symphonies** implies the medium of orchestra and **Chorale preludes** implies organ. If the medium of the resource being described differs from the implied one, it is given as a parenthetical qualifier (e.g., **Symphonies (Band)**). The absence of such a qualifier means that the implied medium can be mapped to the appropriate LCMPT term. Thus, the algorithm must take into account not only literal values within the source data, but also implied values therein based on knowledge of the subject heading system.
- 2) Medium of performance statements are often truncated and generalized, in cases such as ensembles with more than nine parts, accompanying ensembles, vocal soloists within ensembles that include chorus, and many others. In such cases, a machine algorithm can only perpetuate the low level of granularity present in the source LCSH data.
- 3) Art music and geographic biases are inherent to LCSH music headings. Works of art music are often described only in terms of their medium of performance; the art music aspect is implied. Secondly, geographic subdivisions tend to be applied much less frequently when the music

originates in the country of the cataloging agency than when the music originates elsewhere. An algorithm can be designed to account for the former, though the latter is much more difficult to mitigate.

Work is underway within the Music Library Association's Vocabularies Subcommittee to design and test algorithms to perform this decoupling and mapping. As of this writing, provisional algorithms for generating 046, 370, 382, 655, 385, 386, 388 and other fields from LCSH headings have been devised and are being refined. Gary Strawn (Northwestern University) has lent his programming and testing expertise to the project. Extensive testing and community feedback on these algorithms will be needed before they can be implemented in production databases. It is envisioned that they can be made available for local bibliographic and authority files (similar to RDA conversion programs recently authored by Strawn).

Bibliographic Records--Literature

As previously noted, genre/form information can be found in LCSH headings in 650 \$a or \$v, in \$v in other 6XX headings, and in some fixed and control fields. 650 fields can be problematic to parse as stated above. Literature works are especially difficult because it has been a longstanding policy not to add subject headings to individual works of literature. Before \$v was implemented, form subdivisions were coded in \$x, a designation shared with topical subdivisions. XX. But with the addition of \$v, form subdivisions are included with true "subject" headings in the 650 field. This creates a problem for retrospective implementation of genre/form terms literature since headings that might be used to generate 655 fields are not present in the records.

Another problem is there has been some confusion on the proper use of genre/form fields in literature records. The 655, which was defined to be used for genre/form, was not as widely used as it could have been. Today's practices result in these terms being specified in both 650 and 655 fields.

Most records created before current best practices do not have a 655 field to specify genre, relying instead on including genre information in the 650 field. This creates a serious problem for automated retrospective conversion because computer systems or programs have a very hard time differentiating between genre terms specifying the form and the same terms describing what the work is about. For the reasons outlined above, many records lack primary sources of information for retrospective conversion.

Another issue is that literature was one of the last disciplines planned to be added to LCGFT with the majority of the terms added between May and September 2015. So without literature terms to use, many people used LCSH terms in the 655.

Many terms in LCSH (the most widely used subject thesaurus) are not the same as those used in LCGFT; and many of these different terms do not have a see reference for the form used in LCSH. For example, the LCGFT term Detective and mystery fiction does not have a cross reference for the LCSH term Detective and mystery stories. This will make it more difficult to automatically update 655 fields containing LCSH headings to the correct LCGFT term. This will also make it difficult to automatically use the 650 fields with LCSH terms to add 655s with correct LCGFT terms. Separate tables and possibly complicated algorithms will be most likely be needed to define the matches for the needed LCGFT terms.

The issue with retrospective conversion of existing headings is developing logic to separate the information in the heading into the pieces that accurately separate LCGFT from the other information. The heading Children's stories, Argentine consists of three aspects: genre (stories), audience (children), and nationality (Argentine). The first aspect would need to be matched to LCGFT and the other two to LCDGT. The above section on general issues demonstrates some of these difficulties in pulling the different types of information out of a single heading to place it properly in a record and makes suggestions for how this could be accomplished.

The most common place for genre/form information is in headings with a \$v form subdivision. While many of the LCSH form terms can be mapped directly to LCGFT such as Fiction, Drama, and Poetry, there are many others that will need special handling in order to add an appropriate 655 field. Like main subject headings, subfield \$v terms can also include aspects that are out of scope for LCGFT and will have to be parsed in the same way as the main headings.

Subfield \$v may also be used with other actionable information in the record, like fixed fields, in order to get the best term added in 655 and to add audience and creator aspects in the appropriate fields. For example:

If a record has

650 _0 \$a ... \$v Juvenile fiction

Then the following could reasonably be added:

385 __ \$a Children \$2 lcdgt
655 _7 \$a Fiction \$2 lcgft

But if the record has

000/06=a and
008/22=a – Preschool and
650 _0 \$a ... \$v Juvenile fiction

Then the following with a more specific audience term could be added:

385 __ \$a Preschool children \$2 lcdgt
655 _7 \$a Fiction \$2 lcgft

There is some fixed field information that can be used to generate genre/form information. The 008 position 33 in the book format (Leader/06=a or t) is for Literary form. There are not very many codes but these codes can be used to generate at least a general genre term in a 655 field with \$2 lcgft.

The following codes in the 008 have direct equivalents in LCGFT and could be used to generate a 655 with a term from that vocabulary:

| | |
|------------------------|----------------------------------|
| 008/33=0 - Not Fiction | do not add 655 |
| 008/33=1 - Fiction | add 655 _7 \$a Fiction \$2 lcgft |

| | |
|---------------------------------|--|
| 008/33=d - Dramas | add 655 _7 \$a Drama \$2 lcgft |
| 008/33=e - Essays | add 655 _7 \$a Essays \$2 lcgft |
| 008/33=f - Novels | add 655 _7 \$a Novels \$2 lcgft |
| 008/33=h - Humor, satires, etc. | add 655 _7 \$a Humor \$2 lcgft |
| 008/33=j – Short stories | add 655 _7 \$a Short stories \$2 lcgft |
| 008/33=p – Poetry | add 655 _7 \$a Poetry \$2 lcgft |
| 008/33=s – Speeches | add 655 _7 \$a Speeches \$2 lcgft |

The 008/33 also has a code 'i' for Letters, but LCGFT does not have a general term that would correspond to this code. The Working Group recommends adding the general term **Letters** to LCGFT to serve as broader terms to **Business correspondence**, **Personal correspondence**, etc.. This would allow 008/33 code 'i' to be included in the above list.

Genre/form terms for literature added retrospectively to legacy data by automated processes will most likely be more general than those that could be determined manually with the work in hand. Despite the difficulties, such retrospective processing will provide at least some level of access to literary works by genre/form and is worth pursuing.

Bibliographic Records--Law Resources

The American Association of Law Libraries (AALL) Classification and Subject Cataloging Advisory Working Group (CSCAG) started the project of retrospective implementation of law genre/form terms about three years ago. In some cases the retrospective application was successful. However, CSCAG has been facing a number of significant challenges that has forced the group to carefully re-examine its searching methods for identifying appropriate sets of records.

One easily identifiable set of records consisted of records to which LCSH headings had been traditionally applied in order to indicate not what the resource is *about*, but rather what the resource really *is*. In fact materials such as constitutions, law reports and digests, and law reviews have the LCSH headings **Constitutions**, **Law reports**, **digests**, etc. and **Law reviews**, respectively. These were easy targets.

However, identifying most of the other law genres and forms is proving to be a lot more complicated. Focusing on one law LCGFT term at a time, CSCAG members construct Boolean searches that will identify and capture candidate bibliographic records suitable for that LCGFT term. As all members of the Working Group employ Innovative Interfaces software in their online public access catalogs (OPACs), they are able to experiment with the same searches. After reviewing the search results and filtering those records that seem inappropriate, catalogers apply the suitable LCGFT term. At the end of the process a cumulative list of MARC records will be compiled, accompanied with a crosswalk indicating OCLC number and the corresponding LCGFT term(s). Such lists will then be used to add the appropriate LCGFT terms to WorldCat records.

CSCAG members are acutely aware that the crosswalk list will never be fully comprehensive or 100% accurate, yet it will allow reaching the critical mass of records that will retrieve a sizable portion of bibliographic resources described by genre/form terms. The list of terms analyzed so far is: **Treaties**, **Casebooks**, **Constitutions** and **Legislative hearings**.

Other terms currently under scrutiny are: **Administrative decisions, Administrative regulations, Attorneys general's opinions, Bar journals, Court decisions and opinions, Court rules, Hornbooks (Law), Law digests, Law reviews, Legal forms, Legislative Histories, Official gazettes, Restatements of the law, Session Laws, Trial and arbitral proceedings, and Statutes & Codes.**

Here are some examples of strategies adopted so far:

Constitutions proved to be an easy case for the reasons explained above, yet a number of considerations needed to be taken into account to clean up the list.

Search strategy:

110:1#: exists
AND
240:10: has constitution
OR
240:10: has anayas
OR
240:10: has constit (or any other variation in several languages)

Caveat

The end result included some constitutional amendments, which do not belong to the list of records expected to be good candidates for the application of the genre term **Constitutions**. Therefore the list of records captured needed to be refined by identifying and eliminating those records that had the LCSH heading **Constitutional amendments**.

The term **Treaties** seemed initially as straightforward as **Constitutions**. However it proved to be a bit more complex.

Search strategy:

240 HAS treaty
OR 240 HAS treaties
OR 130 HAS treaty
OR 130 HAS treaties
OR 240 HAS protocol
OR 130 HAS protocol
OR 240 HAS convention
OR 130 HAS convention
OR 240 HAS agreement
OR 130 HAS agreement
OR 240 HAS accord
OR 130 HAS accord
OR 650 HAS treaties \$v collections
OR subject HAS \$v treaties
OR content=z

Option: Records with treaties in 730 and 710\$† may have text of treaties in appendix only, but may also be for works ABOUT treaties with full text in appendix only.

Caveat

The results included records describing congressional material with the SuDoc numbers starting with Y: those needed to be excluded. The results list also included congressional hearings which normally have class number KF24.8-KF27.5. These were excluded as well as other records containing the word hearing or hearings in the 245 field.

Records with SuDoc numbers starting with S (Department of State) need to be closely examined since they may well be texts of treaties.

Some records with the form subdivision Treaties might also be *about* treaties and should be attentively examined.

Records with field 110 and 1st indicator 1 (jurisdiction) could legitimately be treaties. However those with a 100 field indicating a personal author should be carefully scrutinized, since they are often works *about* treaties rather than the actual text of treaties.

Authority Records--Use of LCMPT, LCGFT, and LCDGT

These three new vocabularies are very valuable for adding descriptive elements to authority records. LCGFT, LCMPT, and LCDGT terms are used in work and expression authority records; LCDGT terms are also used in authority records for persons.²⁶

Use of LCMPT and LCGFT Terms in Authority Records

With the addition of more machine-actionable fields to the MARC 21 Authorities format there are now places to record more descriptive attributes for work and expressions in a way they can be made discoverable by today's systems. The new Heading Information fields that can use terms from these vocabularies include:

- 380 – Form of Work
- 382 – Medium of Performance
- 385 – Audience Characteristics
- 386 – Creator/Contributor Characteristics

Before the addition of these fields, there were few or no places to record this information except for the 6XX note fields which are very difficult for an automated system to parse and so not useful for discovery.

²⁶Post-MARC record scenario: At some point, the MARC formats will be replaced, and “records” will be superseded at least in part by clusters of discrete Linked Data statements (“triples”). In this paper, we refer to MARC 21 “records”, with the understanding that LCDGT terms and codes (or URIs representing the same) will still be useful for discovery purposes, whether or not the information is coded in MARC 21 and/or in discrete records.

Medium of Performance

Terms for medium of performance have long been employed in name authority records (NARs) for musical works and expressions, in \$m of the access point field. With the addition of the 382 field, this information can be separated out from the rest of the authorized access point where it can be more readily manipulated by a discovery system. The LCMPT vocabulary and the 382 field were designed to work together to create the specificity needed to uniquely identify musical works and expressions and provide greater flexibility and accuracy in describing them, which will in turn make them more discoverable, especially with Linked Data. It will be immensely beneficial for users to be able to identify works and expressions using particular mediums of performance.

Because there is a specific subfield where medium of performance has been encoded, it should be relatively easy to retrospectively add a 382 to authority record that have a \$m in the 1XX. It should be possible to write algorithms that can pull the terms from \$m, match them against LCMPT and put the correct term in a 382 field with \$2lcmpt. The algorithm would need to take into account the things that can be in \$m that are not really a medium of performance term, such as the words “accompaniment” and “unaccompanied,” as well as adding information beyond the term in the other subfields such as \$e for number of ensembles of the same type and \$n for number of performers of same medium. For example (emphasis added):

100 1# \$a Beethoven, Ludwig van, \$d 1770-1827. \$t Sonatas, \$m piano, \$n no. 14, op. 27, no. 2, \$r C# minor

382 ## \$a piano \$2 lcmpt

Or

382 ## \$a piano \$n 1 \$s 1 \$2 lcmpt

Since terms used in \$m are not always from a controlled list there may need to be further evaluation of terms that don't appear in the LCMPT records.

Form of Work

Before the creation of the 380 field for Form of Work, genre/form information were not widely used in authority records for works or expressions. There were few places where genre or form could be recorded.

Most often this information is found as a parenthetical qualifier after a title, mostly for differentiation purposes. These terms are in singular form and are not necessarily taken from a controlled vocabulary, although they are usually based on LCSH. At first, singular terms were also used in 380 and so the qualifying term in the 1XX was also used in the 380. LC and PCC practice now is to use a controlled vocabulary term in the 380 field. Genre/form terms in both LCSH and LCGFT are usually established in the plural form. It may still be possible to use the parenthetical qualifier to retrospectively create a 380 field by taking the plural form of the term and matching to LCGFT. For example:

130 ## \$a War of the worlds (Television program)

380 ## \$a Television programs \$2 lcgft

A few genre/form terms can be found in the subfield \$k in titles including Protocols, etc., and Selections. These may be used to retrospectively add a 380 as shown:

1XX __ ... \$k Selections
380 \$a Excerpts \$2 lcgft

1XX __ \$k Protocols, etc.
380 \$a Treaties \$2 lcgft

At ALA Annual 2015 the ALA/ALCTS/CaMMS/Subject Analysis Committee approved a proposal to allow recording subject and genre/form relationships in work and expression authority records using 5XX fields.²⁷ This would entail a change in PCC/NACO policy to allow 5XX fields to refer to thesauri outside of LC/NAF, and so it was forwarded to the PCC Policy Committee for further approval. The proposal would allow relationships such as the following to be recorded in work/expression authority records:

010 ## \$a nr2001025065
130 #0 \$a California biography series
550 ## \$w r \$i Subject: \$a Pioneers \$z California \$v Biography \$2 lcsb
555 ## \$w r \$i Genre/form: \$a Biographies \$2 lcgft
585 ## \$w r \$i Genre/form: \$v Biography \$2 lcsb

010 ## \$a no2012148566
100 1# \$a Netter, Frank H. \$q (Frank Henry), \$d 1906-1991. \$t Atlas of human anatomy
550 ## \$w r \$i Subject: \$a Human anatomy \$v Atlases \$2 lcsb
550 ## \$w r \$i Subject: \$a Anatomy \$2 mesh
555 ## \$w r \$i Genre/form: \$a Atlases \$2 lcgft

010 ## \$a no2014166453
100 1# \$a Card, Orson Scott. \$t Speaker for the dead. \$l Polish. \$h Spoken word
500 1# \$w r \$i Translator: \$a Cholewa, Piotr W. \$2 naf
500 1# \$w r \$i Narrator: \$a Siemianowski, Roch, \$d 1950- \$2 naf
555 ## \$w r \$i Genre/form: \$a Audiobooks \$2 lcgft
555 ## \$w r \$i Genre/form: \$a Translations \$2 rbgenr

This seems like a promising approach for recording form/genre information, particularly in the light of Linked Data, since 380 is not a linking field, and is not appropriate for use in records describing expressions. However, because of technical issues the proposal was passed on from the PCC Policy Committee to the Library of Congress, where it awaits action after nearly two years.

Recommendation: Urge the Library of Congress to proceed with overcoming the technical issues in implementing the CAMMS/SAC proposal.

²⁷ CaMMS Committee and Interest Group Reports, Annual 2015, available online at <http://www.ala.org/alctsnews/reports/ac2015-camms>

LCDGT Terms

LCDGT may be used to describe the audience characteristics of resources (MARC field 385), as well as the intended creators of, and contributors to, resources (MARC field 386). They may also be used in personal name authority records to express the occupation of the person (MARC field 374) and the gender of the person (MARC field 375). Gender of the intended audience or of the creator/contributor may also be coded into work or expression authority records in fields 385 and 386, respectively.

Adding LCDGT Terms Only to Authority Records, Only to Bibliographic Records, or to Both?

Personal name authority records are commonly created by PCC libraries for personal names associated with bibliographic resources. But this is only a subset of all personal names appearing in RDA English-language bibliographic records. A smaller subset of works and expressions are also represented by authority records, since many PCC libraries create these for series, or when variant access points are needed (such as in the case of works known by more than one title, or translations), or in the case of music, whenever an access point is needed in bibliographic record. Therefore, in the current environment, it may be beneficial to add these terms to both bibliographic and authority records. Once authority records are created for all personal names and works and expressions, then it should suffice to add these terms just to authority records. The authority records would be linked to the appropriate bibliographic records. The catalogers at the American Association of Law Libraries have discussed this issue and have expressed the preference of limiting the use of LCDGT terms to authority records.

Examples:

The examples below will illustrate some of the potential value of LCDGT terms in authority records to aid in discovery.

1. A policeman or detective is looking for a basic textbook on the Spanish language that is focused on his or her needs in law enforcement. If an authority record has been created for the work: Essential Spanish for policemen, lawyers, and judges, and it contains the appropriate audience terms, then all manifestations of the work could be retrieved by doing a subject or keyword search for "Spanish language", and then faceting by the audience term "Police", "Lawyers", or "Judges."

100 1_ Alcorta, Joe H. \$q (Joe Hermenegildo), \$d 1939- \$t Essential Spanish for policemen, lawyers, and judges

385 __ Police \$2 lcdgt

385 __ Lawyers \$2 lcdgt

385 __ Judges \$2 lcdgt

385 __ English speakers \$2 lcdgt

2. A reader wants to see if there are any collections of poetry written by male poets from Northern Ireland. A name authority record with the following Creator/Contributor fields in the authority record could retrieve the relevant resources.

100 1_ Longley, Michael, \$d1939- \$t Poems. \$k Selections
386 __ Northern Irish \$2 lcdgt
386 __ Males \$2 lcdgt

3. An French language reader is looking for a resource by a Quebecois female author for children.

100 1_ Watt, Mélanie, \$d 1975- \$t Scaredy Squirrel goes camping. \$l French
385 __ Children \$2 lcdgt
386 __ Quebeckers \$a Females \$2 lcdgt
400 1_ Watt, Mélanie, \$d 1975- \$t Frisson l'écureuil en camping

The advantage of having the appropriate LCDGT terms in the authority record for the work rather than the bibliographic record(s) is especially apparent in the case of works published in multiple manifestations; the terms would usually only need to be added to one record rather than many. In cases where the work exists in multiple expressions, it would also be useful to create authority records for the individual expressions, and to include LCDGT terms in those records as well, whenever the contributors to the individual expressions differ (e.g., different translations of a work). To take this concept one step further, in the case of works by a single creator, terms for creator characteristics would only need to be added to the authority record for that person; in doing so, these characteristics would not have to be repeated in every record for a work created by that person. Terms for audience characteristics, on the other hand, could be added to authority records for all works and expressions as applicable.

Realizing the full benefits of LCDGT terms in authority records for discovery purposes

The value will never be fully realized until the following conditions are met:

1. Authority records should be routinely created for all works, and for all expressions for which it is practicable to formulate a distinct access point.²⁸ Ideally, minimal level work and expression authority records would be generated automatically at the time of cataloging a resource. Generating NARs for all new personal names would also be helpful. Once authority records are created for all personal names, works, and expressions and are fully populated with the appropriate demographic group terms, it would no longer be necessary to add these terms to bibliographic records.²⁹
2. PCC policy should require or at least strongly recommend the addition of any applicable LCDGT gender and occupation terms in all personal name authority records. Additionally, LCDGT terms for creator/contributor characteristics should be added to work and expression authority records for aggregate works; works by a single creator could “inherit” their creator’s

²⁸ In a post-MARC environment, the access point consideration may no longer be relevant. That is, records/graphs/descriptions could be created for all expressions.

²⁹ The same holds true for LCGFT terms. If these are added to the authority record for the work/expression, it would be similarly unnecessary to add them to each bibliographic record embodying that work/expression.

characteristics as indicated in that person's authority record. LCDGT terms for audience characteristics, on the other hand, should be added to all work and expression authority records whenever applicable.

3. The *Demographic Group Terms Manual* needs to be made available in "final" form, and kept up to date. Widespread training via webinars, conference programs, etc. will be needed.
4. Authority records for all works and expressions that are readily identifiable in bibliographic metadata should be created retrospectively by automated means. Those records may be necessarily minimal in detail³⁰, but could also be populated with attributes common to a preponderance of bibliographic records for resources embodying the work or expression. Person authority records may also be amenable to this process.
5. Retrospective addition of LCDGT terms should be added to as many work and expression level authority records as possible, as well as to personal name NARs.
6. State-of-the-art discovery systems need to be widely available. They should offer fully functional faceting and/or searching by LCDGT terms.
7. Related to number 6, national and/or international standards for discovery systems and faceted searching are needed.

Display, indexing and searching of faceted non-topical attributes

The Working Group is convinced that moving toward faceted vocabularies will result in significant improvements in user experience, but acknowledge that these improvements will only come if retrieval mechanisms are significantly improved at the same time as current and retrospective implementation. Many of the recommendations in this paper entail replacing genre/form and related data currently encoded in 650 fields with faceted data divided among many fields. Researchers used to earlier ways the data were recorded will undoubtedly have difficulty finding information unless retrieval mechanisms are devised to help them. Consider the following cases:

100 1_ Haas, Georg Friedrich, \$e composer.
245 10 LAIR : \$b für Streichquartett (2014) / \$c Georg Friedrich Haas.
650 _0 String quartets.

becomes

100 1_ Haas, Georg Friedrich, \$e composer.
245 10 LAIR : \$b für Streichquartett (2014) / \$c Georg Friedrich Haas.
382 01 violin \$n 2 \$a viola \$n 1 \$a cello \$n 1 \$s 4 \$2 lcmt
655 _7 Chamber music. \$2 lcft

Users are accustomed to finding this sort of music using the term "string quartets." In the case of some string quartets both words might remain in the title of the piece and the word "quartets" might remain in the authorized access point for the work, but not in this case, and neither word remains in the genre/form or related data. Under the faceted data approach standard string quartets are intended to

³⁰ To cite a relevant precedent, a project to generate minimal name-title records for musical works and expressions (so-called "machine-derived authority records" or MDARs) was undertaken by the Library of Congress and OCLC in the 1990s. See <http://www.loc.gov/catdir/mdar.html> for more information.

be retrieved by the presence of the words “violin,” “viola” and “cello” in subfields a of the 382 field and the presence of the numeral “4” in subfield \$s and the numeral “2” in subfield \$n following the term “violin”.

041 1_ eng \$h rus
245 00 Pre-Revolutionary Russian science fiction : \$b an anthology : (seven utopias and a dream) / \$c edited and translated by Leland Fetzer.
650 _0 Science fiction, Russian \$v Translations into English.

becomes

041 1_ eng \$h rus
245 00 Pre-Revolutionary Russian science fiction : \$b an anthology : (seven utopias and a dream) / \$c edited and translated by Leland Fetzer.
386 Russians. \$2 lcdgt
655 _7 Science fiction. \$2 lcgft

Under the faceted data approach Russian science fiction translated into English is retrieved by (1) the coding of 041, which encodes the information that the resource is a translation from Russian into English; (2) the presence of the 386 field, which encodes the information that the creators were Russian; and (3) the presence of “Science fiction” in the 655 field. The terms “Russian”, “English”, and “translations” no longer appear anywhere in the genre/form and related data.

Even the most sophisticated of searchers might very well be stymied in both of these cases, and most of our users are not sophisticated. It is therefore crucial that, as the recommendations of this white paper for retrospective transformation of the data proceed, intuitive search and retrieval mechanisms be developed that will help users find the newly-encoded information.

Recommendation: *Urge utilities such as OCLC and SkyRiver, vendors such as III, Sirsi, and others, and national databases such as LC to quickly develop easy to use search and retrieval mechanisms designed with unsophisticated users in mind to help searchers find materials using faceted genre/form and related vocabularies.*

Display of the new fields designed to accommodate faceted information is also an issue. Sensible user displays in the traditional "labeled" environment are essential.

The simpler fields, such as 385 or 386, may be straightforward, so long as only one term is included in a given field:

386 Russians \$2 lcdgt

could display:

Creator category: Russians

385 Children \$2 lcdgt

could display:

Audience: Children

Display of 382 is more complex.

382 01 violin \$n 2 \$a viola \$n 1 \$a cello \$n 1 \$s 4 \$2 lcompt

could display

Performers: violin (2), viola (1), cello (1); total performers: 4.

Such a display might make sense to database users (though the singular "violin" might appear odd), but early experimentation in a Sirsi system has demonstrated that it is not easy to achieve this display from the contents of the 382 field.

Another important indexing and display issue is that of authority records. This document has noted the desirability of migrating some of this data out of bibliographic records into authority records and not duplicating it between the two types of records. If this happens it will be essential that the information in authority records both display to the public and be included in the public indexes of the database. Even if the inefficient practice of duplicating the information in bibliographic and authority records continues, it remains an important outcome that the rich data in RDA authority records display and index. This also means that authorized access points and preferred terms in authority records must be linked to their associated access points in bibliographic records in a machine-actionable way (i.e., encoding a stable identifier in the bibliographic metadata instead of or in addition to a textual string). Further, it is imperative that authority records be re-envisioned and repurposed as descriptions of *first-order entities of interest* capable of leading the user to related entities of interest, whether these are library resources, persons, works, concepts, or otherwise.

Recommendation: *Encourage vendors and utilities such as OCLC to use data from authority records when fulfilling user searches of the database, and find ways of displaying authority data to the public.*

The Working Group acknowledges that the details of display and indexing specifications are out of scope of the present paper. However, a group consisting of user experience design experts and metadata experts should be formed (as soon as possible) to study the issue further. The present paper outlines the "what," whereas a future report from that group could articulate the "how," and serve as a set of functional requirements upon which library system vendors and developers of open-source tools can base their design decisions.³¹ It must be emphasized that the status quo across the majority of library discovery environments is ill equipped to support the recommendations outlined here.

Conclusion: The Argument for Full-Scale Implementation

³¹ Recommendations of this nature for music resources in the OCLC WorldCat Discovery environment were recently promulgated by a joint MLA/MOUG task force. The report is available at: <http://www.musiclibraryassoc.org/blogpost/1258266/252550/OCLC-Search--Discovery-Task-Force-Report-available> (viewed June 8, 2017).

As should be evident from the complexities described above, full-scale implementation will require a broad cooperative effort among many stakeholders in order to achieve the promise of a fully-faceted discovery environment. Specifically, the Working Group strongly advocates for the following components as a minimum:

- Comprehensive faceted vocabulary training for catalogers working in shared environments; revision of metadata application profiles (e.g., BIBCO Standard Record) and other relevant documentation to require or strongly encourage faceted vocabulary terms where applicable, in both authority records and bibliographic records (or Work descriptions in a possible future BIBFRAME environment)
Required participants: Library of Congress (LC), Program for Cooperative Cataloging (PCC), specialist cataloging communities
- Routine creation of work-level (and in many cases expression-level) authority records for works (or expressions) which are embodied in or are likely to be embodied in multiple manifestations³²; inclusion of attributes described by faceted vocabulary terms that ideally will, eventually, no longer need to be replicated in each manifestation description (i.e., MARC bibliographic record)
Required participants: Library of Congress (LC), Program for Cooperative Cataloging (PCC), specialist cataloging communities
- Retrospective implementation of faceted vocabulary terms using algorithms developed, vetted, and tested by expert communities. Though such algorithms will never be perfect, machine generation of terms with a reasonable level of confidence should be pursued, and records should be flagged for human review as appropriate; updates to the MARC formats may be needed in order to properly encode such meta-metadata
Required participants: Library of Congress (LC), OCLC, library consortia, MARC Advisory Committee (MAC), specialist cataloging communities
- Display and granular indexing of all faceted data, including (but not limited to) MARC bibliographic fields 046, 370, 382, 385, 386, 388 and 655 (or their equivalents in other encoding standards)
Required participants: ILS/discovery layer vendors, open-source communities, expert user experience (UX) design working group
- Display and granular indexing of authority data, including attributes (including MARC fields 046, 370, 372, 374, 375, 380, 382, 385 and 386) and syndetic structures; this should be seen as a basic level of functionality in discovery layers rather than a special enhancement
Required participants: ILS/discovery layer vendors, open-source communities, expert user experience (UX) design working group

³² This is already a standard practice among LC music catalogers and NACO-Music Project participants; name-title records are routinely created for musical works as needed and as resources allow. The Working Group suggests formalizing and expanding this practice in other cataloging communities.

Programmatically generated, broadly-applied implementation of faceted vocabularies in existing bibliographic and authority metadata is possible and will contribute to positioning libraries strategically for the imminent embrace of ontologies and Linked Data.

Generation of terms from the new vocabularies using existing bibliographic data is one important step in converting legacy library data into Linked Data. Linked data conversion will be facilitated by the presence of these more granularly organized vocabularies, with their explicit encoding of descriptive attributes. As has been mentioned, the new vocabularies offer greatly enhanced specificity over more cumbersome vocabularies like LCSH; this specificity will be much more effective if the process is executed on a large scale.

To be most useful, and to serve the purposes of better access and Linked Data, LCGFT, LCMPT, and LCDGT must be implemented on large corpora of data--specifically, the WorldCat database and the LC/NAF. It would be most beneficial if the automated generation of this data is produced and rolled out in a process similar to the one used to update the LC/NAF and other library data to RDA standards: a program was made available free-of-charge and could be run on any library's database to update specific pieces of data to RDA compliance.

This Working Group recommends that OCLC, the Library of Congress, and the Program for Cooperative Cataloging support this endeavor. Such backing will demonstrate the importance of implementation to other library administrators and will encourage the needed wide-scale participation that will make this data effective. Endorsement from these leading organizations will help garner the backing of institutions across the spectrum.

Appendix: Working Group Charge

The Working Group will draft a white paper, whose intended audience will include OCLC, the Library of Congress, the Program for Cooperative Cataloging, and other constituencies, including ILS and authority control vendors, advocating for the full implementation (current and retrospective) of new and emerging faceted vocabularies, most notably the *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT), the *Library of Congress Medium of Performance Thesaurus for Music* (LCMPT), and the *Library of Congress Demographic Group Terms* (LCDGT). The paper should at minimum address the following points:

- The broad rationale for full implementation, including historical problems with the use of Library of Congress Subject Headings (LCSH) to describe what a resource *is* (as opposed to what it is *about*)
- Recommendations on how to make use of LCDGT very broadly, to better describe entities, when appropriate, by taking advantage of this vocabulary that is purpose-built to describe demographic characteristics
- The need for retrospective implementation in legacy bibliographic records, i.e., machine-mediated generation of faceted terms in MARC bibliographic fields such as 382, 385, 386, 388, and 655, based on the presence in those records of LCSH headings and data in other MARC fields
- The need for retrospective application of these faceted vocabularies in LC/NACO authority records and full implementation in newly created name authorities
- Search and display improvements needed to fully exploit the terms, their syntax, and syndetic structure; these should be expressed as broad functional requirements rather than suggestions to improve particular discovery environments
- Training needs for broad(er) implementation in current cataloging
- Recommendations on any revisions needed in LC, PCC, OCLC, or other cataloging policy and practice documentation, for example, the *Descriptive Cataloging Manual Z1*, *CONSER Cataloging Manual*, *BIBCO Participants' Manual*, *OCLC's Bibliographic Formats and Standards*, etc.

The Working Group should consult and seek input from other constituencies and committees with an interest in these issues, such as the PCC Standing Committee on Automation, the Music Library Association, the American Association of Law Libraries, etc.

The Working Group will present a preliminary version of the paper for discussion at the ALA Midwinter Meeting in January 2017. The final version shall be presented to the subcommittee by the end of May 2017.