

2016 Satellite meeting - *Subject Access: Unlimited Opportunities*
11 – 12 August 2016
State Library of Ohio, Columbus, Ohio, USA

Unlimited Opportunities for Enhanced Access to Resources: The Library of Congress' Faceted Vocabularies

Janis L. Young

Policy and Standards Division, Library of Congress, Washington, D.C., United States

E-mail address: jayo@loc.gov



Library of Congress by Janis L. Young. To the extent possible under law, the person who associated CC0 with Library of Congress has waived all copyright and related or neighbouring rights to Library of Congress. See <http://creativecommons.org/publicdomain/zero/1.0/>

Abstract:

Library of Congress Subject Headings (*LCSH*) is the most widely used subject vocabulary in the world. It includes headings that represent every discipline, as well as headings describing the genres and forms of resources, the characteristics of their creators and their intended audiences, and, in the case of musical works, their mediums of performance. The benefits of *LCSH*'s comprehensiveness are offset, though, by inconsistently formulated headings that inhibit both intellectual understanding and algorithmic manipulation.

In 2007, the Library of Congress began to develop three vocabularies that will enhance retrieval by providing direct access to non-topical aspects of resources: Library of Congress Genre/Form Terms for Library and Archival Materials, the Library of Congress Medium of Performance Thesaurus for Music, and Library of Congress Demographic Group Terms.

This paper discusses why the exclusive use of *LCSH* is insufficient for today's resource discovery needs. It will then describe the scope and structure of each of the three new vocabularies, and conclude by showing how the new vocabularies will provide enhanced access to library materials when used alongside *LCSH* or another subject vocabulary.

Keywords: *Library of Congress Subject Headings; Library of Congress Genre/Form Terms for Library and Archival Materials; Library of Congress Medium of Performance Thesaurus for Music; Library of Congress Demographic Group Terms*

INTRODUCTION

Library of Congress Subject Headings (LCSH) is the most widely used subject vocabulary in the world, either in the original English or in translation. It was first published in parts beginning in 1909 under the title *Subject Headings Used in the Dictionary Catalogues of the Library of Congress* and has been continually updated ever since, with thousands of new and revised headings approved every year. In January 2016, it included approximately 340,000 individually established headings, but many millions more can be assigned on-the-fly by catalogers who follow rules to add subdivisions to main headings, thereby creating unique heading strings.¹

The popularity of LCSH is due in large part to its breadth of coverage. It includes headings that represent topics in every discipline, as well as headings that represent other facets (that is, other identifiable aspects) of resources. Some describe the genres and forms of resources – what they *are*, instead of what they are *about*. Others describe the characteristics of the creators and intended audiences of resources, while still others describe mediums of performance for musical works (e.g., a work for piano and soprano voice).

Its comprehensiveness makes LCSH applicable to a wide variety of settings, and the use of a single vocabulary to serve many needs is convenient for users and catalogers alike. It can also limit the ability of libraries to provide effective access to materials, particularly when clear distinctions in heading format and style are not maintained between and among headings that serve different purposes. LC subject headings do not maintain those distinctions.

Genres and forms, mediums of performance, and the characteristics of creators and audiences can all be indicated through main headings. In LCSH, such headings are collectively referred to as *form headings*, as opposed to *topical headings*, although most do not indicate the form of resources. (*Form* more properly refers to the format and/or purpose of resources.) Some typical form headings, along with the facets they describe, are shown in Table 1 below.

Table 1: Examples of LCSH Form Headings

<i>Heading</i>	<i>Facet of the resource</i>
Banjo music	Medium of performance
Braille books	Intended audience
Cancer patients' writings	Creator
Pastoral literature	Genre
Short films	Form

Other form headings combine multiple facets. For example, **Romance fiction, Zimbabwean (English)** describes the genre and form (*Romance fiction*) and the nationality and language of the creator (*Zimbabwean* and *English*, respectively). In the heading **Children's plays, Chinese, children's** describes the audience; *plays*, the form; and *Chinese*, the language of the author. **Concertos (Piano, trumpet, viola)** indicates the form (*concertos*), while the mediums of performance are provided in the parenthetical qualifier. The heading **Piano music (Blues)** is formulated in the opposite way, with the medium of performance in the initial position and the genre in the qualifier.

Sometimes the non-topical facets of a resource appear in subdivisions instead of in main headings, a practice that relegates them to secondary position. The heading **English language—Textbooks for foreign speakers** begins with a topic, and the subdivision indicates the form of the resource (*textbook*) and the intended audience (*foreign speakers*). Likewise, the heading **Libraries—Juvenile literature**

begins with the topic and the resource's intended audience, *children*, is reflected in the subdivision. **Japanese fiction—Buddhist authors** refers to the form (*fiction*) and the characteristics of the authors (*Japanese* and *Buddhists*). In **Trumpet—Methods (Jazz)**, *trumpet* is the medium of performance, *methods* is the form, and *jazz* is the genre.

The inconsistent combination of multiple facets into a single heading can cause confusion, but LCSH form headings also hinder resource discovery in two other ways. First, identically structured headings may describe different facets of resources. For example, the heading **Children's diaries** is used to describe diaries *by* children, while **Children's films** describes films *for* children. Second, and finally, most form headings can also be used to describe resources *about* the forms, genres, etc.

In summary, LCSH combines the topical, genre/form, creator, audience, and medium of performance facets in contradictory and sometimes unpredictable ways, and even headings that are similarly formatted may denote quite different facets. Those problems are only exacerbated by the fact that form headings can also usually be used as topics.

These idiosyncrasies make it impossible to algorithmically determine which headings describe resources *of* a particular genre or form, or *by* particular class of authors, or *for* a particular audience or medium of performance, and which headings describe resources *about* those resources. The result is twofold: subject browse lists interfile headings of all types, requiring users to discern between hits that are and are not pertinent to their needs; and it is not possible to take full advantage of the new discovery interfaces, which often provide options for limiting result sets to a genre or form, a particular audience, a creator group, or a medium of performance.

The Library of Congress is developing three vocabularies that will enhance retrieval and precision by providing direct access to several non-topical facets of resources. They are the *Library of Congress Genre/Form Terms for Library and Archival Materials*, the *Library of Congress Medium of Performance Thesaurus for Music*, and *Library of Congress Demographic Group Terms*; the last describes the characteristics of creators and intended audiences. Each of these vocabularies is structured to take advantage of the new discovery interfaces, but will also enhance access in traditional library catalogs.

DEVELOPMENT AND STRUCTURE OF THE NEW VOCABULARIES

The new vocabularies will serve a twofold purpose: to simplify metadata creation and to provide a better discovery experience for users. This section of the paper will describe the scope and purpose of each of the vocabularies, and briefly explain their structures.

Development of *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT) began in 2007, and it initially consisted only of terms describing the genres and forms of moving images (primarily films and television programs). In recent years, though, it has been expanded to include terms describing spoken-word sound recordings, cartographic, legal, and religious resources, music, and literature, as well as “general” terms that are not specific to a discipline (e.g., dictionaries, encyclopedias). As of January 2016, it included over 1,800 authorized terms.²

Since LCSH form headings describe audiences, creators, and mediums of performance in addition to genres and forms, the Library of Congress knew that vital information would be lost when LCGFT is fully implemented unless accommodation for the other facets could be devised. For example, a sonata for clarinet and piano would normally be assigned the LC subject heading **Sonatas (Clarinet and piano)**, but eventually the genre/form term **Sonatas** will be assigned instead. Since mediums of

performance are not genres and forms, they cannot be included in LCGFT, would not therefore be included in the cataloging record.

After exploring various options, the Library of Congress and the Music Library Association decided to develop a separate controlled vocabulary specific to musical mediums of performance. Development of the *Library of Congress Medium of Performance Thesaurus for Music* (LCMPT), a project LC undertook jointly with MLA, began in 2009. By January 2016, LCMPT included 827 authorized terms.³

Similar concerns drove the decision to create a separate controlled vocabulary to describe the creators and intended audiences of resources, which, like mediums of performance, are not eligible for LCGFT. After the literature terms in LCGFT are implemented, LCSH form headings such as **Cancer patients' writings** and **Children's fiction** will no longer be assigned to describe the creators and intended audiences of resources, respectively.

In 2013, LC decided that a separate vocabulary for terms describing creators, contributors, and audiences – all of which are demographic groups – will replace LC form headings denoting those facets of resources. The first terms for *Library of Congress Demographic Group Terms* (LCDGT) were approved in June 2015. Still in the early stages of development, LCDGT included 827 terms in January 2016.⁴

STRUCTURE: RELATIONSHIPS

The structure of the three vocabularies follows the principles presented in ANSI/NISO standard Z39.19-2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*,⁵ to the extent practicable. Each authorized term represents a single concept, and each concept is represented by a single term. Equivalence relationships are controlled by Used For (UF) references, which lead users to the preferred terms. Associative relationships are represented through Related Terms (RTs) and reflect terms that are neither equivalent nor hierarchical, but are so closely associated that an explicit link between them may suggest useful terms. Hierarchical relationships between terms are represented by Broader Term (BT) and Narrower Term (NT) references, and hierarchical references are made only to the next-higher concept.

While the rules for determining the authorized terms and the equivalence and associative relationships are consistent, with only the variations that one would expect across three vocabularies with different scopes, the rules governing hierarchy differ.

The terms in LCGFT exhibit two of the three relationship types defined by the ANSI/NISO standard: class/class member and whole/part (the third, which is not represented, is the instance relationship). Most terms in LCGFT are members of a class. **Municipal ordinances** is a member of the class **Statutes and codes**, for example. A small portion of the terms displays the whole/part relationship instead, as in **Film clips**, which is an NT of **Motion pictures**; and **Radio outtakes**, an NT of **Radio programs**.

The medium of performance terms display only the class/class member relationship. Instruments are grouped according to the way sound is produced; by air or by strings, for example. The term **jug** is a member of the class **wind instrument**.

In LCGFT and LCMPT, every authorized term except the top term is a member of a hierarchy. In LCGFT, the top terms generally represent disciplines, although in some cases they represent a broad type of material (e.g., **Literature**; **Religious materials**; **Discursive works**; **Informational works**). All of the terms in LCMPT are part of one of three hierarchies: **ensemble**, **performer** (which includes

instruments), and **visual**. **Terms can occasionally be part of two or more hierarchies. For example, the genre/form term Sermons is an NT of both Discursive works and Religious materials**, and the medium of performance term **carillon** has two BTs, **bell** and **keyboard instrument**.

The hierarchical structure of the demographic group terms is quite different from that of the other two vocabularies. Although LCDGT contains some formal BT/NT relationships (e.g., **Bedouin** is an NT of **Arabs**), hierarchical relationships are relatively rare because of the issues inherent in classifying people. A hierarchical reference is made only when a demographic group can be said to *always* be a subclass of a larger group. For example, people described as being deaf-blind are not always completely blind nor completely deaf, so the term **Deafblind** is a member of the class **People with disabilities**, not the classes **Deaf** and **Blind**.

Rather than hierarchies, LCDGT relies on categories as its basic organizing scheme. LCDGT categories are similar to node labels, which the ANSI/NISO standard defines as, A “dummy” term, often a phrase, that is not assigned to documents when indexing, but which is inserted into the hierarchical section of some controlled vocabularies to indicate the logical basis on which a class has been divided. *Node labels may also be used to group categories of related terms in the alphabetic section of a controlled vocabulary [emphasis added].*⁶

Each term is a member of at least one of eleven categories, which serve as the highest-level collocation points in the vocabularies. **Females** is a member of the *gender* category, for example.¹ Since individual terms are not usually brought into conceptual proximity through hierarchies, copious UF references are made to bring terms from various categories into alphabetical proximity. Take for example the following terms.

<i>Reference</i>	<i>Category</i>
Students, Anthropology USE Anthropology students	Social
Students, Catholic school USE Catholic school students	Social
Students, College USE College students	Educational level
Students, Eighth grade USE Eighth grade students	Educational level
Students, Law USE Law students	Social

The terms for students are organized into two different categories because some are educational levels, while others describe either students within a discipline or engaged in a type of schooling; the latter two are not limited to an educational level. For example, anthropology may be taught at any level, and the level and type of education that aspiring lawyers receive is specific to individual countries. Since students in general – that is, students of all disciplines, at all educational levels, and in all types of educational system – are unlikely to be either the creators of an individual resource or the intended audience of a resource, LCDGT does not include the term *Students*. Instead, terms for types of students may be browsed in an alphabetical list by virtue of their UFs.

¹ The categories are *age; educational level; gender; language; medical, psychological, and disability; national/regional; occupation/field of activity; religious; sexual orientation; and social*. Social is used for identifiable demographic groups that cannot be included in one of the other categories.

STRUCTURE: FACETS

All three of these vocabularies are faceted: each has a specific and well-defined scope, and those scopes do not overlap. Genre/form terms are not found in the demographic group terms vocabulary, for example.² Vocabularies with demonstrably different scopes are essential for improving access, as will be discussed below.

The scope of individual terms within each of the vocabularies is also unique. Every term describes a single facet of a resource, and the terms do not overlap with each other. For example, the demographic group terms vocabulary includes the terms **African Americans**, **Chinese Americans**, and **Dentists**, but not *African American dentists* or *Chinese American dentists*. The medium of performance thesaurus includes the individual terms **violin** and **viola**, but not *violin and viola* as a single phrase, although compositions for violin and viola are common.

Faceting is approached differently in the genre/form thesaurus. When LCGFT was first being developed, it seemed to be important to replicate LCSH style as closely as possible. The first groups of terms to be approved were therefore identical in form and scope to the analogous form headings in LCSH. For example, the genre/form term **Fantasy films** is analogous to the LC subject heading **Fantasy films**, the genre/form term **Fantasy television programs** matches the form and scope of the subject heading **Fantasy television programs**, and so on. In LCGFT the film, television program, and radio program terms are in parallel hierarchies, so the parallel structure seemed reasonable.

LCGFT was not faceted until the terms for cartographic resources – generally speaking, maps, atlases, and globes – were added in 2010. Atlases consist of multiple maps, and globes can be considered “maps in the round.” Instead of three parallel terms, *Topographic maps*, *Topographic atlases*, and *Topographic globes*, for example, the three terms **Topographic maps**, **Atlases**, and **Globes** were authorized, and combinations of the terms are assigned as necessary to bring out the form of the resource being described. An atlas consisting of topographic maps is assigned the terms **Topographic maps** and **Atlases**. The redundancies that parallel structure caused in the moving image and radio program terms are thereby avoided.

The literature terms were added after the terms for cartographic materials, but they were impacted by the legacy of the parallel structure nonetheless. To continue the above example, LCGFT now includes the terms **Fantasy comics**, **Fantasy drama**, and **Fantasy poetry** in addition to **Fantasy films**, **Fantasy television programs**, and **Fantasy radio programs**. The terms for musical works were also impacted, but to a somewhat lesser extent. LCGFT includes, for instance, the terms **Christmas films**, **Christmas music**, **Christmas radio programs**, and **Christmas television programs**, as well as **Christmas sermons**, a term in the religious materials hierarchy. Whether this parallel structure should be retained is a matter for future discussion.

Assignment of the genre/form, medium of performance, and demographic group terms is straightforward. Multiple terms are assigned within and across the vocabularies to bring out significant aspects of resources, and terms authorized by the three new vocabularies are never subdivided. LC subject headings continue to be assigned in the usual way, including the use of subdivisions.³ A novel-

² A limited number of genre/form terms, such as **Children’s films** and **Cartographic resources for the visually impaired**, do denote audiences. Those terms were approved during the initial development of LCGFT in order to provide vital access to resources for those audiences. The decision to develop a separate demographic group terms vocabulary had not been made at that time, and LC intends to cancel those terms after LCDGT is implemented.

³ LC subject headings and subdivisions that represent genres and forms, mediums of performance, and the characteristics of creators and intended audiences will continue to be assigned as usual until the new vocabularies

length thriller about a kidnapping that was written by a Vietnamese American man and is intended for teenagers would be assigned these terms:

Genre/form	Thrillers (Fiction) Novels
Audience	Teenagers
Creator	Americans Vietnamese Americans Men
Subject	Kidnapping—Juvenile fiction

The score of a concerto for a bass trombone accompanied by a band that was composed by a Hungarian woman would be assigned the following array.

Genre/form	Concertos Scores
Medium of performance	bass trombone band
Creator	Hungarians Women

As these examples show, each facet that is described by the three new vocabularies is assigned its own term as opposed to a compound phrase or subdivided string consisting of multiple facets.

POSSIBILITIES FOR ENHANCED ACCESS

The three new vocabularies support traditional browsing, but their structure also empowers users to easily combine terms within and across the vocabularies to find resources that are directly pertinent to their needs. This enhanced access is made possible by computer-coding each vocabulary uniquely, which allows computers to separately manipulate the terms in each vocabulary.

The thesaural structure of the vocabularies allows for traditional browse searches. Through *see* references, users can be directed from non-preferred terms to those that are authorized, and they can also widen and limit searches by following broader and narrower term references. Additionally, associative references allow users to jump between hierarchies when terms are neither equivalent nor hierarchical, but are highly associated with each other.

When browsing in the traditional way, users would need to indicate which type of search they want to perform: genre/form, medium of performance, creator characteristics, or audience characteristics. These search options would be provided in addition to the standard name, title, and subject searches of today. Each vocabulary would be searched separately. The benefits of being able to search directly for additional facets of library resources would be partially offset by the silo effect – users would not be able to mix-and-match terms from multiple vocabularies to find materials directly pertinent to their needs.

are fully implemented and retrievable in online catalogs, so that access is not lost during the transition. Revisions to LCSH policy will be considered in the future.

The full utility of the genre/form, medium of performance, and demographic group vocabularies will be realized through the deployment of faceted discovery interfaces. Users will input a simple search and be given a list of options for narrowing it, based on the metadata of the resources retrieved. For example, a user who searches for *novels* might retrieve thousands of records. A list of facets similar to the one in Figure 1 below could appear as either an intermediate screen (before any results are returned), or as a panel on the results screen.

Figure 1: Sample of facets associated with the search *novels*

You searched for <i>novels</i> and retrieved 981,007 results		
You can narrow your search by selecting one or more of the following categories		
<u>Audiences</u>	<u>Genres and forms</u>	<u>Topics</u>
Children	Detective and mystery fiction	Great Britain. MI5—Fiction
Teenagers	Epistolary fiction	Lawyers—Fiction
	Legal fiction	Middle East—Fiction
<u>Creators</u>	Romance fiction	Murder—Fiction
Americans	Thrillers (Fiction)	Detectives—Fiction
English	Western fiction	Mafia trials—Fiction
Lawyers		

These choices can serve as a form of reader’s advisory, showing users what options are available. The user may decide to retrieve (novel-length) **Detective and mystery fiction** that is written by **Lawyers** who are **Americans**. Another user doing the same initial search for *novels* may choose **Teenagers** as the audience and **Epistolary fiction** as the form and retrieve very different results.

The same procedure is pertinent to nonfiction. The options presented in Figure 2 could be returned for users who search for the subject *dementia*.

Figure 2: Sample of facets associated with the search *dementia*

You searched for <i>dementia</i> and retrieved 391 results		
You can narrow your search by selecting one or more of the following categories		
<u>Audiences</u>	<u>Genres and forms</u>	<u>Topics</u>
Caregivers	Biographies	Dementia—Patients—Care
Nurses	Devotional literature	Home care services
Nursing students	Documentary films	Nursing homes
Physicians	FAQs	Older people
	Handbooks and manuals	
<u>Creators</u>	Textbooks	
Caregivers		
Geriatricians		
Nurses		

Users who require materials about personally caring for a relative with dementia could choose **Caregivers** as the audience. **Caregivers** could also be chosen as the creator group if the users want to

know about the experiences of others in a similar situation to their own, or alternately, resources by a **Geriatrician** – a physician who specializes in the diseases of older people – might be more useful if a more clinical approach were desired. Alternately, the potential caregiver may choose to view the list of **Handbooks and manuals** or **Documentary films** available at the library, as opposed to **Devotional literature** for caregivers.

This level of specific and direct access is not available with LCSH alone. LCSH form headings that denote the genres and forms, mediums of performance, and demographic groups of the creators and intended audiences of resources are inconsistent in form and application and create a barrier to access. In addition, most nonfiction materials – excluding works for children – are not assigned access points to indicate the intended audience, and even fewer have access points denoting the demographic group of their creators, although both are often vitally important to users.

Library of Congress Genre/Form Terms for Library and Archival Materials, Library of Congress Medium of Performance Thesaurus for Music, and Library of Congress Demographic Group Terms will together provide opportunities for enhanced access to library materials when they are implemented in faceted discovery systems.

References

- [1] “Introduction to Library of Congress Subject Headings,” accessed May 23, 2016, <http://www.loc.gov/aba/publications/FreeLCSH/lcshintro.pdf>.
- [2] “Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials,” accessed May 23, 2016, <http://www.loc.gov/aba/publications/FreeLCGFT/gftintro.pdf>.
- [3] “Introduction to Library of Congress Medium of Performance Thesaurus for Music,” accessed May 23, 2016, <http://www.loc.gov/aba/publications/FreeLCMPT/mptintro.pdf>.
- [4] “Introduction to Library of Congress Demographic Group Terms,” accessed May 23, 2016, <http://www.loc.gov/aba/publications/FreeLCDGT/dgtintro.pdf>.
- [5] *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies: ANSI/NISO Z39.19-2005*. (Bethesda, Md.: National Information Standards Organization, 2005).
- [6] *Ibid.*, 10.