THE LIBRARY OF CONGRESS

Decimal Classification Division

To: Deane Zeeman, Chair
      Decimal Classification Editorial Policy Committee

Cc: Members of the Decimal Classification Editorial Policy Committee
      Dennis M. McGovern, Chief, Decimal Classification Division

From: Joan S. Mitchell, Editor in Chief
      Dewey Decimal Classification
      OCLC Online Computer Library Center, Inc.

Re: Dewey Numbers in Authority Files (Discussion Paper)

Since the early 1990s, we have used various methods to identify and associate Library of Congress Subject Headings (LCSH) (and more recently, Medical Subject Heading (MeSH] and Canadian Subject Headings [CSH]) with Dewey numbers. We have employed a number of mapping strategies that include editorial mappings (EM), statistical mappings with sampled review (PPT and SHC), statistical co-occurrence (SM), and straight frequency (the former FM) statistical mappings. In the editorial mapping process, we identify the Dewey number or numbers with which a heading is likely to be used. We often have only the main heading available for mapping purposes; in actual assignment, the heading might be used with a free-floating subdivision when applied to a record with a particular Dewey number.

The relationship between the subject heading and the Dewey number is based on the likelihood of use of the heading with the number. There is no explicit definition of the relationship represented by the current LCSH, MeSH, and CSH mappings beyond likelihood of concurrent use.

As interest builds in including Dewey numbers directly in subject heading records, we need to explore the differences between mapping terminology to Dewey numbers and loading a Dewey number into an authority record. Our mapping practice is based on likely co-occurrence; storing a Dewey number explicitly in an authority record for a subject heading means that the subject entity represented by the heading equals or approximates the whole of the number, or is in standing room at that number. The definition of the actual relationship between the Dewey number and the subject heading is found in the Dewey authority record.

Proposed Guidelines

In developing proposed guidelines for the inclusion of Dewey numbers in authority files, we first consulted the rules found in H365 (*Subject Cataloging Manual: Subject Headings*) for the inclusion of Library of Congress Classification (LCC) numbers in subject authority records. H365 contains the following basic instructions for when to load an LCC into a subject authority record:

- The caption for the number is identical or nearly identical in scope, meaning, and language to the subject heading.
- The topic is explicitly mentioned in an "Including" note under the caption for the number.
- The topic belongs to a category for which it is standard LC classification policy to classify works at a level that is broader than the subject headings assigned.

We propose the following general guidelines for the inclusion of Dewey numbers in LCSH authority records:

1. The subject entity represented by the LCSH equals or approximates the whole of the Dewey class.
2. The subject entity represented by the LCSH is explicitly in standing room at the number.
3. The geographic entity represented by the LCSH has an implicit relationship to the Dewey class.
4. The genus/species represented by the LCSH has an implicit relationship to the Dewey class.
5. If the subject entity represented by the LCSH matches more than one Dewey number according to the aforementioned rules, multiple Dewey numbers may be added to the authority record.

It is important to note that the subject entity represented by the heading must have the same meaning as its representation in the Dewey class. For example, Dewey uses the term "clergy" broadly to mean any kind of religious leader, while the LCSH "Clergy" is limited to Christian religious leaders. Therefore, only the Dewey number associated with Christian clergy can be included in the subject authority record for the LCSH "Clergy."

Examples
Example 1.1: Equals or Approximates the Whole
The subject entity represented by the LCSH "Education" equals the whole of classes 370 Education and T1—071 Education.

150 ## $aEducation
083 00 $a370

180 ## $aEducation
083 00 $z1$a071

Example 1.2: Equals or Approximates the Whole

The subject entity represented by the LCSH "Sociology in literature" approximates the whole of class T3C—3552 Sociology and anthropology. Class T3C—3552 is used to represent topics in literature and contains a note that standard subdivisions may be added for either or both topics in the heading.

150   ## $aSociology in literature

083   00 $z3C$a3552

Example 1.3: Equals or Approximates the Whole

The subject entity represented by the LCSH "Emotional intelligence" approximates the whole of the class 152.4 Emotions. "Emotional intelligence," a term with the same meaning as the LCSH, is in a class-here note at 152.4.

150   ## $aEmotional intelligence

083   00 $a152.4

Example 1.4: Equals or Approximates the Whole

The subject entity represented by the LCSH "Khoisan (African people)" equals the whole of class T5—961 Khoikhoi and San. "Khoisan (African people)" is a Relative Index term at T5—961 and is an umbrella term for the Khoikhoi and San.

150   ## $aKhoisan (African people)

083   00 $z5$a961

Example 2.1 Explicitly in Standing Room

The subject entity represented by the LCSH "Violence in literature" is explicitly in standing room at T3C—3552 Sociology and anthropology. Class T3C—3552 is used to represent topics in literature, and "violence," a term with the same meaning as the LCSH, is in an including note at T3C—3552.

150   ## $aViolence in literature

083   00 $z3C$a3552

Example 2.2 Explicitly in Standing Room

The subject entity represented by the LCSH "Pipil language" is explicitly in standing room at T6—97452 Aztecan languages. "Pipil language" (a specific Aztecan language) appears as a Relative Index term at T6—97452.

150   ## $a Pipil language

083   00 $z6$a97452

Example 3: Implicit Geographic Relationship

The geographic entity represented by the LCSH "Oakland (Garrett County, Md.)," a town in Garrett County, is <u>implicitly</u> in standing room at T2—75297 Garrett County.

151   ## $aOakland (Garrett County, Md.)

083   00 $z2$a75297

Example 4: Implicit Biological Relationship
The species of halibut represented by the LCSH "Pacific halibut" is <u>implicitly</u> a member of the Dewey class 597.695 Hippoglossus (Halibuts). "Pacific halibut," whose scientific name is Hippoglossus stenolepis, approximates the whole of the class by virtue of the note "Subdivisions are added for the genus as a whole and for individual species."
150 ## $aPacific halibut
083 00 $a597.695

Example 5: Multiple Numbers Added to Record
The subject entity represented by the LCSH "Cookies" equals the whole of the home preparation number for cookies, 641.8654 Cookies, and is explicitly in standing room in the commercial processing number for pastries, 664.7525 Pastries.
150 ## $aCookies
083 00 $a641.8654
083 00 $a664.7525

Discussion
We have proposed guidelines for loading Dewey numbers in LCSH records that go beyond the "likelihood of use of the heading with the number" guideline that is the simple test for our mappings. In the case of loading Dewey numbers into subject heading authority records, "likelihood of use" is a necessary, but not sufficient, test. The subject entity represented by the LCSH must equal or approximate the whole of the DDC class, or be in standing room at the number.

Two questions arise for discussion:
1. Are the guidelines generalizable to other subject heading systems?
2. Do we need additional definitions of relationships within Dewey number records?

1. Are the guidelines generalizable to other subject heading systems?
If the subject heading system has not been previously used with Dewey, we may not be able to determine the initial "likelihood of use" test before applying the guidelines. As in the case of LCSH, the subject entity represented by the heading must equal or approximate the whole of the number, or be in standing room at the number.

As an immediate test, we will ask H.W. Wilson to review the proposed guidelines against their internal rules for the inclusion of abridged Dewey numbers in the subject authority records for Sears List of Subject Headings.

At the Dewey Translators Meeting at the World Library and Information Congress (IFLA 2006), representatives of the Italian and German translation teams reported on projects to load Dewey numbers in subject authority records. In "Linking DDC Numbers to the New 'Soggettario Italiano'" (http://www.oclc.org/dewey/news/conferences/ddc_and_soggetario_ifla_2006.ppt), Federica Paradisi (National Library of Florence) reported that only one number is currently permitted in a subject authority record. In "CrissCross" (http://www.oclc.org/dewey/news/conferences/crisscross_ifla_2006.ppt), Yvonne Jahns

(Deutsche Nationalbibliothek) reported on a project to link Dewey numbers with Schlagwortnormdatei (SWD). The Italian guidelines are a bit more restrictive than those proposed by the German team and the DDC editors. In the proposed Italian guidelines, only one number is currently permitted in a subject authority record. The German guidelines include three levels of "precision" to represent the relationship between the heading and the Dewey number.

2. Do we need additional definitions of relationships within Dewey number records?
We have stipulated that the definition of the actual relationship between the Dewey number and the subject entity represented by the subject heading is found in the Dewey class record. When the subject entity represented by the subject heading is found in Relative Index alone, its status as "equals or approximates the whole" or "standing room" may not be immediately obvious without further research. For example, one had to consult an outside source for the definition of the Relative Index entry "Khoisan (African people)" to determine that it actually equals the whole of the class "T5—961 Khoikhoi and San" (see example 1.4). An immediate recommendation is that we develop a mechanism to identify the status of all Relative Index terms whose relationship to the class is not explicitly provided in the class record itself. For example, we would provide a designation (in a yet-to-be-determined form) for the index entry "Khoisan (African people)" as "equal to or approximating the whole" of T5—961. Likewise, we would provide a designation for the Relative Index entry "Pipil language" as in "standing room" at T6—97452 (see example 2.2).

A secondary question is, do we need finer designations of relationships beyond "equals or approximates the whole" and "standing room" in Dewey classes? As we look to machine-assisted classification and automatic derivation of relationships for other applications, we may needed finer designations for the relationships of subject entities in Dewey classes than currently provided in Dewey class records. Appendix 1 contains a brief description of some definitions of relationships proposed by others as part of mapping knowledge organization systems.

Appendix 1: Definitions of Mappings
The current mapped terminology serves as additional entry vocabulary into Dewey, but the authority records for both the subject heading and Dewey number must be consulted before using one to apply the other. In the mid-1990s, we experimented with the following simple definitions of mappings between editorially mapped LCSH and Dewey numbers:

1. This heading points to this number exclusively
2. This heading maps to this number and others
3. Other

We captured these categories as "confidence levels" in the records for editorially mapped headings, but discontinued the practice after a brief experiment. We have not experimented further with adding definitions to the mapped relationship beyond the general "kind of mapping" designation itself (e.g., EM).

In 1995, Iyer and Giguere proposed seven categories to describe mappings between the Mathematics Subject Classification (MSC) and Dewey:

1. Exact matches
2. Specific to general
3. General to specific
4. Many to one
5. Cyclic mapping strategies
6. No matches
7. Specific and broad class mapping

Source: Iyer, H., and M. Giguere. 1995. Towards designing an expert system to map mathematics classificatory structures. *Knowledge Organization* 22 (3-4):141-47.

The Renardus Service (http://www.renardus.org/) includes five levels of mapped relationships:

1. *Fully equivalent* means that the subject content of the local page you are jumping to is basically the same as the subject indicated on the Renardus browsing page.
2. *Narrower equivalent* means that the subject content of the local page is a true subset of the browsing page.
3. *Broader equivalent* is the inverted case of narrower equivalent: the local page contains all of the subject content of the Renardus browsing page.
4. *Major overlap* means that the content of the local page represents a large part of the browsing page plus other related subjects.
5. *Minor overlap* indicates equivalence to a minor part of the browsing page plus other related subjects.

There is also a set of mapped relationships currently under discussion in association with SKOS (http://www.w3.org/2004/02/skos/mapping/spec/). These relationships are similar to those used in the Renardus Service:

1. broadMatch: *If 'concept A has-broad-match concept B' then the set of resources properly indexed against concept A is a subset of the set of resources properly indexed against concept B.*

2. exactMatch: *If two concepts are an 'exact-match' then the set of resources properly indexed against the first concept is identical to the set of resources properly indexed against the second. Therefore the two concepts may be interchanged in queries and subject-based indexes.*
3. majorMatch: *If 'concept A has-major-match concept B' then the set of resources properly indexed against concept A shares more than 50% of its members with the set of resources properly indexed against concept B.*
4. minorMatch: *If 'concept A has-minor-match concept B' then the set of resources properly indexed against concept A shares less than 50% but greater than 0 of its members with the set of resources properly indexed against concept B.*
5. narrowMatch: *If 'concept A has-narrow-match concept B' then the set of resources properly indexed against concept A is a superset of the set of resources properly indexed against concept B.*