EPC Exhibit 137-28.3

March 21, 2014

THE LIBRARY OF CONGRESS

Dewey Section

To: Jonathan Furner, Chair

 Decimal Classification Editorial Policy Committee

Cc: Members of the Decimal Classification Editorial Policy Committee

 Karl E. Debus-López, Chief, U.S. Programs, Law, and Literature Division

From: Rebecca Green, Assistant Editor

Dewey Decimal Classification

OCLC Online Computer Library Center, Inc.

Via: Michael Panzer, Editor in Chief

 Dewey Decimal Classification

 OCLC Online Computer Library Center, Inc.

Re: Topic-enhanced classes

This exhibit is part of a long-term effort to develop a version of the DDC that will support automatic applications. Key elements in that version will be the isolation of topics as conceptual units within the system and the identification of relationships between topics. Topics will be represented in authority records through a one-to-many relationship between topics and Relative Index headings (that is, a topic may be expressed by more than one Relative Index heading, but a Relative Index heading should express only one topic); the expression between topics and how the topic is reflected in the text of the category description will also be captured.

This exhibit builds upon a series of papers and EPC exhibits, including others that are part of EPC Meeting137:

|  |  |
| --- | --- |
| **EPC exhibit** | **EPC exhibit title [paper citation]** |
| EPC 137-28.1 | Relative Index: Managing synonyms |
| EPC 137-28.2 | Facet detection [Rebecca Green. “Facet Detection Using WorldCat and WordNet.” To be published in: *Knowledge organization in the 21st century: between historical patterns and future prospects:* *proceedings of the Thirteenth International ISKO Conference, 23–26 February 2010, Rome, Italy*] |
| EPC 136-25 | Modeling topic relationships [Rebecca Green; Diane Vizine-Goetz; Marcia Lei Zeng, Maja Žumer. From modelling to visualization of topic relationships in classification schemes. In: *Classification & visualization: interfaces to knowledge: proceedings of the International UDC Seminar 24–25 October 2013, The Hague, The Netherlands*. Edited by Aida Slavic, Almila Akdag Salah, Sylvie Davies. Wurzburg: Ergon Verlag, 2013, 177–190.] |
| EPC 135-23.3  | Modeling topics in the DDC |
| EPC 134-24.1 | Relationships in the notational hierarchy of the Dewey Decimal Classification [Rebecca Green and Michael Panzer. “Relationships in the notational hierarchy of the Dewey Decimal Classification.” In: *Classification & ontology: formal approaches and access to knowledge: proceedings of the International UDC Seminar, 19–20 September 2011, The Hague, The Netherlands,* edited by Aida Slavic & Edgardo Civallero, 161–176. Würzburg: Ergon Verlag, 2011.] |
| EPC 134-24.2 | See-also relationships in the Dewey Decimal Classification [Rebecca Green. “See-also relationships in the Dewey Decimal Classification.” In: *Proceedings from North American Symposium on Knowledge Organization* *2011, June 16–17, Ryerson University, Toronto, Canada,* vol. 3, edited by Richard P. Smiraglia, 41–51. Toronto, Canada, 2011. ] |
| EPC 132-37.1 | The ontological character of classes in the DDC [Rebecca Green and Michael Panzer. “The ontological character of classes in the DDC.” In: *Paradigms and conceptual systems in knowledge organization: proceedings of the Eleventh International ISKO Conference, 23–26 February 2010, Rome, Italy*, edited by Claudio Gnoli and Fulvio Mazzochi , 171–179. Würzburg: Ergon Verlag, 2010.] |
| EPC 131-37.1 | Relative Index Project |
| EPC 131-37.4 | Explicit relationships in the DDC |
| EPC 130-37.1 | Relative Index Project |
| EPC 129-36.1 | Relative Index Project [Rebecca Green. “Making visible hidden relationships in the Dewey Decimal Classification: How Relative Index terms relate to DDC classes.*”* In: *Culture and identity in knowledge organization:* p*roceedings of the Tenth International ISKO Conference, 5–8 August 2008,Montréal, Canada*, edited by Clément Arsenault and Joseph T. Tennis, 8–14. Würzburg: Ergon Verlag, 2008.] |

In the EPC Meeting 135 presentation, EPC 135-23.3 Modeling topics in the DDC, recommendations were made to add two local subfields to the 083 fields in the authority records for Relative Index (RI) headings; these are the fields in which links between class records and authority records are encoded. One of the new subfields ($9 rt=) would record the text in the class description that reflects the topic named by the RI heading (e.g., $9 rt=Libraries for children); the other subfield ($9 mt=) would give the MARC tag in which the reflecting text is found (e.g., $9 mt=153). A third subfield ($9 eq=) would record the equivalence status of the RI heading, by indicating whether the topic approximates the whole of the number indexed by the topic ($9 eq=AW), or is standing room there ($9 eq=SR); this subfield has been recommended in other discussions of the editorial team.

This table summarizes the coding of $9 subfields of 083 fields seen in the examples below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Subfield** | **Mnemonic for** | **Value** | **Meaning** |
| $9 as= | approval status | AP | approved |
| PR | proposed |
| $9 ps= | print status | PE | print and electronic |
| EO | electronic only |
| $9 rt= | reflecting text |  | text in class description that conveys the topic of an RI |
| $9 mt= | MARC tag |  | 3-digit tag for field in which reflecting text is found |
| $9 eq= | equivalence status | AW | topic approximates the whole |
| SR | standing room topic |

We have recently undertaken a study in 020 Library and information sciences to verify the feasibility of this approach, to examine issues that arise, and to propose alternatives/solutions. This exhibit reports the findings of the study and proposes next steps forward.

Approximately half of the topics named by RI headings in the 020s could be matched automatically to text in the category description with a high degree of confidence. For example, the including note at 023.3 Technician positions reads “Including library aides, assistants, clerks, paraprofessionals, technicians.” The phrase “library aides” clearly reflects the RI heading “Library aides.” This topic is readily identified as a standing-room topic, since the reflecting text is found in an including note. The 083 field would be coded in this manner:

083 04 $a 023.3 $0 (OCoLC-D)ocd00117461 $2 23 $9 as=AP $9 ps=PE $9 rt=library aides $9 mt=680 $9 eq=SR

150 ## $a Library aides

For purposes of matching in the initial pass, subheadings were ignored, because they typically refer to the discipline or subdiscipline, which may or may not be explicitly referred to in the text of the corresponding category description. For example, libraries are the context of all of the 020s. Therefore, when matching the RI heading “Public relations $x libraries” to the class-here note at 021.7 Promotion of libraries, archives, information centers, which reads “Class here public relations,” it is not necessary for the word “libraries” to appear in the note for us to have full confidence in the following coding:

083 04 $a 021.7 $0 (OCoLC-D)ocd00117443 $2 23 $9 as=AP $9 ps=PE $9 rt public relations $9 mt=153 $9 eq=AW

150 ## $a  Public relations $x libraries

In some circumstances, however, an RI subheading may fill a role other than providing disciplinary or subdisciplinary context. The RI main heading and subheading may match text from the same part of the category description. Or the subheading may match text in one part of the category description, while the main heading matches text from another part of the description. The first pattern can be seen in 027.42 Library outreach programs, which has the note, “Class here library extension, the use of bookmobiles (mobile libraries) in public librarianship,” and is indexed by the RI heading, “Bookmobiles $x public library use.” Both the main heading “Bookmobiles” and the subheading “$x public library use” are drawn (the one directly, the other indirectly) from text in the class-here note. The second pattern can be seen with the RI heading “Bibliographic centers $x cooperation” and its relationship to 021.64 Cooperation, which has the including note, “Including bibliographic centers.” The main heading topic is mentioned in the including note, while the subheading topic is (in) the caption. One might argue that the caption topic, “Cooperation,” provides context in the same way that the caption at 021.7 Promotion of libraries, archives, information centers provides context. But the real parallel between the captions at 021.64 and at 021.7 is between “Cooperation” and “Promotion [of . . .]”, both of which specify part of the topic, as opposed to providing disciplinary context. If we do not automatically ignore subheadings, we might arrive at the following 083 encodings:

083 04   $a 027.42 $0 (OCoLC-D)ocd00117775 $2 23 $9 as=AP $9 ps=PE $9 rt=the use of bookmobiles [(mobile libraries)] in public librarianship $9 mt=680 $9 eq=AW

150 ##   $a Bookmobiles $x public library use

In this case, we use square brackets around text (“(mobile libraries)”) that is not considered reflecting text. Although “mobile libraries” and bookmobiles are synonyms, “mobile libraries” could be deleted from the note without affecting the note’s relationship to the RI heading.

083 04  $a 021.64 $0 (OCoLC-D)ocd00117440 $2 23 $9 as=AP $9 ps=PE $9 rt=bibliographic centers $9 mt=680 $9 rt=Cooperation $9 mt=153 $9 eq=SR

150 ## $a Bibliographic centers $x cooperation

In this case, we give two $9 rt= and $9 mt= values, because the topic is not fully reflected in the category description except by taking into account both the caption and the 680 field. But only one $9 eq= subfield is given. The match between subheading and caption does not transform the RI heading into a topic that approximates the whole of the class; it is in standing room.

(EPC Exhibits 129-36.1, 130-37.1, and 131-37.1 reported on the findings of a research project to determine if RI headings approximate the whole of the class they index or are in standing room in the class. While the matching process used there did not result in the identification of specific text in the category description reflecting the topic of the RI heading, it did take into account the various complications with subheadings mentioned here. On the one hand, subheadings were included in the matching process there; on the other hand, subheadings were matched against words and phrases in captions, variant-name notes, and class-here notes in the immediate class and in classes in the upward hierarchy of the immediate class. The deeper analysis used in that project will be incorporated into the current project as it progresses.)

Several circumstances account for most instances in which reflecting text could not be identified automatically in an initial pass:

1. In syntactic constructions including the conjunction *and* and either ((two adjectives and one noun) or (one adjective and two nouns)), the scope and possible distribution of modification is often not altogether clear; after all, we don’t use parentheses in natural text to make groupings clear, in contrast to what was done at the beginning of this sentence. For example, does the caption for 020 Library and information sciences mean (a) (Library) and (information sciences), (b) Library-and-information sciences, or (c) (Library science) and (Information science)? The presence of the two RIs, “Information science” and “Library science,” and the absence of an RI heading, “Library and information sciences” support interpretation #3. If the caption remains as it currently is, the following 083[[1]](#footnote-1) would be proposed for the RI heading “Library science”:

083 04 $a 020 $0 (OCoLC-D)ocd00117305 $2 23 $5 OCoLC-D $9 as=AP $9 ps=PE $9 rt=Library [and information] science[s] $9 mt=153 $9 eq=AW

150 ## $a Library science

Other options might also be considered: mention of the full form of all indexable topics, e.g., 020 Library science and information science; modification of the caption to recognize a single, unified topic, e.g. Library and information science; and/or addition of an RI heading, “Library and information science.” (Common use of the phrase “library and information science” keeps this example from being thoroughly generalizable. Typically we could count only on the first option—that of making mention of the full form of all indexable concepts—to be available.)

1. How topics are phrased in category descriptions and how they are phrased in RI headings sometimes diverge through the use of morphologically related forms or through syntactic variation. For instance, prepositions are much more likely to be used in the text of a category description than in the wording of an RI heading. Therefore, the manners in which a given topic is expressed (a) in a category description and (b) as an RI heading may be related as morphosyntactic paraphrases (i.e., paraphrases that do not depend on lexical semantic relationships; for those that do depend on lexical semantic relationships, see [3]). For example, the RI heading “Library-community relations” and the caption at 021.2 Relationships with the community are morphosyntactic paraphrases of each other.

Akin to the options offered in (1) above, several possibilities could be considered for such situations: (a) change the caption to Library-community relations, (b) change the RI heading to “Relationships with the community,” or (c) maintain a discrepancy between the two expressions of a single topic.

1. Classes may have RI headings whose relationship to their reflecting text depends on synonymy. A few of the examples found in the 020s include the following:

|  |  |  |
| --- | --- | --- |
| **Class** | **RI heading** | **Reflecting text** |
| 025.02 | Technical processes (Libraries)[[2]](#footnote-2) | Technical services |
| 025.04 | Online information systems | Information storage and retrieval systems |
| 025.0425 | Information discovery | Search and retrieval |
| 025.84 | Mending books | repair and restoration |

Often indexable concepts in the category description will have a best-match RI heading, while other RI headings for which the same reflecting text is designated should have appropriate references made between them (the relationship will often be that of equivalence [see EPC 137-28.1 Managing synonyms for a proposal on how equivalence relationships should be recorded], but may also be a hierarchical relationship, where the topic of the RI heading could be more specific than, or could be an instance of, a topic in the category description.

1. It has not been uncommon, when Library of Congress subject headings (LCSHs) have been mapped to a DDC class, that an RI heading expressing the same topic as the LCSH has also been added. For example, the LCSH “Cataloging of continuing resources” has been editorially mapped to 025.3432 Serials. Although continuing resources are not explicitly mentioned in the category description, an RI heading “Continuing resources $x cataloging” has been added for this class (the RI appears only in WebDewey). This phenomenon represents one example of the general case in which the topic of an RI heading (in this example, continuing resources) is more specific than the closest corresponding reflecting text (in this example, serials). (The general case is also exemplified by the situation described in [6].)

Identifying reflecting text automatically is not the only aspect of topic enhancement with some complications: the equivalence status of an RI heading relative to a topic in the category description can also face uncertainty.

1. Editorial policy requires that multitopic captions be accompanied by (standard-) subdivisions-are-added notes if any proper subset of the topics approximates the whole of the class. The topic expressed in a single-topic caption can be assumed to approximate the whole of the class. Similarly, the overall topic of a multitopic caption (that is, the set of topics taken as a whole / as a single topic) is assumed to approximate the whole of the class. Absent such a note, the individual topics of a multitopic caption must, however, be interpreted as standing room topics. Unfortunately the absence of such a note sometimes appears to reflect only the absence of a conscious decision. For example, consider the multitopic caption at 025.342 Clippings, broadsides, pamphlets, which is not accompanied by a standard-subdivisions-are-added note, implying that clippings, broadsides, and pamphlets are standing room topics at 025.342. The class has one note: “Class here contents of vertical files, printed ephemera.” These topics, by virtue of their presence in a class-here note, approximate the whole of the class. To have clippings, broadsides, and pamphlets designated as standing room topics, while contents of vertical files and printed ephemera are designated as topics that approximate the whole of the class appears to be inconsistent. A standard-subdivisions-are-added-for-any-or-all-topics-in-heading note should probably be added.
2. When a caption topic approximates the whole of the class, it is not always clear whether subclasses and instances of the topic concept also approximate the whole of the class or are in standing room. Consider, for example, 027.663 Libraries for people with disabilities, which has the note, “Including libraries for people with visual impairments.” We learn from this including note that libraries for people with specific disabilities (as expressed, for example, in the RI heading Hearing-impaired people $x library services) are in standing room, while libraries for people with any disability (as expressed, for example, in the RI heading Disabled people $x libraries for) approximate the whole of the class. While it is easy enough for a human to reason this way, the DDC infrastructure does not yet support such reasoning. Indeed, it is the long-term goal of the work we are doing with topic relationships to support such inferencing.

A number of general recommendations emerge from these findings:

* Distributed expression of topics should be eliminated from the captions and notes of category descriptions. That is, no word or phrase should have a syntactic scope that makes it part of the expression of more than one topic. For example, the note at 025.81 that currently reads “Including closed and open stacks, integrated shelving of materials in different formats” should be revised to read “Including closed stacks, open stacks, integrated shelving of materials in different formats.”
* The expression of topics in category descriptions should be amplified so that each topic is complete within the note in which it appears; that is, the reflecting text for a topic should not be drawn from multiple MARC fields. For example, the category description for 025.6 Circulation services includes the note “Class here document delivery.” The RI corresponding to the topic in this note is “Document delivery services.” Either the topic “document delivery” should be considered to incorporate within it the notion of service, in which case the RI should be worded simply as “Document delivery,” or if it does not, the note should be worded “Class here document delivery services.”
* Captions should be worded so that commas, semicolons, and the conjunction *and* can be interpreted as topic boundary markers. For example, the caption at 020 Library and information sciences should be reworded “Library science and information science.” (And since LIS is arguably a single topic, it could be added to the class-here note: “Class here archival science, library and information science (LIS).”
* All topics in MARC fields for which topic subfields are defined should be coded individually. (This can be done only if the distributed expression of topics has first been eliminated as detailed above, since each topic must be expressed fully by its own sequence of adjacent words.) For example, the reworded class-here note from the previous bullet would be coded:

 $i Class here $t archival science $i , $t library and information science $i ($t LIS $i )

* Notes that currently include text in parentheses should be revised, if necessary, in one of two ways. One alternative is to revise the note so that the material in parentheses appears at the end of the topic expression and can be coded as a separate topic (as in the previous bullet). Another alternative is to revise the note in a way that eliminates the parentheses; for example, the note that currently reads “Class here library extension, the use of bookmobiles (mobile libraries) in public librarianship” could be revised to read “Class here library extension, the use of bookmobiles in public librarianship, the use of mobile libraries in public librarianship.” (Note: RI headings for 027.42 Library outreach programs include both “Bookmobiles $x public library use” and “Mobile libraries $x public library use.”)
* Statements about the classing of comprehensive works on a topic and of interdisciplinary works on a topic pose additional issues. Should we consider the words “comprehensive works on” and “interdisciplinary works on” part of the topic? If so—and in the coding of MARC records, these phrases are assuredly now part of $t subfields—should the words comprehensive works on” and “interdisciplinary works on” be repeated, or is the repetition of the preposition “on” sufficient? Alternatively, might we consider separate including and class-here notes for comprehensive works and interdisciplinary works? For example, at 025.04, we might have the following class-here notes (if useful, the second indicator could be defined locally to make the regular works / comprehensive works / interdisciplinary works distinction):

680 1# $i Class here $t search and retrieval in information storage and retrieval systems $i ; $t front-end systems

680 1# $i Class here comprehensive works on: $t online catalogs integrated with information storage and retrieval systems $i , $t automated storage, search, retrieval of information $i , $t information filtering systems

680 1# $i Class here interdisciplinary works on: $t databases $i , $t query languages

(Since we already use separate class-elsewhere notes and see references for each number referred to, further proliferation by separating out comprehensive works and interdisciplinary works in those kinds of notes seems counterproductive.)

* Comprehensive works topics are not infrequently compound or complex topics, as, for example, comprehensive works on cataloging and classification (025.3), comprehensive works on online catalogs integrated with information storage and retrieval systems (025.04), comprehensive works on subject indexing and cataloging using natural language and controlled vocabulary (025.47), comprehensive works on patients' and medical libraries (027.662). Compound and complex topics of this sort should not be considered indexable, although the underlying simple topics that go into them generally should be.
* All indexable topics mentioned in category descriptions should have corresponding RI headings with equivalent semantics. It is not necessary that topics be expressed identically in the category description and in the RI heading: morphosyntactic paraphrases should be allowed; the $9 rt= subfield will capture the correspondence. For example:

083 04 $a 021.2 $0 (OCoLC-D)ocd00117434 $2 23 $9 as=AP $9 ps=PE $9 rt=Relationships with the community $9 mt=153 $9 eq=AW

150 ## $a Library-community relations

083 04 $a 025.042 $0 (OCoLC-D)ocd00117484 $2 23 $5 OCoLC-D $9 as=AP $9 ps=PE $9 rt=World Wide Web $9 mt=153 $9 eq=AW

150 ## $a World Wide Web

083 04 $a 025.042 $0 (OCoLC-D)ocd00117484 $2 23 $5 OCoLC-D $9 as=AP $9 ps=PE $9 rt=World Wide Web $9 mt=153 $9 eq=AW

150 ## $a WWW (World Wide Web)

* As seen in the last two examples, multiple RI headings may have the same reflecting text value. Often, as in the case above, the RI headings are bound by an equivalence relationship. Exhibit 137-28.1, which presents a proposal for transforming how such relationships are recorded, acknowledges that cases exist in which a semantic relationship exists between two RI headings, but where no relationship has been recorded in their authority records. Shared reflecting text would provide a means of identifying some of the RI heading pairs that currently lack indication of semantic relationship.
* As noted previously (see [4] and [6]), we anticipate circumstances in which the RI for a class will be more specific than the closest corresponding reflecting text. Should every RI heading have a $9 rt= subfield value recorded, or should reflecting text be understood to apply only where the semantics of the text of some portion of the category description and of the RI heading are (more-or-less) equivalent? (In either case, relationships between RIs with the same closest corresponding text should be captured in 550 fields.) We believe a reflecting text value should always be recorded.[[3]](#footnote-3) If this is done, when a proposal to change any text is considered, one can then be aware of all related RI headings, in case any should be moved or modified.[[4]](#footnote-4)

083 04  $a 025.3432 $0 (OCoLC-D)ocd00117509 $2 23 $5 OCoLC-D $9 as=AP $9 ps=EO $9 rt=Serials $9 mt=153 $9 eq=AW ​

150 ## $a Continuing resources $x cataloging

083 04 $a 027.663 $0 (OCoLC-D)ocd00117796 $2 23 $5 OCoLC-D $9 as=AP $9 ps=PE $9 rt=Libraries for people with disabilities $9 mt=153 $9 eq=SR

150 ## $a Hearing-impaired people $x library services

* Indicating the DDC class range over which subheadings expressing disciplinary context is potentially operative would be useful, since it helps determine how much of an RI should be expressed in its reflecting text. Such subheadings already have their own authority records. This, for example, is the record for the subheading “libraries”:

LDR nz###n## ​

001 och00004828 ​

003 OCoLC-D ​

005 20101005234204.7 ​

008 100206|||d|z||||||##########||#a||#####d ​

040 ##  $a OCoLC-D $b eng $c OCoLC-D $d OCoLC-D $f ddcri ​

180 ##  $x libraries ​

. . . ​

983 0#  $2 20 ​

983 0#  $2 21 ​

983 0#  $2 22 ​

983 0#  $2 23 ​

 780 #0  $x Libraries $0 (DLC)sh2002006394 $5 OCoLC-D

The 983 field, a locally defined field, is used to indicate which editions the subheading is valid in. Since the class range over which the subheading expresses disciplinary context could differ across editions, one possibility is to add this functionality to the 983 field. For example:

LDR nz###n## ​

001 och00004828 ​

. . .​

180 ##  $x libraries ​

. . .

983 0#  $2 23 ​$a 02

 780 #0  $x Libraries $0 (DLC)sh2002006394 $5 OCoLC-D

Here the 983 field is meant to show that the subheading “—libraries” (i.e., the value recorded in 180 $x) is valid in Edition 23 and expresses disciplinary context for 02, that is, the 020s. Ranges would be expressed in the standard fashion through the combined use of $a and $c subfields, e.g., $a 004 $c 006.

1. EPC 135-23.3 suggested treating the reflecting text in such cases in this way:

 $9 rt=Library and information sciences / Library . . . science

but the use of the plural “sciences” in the caption and of the singular “science” in the RI heading renders this option somewhat less satisfactory. [↑](#footnote-ref-1)
2. Some RI headings in the 020s render as a qualifier content better treated as a subheading. These headings will need to be revised. [↑](#footnote-ref-2)
3. This is so, even when the closest reflecting text is not an indexable topic. For example, “Other languages,” the caption at T6—9, would be designated as the reflecting text for the RI Oriental languages. [↑](#footnote-ref-3)
4. This line of argument would lead to recording closest reflecting text and MARC tag values for mapped headings in addition to RIs. Our current plan is to start only with Relative Index headings; if including mapped headings appears sensible and feasible thereafter, the effort could then be expanded. [↑](#footnote-ref-4)