

European DDC Users Group / Napoli, Italia / 17 April 2015

Upcoming changes to DDC data:

- synonyms
- history notes

Rebecca Green

Assistant Editor, Dewey Decimal Classification

Standard thesaurus practice

- Identify a set of (quasi-)synonymous terms
- Designate one of these terms as the authorized heading; use it for all indexing
- All other terms are unauthorized terms; become part of lead-in vocabulary
- See references lead from unauthorized terms to authorized terms

Relative Index practice

- Identify a set of (quasi-)synonymous terms
- If terms associated with 4+ subentries, designate one of the terms as the preferred heading
- Preferred term gets full indexing (i.e., for itself and for all subheadings)
- Non-preferred term gets indexing only for itself
- See also references lead from non-preferred terms to preferred terms

Complications in RI practice

- Designation of equivalence relationships driven by printing concerns rather than semantic concerns
- Synonymy not treated if there are three or fewer subentries
- If interdisciplinary works number not designated, no indexing for term itself; see also reference does not display

Standard MARC practice

- Authority format fields
 - 260 – Complex see reference
 - 360 – Complex see also reference
 - 4XX – See From Tracing field
 - 5XX – See Also From Tracing field
- 260 and 4XX lead from an unauthorized heading to an authorized heading
- 360 and 5XX lead from one authorized heading to another authorized heading
- 260 and 360 fields carry explanatory text

Relative Index practice

- 260 – Used for designated equivalence relationships, but using explanatory text field to specify “see also” text
- 360 – Used for related term relationships, using explanatory text field to specify “see also entries beginning with” text
- 5XX – Used for both ends of all other relationships

Complications in RI practice

- Using 260 for equivalence relationships, not because they are complex, but so they can be associated with “see also” text
- Equivalence relationships where 3 or fewer subentries exist not reflected in data
- Three-sevenths of recorded see-also references undifferentiated as to type of relationship

Proposal

- Identify sets of (quasi-)synonymous headings across full set of Relative Index headings
- For all sets of (quasi-)synonymous headings, designate preferred heading
- Adopt more standard use of MARC fields:
 - Every Relative Index term represented by authority record with term in 1XX field
 - In records for preferred headings, use 4XX fields for all non-preferred headings (*not unauthorized*)
 - In records for non-preferred headings, use 5XX fields for corresponding preferred heading

Identifying synonymous RIs

- Proof-of-concept study in 000s
- Sources used:
 - DDC data
 - RI relationships
 - Notes, etc.
 - Heuristics
 - LCSH data
 - Mappings
 - LCSH relationships
 - WordNet

RI equivalence relationships

- Sample recorded equivalence relationships:
 - Data banks, Databanks, Data bases → Databases
 - Desktop computers, Digital microcomputers, Home computers, Microcomputer workstations, Microcomputers, Microprocessors, Workstations (Microcomputers) → Personal computers
- But some are inaccurate:
 - Flying (Aeronautics) ≠ Flight (Aeronautics)

Synonym qualifiers

001.942 Unidentified flying objects (UFOs)
Variant name: flying saucers
Class here human-alien encounters

Variant names

001.942 Unidentified flying objects (UFOs)

Variant name: flying saucers

Class here human-alien encounters

005.746 Data compression

Variant names: data compaction, file compression

*For the compression of sound files, see 006.5;
for the compression of graphic files, see 006.6*

Hyphenation differences

- Hyphen → space
 - None found in 000s
- Hyphen → nothing
 - Multi-core processing = Multicore processing
 - Multi-core processors = Multicore processors
 - Full-text indexing = Fulltext indexing
 - Non-governmental organizations = Nongovernmental organizations

Acronyms

- Assumes acronyms start with 2+ upper-case letters
- Matches letters of acronym against other RIs, favoring matches at beginnings of words
 - EDP (Computing) = Electronic data processing
 - PIM (Computer software) = Personal information management programs
 - MARC format = Machine-Readable Cataloging—format
 - PRECIS indexing = Preserved Context Indexing System

LCSH equivalence relationships

- RIs at 025.434:
 - 750 #7 \$a Bibliographic Classification (Bliss)
 - 750 #7 \$a **Bliss Bibliographic Classification**
 - 750 #7 \$a **Bliss Classification**
- From LCSH record:
 - 150 ## \$a **Bliss Bibliographic classification**
 - 450 ## \$a Bibliographic classification (System)
 - 450 ## \$a **Bliss classification**
 - 450 ## \$a Bliss's Bibliographic classification
 - 450 ## \$a Classification, Bibliographic \$w nne

WordNet synonyms

- Lexical database of general English; includes > 155,000 words (mostly nouns)
- Captures variety of semantic relationships, including synonymy
 - Bigfoot = Sasquatch
 - Electronic computers = Computers
 - Systems programs = Systems software
 - Annuals = Yearbooks
 - Reference books = Reference works

Extending equivalence relationships

- If A and B are equivalent, then A—C and B—C are equivalent
 - Punch cards—nonelectronic data processing = Punched cards—nonelectronic data processing
 - GUI (User interface)—systems programs = Graphical user interfaces—systems programs
 - Annuals (Publications)—publishing = Yearbooks—publishing

Display of history information (1)

- Past practice in print:
 - Old location
 - 005.74 Data files and databases
 - Data mining relocated to 006.312
 - 006.42 Optical pattern recognition
 - Interdisciplinary works on bar coding relocated to 006.242
 - New location
 - 006.242 Bar coding [*formerly* 006.42]
 - 006.312 Data mining [*formerly also* 005.74]

Display of history information (2)

- Current practice in WebDewey:

006.312 *Data mining [formerly also 005.74] ▼

[000](#) Computer science, information & general works

[000](#) Computer science, knowledge & systems

[004-006](#) Computer science; computer programming, programs, data; special computer methods

[006](#) *Special computer methods

[006.3](#) *Artificial intelligence and natural computing

[006.3/1](#) *Machine learning

006.3/12 *Data mining [formerly also [005.74](#)]

History ▼

Data mining formerly also located in [005.74](#)
2008-08, Edition 22

expanded from [006.31](#)
2003, Edition 22

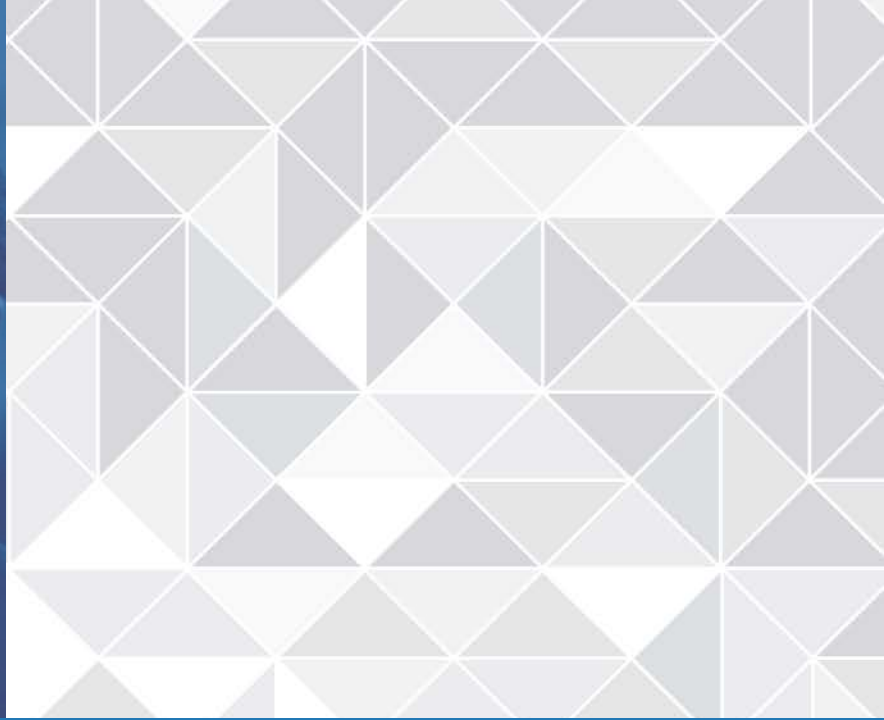
Display of history information (3)

- Current practice in WebDewey:

Notes	History
<p>Standard subdivisions are added for either or both topics in heading</p> <p>Including data validation in file processing, comprehensive works on file sharing</p> <p>Class here data file processing, data file and database management</p> <p>Class comprehensive works on data validation in 005.72</p> <p>Class interdisciplinary works on computer and information science aspects of databases in 025.04</p> <p>Class data files and databases with regard to their subject content with the subject, e.g., encyclopedic databases 030, nonbibliographic medical databases 610</p> <p>For file sharing using specific means, see the means, e.g., file sharing using peer-to-peer networks 004.652, file sharing using an Internet hosting service 004.678</p> <p>For specific types of data files and databases, see 005.75</p> <p>See Manual at 005.74 vs. 005.436</p> <p>See Manual at 025.04, 025.06 vs. 005.7, 005.74</p> <p>Data mining relocated to 006.312</p> <p>*Use notation T1-019 from Table 1 as modified at 004.019</p>	<p>Data mining relocated to 006.312 2008-08, Edition 22</p> <hr/> <p>005.74 *Data files and databases</p> <p>000 Computer science, information & general works</p> <p>000 Computer science, knowledge & systems</p> <p>004-006 Computer science; computer programming,</p> <p>005 *Computer programming, programs, data</p> <p>005.7 *Data in computer systems</p> <p>005.74 *Data files and databases</p> <p>005.74015113 Mathematical logic</p> <p>005.740285 Computer applications</p> <p>005.7406 Organizations and management</p> <p>005.74/1 *File organization and access methods</p> <p>005.74/2 *Data dictionaries and directories</p> <p>005.74/3 *Database design and architecture</p> <p>005.74/5 *Data warehousing</p> <p>005.74/6 *Data compression</p>

Display of history information (4)

- For next edition (2016?):
 - No relocation or discontinuation notes at old location
 - No use of “[*formerly . . .*]” at new location
 - All history information relegated to history box
 - Display of history box could be made optional / managed as a preference
 - Additional history information to be recovered back to DDC 20



Explore. Share. Magnify.

