International Dewey Breakfast
OCLC WELCOMES IFLA WLIC 2016

#WLIC2016
Agenda

- **EPC Meeting 139** – Alex Kyrios, Editor, Dewey Decimal Classification, OCLC

- **Data-driven development** – Rebecca Green, Dewey Editorial Program Manager, Dewey Decimal Classification, OCLC

- **Linking FAST to Wikipedia and Wikidata** – Diane Vizine-Goetz, Senior Research Scientist, OCLC Research

- **Principles underlying the EDUG recommendations for mapping involving Dewey** – Unni Knutsen, Section Mgr, University of Oslo, Humanities & Social Sciences Library

- **PANSOFT software developments** – Peter Werling CEO, PANSOFT GmbH
Alex Kyrios
Editor, Dewey Decimal Classification,
OCLC
T2—128 Subtropics

- New expansion for subtropical climates in Table 2
004–006 Computer science

• “Hot topics” in LCSH
• 005.8 Data security → Computer security
• Ability to distinguish between threats and countermeasures
• Provisions for data security, internet governance
320.52 Conservatism

- Clarity on how to classify ideologies that are both conservative and liberal
- Helps sort like ideologies across political systems where terms are used differently
New chemical elements

- Elements 113, 115, 117, and 118 have their places in the schedule.
- Official names not due until November.
Dementia

• Expansions to allow more detailed representation of dementia topics, including comprehensive works

• New coverage for specific types of dementia, such as frontotemporal, Lewy body, and vascular

Alois Alzheimer
Sports biographies

• Clarity on how to classify
• Replaces the Manual note at 796.092
• No more guessing what makes a sport!
796.8 Combat sports

- New developments for types of armed combat sports
- Distinguishing geographic treatment of martial arts vs. martial arts from certain areas
Period notation in the 900s

956.91 *Syria

956.910 2 640–1516

956.910 3 Period of Ottoman Empire, 1516–1920

956.910 4 1920–

956.912-.914 Localities of Syria

Add to base number 956.91 the numbers following —5691 in notation —56914–56914 from Table 2, e.g., City of Damascus 956.9144; then add further as follows:

001–009 Standard subdivisions

Add to 00 the numbers following 00 in notation 001–009 from table under 930–990, e.g., ethnic and national groups 004

02–04 Historical periods

Add to 0 the numbers following 956.91 in 956.9102–956.9104, e.g., City of Damascus during period of mandate 956.9144041
DATA-DRIVEN DEVELOPMENT

Rebecca Green
Dewey Editorial Program Manager,
Dewey Decimal Classification, OCLC
Moving forward

• In the past . . .
  – Each subject area reviewed during 7-year print cycle
• In the present and future . . .
  – Development efforts driven by objective data to focus on specific areas needing attention
Initial data sources (1)

- DDC 23 numbers assigned to WorldCat records
  - Frequency of number + frequency of numbers built with first number as base
  - Notation is not well developed / notation has few explicitly defined subordinate numbers /
  - Instruction to add further at notation is lacking or has not been used often
Top development candidates (1)

• 158.1 Personal improvement and analysis
• 006.3 Artificial intelligence
• 005.1 (Computer) Programming
• 248.4 Christian life and practice
• 658.4092 Executive leadership
Initial data sources (2)

- LCSHs (from most recent 5 years) assigned to WorldCat records
  - Frequency of assignment in WorldCat
  - Frequency of co-assignment with DDC number
  - Density of high-frequency assignment and co-assignment within same schedule area
Top development candidates (2)

• [004.16 Personal computers]
• 005.276 Programming for distributed computing
• 306.7 Sexual relations
• 341.4 Jurisdiction over physical space; human rights
• 341.6 Law of war
• [345.02 Criminal offenses]
Top development candidates (3)

• 363.3 Other aspects of public safety (e.g., terrorism, disasters)
• 364.1 Criminal offenses
• 572.8 Biochemical genetics
• 616.8 Diseases of nervous system and mental disorders
• 618.92 Pediatrics and geriatrics
• 741.5 Comic books, graphic novels, fotonovelas, cartoons, caricatures, comic strips
OCLC is involved with several key schemes

- **DDC**
  - Library world’s most-used classification scheme
    - English + translations from partners
- **FAST**
  - Faceted, general subject heading system
    - Derived from LCSH
- **VIAF**
  - Web-scale hub of library name authority data
    - Combines multiple name authority files
FAST (Faceted Application of Subject Terminology)

- Faceted vocabulary
  - Eight, distinct, non-overlapping facets or entities
- Available as an authority file
  - Unique identifiers for all headings
  - Relationships with LCSH are expressed
- Tools for application
- Published as Linked Data
Why linked data

• **Efficiency & Quality**
  – Facilitates *efficient* creation of *quality* metadata

• **Connectivity**
  – Helps to *connect* library data to the networked environment

• **Creativity**
  – Opens the door to *creative* reuse of library data
# Links to External Data Resources in the FAST File

<table>
<thead>
<tr>
<th>Facet Type</th>
<th>Persons</th>
<th>Organizations</th>
<th>Events</th>
<th>Titles of Works</th>
<th>Topic</th>
<th>Geographic places</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>698,103</td>
<td>362,382</td>
<td>12,461</td>
<td>63,071</td>
<td>407,350*</td>
<td>177,959</td>
</tr>
</tbody>
</table>

**Links to External Files**

<table>
<thead>
<tr>
<th></th>
<th>LC Subject Headings</th>
<th>LC Name Authority</th>
<th>VIAF</th>
<th>Wikipedia**</th>
<th>GeoNames</th>
<th>Total Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons</td>
<td>22,945</td>
<td>675,157</td>
<td>669,902</td>
<td>160,675</td>
<td>0</td>
<td>1,528,679</td>
</tr>
<tr>
<td>Organizations</td>
<td>8,414</td>
<td>353,953</td>
<td>352,111</td>
<td>5</td>
<td>0</td>
<td>714,485</td>
</tr>
<tr>
<td>Events</td>
<td>5,422</td>
<td>6,916</td>
<td>6,901</td>
<td>89</td>
<td>2</td>
<td>19,328</td>
</tr>
<tr>
<td>Titles of Works</td>
<td>85</td>
<td>62,986</td>
<td>62,923</td>
<td>1</td>
<td>0</td>
<td>125,995</td>
</tr>
<tr>
<td>Topic</td>
<td>217,569</td>
<td>0</td>
<td>62,923</td>
<td>75,935</td>
<td>0</td>
<td>293,502</td>
</tr>
<tr>
<td>Geographic places</td>
<td>46,288</td>
<td>121,923</td>
<td>121,239</td>
<td>64</td>
<td>85,411</td>
<td>374,925</td>
</tr>
</tbody>
</table>

*181,821 headings are non-subdivided;  
**Wikipedia links to non-topical facets were extracted from VIAF
• The application of Linked Data principles has the potential to improve the quality and usefulness of library metadata
• Unique identifier for all headings
• Identifier + base URL provides access to HTML or RDF description of the concept
Links among datasets enable people and software to navigate between resources and to discover and use additional resources.
Links from FAST to Wikipedia *connect* library data to the networked environment
Wikidata links provide access to other language versions of Wikipedia, e.g., German Wikipedia, French Wikipedia, etc.
Principles underlying the EDUG recommendations for mapping involving Dewey

Unni Knutsen, Oslo University Library

Mapping to Norwegian WebDewey
Information and documentation — Thesauri and interoperability with other vocabularies —

Part 2: Interoperability with other vocabularies
Workshop in Naples
EDUG recommendations

EDUG’s recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)

(Version 20160616)
Policy statement for the project
Mapping to Norwegian WebDewey
at the University of Oslo Library

The policy statement provides guidelines for the activities in the project at a general level. It expresses the aim and purpose for the project, as well as the principles and general conditions for how the project work will be conducted. Decisions made as part of the project should be based on the policy statement. Key concepts are defined in Section 4.
Hub model

Voc W

Voc X

Voc Y  Voc Z
Independent mappings

**Relationship types**
- **Exact equivalence** (\(=\)EQ)
  - Source vocabulary: Industry
  - Target vocabulary (WebDewey): 338 Production
- **Inexact equivalence** (~EQ)
  - Source vocabulary: Industry
  - Target vocabulary (WebDewey): 322.3 Business and industry
- **Broader mapping** (BM)
  - Source vocabulary: Industry
  - Target vocabulary (WebDewey): 343.07 Regulation of economic activity
- **Related mapping** (RM)
  - Source vocabulary: Industry
  - Target vocabulary (WebDewey): 333.7965 Energy for industrial use
Further examples

Relationship types

Source vocabulary | Target vocabulary (WebDewey)

Exact equivalence (=EQ)

Social medicine | 306.461 Medicine and health

Inexact equivalence (~EQ)

Home improvement | 643.7 Renovation, improvement, remodeling

Broader mapping (BM)

Fortune-telling by dice | 133.3 Divinatory arts

Related mapping (RM)

Living arrangements | 643.1 Housing
Join in the discussions!
PANSOFT SOFTWARE DEVELOPMENTS

Peter Werling
CEO, PANSOFT GmbH
Questions?

Alex Kyrios
Editor, Dewey Decimal Classification
kyriosa@oclc.org

Rebecca Green
Dewey Editorial Program Manager, Dewey Decimal Classification
greenre@oclc.org

Diane Vizine-Goetz
Senior Research Scientist, OCLC Research
vizine@oclc.org

Unni Knutsen
Section Mgr, University of Oslo, Humanities & Social Sciences Library
unni.knutsen@ub.uio.no

Peter Werling
CEO, PANSOFT GmbH
peter.werling@pansoft.de