Local Metadata management in a global environment

Daniel van Spanje
Global Productmanager Metadata Services
OCLC
“…metadata has become the structure on which we’re building information services.”

_in: Metadata everywhere: the catalog is out of the box, p. 5. (Nextspace, no. 15, p. 4 - 9)_
Multivariate Analysemethoden: Eine anwendungsorientierte Einführung
Klaus Backhaus

Local libraries

Hertie School of Governance (2 km)
FRIEDRICHSTRASSE 180
10117 Berlin, Germany

European School of Mgmt & Tech Information Center (2 km)
SCHLOSSPLATZ 1
10178 Berlin, Germany

Staatsbibliothek Zu Berlin (3 km)
PRENBISCHER KULTURBESITZ, POTSDBERG STR 33
The catalogue is a mash-up
Dewey T2—94 Australia displayed with Google Earth
The paradigm shift in cataloging

Main Entry: ²catalog
Variant(s): or catalogue
Function: verb
Inflected Form(s): cataloged or catalogued; catalog·ing or catalogu·ing
Date: 1598

Transitive verb
1 : to make a catalog of
   2 a : to enter in a catalog b : to classify (as books or information) descriptively

Intransitive verb
1 : to make or work on a catalog
2 : to become listed in a catalog at a specified price <this stamp catalogs at $2>
   — cat·a·log·er or cat·a·logu·er noun

(http://www.merriam-webster.com/dictionary/cataloging)
The catalog as the direct result of cataloging

Cataloging changes into The OPAC changes into a mash-up and discovery experience

Index engines search algorithms relevance ranking

Structured storage of metadata management

Efficiencies in relevance ranking

Content Integration into metadata management
The user changed

“No, you weren't downloaded. Your were born.”
STATEMENT OF INTERNATIONAL CATALOGUING PRINCIPLES

Introduction

Over forty years later, having a common set of international cataloguing principles has become even more desirable as cataloguers and their clients use OPACs (Online Public Access Catalogues) around the world. Now, at the beginning of the 21st century, an effort has been made by IFLA to produce a new statement of principles that are applicable to online library catalogues and beyond. The first principle is to serve the convenience of catalogue users.

2. General Principles

Several principles direct the construction of cataloguing codes. The highest is the convenience of the user.  

2.1. Convenience of the user. Decisions taken in the making of descriptions and controlled forms of names for access should be made with the user in mind.

2.2. Common usage. Vocabulary used in descriptions and access should be in accord with that of the majority of users.
Key findings:

• End users bring their expectations from popular Web sites to online catalogs

• The end user’s *delivery* experience is as important, if not more important than the *discovery* experience

• Most important for analog materials: summaries, tables of contents, etc.

• Most important for e-content: linking to the content itself

http://www.oclc.org/reports/onlinecatalogs/default.htm
Meet the “user” - primary personas of worldcat.org

Cliff Bungalow
Age: 52
Technology
- Has a 5 year old desktop computer at home, and recently got broadband as a benefit of moving into the condo which is new. He has a cell.

Retail/consumer habits
- He is particularly interested in the organization when it's a Westmont Police cheerleader.

Andy
Age: 25
General

“I just got my third article published...I can’t wait to tell my advisor.”

Graduate Student

Thomas F. Stephens III
“Doc Stephens”
Age: 61
(P) BOOMER
Technology
- Lovers radio shack
- Buys from Amazon
- Skype account
- Shops at Macy’s but buys socks at JC. Penny b/c they’re cheaper
- Wears a bike helmet every day
- Vacations are usually to visit family
- Drives an old yugo with all the school stickers on the back

Financial status
- NRI salary—he’s tenured, published enough to be respectable, well-liked by students
- Lives in a small bungalow, has basement"
- Lives within means, sees himself as solidly middle class

Hobbies
- Tennis (he sponsors the tennis club)
- Plays a mean blues harmonica (learned at NYU)

Info sources/influencers
- Colleagues, family
- Newspapers/Web sites
- The TV

Relationship with library
- Loves it deeply
- Doesn’t use it much day-to-day
- Finds an excuse to go in every couple of years, just in case something new has happened
- Frequently uses Google Scholar

“I want to ignite the next generation of math scholars—and continue my own research.”

Faculty/teacher

Goal
- To ignite the next generation of math scholars
- To push the department just a little bit each year with his somewhat quirky ideas
- To keep up with colleagues all over the world
- Read and gets every math journal there is...

Meet the “user” - primary personas of worldcat.org
### Metadata Sources are also changing

<table>
<thead>
<tr>
<th>Category</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professionally produced</td>
<td>• Bibliography – cataloging; abstracting and indexing services</td>
</tr>
<tr>
<td></td>
<td>• Authority and classification data</td>
</tr>
<tr>
<td></td>
<td>• Terminologies</td>
</tr>
<tr>
<td></td>
<td>• Publication supply chain data</td>
</tr>
<tr>
<td>Author/User contributed</td>
<td>• Institutional repositories</td>
</tr>
<tr>
<td></td>
<td>• Scholarly portals (e.g., arXiv.org)</td>
</tr>
<tr>
<td></td>
<td>• Tags, reviews, lists, etc.</td>
</tr>
<tr>
<td>Mined</td>
<td>• WorldCat Identities</td>
</tr>
<tr>
<td></td>
<td>• VIAF</td>
</tr>
<tr>
<td></td>
<td>• Facets for topics, places, events …</td>
</tr>
<tr>
<td></td>
<td>• FRBR Work Sets</td>
</tr>
<tr>
<td>3rd party content</td>
<td>• Databases and article content,</td>
</tr>
<tr>
<td></td>
<td>• Evaluative content</td>
</tr>
<tr>
<td></td>
<td>• Full text</td>
</tr>
</tbody>
</table>
Globalization

James Rosenau (NWU) 1924 - ..
Web page at GWU

People count (2007)

Distant proximities (2003)
Rosenau, James N.

Overview

Works: 146 works in 246 publications in 8 languages and 21,269 library holdings
Roles: Redacteur, Overig, Schepper
Classifications: xx1396, 327

Publication Timeline

Most widely held works about James N Rosenau

- The Truman doctrine and public opinion an application of a conceptual framework of James N. Rosenau by Gary J Buckley (Book)
- Governmental and societal influences on foreign policy: a partial examination of Rosenau's adaptation model by David William Moore (Book)
- Mexico, Cuba, and the Organization of American States: a case study of the idiosyncratic variable in international relations by Fred M Shaver (Book)
- James Rosenau his place in the debate over international relations theory by Brian McCormack (Book)
- Linkages in the Rosenau paradigm: a propositional inventory by Thomas Jeffrey Yehl (Book)
“Outward Integration”—Into the Cloud

“Integration should be outward rather than inward, with libraries seeking to use their components in new ways”

--Interviewee for LC report on future of the catalog (Karen Calhoun)
Create system-wide efficiencies in library management

WorldCat: an inclusive platform

Arabic
Bengali
Chinese
Cyrillic
Devanagari
Greek
Hebrew
Japanese
Korean
Latin
Tamil
Thai

MARC
ONIX
Dublin Core
FRBR
Full Text
New Contributors
Links
Unicode
Graphics, sound, motion
Create system-wide efficiencies in library management

WorldCat Growth since 1998

Millions of records

- 1998: 39
- 1999: 41
- 2000: 44
- 2001: 47
- 2002: 50
- 2003: 52
- 2004: 55
- 2005: 61
- 2006: 67
- 2007: 86
- 2008: 108
- 2009: 139
- 2010: 183
Create system-wide efficiencies in library management

Multilingual WorldCat

<table>
<thead>
<tr>
<th>Total Records</th>
<th>1998</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>23.9 m</td>
<td>64.3 m</td>
</tr>
<tr>
<td>German</td>
<td>2.2 m</td>
<td>17.9 m</td>
</tr>
<tr>
<td>French</td>
<td>2.3 m</td>
<td>16.3 m</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.6 m</td>
<td>4.5 m</td>
</tr>
<tr>
<td>Chinese</td>
<td>.7 m</td>
<td>4.3 m</td>
</tr>
<tr>
<td>Dutch</td>
<td>.2 m</td>
<td>2.9 m</td>
</tr>
<tr>
<td>Japanese</td>
<td>.8 m</td>
<td>2.8 m</td>
</tr>
<tr>
<td>Russian</td>
<td>.8 m</td>
<td>2.3 m</td>
</tr>
<tr>
<td>Italian</td>
<td>.7 m</td>
<td>2.1 m</td>
</tr>
<tr>
<td>Latin</td>
<td>.3 m</td>
<td>1.9 m</td>
</tr>
<tr>
<td>Hebrew</td>
<td>.3 m</td>
<td>1.2 m</td>
</tr>
<tr>
<td>Portuguese</td>
<td>.2 m</td>
<td>1.1 m</td>
</tr>
</tbody>
</table>

Percentage of Non-English Records

- 1998: 36%
- 2009: 53.8%
1.9 billion items and growing!

<table>
<thead>
<tr>
<th>Physical holdings in WorldCat</th>
<th>Licensed digital content in library collections</th>
<th>Local library content being digitized</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 million bib records</td>
<td>325 million electronic database records</td>
<td>30 million items (Google, HathiTrust, OAIster)</td>
</tr>
<tr>
<td>3.6 million digital items</td>
<td>JSTOR: 4.5 million records</td>
<td></td>
</tr>
<tr>
<td>1.5 billion holdings</td>
<td>Gale: 100 million records</td>
<td></td>
</tr>
</tbody>
</table>
More metadata: building the knowledge base

Core resources are the same

• EBSCO

KB release 87 (June 2010) totals:

119 providers
1721 collections
1383 full-text collections

4,317,813 records:
• 949,348 full-text records
• 1,102,359 records for free content
• 2,285,148 ebook records
WorldCat knowledge base: library input

- Each library can provide us with a feed of the holdings from their existing link resolver / knowledge base system.
- The data is matched to the WorldCat knowledge base and holdings are set on records that are already present in our knowledge base data.
The OCLC strategy part B: make OCLC Web Services a valued part of library operations

- WorldCat Search API
- WorldCat Registry API’s
- xISBN
- xISSN
- WorldCat Identities
- Dewey.info
- WorldCAT Knowledge base API
- Google Book Syndication
- SRU-based metadatasynchronization
- Experimental space for linked DDC data
- Available now: Linked data version of the DDC Summaries in ten languages
- Features include:
  - Access to the top three levels of the DDC in eleven languages
  - Actionable URIs for every class of the DDC Summaries
  - Classification semantics encoded in SKOS
  - Representations for machines and for humans
  - Data is reusable under a Creative Commons BY-NC-ND license
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Description</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>640</td>
<td>Home &amp; family management</td>
<td><a href="http://dewey.info/schema/e22/">http://dewey.info/schema/e22/</a></td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>Food &amp; drink</td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>640</td>
<td>Hauswirtschaft</td>
<td><a href="http://dewey.info/schema/e22/">http://dewey.info/schema/e22/</a></td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>Essen und Trinken</td>
<td></td>
</tr>
<tr>
<td>fr</td>
<td>640</td>
<td>Gestion de la vie familiale</td>
<td><a href="http://dewey.info/schema/e22/">http://dewey.info/schema/e22/</a></td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>Nourriture et boisson</td>
<td></td>
</tr>
<tr>
<td>sv</td>
<td>640</td>
<td>Hem &amp; hushåll</td>
<td><a href="http://dewey.info/schema/e22/">http://dewey.info/schema/e22/</a></td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>Mat &amp; dryck</td>
<td></td>
</tr>
<tr>
<td>pt</td>
<td>640</td>
<td>Gerenciamento da casa e família</td>
<td><a href="http://dewey.info/schema/e22/">http://dewey.info/schema/e22/</a></td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>Alimentos e bebidas</td>
<td></td>
</tr>
</tbody>
</table>
Putting it together: multilingual and mobile
<table>
<thead>
<tr>
<th>Call Number</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>360 E475</td>
<td>The Social Security fraud.</td>
<td>Ellis, Abraham, 1911-</td>
</tr>
<tr>
<td>368 LAN</td>
<td>The insurance maze: how you can save money on insurance—and still get the coverage you need / Kimberly Lankford.</td>
<td>Lankford, Kimberly</td>
</tr>
<tr>
<td>360.014 FEI</td>
<td>Delay, deny, defend: why insurance companies don’t pay claims and what you can do about it / Jay M. Feinman.</td>
<td>Feinman, Jay M.</td>
</tr>
<tr>
<td>368.3 STA</td>
<td>Marketing life insurance: its history in America [by] J. Owen Stalson.</td>
<td>Stalson, J. Owen, 1900-</td>
</tr>
<tr>
<td>368.38 HOG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WorldCat Knowledge Base API model

WorldCat knowledge base API:
- Designed to expose KB metadata
- Includes read access
- Includes write access
- Based on OpenURL 1.0 syntax
- Documented protocol syntax
- Same API used for internal and external services
- Supports service-neutral principles

All OCLC Discovery, Delivery, Metadata & Management System Services

KBWC API

Web-scale Management Services
- WMS ERM
- WMS License Manager
- Collection Builder

Librarians

Local Service

Third Party Service
Create system-wide efficiencies in library management

Embedded Google Book Search content

Preview this item

Text excerpt (Dial-a-Book)
More collaboration
More institutions
More Web-scale
More innovation
More synchronization
Create system-wide efficiencies in library management

**Synchronization: real-time update**

- Records added automatically to WorldCat
- Record id’s sent back to original database

- 4 million holdings contributed to date
- SRU based (Search and Retrieve via URL / Online process)

<table>
<thead>
<tr>
<th>NCC (Dutch Union Catalogue)</th>
<th>this FY</th>
<th>prev this FY</th>
<th>this mth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total records to Wcat this FY</td>
<td>350.937</td>
<td>338.358</td>
<td>12.579</td>
</tr>
<tr>
<td>Total adds to Wcat this FY</td>
<td>268.742</td>
<td>260.274</td>
<td>8.468</td>
</tr>
<tr>
<td>Total merges to Wcat this FY</td>
<td>82.195</td>
<td>78.084</td>
<td>4.111</td>
</tr>
</tbody>
</table>
Operational between NCC (Netherlands) and WorldCAT since March 2010

Supported actions: replace and delete (not: validate, create)

Tags can be added, modified or deleted

Current options: “add only” or “add only if tag not yet present”

Tags involved: UDC, DDC, LCC, NLM, LCSH, LC geographical names, URL’s to additional content

Actions per week between 164.996 in week 11 and 703.105 in week 19

Success rate of the actions between 96.6 % and 97.3 %
Create system-wide efficiencies in library management

Synchronization: real-time 2Way update

SRU statistics GGC (NL) from WorldCAT (SRU2Way): tags added to bib records

<table>
<thead>
<tr>
<th></th>
<th>UDC</th>
<th>NLM</th>
<th>DDC</th>
<th>LCC</th>
<th>LCSH</th>
<th>LCGN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kmc 5000 (045G)</td>
<td>kmc 5040 (045C)</td>
<td>kmc 5010 (045F)</td>
<td>kmc 5030 (045A)</td>
<td>kmc 5500 (044A)</td>
<td>kmc 5501 (044B)</td>
</tr>
<tr>
<td>Week 11 (March)</td>
<td>303</td>
<td>10.714</td>
<td>48.841</td>
<td>74.441</td>
<td>172.845</td>
<td>48.637</td>
</tr>
<tr>
<td>Week 12 (March)</td>
<td>761</td>
<td>16.305</td>
<td>84.145</td>
<td>120.080</td>
<td>293.403</td>
<td>70.194</td>
</tr>
<tr>
<td>Week 13 (March)</td>
<td>222</td>
<td>10.464</td>
<td>57.451</td>
<td>86.632</td>
<td>196.781</td>
<td>37.365</td>
</tr>
<tr>
<td>Week 14 (April)</td>
<td>235</td>
<td>21.044</td>
<td>64.274</td>
<td>98.540</td>
<td>247.242</td>
<td>48.799</td>
</tr>
<tr>
<td>Week 15 (April)</td>
<td>234</td>
<td>12.656</td>
<td>54.770</td>
<td>96.816</td>
<td>208.134</td>
<td>46.909</td>
</tr>
<tr>
<td>Week 16 (April)</td>
<td>198</td>
<td>7.878</td>
<td>32.887</td>
<td>55.732</td>
<td>135.115</td>
<td>25.317</td>
</tr>
<tr>
<td>Week 17 (April)</td>
<td>228</td>
<td>6.592</td>
<td>37.019</td>
<td>60.638</td>
<td>139.762</td>
<td>29.063</td>
</tr>
<tr>
<td>Week 18 (May)</td>
<td>721</td>
<td>17.591</td>
<td>108.470</td>
<td>170.567</td>
<td>559.669</td>
<td>115.988</td>
</tr>
<tr>
<td>Week 19 (May)</td>
<td>772</td>
<td>26.159</td>
<td>135.209</td>
<td>264.316</td>
<td>786.668</td>
<td>165.367</td>
</tr>
<tr>
<td>Week 20 (May)</td>
<td>355</td>
<td>10.471</td>
<td>56.563</td>
<td>291.181</td>
<td>291.181</td>
<td>57.362</td>
</tr>
<tr>
<td>Week 21 (May)</td>
<td>652</td>
<td>8.687</td>
<td>51.521</td>
<td>82.535</td>
<td>396.628</td>
<td>101.342</td>
</tr>
<tr>
<td>Week 22 (June)</td>
<td>507</td>
<td>12.738</td>
<td>53.905</td>
<td>92.738</td>
<td>376.651</td>
<td>73.705</td>
</tr>
<tr>
<td>Week 23 (June)</td>
<td>169</td>
<td>2.459</td>
<td>14.525</td>
<td>36.400</td>
<td>76.537</td>
<td>12.018</td>
</tr>
<tr>
<td>Total</td>
<td>5,357</td>
<td>163.758</td>
<td>799.580</td>
<td>1,530.616</td>
<td>3,880.616</td>
<td>832.066</td>
</tr>
<tr>
<td>Average per week</td>
<td>412</td>
<td>12.597</td>
<td>61.506</td>
<td>117.740</td>
<td>298.509</td>
<td>64.005</td>
</tr>
</tbody>
</table>
### Conclusion: metadata management workflows are changing

<table>
<thead>
<tr>
<th>Before the Web (B.W.)</th>
<th>After the Web (A.W.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• For finding and managing library collections (mostly print)</td>
<td>• For finding and managing many types of materials, for many user communities</td>
</tr>
<tr>
<td>• Catalogue records (well-understood rules and encoding conventions)</td>
<td>• Managed, created for, and consumed by many user communities</td>
</tr>
<tr>
<td>• Shared cooperative cataloguing systems</td>
<td>• Loosely coupled metadata management, reuse and exchange services among multiple repositories</td>
</tr>
<tr>
<td>• Usually created and edited one record at a time</td>
<td>• Multiple batch creation and metadata transfer services</td>
</tr>
<tr>
<td>• Record creation and editing generally a solitary activity undertaken by library-trained professionals</td>
<td>• Metadata creation and manipulation a fundamentally collaborative activity involving various specialists</td>
</tr>
</tbody>
</table>

Create a catalogue in text

Create a database with actionable data
An open cataloging and metadata platform that supports the organization and management of library collections (physical, licensed electronic and digital collections)

Functionality to efficiently create, manage and share metadata both at the collection level and at the level of a single information object

Shared metadata creation and management not only for individual titles and works, but also for people and other types of information objects

Easy, fast ways to ingest an OCLC cataloging subscriber’s data and keep member holdings synchronized with WorldCat

A vibrant, worldwide Expert Community working with OCLC to extend and improve WorldCat quality

Integration with WorldCat Local and WMS, and support for use by external applications for collections and metadata management (Nextspace, no. 15, p. 3)
Thank you!

Frankfurt
16 June 2010