WorldCat in the Global Information Space: Issues and Priorities in Enhancing WorldCat Quality

OCLC Global Council
April 13, 2011

Berndt Dugall
Direktor/Librarian, Universität Frankfurt

Karen Calhoun
Vice President, Metadata Applications
The ideal catalogue (database)

- Expectation of users:
  - Easy to handle (software)
  - Fast results (software and technical infrastructure)
  - Exactly what they want (quality of database)

- Users are primarily focused on content
  - Minor value of different manifestations (hardcover vs paperback; often different editions)

- Users do not want records but copies (content)

- Cataloguers do not want content but records
Problems and questions

- Quality and growth

*Automatic upload*

Dominant effect since 2008

1968 - 1997
1998 - 2007
2008 - 2010
Problems and questions II

• How to improve bibliographic records?
• How to avoid (remove) duplications?
• How to link holdings with the „correct“ record?
Outline

1. WorldCat in the global information space
   - Who uses WorldCat and what do they want?
   - Which WorldCat?
   - Focusing on what we know about the librarian’s view of WorldCat

2. What are the challenges we face today?

3. What is OCLC doing to address the challenges?
Who uses WorldCat? **Librarians do.**

- **Libraries:**
  - 389.3 million items cataloged
  - 57.8 million records added to WorldCat
  - 10.2 million interlibrary loans arranged
  - 68.4 million cataloging records exported

Who uses WorldCat.org? **End users do.**

<table>
<thead>
<tr>
<th>Total End Users</th>
<th>7,583</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students*</td>
<td>28%</td>
</tr>
<tr>
<td>Teacher or professor</td>
<td>22%</td>
</tr>
<tr>
<td>Business professional</td>
<td>19%</td>
</tr>
<tr>
<td>Other</td>
<td>16%</td>
</tr>
<tr>
<td>Retired</td>
<td>9%</td>
</tr>
<tr>
<td>Healthcare professional</td>
<td>6%</td>
</tr>
</tbody>
</table>

69% Students
Teacher/professor
Business/professional

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

Source: **Online Catalogs** study, p. 16
Online Catalogs: What Users and Librarians Want

End users expect online catalogs:

- To look/behave like popular Web sites
- To have summaries, abstracts, tables of contents
- To link directly to needed information

Librarians expect online catalogs:

- To help them carry out work responsibilities
- To have accurate, structured data
- To exhibit library principles of organization

www.oclc.org/us/en/reports/onlinecatalogs/
WHAT WE KNOW ABOUT MEMBERS’ VIEWS OF WORLDCAT
Web-based survey to understand members’ use of WorldCat

- 1,397 respondents
  - 895 North American institutions
  - 360 Institutions outside North America

- Studied satisfaction and perceptions of WorldCat
- Investigated potential for data quality enhancements

- November 2008 = baseline
- Survey being repeated second quarter calendar 2011
Which WorldCat interfaces did respondents use most and how frequently?

<table>
<thead>
<tr>
<th>Interface</th>
<th>Nearly every day</th>
<th>Few times a week</th>
<th>Few times a month</th>
<th>Few times a year or less</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connexion</td>
<td>74%</td>
<td>18%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>FirstSearch</td>
<td>48%</td>
<td>31%</td>
<td>16%</td>
<td>5%</td>
</tr>
<tr>
<td>WorldCat.org</td>
<td>29%</td>
<td>31%</td>
<td>27%</td>
<td>13%</td>
</tr>
<tr>
<td>Z39.50</td>
<td>63%</td>
<td>25%</td>
<td>11%</td>
<td></td>
</tr>
</tbody>
</table>
Satisfaction with WorldCat by region (November 2008)

<table>
<thead>
<tr>
<th></th>
<th>Total (n=1312)</th>
<th>North America (n=869)</th>
<th>Outside NA N=304</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent 10</td>
<td>19%</td>
<td>20%</td>
<td>17%</td>
</tr>
<tr>
<td>9</td>
<td>22%</td>
<td>24%</td>
<td>19%</td>
</tr>
<tr>
<td>8</td>
<td>31%</td>
<td>32%</td>
<td>29%</td>
</tr>
<tr>
<td>7</td>
<td>16%</td>
<td>15%</td>
<td>17%</td>
</tr>
<tr>
<td>6</td>
<td>6%</td>
<td>5%</td>
<td>9%</td>
</tr>
<tr>
<td>5</td>
<td>3%</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>4</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Poor 1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Don't know</td>
<td>1%</td>
<td>0%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Three-fourths of North American respondents rated overall satisfaction as 10, 9 or 8

Versus

Two-thirds of respondents outside North America
“Why are you satisfied or not?” An analysis of the responses by rating

<table>
<thead>
<tr>
<th>Positive comments on the database, good coverage, etc.</th>
<th>Respondents who rated 10 (excellent)</th>
<th>Respondents who rated 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72%</td>
<td>68%</td>
</tr>
</tbody>
</table>

| Positive comments on access, ease of use, reliability | 47%                                 | 43%                     |

| Positive comments on service: Meets needs for cataloging, ILL, patrons | 45%                                 | 32%                     |

| Negative comments on problems: searching, minimal records, duplicates | 3%                                  | 32%                     |
Then came the deluge ...

Unprecedented growth in WorldCat 2008-2010
WHAT ARE THE CHALLENGES?
Issues reported: EMEA WorldCat Quality Breakout Session

- Duplication in WorldCat
- The impact of “parallel records” on search displays and the resulting “holdings scatter” across those parallel records
- Incorrect or missing links to union catalogs and/or library local catalog

Deutsche Nationalbibliothek, Frankfurt, site of EMEARC 2011 Meeting

Details in Calhoun trip report, email to SLT 8 March
Why are duplicates a problem?

- Missed opportunities
  - Discovery: Confusion in searching/browsing
  - Delivery: The copy available to the end user is hidden
  - Cataloging: Increased cost in copy cataloging: i.e., it takes longer to find the “right record”

- Result: Dissatisfaction with WorldCat
Duplicate Detection and Resolution (DDR) of WorldCat bibliographic records

- Reimplementation and expansion of previous software: Now handles all types of material (not just books)

- Fully operational in early 2010 in two separate processes
  - “Walking the database” (Complete September 2010)
  - Selected records from each day’s daily journal files (Ongoing)

- The result is “continuous cleaning” of WorldCat
Cumulative number of duplicates removed, Jan—Sept 2010: 5.1 million

“Walking the Database”

[Graph showing the cumulative number of duplicates removed from January to September 2010, with data points for each month.]

OCLC | The world’s libraries. Connected.
Despite DDR, challenges remain

Parallel records

Language of cataloging

*NOT* language of resource

Reprints and reproductions
245 0 0 Made of light : the art of light and architecture / Mark Major, Jonathan Speirs, Anthony Tischhauser.


300 231 S. b zah


300 231 str. : b ilustr. ; c 30 cm.

500 Bibliografija: str. 231.

Language of cataloging = Slovenian

Parallel records
Reproductions

001 1394475
010 33024226
040 DLC $c MUB $d OCL $d SER $d CHS $d OCL $d OCLCQ $d BUF $d OCLCG $d STF $d NIALS $d OCL
041 1 eng $h ger
050 00 Q141 $b L62 1933a
082 04 509 $b L563X
096 WZ 112 L563g 1933
100 1 Lenard, Philipp, $d 1862-1947
245 10 Great men of science; $b a history of scientific progress, $c by Philipp Lenard ...
translated from the second German edition by Dr. H. Stafford Hatfield, with a preface by E.N. da C. Andrade ...
260 New York, $b Macmillan Co., $c 1933.
300 xx, p., 1 l., 389 p. $c 22 cm.
505 00 $t Pythagoras of Samos — $t
WHAT IS OCLC DOING TO ADDRESS THE CHALLENGES?
From WorldCat Quality “Projects” to a WorldCat Quality “Program”

- WorldCat quality initiatives = a “Red dot” in FY12
- White paper on WorldCat quality—in process
- Gather community input on white paper
- GLIMIR project—underway
- First user-apparent results from GLIMIR in Discovery—WorldCat.org and WorldCat Local
- Regular communications—highly visible WorldCat quality program
What is GLIMIR?
(Global Library Manifestation Identifier)

- GLIMIR is a tool for grouping related records in WorldCat
  - Because of the pace of growth and the global sources of data, we have many things in WorldCat that represent the same bibliographic entity; for example: digital copies, reproductions, different language of cataloging

- GLIMIR improves the view of records for the end user by bringing those records together
  - This allows us to show all of the holdings together (a big benefit to end users)
  - And to show the most appropriate record to the end user (by language or other preference)

Timeline for GLiMIR

<table>
<thead>
<tr>
<th>FY10</th>
<th>FY11</th>
<th>FY12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifications &amp; Design</td>
<td>Infrastructure</td>
<td>GLIMIRize the database</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implement in Discovery</td>
</tr>
</tbody>
</table>

Bonus:
Creates a new identifier that enhances the value proposition of WorldCat (e.g., Hathi Trust)
GLIMIR WILL CLUSTER THESE FROM SIX TO TWO RESULTS.
WorldCat.org & WorldCat Local usage of GLIMIR

• THE MAIN THINGS:

• GLIMIR clusters will greatly improve the accuracy of FRBR worksets, reducing appearances of exact duplicate records

• GLIMIR clusters will improve the clustering of holdings, improving linking from WorldCat.org to individual library catalogs
Discussion