Today’s WorldCat: New Uses, New Data

Ted Fons
Executive Director, Data Services & WorldCat Quality

Good Practices for Great Outcomes:
Cataloging Efficiencies that Make a Difference

DePaul University - Chicago
WorldCat: Happy 40th!

26 August 1971
WorldCat has changed a lot since it was born in 1971

It is growing to suit many different uses

Let’s talk about those uses and what we are doing to preserve the quality and reliability of WorldCat...

GLIMIR Deduplication WMS

Full Text Content More Links WorldCat Local

Global Cooperative Cataloging
FIRST...
WHAT’S IN WORLDCAT?
## Representing the collective collection

### 2.1 billion items and growing!

<table>
<thead>
<tr>
<th>Physical holdings in WorldCat</th>
<th>Licensed digital content/articles in library collections</th>
<th>Local library content being digitized</th>
</tr>
</thead>
<tbody>
<tr>
<td>243 million bib records</td>
<td>526.6 million records</td>
<td>30 million items (Google, HathiTrust, OAIster)</td>
</tr>
<tr>
<td>1.75+ billion holdings</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As of 30 September 2011
<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>189,421,960</td>
</tr>
<tr>
<td>Serials</td>
<td>9,363,065</td>
</tr>
<tr>
<td>Visual materials</td>
<td>7,220,470</td>
</tr>
<tr>
<td>Maps</td>
<td>3,725,484</td>
</tr>
<tr>
<td>Mixed materials</td>
<td>4,072,021</td>
</tr>
<tr>
<td>Sound recordings</td>
<td>8,437,557</td>
</tr>
<tr>
<td>Scores</td>
<td>5,736,058</td>
</tr>
<tr>
<td>Computer files</td>
<td>7,946,335</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>235,822,950</strong></td>
</tr>
</tbody>
</table>

as of July 1, 2011
### Percentage of records for non-English materials

<table>
<thead>
<tr>
<th>Language</th>
<th>Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>97.7m</td>
</tr>
<tr>
<td>German</td>
<td>33.0m</td>
</tr>
<tr>
<td>French</td>
<td>21.3m</td>
</tr>
<tr>
<td>Spanish</td>
<td>9.7 m</td>
</tr>
<tr>
<td>Japanese</td>
<td>7.5 m</td>
</tr>
<tr>
<td>Chinese</td>
<td>5.9 m</td>
</tr>
<tr>
<td>Italian</td>
<td>3.9 m</td>
</tr>
<tr>
<td>Dutch</td>
<td>3.5 m</td>
</tr>
<tr>
<td>Russian</td>
<td>3.2 m</td>
</tr>
<tr>
<td>Latin</td>
<td>3.2 m</td>
</tr>
</tbody>
</table>

**Total Records:** 235.5 m

**Percentage:** 58.5% as of 30 June 2011
Who uses WorldCat?

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

Public, college/university, State and national = 40%

Who uses WorldCat.org?

End Users:
Students,
Teacher/professor,
Business professional

<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>

<table>
<thead>
<tr>
<th>Type of Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate college/Post-graduate</td>
<td>54%</td>
</tr>
<tr>
<td>Undergraduate college/Post-secondary</td>
<td>30%</td>
</tr>
<tr>
<td>High/Secondary school</td>
<td>10%</td>
</tr>
<tr>
<td>Middle/Intermediate/Junior high school</td>
<td>3%</td>
</tr>
<tr>
<td>Other</td>
<td>2%</td>
</tr>
<tr>
<td>Elementary/Primary school</td>
<td>1%</td>
</tr>
</tbody>
</table>

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

Source: Online Catalogs study, PDF p. 16
Where does the content come from?

- Member Libraries: 72.65%
- National Libraries: 21.23%
- Vendor Partners: 1.59%
- LC: 4.53%
WorldCat growth since 1998

What happened here?
WorldCat got very big from a huge variety of sources, formats & cataloging schemes.

We started to look at WorldCat quality from the user’s perspective.
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 staff carry out work responsibilities
- to have accurate, structured data
- to exhibit library principles of organization

End-User Results: Recommended Enhancements

- More links to online content/full text: 36%
- More subject information: 32%
- More tables of contents: 18%
- Add summaries/abstracts: 18%
- More information in the “details” tab*: 16%
- More author information: 14%
- Add sample text/excerpts: 12%
- Remove duplicate records: 12%
- More edition information: 11%
- More selection of non-English language items: 10%
- More cover art: 10%
- More reader reviews: 9%
- More citation information: 9%
- Add recommendations: 7%
- Add editorial reviews from popular publications: 7%
- More publisher information: 7%
- Increase accuracy (e.g., name, subject headings): 7%
- More format/type information: 6%
- Add ratings: 5%
- Other: 5%

Source: Online Catalogs study, PDF p. 51
### Relative Ranking of Data Quality Enhancements

**In Library and End-user Surveys**

*Which of the following enhancements would you recommend? (Library survey)
What changes would be most helpful to you in identifying the item that you need? (End-user pop-up survey)*

<table>
<thead>
<tr>
<th>Comparable Enhancements Choice*</th>
<th>Top Third</th>
<th>Middle Third</th>
<th>Bottom Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merge duplicate records</td>
<td>#1 Choice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More links to online content</td>
<td>#1 Choice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add tables of contents to records</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add summaries to records</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add cover art to results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add more formats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More records for non-English materials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add excerpts to the records</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: *Online Catalogs* study, PDF p. 52

**Basis of 2009-2010 WorldCat Quality Program**
SO, WHAT DID WE DO FOR END USERS?
7.1 million eBooks records added

Some additions

- Project Gutenburg
- Sage
- BioDiversity Heritage Library
- MyiLibrary
- Google Book Search
- HathiTrust
Central index of citations leading to full text

Article Citations

[Graph showing a line chart with data points from June 2006 to January 2011, illustrating the increase in article citations over time.]
<table>
<thead>
<tr>
<th>Date</th>
<th>WorldCat Local Central Index Article Records</th>
<th>WorldCat Local Central Index Journal Titles Indexed</th>
</tr>
</thead>
<tbody>
<tr>
<td>As of June 30, 2011</td>
<td>530,886,124</td>
<td>85,834</td>
</tr>
</tbody>
</table>
Better links: brought to you by the WorldCat knowledge base

- One place to manage a library’s electronic holdings--both ejournals and ebooks--at the network level (i.e., in the cloud)
- Collection or package level management of holdings
- Controls display of journal/book/article level links in WorldCat Local
- Enables resource sharing of e-articles (as licenses allow)
WHAT DID WE DO FOR LIBRARIANS?
Recommendations from librarian survey

Merge duplicate bibliographic records

Enrichment—TOCs, summaries, cover art—work with content suppliers, use APIs, etc.

Make it easier to make corrections to records (fix typos; do upgrades); “social cataloging” experiment—Wikipedia

More emphasis on accuracy/currency of library holdings
MERGE DUPLICATE BIBLIOGRAPHIC RECORDS
“Duplicates” in WorldCat

• Real Duplicates:
  • Delete them
    • ...actually, we merge them so we can keep unique data

• Apparent duplicates
  • Cluster them
Removing Real Duplicates

http://www.oclc.org/worldcat/catalog/quality/ddr
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 2 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
MANAGING APPARENT DUPLICATES & IMPROVING WORKS
GLIMIR = Global Library Manifestation Identifier.

Clusters manifestations and assigns unique identifier to each manifestation

Clusters records for parallel records (differing languages of cataloging for the same manifestation) and for reproductions

Re-clusters FRBR work sets
Apparent duplicates in WorldCat.org

Double Indemnity
by Billy Wilder; Barbara Stanwyck; Edward G Robinson; et al
Language: No Linguistic Content
Publisher: [S.l.]: BOC 2, 1944.

Double Indemnity
by Raymond Chandler; Billy Wilder; Fred MacMurray; Barbara Stanwyck; Edward G Robinson; Porter Hall; Jean Heather; Tom Bonanova; John Philliber; John F Seitz; Doane Harrison; Miklós Rózsa; Edith Head; Richard Schickel; Lem Dobbs; Nick Redr Richard Crenna; Lee J Cobb; Robert Webber; Samantha Eggar; James M Cain; Paramount Pictures, Inc.; Universal Studios H DVD video: NTSC color broadcast system
Language: English
Publisher: Universal City, CA : Universal Studios Home Entertainment, [2006]

Double Indemnity
by Paramount Pictures; Joseph Santrom; Miklos Rozsa; John F Seitz; James M Cain; Billy Wilder; Fred MacMurray; Barbara Stanwyck, Raymond Chandler, Edward G Robinson
DVD video: Videocassette
Language: English
Publisher: UIP via Roadshow, 1944.
Future Enhancements for Discovery

Non-English browser setting: Show the end user the record in their preferred language.

Improved FRBR worksets will provide a better “Work” view in WorldCat.org (coming in 2012)
Capturing more data through batchloading

July 2009-July 2011

- LC-type call number: 34%
- Dewey-type call number: 41%
- Other call number: 145%
- Contents/Summaries: 87%
- Subject terms: 63%
- URLs: 430%

Comparing data between July 2009 and July 2011.
Enriched content from partners

43 million data elements under contract
- Over 9.5 million book jacket covers, summaries, 1st chapters
- Over 14.4 million ToCs
- Over 360,000 music album covers (added Aug. 2010)
- Over 1.5 million non-US book covers (added Nov. 2010)
Mining WorldCat: Sharing data elements across a FRBR Work Set

- **Work**
  - The novel

- **Expression**
  - Original Text
  - Translation
  - Critical Edition

- **Manifestation**
  - Summary
  - Classification
  - Subject Terms

- **Records**
MAKE IT EASIER TO CORRECT RECORDS: “SOCIAL CATALOGING”
Making it easier to correct the records: WorldCat community maintenance

<table>
<thead>
<tr>
<th>Activity</th>
<th>TOTAL FY10</th>
<th>TOTAL FY11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Community</td>
<td>271,626</td>
<td>304,759</td>
</tr>
<tr>
<td>Database Enrichment</td>
<td>198,084</td>
<td>235,533</td>
</tr>
<tr>
<td>Minimal-Level Upgrade</td>
<td>176,618</td>
<td>194,634</td>
</tr>
<tr>
<td>Enhance Regular</td>
<td>176,491</td>
<td>155,713</td>
</tr>
<tr>
<td>Enhance National</td>
<td>45,451</td>
<td>47,876</td>
</tr>
<tr>
<td>CONSER Authentication</td>
<td>15,705</td>
<td>21,208</td>
</tr>
<tr>
<td>CONSER Maintenance</td>
<td>61,949</td>
<td>57,917</td>
</tr>
<tr>
<td>TOTAL</td>
<td>945,924</td>
<td>1,017,640</td>
</tr>
</tbody>
</table>
WORLDCAT STEWARDSHIP BY OCLC STAFF
## OCLC Staff Maintenance Activity in FY 2010

<table>
<thead>
<tr>
<th>Activity</th>
<th>TOTAL FY10</th>
<th>TOTAL FY11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic Records Replaced</td>
<td>12,511,044</td>
<td>23,305,162</td>
</tr>
<tr>
<td>Records Merged</td>
<td>150,992</td>
<td>301,667</td>
</tr>
<tr>
<td>Authority Records Created</td>
<td>1,977</td>
<td>119</td>
</tr>
<tr>
<td>Authority Records Replaced</td>
<td>94,744</td>
<td>797</td>
</tr>
<tr>
<td>CIP Records Upgraded</td>
<td>16,145</td>
<td>5,742</td>
</tr>
</tbody>
</table>
Other OCLC enhancements to WorldCat - a couple of recent examples

Updated subject headings

- Recent example: *Cookery* changed to *Cooking*
- Over 314,000 records affected
- ~75 new subject headings proposed to Library of Congress

Adding Linking ISSNs (ISSN-L)

- Added to about 800,000 records thus far

Provider-neutral records

- About 9.5 million records touched thus far
Additional OCLC staff enhancements

Adding non-Latin cross-references to authority records
  • Almost 500,000 records affected
  • Non-Latin forms derived from WorldCat records

Authority records for geographic names
  • Indirect subdivision forms added (about 90,000 records)
  • Geographic coordinates added to field 034 (more than 78,000 records)
Looking ahead -

- More automated enrichment of bibliographic records from mining FRBR work set data (summer/fall 2011)
Another quality survey to begin.

WorldCat Quality Whitepaper

Formal definitions of data quality & consistency.

NEW USES FOR WORLDCAT...
New Uses for WorldCat

Progressing from a cataloging and ILL database to...

Discovery  Management System:  Circulation  Acquisitions  Syndication to Google
WorldCat Local more than just books

As of Sept 1, 2011

- Maps: 3.6 million
- Scores: 5.6 million
- Visual materials: 7.7 million
- Sound recordings: 8.5 million
- Serials: 9.5 million
- eBooks: 11.9 million
- Conference proceedings: 12.1 million
- Institutional repository records: 14.3 million
- Archival materials: 16.1 million
- Theses and dissertations: 16.7 million
- Web/Internet resources: 28.2 million

Books: 198 million

Articles: 614 million
Print holdings display based on WorldCat knowledge base holdings
Evaluative content

Goal: Offer a robust collection to enhance discovery
How: Aggregate evaluative content from multiple sources
Status: 44+ million data elements included

- 6.5 million book jacket covers
- 14.4 million ToCs
- 3.1 million summaries
- 294 thousand author biographies
- 291 thousand first chapters
- 398 thousand million album covers
- 1.7 million user reviews
- 360 thousand album reviews
- 12 million ratings
Einstein: his life and universe

Author: Walter Isaacson


Edition/Format: Book: Biography: English

Summary: The first full biography of Albert Einstein since all of his papers have become available shows how his scientific imagination sprang from the rebellious nature of his personality. Biographer Isaacson explores how an imaginative, impertinent patent clerk—a struggling father in a difficult marriage who couldn't get a teaching job or a doctorate—became the locksmith of the mysteries of the atom and the universe. His Read more...

Find a copy in the library

Macalester College
Mobile access included
Quantitative results:
Increased usage of resources

- **ILL books**: Increased by 270%
- **resolver hits**: Increased by 74% (Willamette University), 80% (Portland Community College)
- **consortial borrowing**: Increased by 59% (University of Washington), 21% (Willamette University)
- **articles**: Increased by 58% (Willamette University)

Institutions:
- University of Washington
- Willamette University
- Portland Community College
Management Systems
What’s ahead:
OCLC Web Scale Management Services

The first cooperative management service for libraries.

Lower total cost of ownership
Simplify back-office operations
Reduce support costs for disparate systems
Customize and extend services through an open, extensive platform
Free staff time for high-priority services

Now Available!

Circulation  
Acquisitions  
License Management
New Data for WMS

Circulation
- Patron/Identity
- Item
- Circ History
- Fines
- Holds
- Loan Rules

Acquisitions
- Vendors
- Funds
- Invoices
- Payments

License Management
- License terms
- Knowledge base
What is new?

Circulation
- Transaction data at the network level

Acquisitions
- Cooperatively managed vendor data

License Management
- Cooperatively managed license terms
- Cooperatively managed knowledge base
Local Data Fields:

- Required to record the full bibliographic and copy-level details
- Many libraries require this to provide a complete discovery and management system environment
Storing Local Bibliographic Data via OCLC

How to order local bibliographic data for display to your users

1. In the Bibliographic Batchload Order form question “What OCLC product(s) will be supported by your project” select WorldCat Local and/or Web-scale Management Services.

See complete ordering instructions in Order a Batchload Project in the Batch Services User Guide.

2. In the question “Do you want to display local bibliographic data to your users via WorldCat Local or Web-scale Management Services?” select Yes, and then click to select check boxes for one or more of the fields listed.

Local bibliographic data fields you can select in your order:

<table>
<thead>
<tr>
<th>Uniform title (240)</th>
<th>Local subject added entry—Geographic name (691)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specialized notes (500, 59X)</td>
<td>Added class number (695)</td>
</tr>
<tr>
<td>“With” notes (501)</td>
<td>Local subject added entry—Personal name (696)</td>
</tr>
<tr>
<td>Subject added entry—Personal name (600)</td>
<td>Local subject added entry—Corporate name (697)</td>
</tr>
<tr>
<td>Subject added entry—Corporate name (610)</td>
<td>Local subject added entry—Meeting name (698)</td>
</tr>
<tr>
<td>Subject added entry—Meeting name (611)</td>
<td>Local subject added entry—Uniform title (699)</td>
</tr>
<tr>
<td>Subject added entry—Uniform title (630)</td>
<td>Added entry—Personal name (700, 790, 796)</td>
</tr>
<tr>
<td>Subject added entry—Topical term (650)</td>
<td>Added entry—Corporate name (710, 791, 797)</td>
</tr>
<tr>
<td>Subject added entry—Geographic name (651)</td>
<td>Added entry—Meeting name (711, 792, 798)</td>
</tr>
<tr>
<td>Index term—Genre form (655)</td>
<td>Added entry—Uniform title (730, 793, 799)</td>
</tr>
<tr>
<td>Local subject added entry—Topical term (690)</td>
<td></td>
</tr>
</tbody>
</table>
