Accessing library materials via Google and Other Web Sites

Janifer Gatenby, Strategic Research, OCLC PICA

ELAG 9th May 2007, Barcelona, Spain
Modern users

- Use Google, Yahoo, MSN ++
- Amazon, eBay, Netflix
- Search & find themselves
- Register & order
  - Internationally
  - Request home delivery

If you can’t beat ‘em, join ‘em
Aug '03: Google contract signed

Oct '03: Google harvests 150,000 records

Sep '03: Google given go-ahead to harvest records

Dec.'03: Records begin to appear in Google; 800 inbound-links logged (search-site-originating [SSO])

Jan.'04: 32,000 inbound links logged (SSO)

Mar.'04: 109,000 inbound links logged (SSO)

May '04: Yahoo contract signed

May '04: Yahoo harvests records

May '04: 725,000 inbound links logged (SSO)

Jun '04: Yahoo completes indexing of 2 million WC records

Aug '06 Countries with loading contracts to Google
Australia, China, *Czech Republic, *Denmark, Ireland, Israel, Hungary, Lithuania, *Netherlands, Taiwan, *United Kingdom, *United States

* Significant contributions via WorldCat
How Data is Contributed

- XML record
  - Subset at work level
  - ~75% holdings

- Inktomi Data Interchange Format (IDIF)

- Title, author, publisher, identifiers, electronic location, subject, contents, country holding, holdings count

- Google
  - 4.4 million (3m clusters)

- Google Scholar
  - More: whatever of 67 million maps to its dbase

- Google Books
  - 100% mapping TO

- Yahoo
  - 3.5 (3m clusters)

- MSN
  - 4.5 million growing to 10 million
  - Physical sciences & biomedical
Open WorldCat

Limitations on loading imposed by harvesters – Google only interested in titles with many holdings
Common Problems

• Coverage
• Ranking
  – Conflicting business models
  – Page hits, referrals + $ revenues
  – 2 billion pages 2004; 12 billion 2007
• Matching
• Work level data needed
• Not enough influence
Positive Signs: Statistics

• March 2007
  • Total Referrals: 14,118,777
    - Av. Per day: 455,444
  • Links to library services: 834,866
    - % Conversion to lib = 10.5%

• 03 2006
  • 96% ➤ 363%
Referrals to WorldCat.org pages

Referrals in millions

Month over month

- 2005
- 2006

Exposure
What we measure:

- Impressions
- Click-throughs
- Conversions
WorldCat.org Tracked Statistics at-a-glance

Impressions:
300M to 1B/year

Click-through:
10M/month
120M per year

Conversion:
550k/month
6.6M/year

Conversion rate:
- Overall: 5.5% = “best in class” for web site
- 5.9% in January 07

It’s worth noting that our conversion rate is well above Google’s “typical” range of 1-3%.
Worldcat.org Referrals March / April 2007

- Library sites: 31%
- Education sites: 9%
- Yahoo: 26%
- Google books: 14%
- Google: 9%
- Google Scholar: 3%
- Blogs, CNN, AltV+: 8%

Referral sources <50% of the traffic
Web Page Rankings


<table>
<thead>
<tr>
<th>Rank</th>
<th>Website</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>YouTube</td>
<td>19,652</td>
</tr>
<tr>
<td>10</td>
<td>Wikipedia</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>Amazon.com</td>
<td>10</td>
</tr>
<tr>
<td>42</td>
<td>Flickr</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>19,652 (11,000)</td>
<td></td>
</tr>
</tbody>
</table>
Some Observations

- Need for maximum exposure
  - Sites in addition to Google
  - Google & Yahoo swap places as top referrers
- Users value library information
  - Conversions way above web average
- Alternative means of exposure are effective
4,307 referrals March 2007

Worldcat Websites
People Who Like Worldcat Sites

mazaazam
30 / M
New Jersey USA

castagna
30 / M
England UK

judell
M
New Hampshire USA

marijar
F
Croatia

Featured Worldcat Sites

Try a search [OCLC - Open WorldCat program]
Reviewed by Metamorphosis on Jan 7, 2006 10:23pm
A tool to find books at a library near you, cool.
1 review • worldcat, library-resources, library, books, search

online.wsj.com/article_email/0,,SB110298438601999135-IRjf...
Reviewed by judell on Dec 14, 2004 4:47am
Pignic banking on a concrete database. Google’s service allows consumers to type in their ZIP Code and by libraries near them.

WorldCat [OCLC]
Reviewed by mamabear10 on Feb 25, 6:29am
This looks interesting!
More observations

- **Google Print**
  - 950,765 in March 2007
  - Link from snippets to metadata
  - Copyrighted material searching holdings

- Hard working with search engines
Referrals to Amazon.Com Q1 2007

<table>
<thead>
<tr>
<th></th>
<th>Ger</th>
<th>Can</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referrals</td>
<td>0</td>
<td>1493</td>
<td>7291</td>
<td>71072</td>
</tr>
<tr>
<td></td>
<td>1546</td>
<td>1311</td>
<td>6525</td>
<td>82813</td>
</tr>
<tr>
<td></td>
<td>2553</td>
<td>1314</td>
<td>6461</td>
<td>92631</td>
</tr>
<tr>
<td>total</td>
<td>79856</td>
<td>92195</td>
<td>102959</td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>0</td>
<td>3.82%</td>
<td>8.30%</td>
<td>4.84%</td>
</tr>
<tr>
<td>Conversions</td>
<td>4.72%</td>
<td>1.68%</td>
<td>9.16%</td>
<td>4.26%</td>
</tr>
<tr>
<td></td>
<td>4.07%</td>
<td>2.51%</td>
<td>7.85%</td>
<td>4.05%</td>
</tr>
<tr>
<td>av.</td>
<td>5.65%</td>
<td>6.60%</td>
<td>6.16%</td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>102959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can</td>
<td>92195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>102959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>92631</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orders</td>
<td>57</td>
<td>22</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td></td>
<td>104</td>
<td>507</td>
<td>3747</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>4100</td>
<td>4218</td>
<td>4391</td>
<td></td>
</tr>
</tbody>
</table>

Remember clicks to libraries 834,886
Get It Button

![Image of a web browser window with a table displaying book information]

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Author</th>
<th>Format</th>
<th>Supplier</th>
<th>Fee</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moby Dick or the Whale (Modern Library)</td>
<td>Herman Melville, Rockwell ...</td>
<td>Book</td>
<td>Amazon</td>
<td>$14.93</td>
<td>Purchase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book</td>
<td>Barnes &amp; Noble</td>
<td>$21.95</td>
<td>Purchase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book</td>
<td>Open WorldCat</td>
<td>Free</td>
<td>2 week loan</td>
</tr>
<tr>
<td>On Time and Being</td>
<td>Martin Heidegger</td>
<td>Book</td>
<td>Amazon</td>
<td>$24.03</td>
<td>Purchase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book</td>
<td>Barnes &amp; Noble</td>
<td>$29.85</td>
<td>Purchase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book</td>
<td>Open WorldCat</td>
<td>Free</td>
<td>2 week loan</td>
</tr>
</tbody>
</table>

OCLC PICA

Discovery to Delivery
Delivery Components

- online user registration, (free)
- registry of digital masters (free)
- global library registry (free)
- direct home delivery pilot
- NLA also conducting ambitious DD pilot

The delivery gap will not be solved just by making available new technical components. Libraries must make the necessary policy changes to cooperate on an international scale.
The motivation of exposure has been to attract one’s own user population to the resources of one’s library.

• BUT ALSO
  – Your users find things in other libraries on a global scale
  – Other libraries’ users from anywhere in the world find things in yours
In Conclusion

- We cannot expect the user to come to the library web site anymore
- It doesn’t matter if our sites are found by the back door: the more the merrier
- Now is the time to focus on delivery
  - Focus is on discovery at the expense of Delivery

Just how many of our users who are currently using Amazon would acquire from a library instead if it were just as easy
Thank you very much

Acknowledgments to:
Chip Nilges and Bill Brembeck