Web Scale Discovery

What is it?

...and how does it work?
- Overview of WSD
- Discovery at Harvard
- Open Discovery Initiative
- StackLife
- Questions
what is web-scale discovery?

discovery layer
“next-gen” search interface

index
pre-harvested, indexed content
who:

http://librarytechnology.org/discovery.pl
pre-harvested index

full-text articles, article citations, book chapters, ebooks, digital objects, open source content

optional: local content

harvested using FTP and OAI-PMH

**publishers:** Elsevier, Wiley, Springer, etc.

**aggregators:** Gale, LexisNexis, etc.

**open source:** Hathi Trust, DOAJ, Medline, etc.

**local content:** catalog data, digital/special collections, etc.

**A&I services:** MLA, LLBA, PsycINFO, etc.
software engineering process, and developers and project managers should read through the entire book. The book is also intended for people who design documentation, help systems, and training courses, since these are elements of the “total user interface” just as much as the screen designs. This book is not intended to teach technical writing as such, but it can help writers produce support materials that users will find easier to use.

Furthermore, large parts of the book should be helpful to the users themselves and to computer support managers who need to determine which computer systems and software to recommend to their users. Even though it is fairly rare for customer organizations to perform their own usability testing, there is no reason why a large organization should not use some of the techniques in Chapter 6, Usability Testing, to compare software packages and whole systems before deciding on what to buy. Smaller organizations and individual users can use the definitions in Chapter 2, What Is Usability?, and the usability principles in Chapter 5, Usability Heuristics, as a checklist to consider whether an interface seems usable before buying it. Multinational corporations and other international organizations should benefit from Chapter 9, International User Interfaces, when planning the requirements for their information systems. Finally, user organizations that contract out for software development can use Chapter 4, The Usability Engineering Lifecycle, and Chapter 8, Interface Standards, to help set requirements that will ensure the usability of the product they will eventually receive from their vendor.

**Usability** testing in libraries: methods, limitations, and

**Authors:** Emanuel, Jennifer

**Source:** OCLC Systems & Services; 2013, Vol. 29 Issue 4, p204-217, 14p

**Document Type:** Article

**Subject Terms:** *LIBRARY science -- Research
WEBSITES
INFORMATION technology
LIBRARY education
SOCIAL sciences

**Author-Supplied Keywords:** Limitations
Qualitative research
Quantitative research
Research methods
Usability testing
User-centred design
Web site evaluation

**Abstract:** Purpose — Usability studies are a form of library evaluation that usability is an evaluation method, not a research method. The valid form of scholarly research if certain limitations inherent in Design/methodology/approach — Through evaluating literature

**full-text vs. metadata only**
“blended results:”

- vendor’s discovery layer
  - vendor index
  - content and local content in single result set

“bento box:”

- non-native discovery layer (Blacklight, Solr, Drupal, VuFind, etc.)
- article results (vendor index)
- book results (catalog)
### Resource Collections Activation

Please select resources from the following list and click **Done** to activate them.

**Filter By:**
- **All**
- **Search:**
- Upload resource activations from MetaLib

**Provider Name**

<table>
<thead>
<tr>
<th>Provider Name</th>
<th>Requires Subscription for</th>
<th>Added On</th>
<th>Show Info</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABC-CLIO</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The ABC-CLIO eBook Collection</td>
<td>Delivery</td>
<td>07/06/2011</td>
<td></td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04/06/2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARTstor, Inc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The ARTstor Digital Library</td>
<td>Search , Delivery</td>
<td>04/11/2013</td>
<td></td>
</tr>
<tr>
<td>ASTM International</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessible Archives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adam Matthew Digital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexander Street Press</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Council of Learned Societies (ACLS)</td>
<td>Search</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Geosciences Institute</td>
<td>Delivery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Institute of Physics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Mathematical Society</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Psychological Association (APA)</td>
<td>Search</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• v2 summon

• EDS
Web-Scale Discovery

• Systems now able to index large quantities of records
• Large result sets with potentially equivalent candidates returned for any given query
  – Relevancy
  – Faceting
  – Pre-search scoping/Advanced search
• So, is content king?
  – Features are built on a variety of elements (citation metadata, A&I metadata, full text/transcript)
Web-Scale Discovery

• In some ways, yes.
  – Desire for full collection (ILS data, special collections data, and licensed materials at the unit level (article, e-book, song, etc) to be searched in the system.
  – Easier to determine this for local collection data; harder to evaluate for the “mega-indexes”

• Competitive Marketplace – coverage often seen as key differentiator

• How can libraries understand the differences in coverage among competing services?
Web-Scale Discovery

• Difficult to evaluate coverage based on numbers of items indexed alone

• Breadth
  – Titles
  – Date Span

• Depth
  – Is content indexed at the citation or full-text level?
  – Value Added A&I Data
Web-Scale Discovery

• Important to know whether the discovery service favors the content of any given publisher, intentionally or not intentionally.
• For example, does the relevancy reflect bias or publisher preferences?
• New studies looking at impact on underlying content usage (example 1 and example 2).
Non-Cooperative Scenarios

• Many A&I and specialized resources do not contribute to Discovery services
• Two major players are both publishers and discovery service providers
  – EBSCO – ProQuest
• ProQuest does not provide content to other discovery services
• EBSCO does not provide content to other discovery services
• Issue currently being pressed by Orbis Cascade Alliance.
Open Discovery Initiative

• NISO Work Group to Develop Standards and Recommended Practices for Library Discovery Services Based on Indexed Search
• Informal meeting called at ALA Annual 2011
• Co-Chaired by Marshall Breeding and Jenny Walker
• Term: Dec 2011 – Dec 2013
Balance of Constituents

Libraries

Marshall Breeding, Vanderbilt University
Jamene Brooks-Kieffer, Kansas State University
Laura Morse, Harvard University
Ken Varnum, University of Michigan
Sara Brownmiller, University of Oregon
Lucy Harrison, College Center for Library Automation
(D2D liaison/observer)
Michele Newberry

Publishers

Lettie Conrad, SAGE Publications
Roger Schonfeld, ITHAKA/JSTOR/Portico
Jeff Lang, Thomson Reuters
Linda Beebe, American Psychological Assoc
Aaron Wood, Alexander Street Press

Service Providers

Jenny Walker, Ex Libris Group
John Law, Serials Solutions
Michael Gorrell, EBSCO Information Services
David Lindahl, University of Rochester (XC)
Jeff Penka, OCLC (D2D liaison/observer)

Slide Credit – Marshall Breeding
ODI Project Goals:

• Identify ... needs and requirements of the three stakeholder groups in this area of work.

• Create recommendations and tools to streamline the process by which information providers, discovery service providers, and librarians work together to better serve libraries and their users.

• Provide effective means for librarians to assess the level of participation by information providers in discovery services, to evaluate the breadth and depth of content indexed and the degree to which this content is made available to the user.
## ODI Timeline

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Target Date</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment of working group</td>
<td>Dec 2011</td>
<td>✔️</td>
</tr>
<tr>
<td>Approval of charge and initial work plan</td>
<td>Mar 2012</td>
<td>✔️</td>
</tr>
<tr>
<td>Agreement on process and tools</td>
<td>Jun 2012</td>
<td>✔️</td>
</tr>
<tr>
<td>Completion of information gathering</td>
<td>Jan 2013</td>
<td>✔️</td>
</tr>
<tr>
<td>Completion of initial draft</td>
<td>Jun 2013</td>
<td>✔️</td>
</tr>
<tr>
<td>Completion of final draft</td>
<td>Sep 2013</td>
<td>✔️</td>
</tr>
<tr>
<td>Public Review Period commences</td>
<td>Sep 2013</td>
<td>✔️</td>
</tr>
</tbody>
</table>
ODI Best Practice Highlights

• General
  – Continue the work with a SC or WG
  – Voluntary disclosure by CPs and DSPs to assert conformance with best practices
    • Provide minimum set of metadata to all discovery services
    • Provide library with statement of participation
      – Statement on non-bias
      – Usage reports
      – Content listings
ODI Best Practice Highlights

• Content Providers
  – Metadata elements to provide to DSPs (both minimum metadata, full text, and enhanced metadata)
  – Disclosure of this information to libraries
  – Technical format compliance
ODI Best Practice Highlights

• Discovery System Providers
  – Disclosure of information about content included in index to libraries
  – Guidelines on transparency for linking without bias
  – Technical data transfer guidelines
  – Guidelines on transparency for linking without bias
  – Usage statistics
ODI Best Practice Highlights

• Areas for further discussion
  – APIs
  – “Restricted” content access
  – Additional metrics/reporting
  – and much more
Further Reading

• NISO RP-19-201x, Open Discovery Initiative: Promoting Transparency in Discovery

• Orbis – Cascade
  http://www.orbiscascade.org/index/ebsco-and-ex-libris

• Stakeholders Strive to Define Standards for Web-Scale Discovery Systems
  http://www.thedigitalshift.com/2012/10/discovery/coming-into-focus-web-scale-discovery-services-face-growing-need-for-best-practices/

• Discovering Reciprocity
  http://lj.libraryjournal.com/2013/04/opinion/editorial/discovering-reciprocity/

• Discovery or Displacement?: A Large Scale Longitudinal Study of the Effect of Discovery Systems on Online Journal Usage
  http://www.slideshare.net/MichaelLevineClark/mlc-jdm-jsp-charleston-2013-slideshare-28161600

• Revisiting Plato's Cave, Bruce Heterick, JSTOR
  http://www.slideshare.net/CharlestonConference/20131108revisiting-platos-cavecharleston