Future of the Catalog

OPAC Complainers

There is certainly no dearth of OPAC complainers. You have Andrew Pace (OPACs suck), and Roy Tennant (You Can’t Put Lipstick on a Pig) writing and presenting about the need for change (more simplicity) in the OPAC world. I can appreciate their arguments for a simpler OPAC (not to mention the rest of the system) but other then present their arguments, neither has much in the way of suggestions nor have they sparked a movement among librarians or the automation vendors to do anything about the situation.  
-ACRL Blog entry, 13-Oct-2005

Problems with Existing Catalogs?

- Known item searching works pretty well (sometimes), but …
- Lots of topical searches and poor subject access
  - keyword gives too many or too few results – leads to general distrust among users
  - authority searching is under-utilized and misunderstood
- Relevance = system sort order
- Impossible to browse the collection
- Unforgiving on spelling errors, stemming
- Response time doesn’t meet expectations of web-savvy users

Valuable metadata is buried

- Subject headings are not leveraged in keyword searching
  - they should be browsed or linked from, not searched
- Data from the item record is not leveraged
  - should be able to easily filter based on user’s changing requirements using item type, location, circulation status, popularity

Do we agree?

- "Most integrated library systems, as they are currently configured and used, should be removed from public view."
  - Roy Tennant, CDL

Libraries Today

- Starting points:
  - Technology-driven research, teaching and learning
  - User self-sufficiency (decrease in guided access to content)
  - Global “infosphere”
  - Accelerating shift in information seekers’ preferences for Web-based information and multimedia formats
A New Kind of Information Seeker

- Even more self-sufficient
  - "Most respondents indicated they have not sought help (64 percent) when using library resources"—OCLC report on perceptions of libraries, 2005

- On Web
  - Popular search engine traffic in November 2005: 5.15 BILLION searches (& Google out front)
  - Expect seamless linking & instant gratification

The Decline of the Catalog

- Users taking the bypass
  - 89% of college students say they begin with search engines vs 2% with library Web pages

- One piece of a fragmented library information landscape (and hard to use!)
  - Principle of Least Effort
  - Metasearch in trouble

- Cataloging tradition unsustainable
  - “Just how much do we need to continue to spend on carefully constructed catalog?”—Deanna Marcum, LC Associate Librarian

Challenges Facing Cataloging

<table>
<thead>
<tr>
<th>Affordability and Scalability</th>
<th>Expenses of cataloging</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rapid growth of Web resources and digital assets</td>
</tr>
<tr>
<td></td>
<td>Need more than descriptive metadata</td>
</tr>
<tr>
<td></td>
<td>Interoperability issues</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competition for Resources to Develop New Library Services</th>
<th>Shrinking tech services departments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Streamlining tech services workflows</td>
</tr>
<tr>
<td></td>
<td>Increasing use of external sources of data; automated cataloging methods</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes in Information-Seeking Behavior</th>
<th>Preference for online information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reliance on simple keyword search</td>
</tr>
<tr>
<td></td>
<td>Decline of subject searching</td>
</tr>
<tr>
<td></td>
<td>Expectation of seamless linking</td>
</tr>
</tbody>
</table>

A New Kind of Library

- Build a vision of a new kind of library

- Examine assumptions

- Be more involved with research and learning materials and systems

- Move to next generation systems and services

- Make library collections and librarians more visible

Challenges Facing Cataloging

<table>
<thead>
<tr>
<th>Availability of Catalog Librarians</th>
<th>LIS grade not choosing cataloging</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Greying of the library profession (demographics)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Significance of the Catalog</th>
<th>Catalog is one part of a much larger infoosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Many new types of scholarly information objects not covered by catalog</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Future of Individual Library Catalogs</th>
<th>Less emphasis on one catalog per library</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shift toward multiple catalogs appearing as one catalog; shared catalogs; catalogs interwoven into the Web (Open WorldCat)</td>
</tr>
</tbody>
</table>

Research and learning materials:

- Printed/Reports
- Learning objects
- Courseware
- Multimedia
- Untransferred records
- Freely-accessible web resources

Challenges Facing Cataloging
The Continuing Importance of the Catalog

• Books and serials are not dead, and they are not yet digital
• ARL libraries spent the lion’s share of $665 million on books and serials in 2006

The legacy of the world’s library collections is tied to the future of catalogs

What To Do About It

Revitalize:
1. Develop new uses for catalog data
2. Find new users for the existing product
3. Find new uses and new users

New Users, New Uses

• New users, Existing uses
  – Examples:
    • Programs for freshmen
    • "Push" to course Web pages

• New users, New uses
  – Examples:
    • Mass digitization
    • Large scale integration with other systems
    • Universal access

• Existing users, Existing uses
  – Examples:
    • Minor enhancement to existing catalogs

• Existing users, New uses
  – Examples:
    • E-journal discovery
    • Subject publishers
    • Export to bibliographic management software

Information User and Use Environments for Bibliographic Data

• The consumer environment: end-user of the bibliographic data, the information consumer, and services that are designed to assist the end-user in finding relevant information, from search engines to specialized catalog interfaces
• The management environment pertains to resource collection management.

Consumer Environment

• Three main factors that affect consumer use of bibliographic data
  – system knowledge
  – domain expertise
  – procedural knowledge
• Large majority of users (77%) have low system knowledge and low domain expertise/procedural knowledge ("double novices"),
• At the other end of the scale, only 0.5% of users have high system knowledge and high domain expertise/procedural knowledge ("double experts").

Possible New Tool

• the ability to recognize clusters of knowledge production (persons and subjects),
• the lineage of publications (i.e., how they exist in chronological relationship to each other)
• the ability to make previously unknown connections among resources
• the ability to make serendipitous or unforeseen connections among topics
• identification of the authoritativeness of sources
• the popularity/amount of use of a resource
• the sociology of knowledge, for example the "pedigree" of authors and publishers
Management Environment

- Traditionally, libraries manage inventory/bibliographic resources at the level of bundles, whether an anthology of works, a journal, or a single work.
- This has become insufficient to meet user expectations of more granularity in bibliographic description and to handle an inventory that is increasingly comprised of electronic formats that are more fluid and accessible at a more granular level.
- Providing unified management of these disparate collections is the challenge and interoperability of bibliographic data is important to success.

Problematic Bibliographic Data

- encoded data in the 006, 007, 008 and leader fields,
- uniform titles
- analytics
- multi-language resources
- multiple unique identifiers

What’s the big picture?

- Improve the quality of the library catalog user experience
- Exploit our existing authority infrastructure (aka make MARC data work harder)
- Build a more flexible catalog tool that can be integrated with discovery tools of the future.

Plugging Holes in the System

- Natural language problem
  - LCSH=United States—History—Revolution, 1775-1783
  - keyword=revolutionary war (834 hits)
  - Subject keyword="United States—History—Revolution, 1775-1783" (3081 hits)
- Facets taken out of the free-floating and hierarchical context of LCSH can be misleading
- I’ve followed many a tag cloud, but assuming that browsing is still popular, how does one browse keywords?

Paradox #1

We finally have interesting discovery tools that make use of bibliographic data in ways that show us that the data are not completely adequate for use with the new discovery tools.

“Subject Keywords”

- “[from Recommendations:] ....Abandon the attempt to do comprehensive subject analysis manually with LCSH in favor of subject keywords; urge LC to dismantle LCSH”

Paradox #2

“Subject keywords” should replace the controlled vocabulary from which the keywords themselves are most easily derived.

Let’s build bridges between the mountains of bibliographic description so that we can tear down the mountains.

If not LCSH, then what?

Paradox #3

• Computational (e.g. non-human mediated) creation of subject-based facets will work perfectly once all the full text of every work is available in electronic format.

What does a search and retrieval system for 50 million books and 50 million articles look like?

Bibliographic Control Wish List

• A classification or subject thesaurus system that enables faceted navigation
• A work identifier for books and serials
• Something other than LC Name Authority for “organizations”
• Physical descriptions that help libraries send books to off-site shelving and to patron’s mailboxes
• Something other than MARC in which to encode all of the above
• Systems that can actually use the encoding

Paradox #4: The Ultimate Paradox

• “You’re damned if you do and you’re damned if you don’t.”
• - Bart Simpson

Stones (Boulders) In the Road

• Many are not ready for change of the magnitude required
• Progress toward interoperability is slow
• Copyright law has not caught up with the digital world
• Precedents for large-scale collaboration are few
• There may not be enough money

Vision for Change

• The service model for the catalog will be financially sustainable
• The catalog will evolve toward full integration with other discovery tools
• Shared catalogs and open information systems will radically democratize access to library collections and boost scholarly productivity to new levels
Innovations and Cost Reductions

• Much better linkages: ingest, convert, extract, transfer
• Interoperate
• Simplify & exploit all sources of catalog data
• Eliminate custom practices
• Automate and streamline workflows
• Explore automatic classification, subject analysis; reengineer and automate LCSH practice
• Mine catalog data for new uses; experiment with FRBR (Functional Requirements for Bibliographic Records)

Challenges from a changing environment

1. An expanding information universe
• The role and place of the opac is changing dramatically!
• OPAC one of many peer resources
  – Multiple local collection catalogs: visual materials, GIS, archival collections, social science datasets, plus an opac (and lots of little databases)
  – Licensed external services proliferating
  – Plus internet engines, on-line book stores, etc., etc.

2. Better search systems
• Internet and the explosion of digital information generating tremendous research and innovation in search technology
  – Faster
  – Better results
  – Assist the user
  – Dealing with large retrieval sets
• Hope that OPAC will profit from Internet search innovation

3. Invaders in our domain
• Amazon.com: We’ve been hearing: “it’s so much easier to find books in Amazon – I go there first, than to the catalog”
• And now…Search Inside the Book
A Quote

- “On the other hand, there’s Amazon.com. I’m hardly the first to note that Amazon as a catalog or research tool is easier to use and significantly more productive than conventional academic library catalogs.”
  - Tim Burke (Swarthmore): *Burn the Catalog* (2004)

3. Invaders in our domain

- Google Print and Google Scholar
- * library metadata from OCLC and digital libraries
- * search contents of e-journals (CrossSearch)
- * search contents of books (GooglePrint)
- What next???
- (and there will be something next!!!)

3. Invaders in our domain

- “Why can’t I find journal articles along with books in the catalog?”
- *(That’s what Google is doing…)*
- *A fear*: Opac increasingly ignored for more appealing and powerful services
- *A hope?*: Integrate opac information with other search services
  - (search Google, then find the book location in your library)

4. An unstable environment

- Many, many more players in the information environment
- Enormous amount of experimentation, creativity
- Technology enables new models, services, and players
- Change enormously rapid
- Google is only 6 years old.

4. An unstable environment

- A fear: OPAC will stagnate and become irrelevant
- Why stagnation
  - Opac technical platform not flexible, unable to evolve rapidly
  - opac developments tied to very long development timeframe
  - underlying opac model 20 years old, interfaces 10 year old…
  - ILS vendors turn their attention elsewhere
  - no longer invest resources in opac

5. A role for evaluation/recommendation

- Opacs are consciously non-evaluative… does that serve all users?
- Not all users are the same
- Some want to fend for themselves while others would welcome some assistance
- A fear: The opac will increasingly be a tool for only the sophisticated researcher.
6. Portals are a puzzle: Three questions

- Is recreating the opac in the portal sensible?
- Will portals scale as the number of e-resources grows?
- Can we afford duplicate maintenance of portals and opacs?

7. FRBR

- Functional Requirements for Bibliographic Records (IFLA)
- Hierarchic model for bibliographic data
  - Work, expression, manifestation, item
- Potentially more coherent view of bibliographic holdings than the “unit record” of catalog cards (and MARC records!)

7. FRBR

- A hope: Opacs move beyond “one card per item” model and use the power of the computer to organize complex data
- A fear: The library community will be consumed by FRBR/AACR – III debates and implementation while the information environment moves on without us.

The fearful picture

- Opac is bypassed for more exciting and effective search engines
- Opacs stagnate through neglect
- Opacs feel increasingly rule-bound and obsolete
  - used only by the sophisticated researcher
- Librarians argue about cataloging rules while the larger world moves on…

More quotes from *Burn the Catalog*

- “I think we’d be better off to just utterly erase our existing academic catalogs”
- “lock all the vendors and librarians and scholars together in a room, and make them hammer out electronic research tools that are Amazon-plus”
- (to create) “a catalog that is a partner rather than an obstacle in the making and tracking of knowledge”

Expectations are changing

- 20 years ago the Tim Burkes of the world were wildly enthusiastic about opacs!
The hopeful picture

• The opac becomes more integrated with the larger information environment
  – including metasearch engines
  – and internet engines such as Google
• Opac searching improves in parallel with other search environments
  – including help with larger retrieval sets
• Opacs and portals merge to simplify the environment for both users and librarians

Questions…

• Does the competition matter?
  – let’s use Google and Amazon if that suits their needs!
• Even if it matters, do we have the resources to hold our own in this environment?
  – Google spent $200M in ’04 in R&D (not including stock options…)
  – and expects to increase that by 50% this year

Questions…

• Should we shrink the role of the opac?
  – locating items, organizing deeply complex parts of the collection
  – and shrink the cost of creating it?
• Or separate it from the ils and “modernize” it using a commercial search engine?
  – possibly not “MARC aware”…

Questions…

• Do opacs need expensive, complex metadata in the world of Amazon (simple metadata) and Google (full text searching)?
  – is the world moving towards dumb data, smart engines?

Library of Congress
Working Group on the Future of Bibliographic Control

• Three Guiding Principles
  1. Redefine bibliographic control
  2. Redefine the bibliographic universe
  3. Redefine the role of the Library of Congress

Present findings on how bibliographic control and other descriptive practices can effectively support management of and access to library materials in the evolving information and technology environment
Bibliographic Control

• Describing
• Analyzing
• Organizing
• Managing

• ... to assist discovery, identification, selection, and access

Library Materials

• Books
• Monographs; novels; government reports
• Conferences/papers
• Anthologies/stories, poems
• Essay collections; exhibit catalogs …
• Journals/Articles
• Recordings/Songs, Interviews, Speeches
• Archives/MSS, Letters, Photos
• Physical; licensed; digitized; web-available

Library Materials

• All are potential objects of discovery
  – (chapters, charts; poems, pictures)
• All are potential targets of citations, links
• Ergo, all are subjects for bibliographic control

• Bibliographic Control = Catalog books

Library Materials: in the catalog … and not

• Books
• Monographs; novels; government reports
• Conferences/ papers
• Anthologies/ stories, poems
• Essay collections/ essays
• Journals/Articles
• Recordings/ Songs, Interviews, Speeches
• Archives/ MSS, Letters, Photos

Blended Discovery

• Wikipedia
• Google Books
• JSTOR article
• Smithsonian exhibit
• Guardian book review
• IMDB German TV Miniseries

Guiding Principles another way

• Library materials are both objects and their content
• Bibliographic Control is more than “descriptive practice”
• Bibliographic Control doesn’t just happen in the library catalog
Blended Discovery: Implications

- Many descriptions of the same thing, intermingled
- Reliance on machine-enabled connections
  - Google to WorldCat; Amazon to OPAC
- Emphasis on machine-recognized identifiers
  - ISBN, DOI, OCLC #, Author #; Open URL
- Modular application of standards
- No safe haven

Guiding Principles

- Recommend ways in which the library community can collectively move toward achieving this vision
- Advise the Library of Congress on its role and priorities

Redefining Roles

- We’re all in this together
- “Some animals are more equal than others”
- The members of the Working Group are not economists

1. Increase efficiencies

- Eliminate Redundancies
- Distribute responsibility
- Re-examine economics

Be cognizant of other sources

- Publishers (ONIX)
- Vendors
- Foreign Libraries
- Commercial (IMDB)
- Entrepreneurs
- Folks

Efficiency

- Use as much as we can
- Change as little as possible
- Add what’s most valuable
- Automate the processes
Efficiency cont’d

• “be more flexible …” – the Casalini case
• Do it our way/Have it your way
  – 200% cost difference
  – 80% agreement on what matters
  – But, is it flexible enough?

2. Enhance Access to Hidden Collections

• Make the discovery of rare & unique materials a high priority
• Provide some level of access to all material, rather than comprehensive access to some material and no access at all to other material
• Encourage digitization to allow broad access
• Share access to unique materials
  – LC and flickr

4. Position our Community for the Future

• Extending Beyond Library-Created Data
  – Provide links to appropriate external data
    • Contents, summaries, reviews; Google Books API
  • Integrate user-contributed data
    – Balancing democracy and demography
  • More research into use of computationally derived data (holdings patterns, usage …)

4. Position our Community for the Future (con’t.)

• Subject Analysis
  – Controlled vocabularies are valuable
  – Continue to use LC Subject Headings
    • Simplify the process
    • Get more benefit from results
  – Recognize value and reality of multiple schemes; use and connect them

Putting it all together

• Play well with others
• How to describe them:
  – EAD, DC, MODS, CCO, VRA
• People, places, and concepts:
  – NACO, ULAN, VIAF, LCSH, AAT
• Where they’ll be found:
  – Extended catalog, WorldCat, Aquifer and …
• Google, MSN, Yahoo

Finale

• Principles for creation of surrogate records that have been developing over hundreds of years can be used to catalog (to metadata?) anything!
• It is an exciting time to be a cataloger!
• GO FORTH AND CLASSIFY!