What is the Library’s role in Discovery, including cataloging and discovery systems

David De Lorenzo, Lee Leighton and Brian Quigley

Overview

Lorcan Dempsey, an OCLC Vice-President, states that “discovery of the cataloged collection will be increasingly dis-embedded, or lifted out, from the ILS system, and re-embedded in a variety of other contexts.” The reliance of faculty, students and the general public on major search engines such as Google has brought information discovery to the network level. Libraries – and their catalogs and databases - are no longer considered the only places to search for research materials. Library catalogs are viewed only as ways to locate books and request remote materials found elsewhere; library databases are viewed as second choices after Google and Google Scholar.

It is also true that these perspectives pre-suppose that metadata exists. In any conversation about discovery, we must recognize that backlogs, hidden collections, and proprietary metadata still represent a large portion of the information universe and must be included in our planning.

National trends

- On the record, the report of the Library of Congress Working Group on the Future of Bibliographic Control was issued in January 2008. It discussed LC’s work in developing and maintaining the cataloging rules, the MARC format, the name authority file and the LC subject headings, which are used by American, British, Canadian and Australian libraries. However, LC is not our national library, and it is not funded as such.
- Library holdings and content are becoming increasingly discoverable at the network level (OCLC, Google Books, Google Scholar).
- “Perceptions of the value of library’s gateway function have declined since 2003, especially among scientists and economists, probably best explained by the availability of new e-resources” (Kevin Guthrie). Still, the percentage of faculty rating this function as “very important” is over 60% (and over 50% even in the sciences).
- “Users want a rich pool from which to search, simplicity and satisfaction” (Bibliographic Services Task Force Report).
- “Students prefer the content available in library databases …. However, if discovery and access to library databases is more cumbersome than they expect, they will abandon library resources for the more familiar terrain of Google and Wikipedia” (ProQuest quoted in Yet another wake-up call).
- User studies consistently reflect a desire for “one-stop shopping” and integrated search of all library content-not silos (UC Berkeley User Research).
- Researchers report difficulty finding and accessing library materials, particularly chapters in multi-authored books and special formats.
- Next generation OPACs are being developed and released with such features as content integration (Encore, Primo), visual presentations (Aquabrowser), facets (Endeca, WorldCat Local), reviews (WorldCat Local), tags (PennTags,
LibraryThing for Libraries), and recommender systems (LibraryThing for Libraries).

**Local issues**

Following the pilot, WorldCat Local may replace the Melvyl catalog with a searching tool for the UC libraries that will search across the entire OCLC database and bring together local holdings with those of other institutions. The pilot will also search article databases for which OCLC has obtained metadata and access. Features of the pilot include:

- Very brief views of records
- All fields indexed, but only selected fields displayed
- Editions brought together in FRBRized clusters
- Reliance on the request function, which is very well received by faculty and students

The Library is also in the process of replacing the last locally developed ILS among the ARL libraries with a new ILS from one of the few remaining commercial vendors. The new ILS will make adding record sets from the UC Shared Cataloging Program (SCP) and other sources much faster and efficient. The Library may not maintain a separate OPAC in the future.

The Library and its partners also continue to make non-MARC data and XML-encoded documents much more accessible. These activities include:

- Creation of Finding Aids to archival collections
- Integration of various data sets (MARC/non-MARC) on OCLC’s ArchiveGrid
- OAC redesign to include a Google-like search design, improved searching of digital objects and text-rich documents, creation of social networking, referrals and recommendations.

Many of these activities have been instigated by what searchers want from their search experience. The Bibliographic Services Task Force report identified several issues for the next generation OPAC:

- Enhanced searching capability: relevance ranking, spell check, date searches, etc.
- Amazon.com features: cover art, table of contents, introductory pages, excerpts, index, summary
- One stop shopping
- Social networking tools such as Amazon reviews and recommendations.

The UC libraries are discussing coordinated processing.

- The UC Shared Cataloging Program has received an 18% ($48,000) cut in fiscal year 2008-09.
- In order to mitigate the effects of the cut, the UC libraries are exploring coordinated cataloging of California documents with the State Library
- The UC Shared Print Program will soon begin a project with coordinated selection and cataloging of Catalan materials
- Shared Archival/Museum Collection Management tool
- Centralized EAD Encoding Service
• HOTS and CAMSIG (the cataloging and metadata subgroup) are discussing coordinated processing for CJK, Slavic, Arabic, Judaica and Western European area studies materials.

Starting points

• Select and implement a new integrated library system (ILS) as soon as possible as a foundation to the other starting points.
• Ensure all library holdings, including Affiliated Libraries holdings, are discoverable at the network level (OCLC, Google) in a complete and timely fashion.
  o Upgrade all low-level records and load them into OCLC
  o Negotiate rights to load vendor records into OCLC
  o Ensure all electronic holdings are represented in the ILS and OCLC
  o Support short-term positions for reducing the UCB Library’s enormous backlogs (400,000 low level records for materials stored at the NRLF and an archival backlog of more than 25,000 cartons)
• Expand enhanced metadata, including searchable tables of contents in WorldCat Local and/or our own OPAC through Blackwell Table of Contents Service, ONIX, and/or other similar services.
• Advocate for open metadata with vendors and publishers to facilitate discovery at the network level, cooperative gateways, and text mining.
• Provide a single search interface, based on user preferences, to simply and adequately meet the majority of research needs. Do not let more comprehensive needs be obstacles to simplicity and satisfaction for the majority of uses. Non-exclusive options might include:
  o Partnering with OCLC and CrossRef to pre-harvest CrossRef Search or CrossRef Web Services metadata for WorldCat Local.
  o Partnering with Google to create a UC (or UC Berkeley) “channel” for Google Scholar.
  o Implementing federated search via open source or licensed software (MetaLib) to develop broad subject portals: general/multi-disciplinary, arts & humanities, sciences & engineering, and social sciences.
  o Implementing an Endeca, Encore, Primo, or similar product over our ILS, federated search, and/or open archives tools.
• Continue to support more advanced researchers who recognize the subject brand and value-added features of core databases (Web of Science, PubMed, MathSciNet, SciFinder Scholar, ArtSTOR, etc.). Advocate for value-added features (such as relevance ranking, controlled vocabulary, name authority, and pattern analysis) among database producers. Cancel databases and tools that do not provide unique content and value-added features.
• Create a “Discovery” team comprised of public services, systems, technical services, and user interaction design specialists to implement next generation discovery tools (locally or with partners), to ensure cataloging processes and outsourcing support new discovery methods, to assess usability of discovery tools, and to imagine and work with partners to implement continuous improvements to discovery.

• Design and implement processes that truly recognize the importance of electronic resources and maximize their discoverability.
  o Implement an electronic resource management system (ERMS), with MARC record updates to keep the ILS accurate.
  o Obtain and auto-load as many catalog records as possible via YBP, SCP, and publishers.
  o Consider working with other campuses to expand SCP, perhaps making it a separate entrepreneurial unit. Consider outsourcing all electronic resource processing to SCP.

• Redesign library websites as subject portals focused on discovery.