

“Next-Generation” Library Catalogs

This presentation outlines some of the possibilities for “next-generation” library catalogs. Specifically it describes the technology behind these applications, enumerates ways they can exploit sets of globally networked computers, and posits opportunities for new and improved services within and around them. Librarianship has traditionally been about collection, preservation, organization, and re-distribution. These over-arching goals are still relevant in today’s environment, but the methods the profession can use to achieve them is changing. The venerable library catalog and the thing it seems to be morphing into is just one example. This presentation brings together the how’s and why’s of this issue.

Some questions

It goes without saying that the information landscape has significantly changed with the advent of the Internet. With these changes, it behooves us to re-examine the answers to questions that we have long thought were immutable. One of these questions includes, “What is the library catalog?” To what degree is it an inventory list or a finding aide? What sorts of things is it expected to contain? What functions does it serve and for whom? More fundamentally, what sorts of problems is it expected to solve? Answering these questions, while considering the current environment, is a necessary step when it comes to designing any “next generation” library catalog application.

To my mind, the library catalog is a specialized index – a set of pointers to information resources of particular usefulness to a library’s constituents. As such, it is not necessarily an inventory list. That is only part of its functionality since the globally networked world where we live and work enables us to link and provide access to information resources that we do not own or license.

In an analog environment, information was for the most part embodied in people and a host of recorded materials ranging from cave paintings to motion pictures, from manuscripts to books, from newspapers to journal articles, from musical scores to sound recordings. In a digital environment, the types of information have taken on new forms that still include people, but also include datasets, simulations, blogs, RSS feeds, mailing list archives, Powerpoint presentations, and all of the electronic incarnations of the traditional analog materials.

To what extent is the library catalog expected to contain these sorts of items? If you lean towards the inventory list side of things, then the answer to the previous question is, “Not very much.” More philosophically, the answer to this question is dependent on your collection development policy which will take into account the purpose of your library, the needs of your patrons, and the resources (time, money, and skills) you are able to bring to bear. Put another way, the answers to these questions are less

objective and Platonic and more down to earth and Aristotelian in nature. They can be answered in any way that makes sense for your particular institution at a particular point in time.

They’re all the same

Technically speaking, the current crop of “next generation” library catalogs (Aquabrowser, Blacklight, Evergreen, eXtensible Catalog, Koha, Primo, Scribio, VuFind, etc.) have more things in common than differences.[†] Specifically, they all:

1. normalize and distill metadata of various formats into a single information schema, and
2. provide a search interface against an index of this metadata.

This is why “next generation” library catalogs are increasingly called “discovery” systems. Think index. Think search. Think find. Discovery.

In most of these “catalogs”, content, usually in the form of MARC data, is first saved to a centralized store – a relational database. Next, reports are then written against the database and fed to an indexer – increasingly an open source indexer called Lucene. Finally, some sort of user interface is provided facilitating search, browse, and navigation through the index thus enabling people to identify items of interest. The only significant technical difference between all of these applications are the interfaces provided at either end of the process – the creation/ingestion of the metadata and its subsequent searching.

When looked at from this point of view, the changes we are experiencing in regards to “next generation” library catalogs is definitely incremental or evolutionary, not revolutionary. For the most part, the content of our catalogs is the same – bibliographic data. Technically, we are trading the rather top-down approach of searching relational databases for the more statistical approach of searching indexes. Add a few Web 2.0 thingees – tags, reviews, ratings, RSS syndication, faceted browse, citation creation, cover art, “your” page, a mobile interface, etc. –

[†] Actually, we ought to distinguish between “next generation” library catalogs, “integrated library systems”, and “OPAC’s”. In my mind, integrated library systems represent a suite of software used to automate many library functions: acquisitions, cataloging, circulation, etc. The OPAC – online public access catalog – has traditionally been one of those functions in an integrated library system. The phrase “‘next generation’ library catalogs” is an unfortunate one, and I have always, more or less, equated it the term OPAC. In this way, systems such as Koha and Evergreen are more akin to integrated library systems but they requires some sort of “OPAC” interface.

More questions

Bringing all of these materials together into a central collection is nice, but how is anybody going to be able to access the materials if it is not organized to smaller, more manageable piles. How will they be able to access it if someone does not make the content easily available via the user's networked computer? This is what the acquisitions and cataloging processes are all about. There is a lot of scholarly material available on the Web, just for taking. Use any number of Web spiders, crawlers, harvesters, and robots to copy things locally. Is the Acquisitions Department synonymous with the Purchasing Department? Increasingly, since the Library of Congress initiated the process early in the last century, we rely on shared cataloging practices to accomplish the goal of organization. Think catalog cards, MARC records, OCLC, and authority lists (name, title, and subject). In the current environment, do these processes make the most sense? To what degree are things like XML, folksonomies, and statistical analysis more economical considering the quantities of accessible materials, especially if full text is so abundant?

copying of files into the future. The things that are most wanted will be those things copied forward by users, and the things that aren't are left behind to go to Big Byte Heaven.

The “Next ‘next generation’ library catalog”

[illegible]

Trace an idea through. Summarize. Annotate. Plot on a map. Graph a chat. Do word counts against. Print & bind. Play. Save. Transform. Translate. Find in a library. In my opinion, this is where the growth lies. Let Google do what it does best. Let libraries do what they have always wanted to do – collect, preserve, organize and provide access to content for the betterment of a it's patrons.

Wouldn't it be nice if your library users said, "Not only can I find all the materials I need in my library, but it also provides me with interfaces – people and tools – to help me understand the things I find more quickly and easily. That library saves me time. It is well-worth the cost."