

## The Future of the Catalog Michael Boock

Let me begin by asking you a few questions:

How many of you are from public libraries? Academic? Other?  
Who is your ILS vendor? III, Sirsi/Dynix, Ex Libris/Voyager, VTLS, other?  
Which of you think that our catalogs generally do a good job locating relevant information for patrons?  
Of those that think there opacs do a pretty good job, are there improvements you'd like to see made and are you confident that your vendors will make the constant improvements that are necessary?

Opac sucks video - <http://www.youtube.com/watch?v=tJD-safYEb0>

### Slide 2 – What I'll Cover

I mentioned in a presentation on the future of library technical services a couple months ago that I'm tired of talking about the future of the catalog and said, can't we just get there already. But I've since realized that's the wrong approach. I think we need to be constantly and actively considering our OPAC as it continues, for the time being, to be a key point of access to the library's materials and an important public face for our libraries, especially, I think, for public libraries.

I'm going to talk about whether or not the catalog, or, more specifically, the local OPAC, has a future and talk about what the OPAC is competing with and being measured against as a search tool for library resources.

I'll discuss both short-term changes to our OPACs that I feel we need to make quickly and longer term changes and possibilities.

Finally, we'll take a look at what's out there now in the way of some of our vendor OPACs, open source options, and other search and retrieval tools that we should, in my opinion, be comparing with our OPACs, stealing their functionality and building upon it where possible.

### Slide 3 – Is There a Future for the OPAC?

So, is there a future? The catalog as an inventory management tool used for managing circulation functions and as a module used with other integrated library system functions will likely remain. As a local public interface to the library's holdings, I truly believe that it has a short life-span. In academic libraries in particular, where the article, completely unrepresented in catalogs, has become so much more essential to our researchers because of its greater timeliness, the catalog is already not the primary search tool.

It doesn't contain so many of the resources that our patrons are looking for, not just articles, but also web resources are poorly represented in our catalogs. Many of us have devoted a lot of energy to selectively cataloging and incorporating appropriate web sites into our catalogs. I am one of those initial few dozen or so catalog librarians who began cataloging web sites in 1993 or 94 as part of the OCLC led InterCat project that sought to determine the efficacy of cataloging web resources using MARC/AACR2. Hundreds or thousands of catalogers have since cataloged hundreds of thousands of web resources, but if someone is looking for the department of energy web page or a DOE report for that matter, are they more likely to find it via a web search or via our catalog? And even if we've cataloged it, will they find it given the current state of our OPACs? Where are they more likely to look? They might find it serendipitously via a catalog search, by doing a subject or keyword search for energy for example – so there is that value of including web resources in our catalogs, but on the whole our patrons are finding things on the web just fine on their own, without any librarian mediation.

What about images and digital collections? The same arguments might apply. Even if we harvest metadata for digital collections into our catalogs, do our catalogs really work well enough that the resources will be presented to our users effectively? As effectively as Google?

Our OPACs are digital versions of our card catalogs, holdovers really, without very much added functionality when you get right down to it. Keyword searching is possible but hasn't worked very well because of the lack of good relevancy ranking. Other index searching like issn/oclc no./etc. is rarely used. Because OPACs are incapable of full text searching, we might question if that alone will lead to their demise.

Slide 4 – Other Search and Discovery Options

Many of our users, we've discovered in usability testing sessions and focus groups at OSU, already use google to find relevant materials and then do a title search in our library catalog to find out whether or not we have it. So, because OPACs work so poorly as a search and discovery mechanism, they've already lost that role to tools that do a better job of it like Google Scholar or plain old Google, searching those resources quickly to locate relevant materials and then, if they aren't freely available directly, searching the catalog to see if the library has it. Or, not bothering with the library at all and just using whatever can be found that is freely available.

Libraries for some time have been moving to federated search options that search multiple pools of resources simultaneously. You're probably aware that there is a session happening right now on this topic. OSU has built an open source metasearch product called LibraryFind that searches and displays journal articles from citation and full text databases, materials from our digital collections and institutional repository in a single result list that offers faceted browsing, and 2 click access to full text using a built in openurl resolver. Currently, because III doesn't make our metadata available to us efficiently outside the opac, we have to pass searches to our catalog, continuing to rely on the less than stellar OPAC search capabilities via Z39.50. We also link users to the catalog to find status and specific location information for book results. It is very conceivable that once we get efficient access to bibliographic and holdings metadata, we will provide our own indexing so that searches do not have to take place in the catalog. Location and status information (is a book checked out?) will be available in the federated search results, no longer requiring the catalog for that purpose. Soon, the opac at OSU will be entirely replaced, not necessarily with LibraryFind federated search although that's possible, but possibly with our own search tool built on top of the ILS, with an improved consortial catalog, with Local WorldCat or some other search mechanism that we haven't thought of yet.

The University of Washington and several public and academic libraries in Illinois are piloting what is called Local WorldCat. You may have seen the press release announcing it just this week. What I've been able to glean from conversations with UW staff involved in the project is that UW users will be able to search WorldCat, presenting local holdings prominently to the user but also Summit (the academic library consortia for OR/WA) library holdings, and all of WorldCat. In addition, article citations from OCLC FirstSearch databases will be presented so it functions also as a federated

search. It searches open worldcat, OCLC's FRBRized catalog and takes advantage of the enhanced search capabilities and excellent relevancy ranking algorithms that OCLC continues to develop in Open WorldCat. Search takes place in WorldCat, entirely outside of the OPAC. The OPAC and consortial catalogs remain connected so that users can place holds, request books etc., but the search is entirely outside of the OPAC.

## Slide 5 – Short Term Changes

We should remember when considering the potential for OPACs to be able to change and keep up with information research and search, that what we have now is almost exactly what we had when our first Web OPACs went online around 1994 or 1995. So while northernlight was working on faceted browsing in the late 90s, while amazon enabled value added user reviews, ratings, list creation and user groups; and while google and every other search tool on the Internet enabled amazingly effective keyword searching with advanced (or not even terribly advanced) search algorithms, while Google and Amazon allowed developers access to their services as web services so that we could reuse reviews, ratings and cover images in Amazon and services like search and Google Maps in google. Well, because our catalogs are black boxes, we cannot develop services on top of them. Even if we wanted to, and I'm not saying we've been clamoring for this for more than a decade – we haven't, we'd have to ask our vendor to do these things for us and probably pay for it in some way. Library catalogs couldn't really do many of these things or even take advantage of any of it. We had Syndetic Solutions offering reviews, summaries and book covers but most academics have been strangely uninterested in this whereas public library catalogs look a whole lot better than academics because they usually include these features.

Now that we want them, we've got to pay for them, sometimes twice. We may have to pay our vendor to allow us to load the data and pay the vendor for the data, even though most of it is freely available on the web and if our opacs were open and developable, many of us could do the work ourselves quite simply or could adopt the work of others.

Where we have seen large-scale improvements in OPACs, and it has only happened in the last year or two, is with those offered by open source ILS's

like Koha and Evergreen. We'll take a look at examples of Koha and Evergreen opacs in a few minutes.

So, what can we do to quickly improve our catalogs and to open them up so that we can more quickly take advantage of future web technologies? What do we need to be telling our vendors to do?

I'm going to begin by describing short term changes to the catalog. When people talk about the Next Generation Catalog, I think often they are talking about current generation improvements, or, just getting up to speed with things that the user now expects based on their use of other search interfaces and web tools.

Karen Schneider, in an ALA Tech Source series of blog posts last year provided a long checklist of things that all search engines provide but that are rarely or never provided by library catalogs.

Relevance ranking and field weighting. Every search interface except library catalogs uses relevancy algorithms to present the searcher with the materials that they are most likely to be looking for. Relevancy ranks use field weighting – does the term appear in the title or subject or notes fields. If it appears in the title, then it is more likely to be relevant. Another piece of the algorithm might include log analysis, where number of clicks enters into the calculation. In consortial or union catalogs, number of holdings might be included. Phrase matching may be part of the algorithm.

For the most part, known item searching works fine in our catalogs with title and author searches, just as it did in the card catalog. Resource discovery does not work so well.

- Faceted browsing
- Recommendations - people who borrowed this, also borrowed
- Did you mean?/spell check and automatic correction stemming.  
Automatic stemming is a feature that most of our catalogs do not have. Does everyone know what stemming is?
- Sorting – right now, most catalogs do what's sometimes referred to as a system sort or sort by last in appears first. We should be able to sort by popularity, librarian and user recommendations, ratings, currently available, locally available in a consortial setting
- Search visualization- Aquabrowser, Grokker

– What else?

## Slide 6 – Longer Term Changes

What are some of the effects of the network on the OPAC. In the network, there is more available to our users directly. Via other libraries and via booksellers. Patrons are able to receive an item in 3 days or less from any summit library. Is the fact that a book is located at OSU so important as to exclude everything else a patron is capable of getting in 3 days? Why not just offer a consortial catalog as the default catalog for the user? Or Open Worldcat which provides links to regional catalog holdings. With amazon and other online booksellers, patrons are used to waiting a week or so before getting what they want. We can do better than that with our ILL and consortial systems. Why not show patrons everything they can get within a reasonable amount of time via ILL?

We've been doing ILL in libraries for decades but perhaps haven't always promoted it so well, in part, probably because we didn't want it to be too popular because we didn't have staff to devote to it. But now, thanks to the network, and the amazing search and fulfilment capabilities of online services, we can present more options to our users. OCLC is working towards offering unmediated borrowing and lending and is looking at home delivery. Is it time that we did the same?

Why not offer a resolution service that allows patrons to place holds, request from another library, purchase from the bookstore/online booksellers. Offer these options conveniently at the record level or, better yet, from browse screens? The user should be able to identify, download, locate and take these other actions directly from the browse screen. Users shouldn't be required, unless they need the additional information for identification purposes, to look at a bibliographic record. We need to get away from thinking of the metadata as what's important to the user. Of course its essential for search and retrieval, but the user doesn't need to be aware of it. Obviously, the content is the important thing and our search tools should reflect that.

So what is the value of having a local OPAC? Well, sometimes patrons will just want to see what is within the libraries walls, but that will still be possible in the networked environment. Open worldcat and local worldcat, for example, present materials available in the local and nearby libraries

more prominently but do not exclude other relevant materials that are available via ILL and direct requesting.

As mentioned, many of our users are most interested in journal articles or just interested in finding something as quickly as possible – equating to, is it on the web? For these things or for these purposes, the catalog is obviously not their best option anyway.

What about not just thinking about the catalog as a search and access mechanism but as a tool for using data. What about doing some of the things Amazon does so well, allowing users to contribute to the database by rating, recommending, and reviewing materials? Why not offer an option to purchase the book from Amazon or another online bookseller – your campus bookstore? What about allowing users to export citations in a particular citation format?

The future catalog should use open standards such as xml, oai, openurl, sru/w, open search that are understood by developers and coders so that they can be built on more easily. You should easily be able to offer mashups between your catalog and other web services. If you want to incorporate another source of reviews, as long as those reviews are made available using an open standard such as xml, they can be incorporated. If other services want to use your data for something, they can do so as long as the data is open and it is standardized in a way that other services can understand it (that is, not MARC). For example, if you want Google to be able to provide access to your resources, that can be more easily accomplished as long as the metadata is in a language that can easily be understood by Google (i.e. xml) and easily retrieved out of your system (i.e. not a black box that is closed).

The future catalog will have an open api or programming interface that allows you to revise code as appropriate to your institution. Don't like the way that your catalog is searching or indexing the metadata or presenting it to users, with an open api you can make the changes you need to make. If you don't have the staffing to do it yourself, you can hire someone else to do it. You do not have to wait for your vendor to decide to enhance their product, which can take years with no guarantee that it will ever happen.

Slide 7 – What Are We Gonna Do About it?

We do not have to wait for your vendor to decide to enhance their product, which can take years with no guarantee that it will ever happen. Remember that we're really only recently seeing any prominent changes to the WebPAC since they first appeared 12 years ago. Are we willing to wait that long?

Need to change quickly. Is there any reason to think our vendors will be leading the way, when they are forced to respond to minor tweaks to every other part of the system from every one of the hundreds of other libraries they work with? Not likely. Must be flexible, adaptive.

Implement the short term changes as quickly as possible.

Push vendors to improve the catalog – its all that matters, everything else should be in service to it!

Must be able to tweak relevancy ranks so that your materials are found, not rely on vendor to tweak for every system and then pay for a reindex that takes several weeks between asking for it and it happening.

1. usability, usability, usability
2. User testing
3. Have at least one person on staff with web usability expertise and responsibility. What is more important?
4. Search log analysis and modify based on that information.
5. Circulation usage

Need to determine audience and involve them in the design and functionality. You need it to do what they need it to do, not what you think they need it to do. This can be accomplished through the use of surveys, focus groups and usability studies.

## Slide 8 – Let's Look Around

Let's look at some examples of what we have now and what we are looking for.

Look at sample catalogs: OSU, Summit, endeavor, sirsi (ok), ex libris aleph (ok), good public (dynix?), vtls  
<http://osulibrary.oregonstate.edu/>  
<http://oasis.oregonstate.edu/>

<http://summit.orbiscascade.org/>

<http://lms01.harvard.edu/>

<http://library.ci.corvallis.or.us/corvallis/default.asp>

Good public (OSS): <http://search.athenscounty.lib.oh.us/>

[endeca.phoenixpubliclibrary.org](http://endeca.phoenixpubliclibrary.org)

<http://gapines.org/opac/en-US/skin/default/xml/index.xml>

<http://www.lib.ncsu.edu/catalog/>

Plymouth State

<http://fish4info.org/union/>

Aquabrowser/grokker libraries KCPL

Compare with amazon, ebay, google, oclc, library thing, amg, flickr

<http://www.worldcat.org/>

## Slide 9 – References

- Lorcan Dempsey, Lifting out the catalog discovery experience.  
<http://orweblog.oclc.org/archives/001021.html>
- Lorcan Dempsey, The Library Catalog in the Discovery Environment.  
<http://www.ariadne.ac.uk/issue48/dempsey/>
- Andrea Mercado, Not Your Dad's Search Interface.  
<http://plablog.org/2007/01/not-your-dads-interface.html>
- Terry Reese, Can the Open Source Community Help the ILS Matter.  
<http://oregonstate.edu/~reese/blog/archives/417>
- Karen Coyle, Users and Uses, New Services.  
<http://kcoyle.blogspot.com/2007/03/users-and-uses-new-services.html>
- Karen Schneider, How OPACs Suck.  
<http://www.techsource.ala.org/blog/2006/03/how-opacs-suck-part-1-relevance-rank-or-the-lack-of-it.html>
- Karen Schneider, How OPACs Suck, Part 2: The Checklist of Shame.  
<http://www.techsource.ala.org/blog/2006/04/how-opacs-suck-part-2-the-checklist-of-shame.html>
- Karen Schneider, How OPACs Suck, Part 3: The Big Picture.  
<http://www.techsource.ala.org/blog/2006/05/how-opacs-suck-part-3-the-big-picture.html>
- The OPAC Sucks. <http://www.youtube.com/watch?v=tJD-safYEb0>
- Eric Morgan, A “next generation” catalog.  
<http://www.library.nd.edu/daiad/morgan/musings/ngc/>