

Defining the Next-Generation Catalog

Abstract

This chapter will define the next-generation catalog (NGC) and briefly look at some of the products in the marketplace.

The term next-generation catalog (NGC) first became omnipresent throughout the library industry with the founding of the NGC4Lib mailing list. Eric Lease Morgan of the University of Notre Dame founded the mailing list in order to create a channel for discussion on the topic of the next generation of library OPACs (online public access catalogs). Morgan noted four principles that define the NGC in a posting entitled “Next Generation Library Catalog.”¹ These four principles are the following:

- It is not a catalog.
- It avoids multiple databases.
- It is bent on providing services against search results.
- It is built using things open.

Library industry vendors and open source communities have provided solutions that appear to meet these needs—but as we further analyze the solutions, it is clear that they touch only the surface of these needs. It is clear that the NGC solutions that have been used in libraries fail these four principles. Let’s take a closer look.

- **Principle 1: It is not a catalog.** A typical NGC solution is more than just a catalog—many of these products provide the ability to search more than just the bibliographic records from the ILS, such as digital collections produced by the institution or open-access data culled from open repositories.

Some of these solutions have included the ability to harvest OAI-based repositories to include additional content in the index. However, this is still very siloed and narrowly focused. The NGC just blurs those boundaries making the distinction even more difficult for the general user. Because the solution expands the data set to include content from a few additional external sources, understanding the boundaries of this system is even more confusing for the user.

- **Principle 2: It avoids multiple databases.** While the NGC by and large has avoided multiple databases, many have incorporated federated search to provide a greater level of access. The NGC was thought of as the single-search-box paradigm that libraries have been dreaming of; however, federated search just exacerbated the problem by creating a less convenient and less simple interface—which was one of the key driving factors for the invention of the NGC. A single database is key to providing a simple interface, which brings us back to the failures in Principle 1. Many NGC solutions have attempted to be more than just a catalog by incorporating additional content, but in doing so have integrated federated search, thereby failing to meet Principle 2.
- **Principle 3: It is bent on providing services against search results.** Many NGC solutions have done very well with this principle. The interface and functionality have all been designed around working with the results set and providing services around it. For example, the incorporation of faceted navigation allows the user to modify results through the use of filters. Many NGC solutions provide recommendation functionality as well as the ability to share results in a more social environment

and to expand the research to external entities such as Google Books or Wikipedia. Due to failures with Principles 1 and 2, these services are still fairly myopic—focused on the smaller collections represented within the NGC.

- Principle 4: It is built using things open.** Here is another area where the NGC solutions have shone. Many have been built from open source technology and have incorporated functionality to include open-access content. Two solutions, VuFind and Blacklight, are available under an open source license, allowing them to be downloaded and installed at no cost. Of course, I am referring to direct financial cost and not staffing and resource cost—“free as in kitten, not beer.” Utilizing open source technology is a great way for the vendor of the product to reduce cost and build on a platform that other like organizations are also building on. For example, consider the widely popular Apache Solr and Apache Lucene, a search engine platform and an indexing engine respectively. These two open source products have become extremely popular in the library market and can be found in almost every product in the NGC market. As these technologies continue to evolve and get better, so will the solutions that are built around them. There has been one failure around this principle, however; the NGC has not facilitated the open sharing of content in a convenient manner. No NGC on the market today provides an open sharing process of MARC records. One open source solution, SOPAC, developed by John Blyberg of the Darien Public Library, has taken on the role of being a collaborative engine of social tags. One library with SOPAC can pool and share tags on records in its collections with other libraries that are using SOPAC. This is a great model that seems to have seen little adoption; however, a newer commercial product, BiblioCommons, seems to be trying to push this approach further. This concept of libraries sharing resources and services seems like a highly valuable proposition that deserves further research and investment. Lastly, while a typical NGC uses open content and open source software, it is not able to provide access to all of the vast collections of open-access content.
- Figure 6, a diagram drawn by Morgan in 2006, depicts the architecture of a next-generation

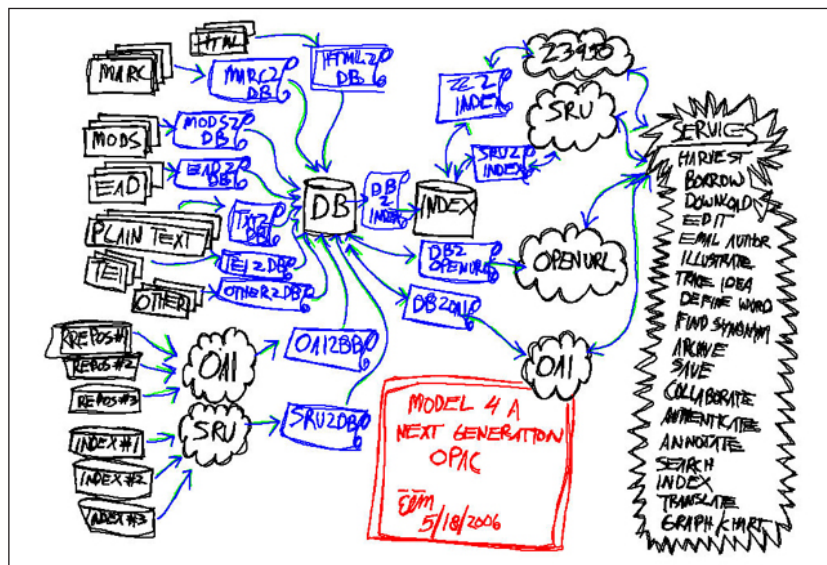


Figure 6
Eric Lease Morgan’s diagram of the architecture of a next-generation catalog.

catalog. This diagram proves to still be very relevant today. However, there are three services that are missing from what is listed on the right-hand side to allow the NGC to better meet user expectations. These are recommend, browse, and relate.

Recommendations are becoming part and parcel of discovery systems. Amazon.com has been known for using this approach to help increase the visibility of its products and sales; similarly, libraries have been adopting this model to broaden the exposure of their collections. VuFind provides recommendations based on common elements.

In figure 7, we can see a view of the record for *The Cathedral and the Bazaar*—a popular book about open source software. On the right-hand side, we see similar items that are recommended to the user. Below that, the Other Editions box provides a link to the first edition of the book.

Browsing is also a highly valuable approach to website navigation, and the faceted navigation model makes that highly intuitive and greatly increases precision of the search results. Many sites that adopt the faceted navigation model, such as e-commerce sites like Bestbuy.com or Shopper.com, allow the user to start not by searching, but by browsing the collection starting from a list of facet values. If I am searching for a new television on the Best Buy website, for example, I start with TV & Video, then TVs, then LCD TVs (see figure 8). This path allows me to browse through the product line and get directly to what I am looking for. I don’t have to think of search terms up front but am able to browse the taxonomy of terms in a hierarchical manner to find exactly what I want in a very intuitive way.

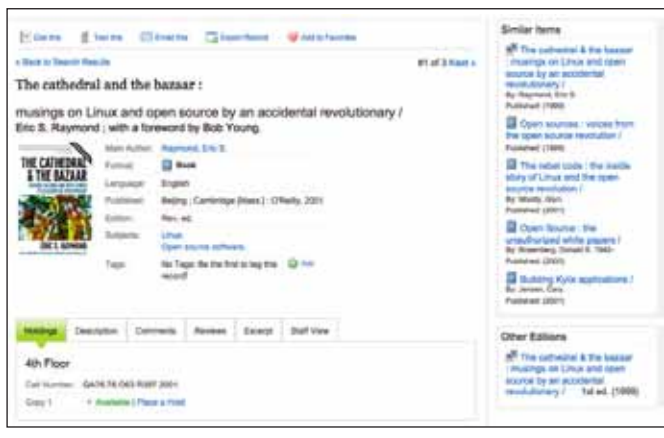


Figure 7
VuFind page on *The Cathedral and the Bazaar*.

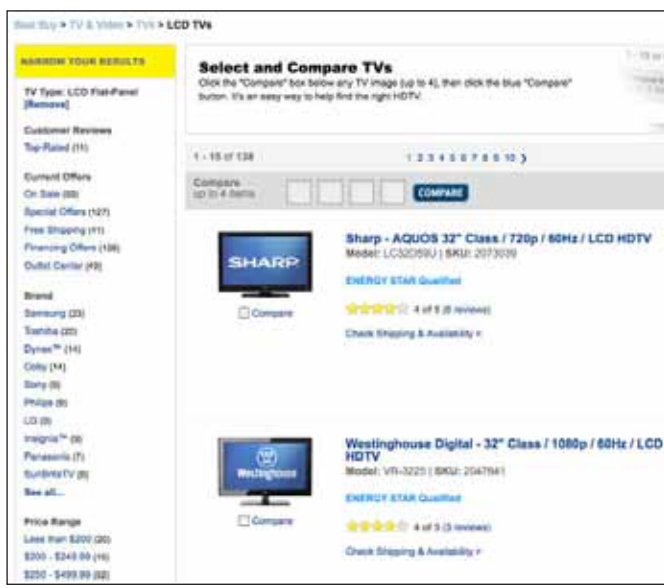


Figure 8
Best Buy website page showing flat screen LCD TVs.

There is a growing need in the information industry to provide the ability to relate. With the advent of the Semantic Web, building relationships between entities will allow the researcher to understand more about the content that is being studied. Libraries have the ability to help the Semantic Web take shape. By participating in the Semantic Web and evolving cataloging practices, libraries can foster and define these relationships. A next-generation solution can be the tool that allows libraries to do this. The library catalog is an authoritative source on materials held by the library, and other sources are authoritative on subject terms, authors, and call numbers. When these connections are made, the researcher can be better equipped to browse at a more macroscopic level through this notion of the Semantic Web.

Morgan's assumptions from 2006 are quite visionary and depict a future that goes beyond the NGC. What Morgan has described is what is being adopted today by libraries as the next step in discovery and access, the web-scale discovery solution.

Products

The NGC market has grown over the past five years with a multitude of options, including both commercial and open source options, full turn-key solutions and those that require local development efforts. Here is a sampling of some of the products in the marketplace.

AquaBrowser

Medialab Solutions BV, founded in 2000 in Amsterdam, the Netherlands—a small company at the time—set out to create a search engine solution that could be customized to the collections of commercial companies, nonprofits, and governments. It quickly found a successful channel working with public, academic, corporate, and government libraries with its AquaBrowser library solution (see figure 9). By 2010, over 800 libraries around the world used AquaBrowser as the search solution.

Encore

Encore was first announced in the summer of 2006 and released in the summer of 2007. The announcement by Innovative Interfaces (see figure 10) said, “patrons will be able to see everything the library has to offer, in terms of services and content, with minimal effort.”²

Endeca

While Endeca is not precisely an NGC, this company and product are worth mentioning. Endeca is a solutions company that provides search engine technology. This widely adopted technology has found a home in the library world. Its first use was by North Carolina State University (see figure 11), and it has expanded from there to libraries that are seeking a highly tailored search solution. This solution requires the library to build its own front-end interface, but its back end is very rich with features and highly scalable.

Primo

Primo (see figure 12) was first announced in the summer of 2006 and released in summer of 2007. Ex Libris announced Primo as “a single unified solution for the

discovery and delivery of all local and remote scholarly information resources, including books, journals, articles, images, and other digital content.”³

VuFind

VuFind, an open source solution first released in the summer of 2007 by Villanova University, was intended to provide a leading-edge interface allowing library patrons to discover the library’s collection in the same manner that they are used to when using the open web every day. A product that was developed by libraries for libraries, it made a big splash when the first production installation of the software was deployed by the National Library of Australia in May 2008. VuFind is not the only open source NGC solution available. The number is growing; some of the others are Blacklight, SOPAC, Scriblio, and Summa. Today, many libraries around the world have adopted VuFind and have deployed it as the central point for research on the library website.

As you can see from this sampling of products, there is a common thread—they all employ faceted navigation. The idea behind this style of navigation fits the search-and-refine user behavior model, a search behavior that is popular with the Google approach to searching. Users search on a term or set of terms that are relevant to their topic. They then analyze the results and refine the search terms based on the results presented. Faceted browsing is an approach that makes this model more effective by presenting users with faceted values of the search results that can then be applied as filters. A user can start with a broad topic, for example “green energy,” and then narrow the results down to something more specific. Faceted navigation has been researched heavily by professor Marti Hearst at the UC Berkeley iSchool, who notes, “Faceted navigation is a proven technique for supporting exploration and discovery and has become enormously popular for integrating navigation and search on vertical websites.”⁴

Open Source versus Commercial Solutions

A library that is looking to implement a commercial solution has a different set of needs from one that is looking to implement an open source solution. While open source may be attractive due to a perceived low cost when compared to commercial solutions, one must remember that open source is “free as in kittens, not beer.”



Figure 9
AquaBrowser on the Queens Library website.



Figure 10
Encore on the Grand Valley State University website.



Figure 11
Endeca on the NCSU Libraries website.



Figure 12
Primo on the University of Tennessee website.

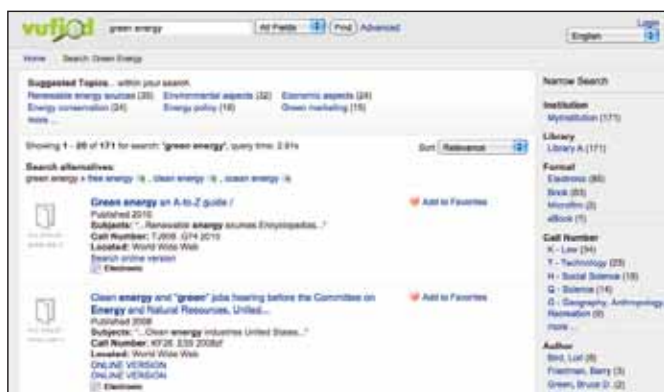


Figure 13
VuFind.

While a free kitten is cute and cuddly, it needs lots of love and attention in order to keep it healthy. It also needs care over the years to retain its health, an indirect cost that is associated with its adoption. A free beer is delicious and free—it needs no love and care, just quick consumption. Open source must be viewed as a free kitten: it needs direct involvement to get the solution installed, set up, configured, customized, and launched. Paying for support and maintenance is also an ongoing indirect cost. However, this indirect cost can vary from organization to organization. If you have a software developer in your team, your cost might be lower than the cost to an organization that needs to hire a developer to do the initial installation and maintenance over time. Organizations that have the resources in place and that are already familiar with open source solutions will find that an open source NGC can be a great fit. They can download and install various available solutions in a relatively short time, then test and evaluate each solution for little or no cost. For example, VuFind and Blacklight share many common technologies. Both use the open source Apache Solr for their

underlying search engine, and both use the open source SolrMarc tool for loading MARC records into the index. A library can download and install both and try them out at the same time without having to create two different environments in which to install the applications. The open source Evergreen ILS has even created a snapshot of an operating system with the product already set up and loaded with sample data for immediate deployment into a virtualization application.⁵ These organizations can communicate with existing users of the software in open collaborative communities to get more insight into the strengths and weaknesses of the product. Evaluating and talking with existing clients of commercial software is not as easy. Of course, commercial software has its strengths—support from product experts, a company that needs to keep the product active and in development, a financial investment in the future of the product. And of course, there is someone to sue when something goes wrong—an actual statement that I heard from a librarian.

In every marketplace, there is a fit for open source software and there is a fit for commercial software. Both have their strengths and weaknesses. There is no one right or wrong answer to choose one over the other.

Notes

1. Eric Lease Morgan, “Next Generation Library Catalog,” Infomotions website (originally published on the LITA blog [www.litablog.org]), June 2, 2006, updated Dec. 27, 2007, <http://infomotions.com/musings/ngc/index.shtml>.
2. Innovative Interfaces, “Innovative Announces Encore” (press release), May 26, 2006, *Library Technology Guides* website, <http://www.librarytechnology.org/ltg-displaytext.pl?RC=12014>.
3. Ex Libris, “Vanderbilt University and University of Minnesota Partner with Ex Libris to Deliver Primo—The Next-generation, User-centric Discovery and Delivery” (press release), June 19, 2006, http://www.exlibrisgroup.com/default.asp?catid={EEF2DEB0-987D-45F4-9069-7D1B4178196F}&details_type=1&itemid={8FDC2D12-51A0-4447-B34E-B3FD63614ACF}
4. Marti A. Hearst, “UIs for Faceted Navigation: Recent Advances and Remaining Open Problems” (paper presented at the Workshop on Human-Computer Interaction and Information Retrieval, HCIR 2008, Redmond, WA, Oct. 23, 2008), 1, <http://people.ischool.berkeley.edu/~hearst/papers/hcir08.pdf>.
5. Evergreen, “Evergreen Downloads” (see paragraph under “Evergreen Virtual Images”), http://open-ils.org/downloads.php#evergreen_vm.