

Differentiators and A Final Note

Abstract

The previous chapters introduced web scale discovery and profiled a majority of the key players engaged in this space as relates to the library environment. While similarities abound, differentiators are present as well. This chapter highlights some of the differences in the areas of content coverage, metadata and relevancy, pricing, integration with other systems, and the interface. As evidenced throughout this report, each service continues to evolve at an extremely rapid pace in terms of content covered, and the features, functionality, and flexibility of the interface. While these services each hold great potential, a final note observes that web scale discovery services, at least at their present stage of development, are not the “final word” for the library discovery environment.

Web scale discovery platforms customized to the library environment, handling local library and remotely hosted aggregated publisher content are in their extreme infancy. As observed in the first chapter, features, functionality, and content scope are changing—expanding—rapidly for all players. Press releases occur often, and annual library conferences provide a showcase forum for vendors to introduce their products to potential new customers and highlight enhancements to existing customers. Vendors host presentations and panel sessions discussing the merits of their discovery service and oftentimes provide information on why they feel their offering is the best on the market. From the preceding chapters, readers will note many similarities among the discovery services, and this observation is indeed valid. Given the extremely rapid cycle of development combined with the growing openness of such platforms, this issue of *Library Technology Reports*

wasn't constructed as a compare-and-contrast product survey. Things change through enhancement cycles as vendors progress beyond version 1.0 and customers request new features.

Customers also create their own innovations facilitated by the openness of these platforms; as platforms become more open, libraries with technical staffing can truly customize these tools to their local environments and include additional functionality. Consider just a few examples. Claremont Colleges Library, with its Sherlock Search, has brought two discovery services together—a front-end Primo interface with harvested local resources, blended with commercial content populated through the Summon index service in the background. North Carolina State University has developed its own front-end interface, QuickSearch, which pulls content from a multitude of services. For a single search, this custom interface returns organized results including (but not limited to) commercial content—such as articles—from the Summon index, book and media materials from NCSU's Endeca-based catalog, results from a library website search, and Did You Mean? suggestions utilizing Yahoo! Web Services.

So, acknowledging the power and creativity opening doors as never before, what are some things to keep in mind for new customers that have yet to embrace a Web scale discovery service? Here are some broad factors to consider.

Content

The ultimate goal of any discovery service, bar none, is to place content in the hands of the user or, more specifically, to discover, present, and deliver relevant content in a convenient, intuitive manner to today's

Claremont Colleges Library Sherlock Search

http://chipri04lsna.hosted.exlibrisgroup.com:1701/primo_library/libweb/action/search.do?vid=CLA&fromLogin=true

NCSU Libraries QuickSearch

www.lib.ncsu.edu

(Note: Choose Search All.)

More information on QuickSearch

www.lib.ncsu.edu/search/about.html

researcher. As far as content scope goes, the overall volume of content natively indexed by each service remains a differentiator, but the difference is rapidly shrinking. Each vendor is busy inking agreements for access to publisher metadata and, preferably, full text for purposes of discovery; each completed agreement can open up thousands if not millions of new items that can be included in the discovery service's index. Publishers are aware that libraries look at click statistics and usage of their content. Unused resources are ripe for the chopping block in tough—or any—economic times. A growing number of publishers are willing to participate with Web scale discovery services.

Some vendors developing Web scale discovery platforms may be entering into exclusive agreements for content from particular publishers; others may refuse to enter into exclusive agreements, believing content should be open to any discovery service. Some vendors indicate they are “content-neutral” and, since they are not themselves native providers of content, suggest that returned query results utilizing their services are free of any potential bias related to provider or source of content. They posit that content neutrality holds a potential for rich future publisher agreements. Given that there is no conflict of interest—the discovery service isn't owned by a parent company that's also a content provider (competitor) itself—they suggest more publishers may be willing to enter into agreements for purposes of having their content centrally indexed. To be fair, other vendors deny any hints or suggestions from competitors that query results from their products are biased or that their publisher agreements are lacking compared to other services. The author is not presenting a view or opinion one way or another on this question but raises the concept of content neutrality for the reader to consider. It's a touchy philosophical subject.

No matter what content is covered in the central index, it's important for individual libraries—potential customers—to work with vendors to conduct content overlap analyses to see what amount of that library's licensed or purchased content is included in each vendor's centralized index. Ideally, a lot of what

the library subscribes to and researchers are interested in—online, full-text, and 24/7-available electronic content—will be included and discoverable in the central index and, through a link resolver or similar mechanism, accessible from start to finish for the researcher. Libraries can choose how exhaustive they wish an overlap analysis to be. At one extreme, the library could choose to provide to discovery vendors a full set of electronic journal titles and publisher packages with holdings information and ask to what degree the central index encompasses such content. Or the analysis may be more streamlined—such as a library determining the top 100 or 500 journal or newspaper titles and asking vendors to provide an overlap analysis with this information. Always remember that all vendors are working aggressively with publishers to ink additional agreements and expand the content coverage of their services. Some vendors have focused on scholarly article-type content, others have greater e-book content, and some have greater coverage of newspaper content. Vendors are alike in that their initial focus has been on academic customers, who often have richness and depth of article-level subscriptions with publishers and aggregators.

Metadata and Relevancy

Metadata (amount and quality), sound indexing, and relevancy-ranking algorithms are all crucial in best matching items to a user's search. Different vendors have varied viewpoints on what constitutes sound metadata and the source of that metadata and talk about why they feel their approach is the ideal solution. Metadata conversations encompass “thin metadata”—a few record fields, perhaps a table of contents—and “thick metadata”—covering more fields, including additional abstracting and indexing by dedicated staff, or including author-supplied subject headings and abstracts. Some vendors have access to complete and comprehensive metadata from well-established content databases. Several vendors utilize a super or merged record, where different fields or levels of metadata for the same item—received from multiple content providers—are joined through common matchpoints and, through normalization and deduplication processes, result in a rich, accurate, highly discoverable and relevant record. Reading between the lines, 100 percent coverage of a particular resource from one vendor may not be precisely the same as 100 percent coverage of that same resource from another vendor. More specific statements are difficult, given the fact that thousands and thousands of indexed titles exist, and detailed studies would be needed to judge the accuracy of one vendor's point of view—and facts—versus another's and are outside the scope of this report.

Each vendor has developed its own proprietary relevancy algorithms. Some indicate that they take into account publishers' own relevancy ranking for materials provided by that publisher. Each offers a strategy for how to prevent items with thin metadata from being lost among items with thick metadata; however, no system will *ever* be perfect for all searches by all users. Some services allow the local library to influence the algorithm or otherwise promote or boost items within search results, and, depending on the service, this boost may be at the item level, collection level, or database level. Some vendors may place greater emphasis on currency, some on full text, some on subject headings. Some fields may factor heavily into one service's algorithm and carry less weight in another service's. Such factors can vary by item type, regardless of service. It's up to the local library to question vendors, conduct sample searches, and gauge what level of satisfaction they have with the vendor's approach.

Price

Each vendor has its own pricing model, and while some similarities exist, differences are also present. Some pricing models include, among other factors, references to the number of local records harvested. Some focus on institutional FTE or level of degree granted by academic customers. Most, if not all, vendors are willing to discuss consortial and multiyear discounts or to give price breaks if other products they market are also purchased or subscribed to.

Staffing is also a pricing consideration. All vendors offer completely hosted versions of their discovery service—providing the hardware, maintaining backups, and hosting the interface and centralized index. Such a scenario relieves local staff from maintaining hardware and performing backups. Some services allow the library to host the hardware and the interface (whether the vendor's or one developed locally). In some cases, hosting the hardware locally may provide even greater flexibility in customizing the service. In all cases, the preaggregated central index is hosted and accessed remotely. That said, response times for all the services are outstanding and similar to a Google search; the only (short) lag noticed may be with the real-time status check for items in the library's ILS.

Integration with Other Systems

A fundamental shift occurred several years back with the advent of next-generation discovery layers (e.g., Ex Libris Primo, Innovative Interfaces Encore, Serials Solutions AquaBrowser); such discovery layers added new features and functionality on top of the traditional

ILS online public-access catalog and were agnostic to the underlying ILS. Several of the new Web scale discovery services are built on top of these next-generation discovery layers from a few years back. One Web scale discovery service, Summon, was built from the ground up. How well and to what degree a particular web scale discovery service may integrate with a given ILS from a particular vendor may vary for purposes of placing holds, seeing which items are checked out, and so on. A library with a Web scale discovery service and underlying ILS from the same vendor may find tighter integration, such as easily enabling the same student account to be used for both systems, enhanced information display capabilities, and so on. A critical step for any library considering a Web scale discovery service is to ask the vendor detailed questions about integration with the underlying ILS (and other information repositories). It's important to understand what discovery services may require a jump to the underlying ILS for traditional OPAC functions (holds, requests, ILL) and which ones can accommodate such functions from directly within the discovery service interface. Just as important, libraries should ask if any existing customers using the prospective library's ILS have gone live (such examples likely exist). Potential customers can take a look at the live site and contact the live library for its experiences and observations. All of this aside, keep in mind that the pool of traditional library holdings—physical items cataloged into the ILS—is not the shining star and chief selling point for Web scale discovery, and so level of integration with the underlying ILS shouldn't necessarily be a strong area of scrutiny. Rather than focusing on local content, numbering in the thousands to several million items, depending on library, Web scale is focused on the *hundreds of millions* of items not present in the ILS—the massive, current, growing body of journal articles, newspaper articles, conference proceedings, and so on. The beauty of these Web scale discovery services is their ability to host, search, combine, and deliver content from both content pools, local and remote.

Some systems may presently handle consortial-level implementations better than others. This is an interesting topic for some but not all libraries and was left out of this issue of *Library Technology Reports*. It's fair to say that some systems (e.g., WorldCat Local) are built upon systems with extensive knowledge of other libraries' materials and have integrated mechanisms and available workflows in place to facilitate things like ILL requests. Another system (Primo) can search the local indexes (ILS records and other information repositories) of other Primo sites; a Summon search can include the digital collections and institutional repository materials from other Summon sites. WorldCat Local tiers can be constructed, scoping the search from the local library to an extensive consortium. Not to be left out, Ebsco has features facilitating

consortial installations as well, and every system offers additional consortial options not mentioned here. If libraries are interested in a consortial purchase of a discovery service, they will be well served to ask each vendor about how its service can fit into their consortial environment. Questions about staff workflows, integration with (or accommodation of) consortial catalogs, branch or site branding, and the ability of the discovery service to be scoped to the particular hard-copy and electronic holdings of disparate consortial members are all relevant.

Different vendors offer optional add-on products, usually for an additional cost. All Web scale discovery service vendors also offer federated search products in addition to their preaggregated central index. Vendors generally agree that it is likely that some content of interest to libraries will always be missing from the central index. This statement is definitely true for the present and at least the short-term future. Several vendors indicate that federated search can help plug the gaps and will be part of the discovery landscape if one wants to conduct a search that's as inclusive as possible; others indicate that combining federated search with Web scale discovery via a central index can be confusing and difficult and that the traditional problems of federated search remain—problems such as slow delivery times, poor relevancy-ranking capabilities, limited query returns, and results lost within the larger aggregate of centralized content. Different vendors have taken different approaches, and each has arguments for why it feels that its approach is best. Potential library customers should learn how federated search fits—or doesn't fit—into the overall discovery solution by conducting their own research, talking to existing customers, and having detailed questions prepared for vendor visits and conversations. One service, Primo, offers the optional bX Recommender service (described in chapter 5), which merits investigation.

Parallel with the integration discussion arises an efficiency discussion (and, on the flip side, the discussion of reliance on a single vendor). The highlight of Web scale discovery lies with exposing the huge amount of published content subscribed to or purchased by libraries. As all libraries are aware, collection development, rights management, and maintenance of electronic content are significant tasks. Different vendors offer different products in their portfolio, which, when taken in sum, could often be seen as a complete library solution (similar to turnkey ILS systems awhile back). Such products can include an ILS (or components of an ILS), an electronic resource management system, MARC record services, enrichment content services, link resolver or similar knowledge base for rights management, A–Z title lists, proxy server, and so on. Products from a single vendor are often designed to integrate well and can foster staff efficiencies. Fiscal efficiencies may also result through

bundled product purchases or associated annual licensing and maintenance. Reliance on a single vendor has some potential downsides—competing products may offer what the library deems are must-have features; a vendor could choose to inordinately raise maintenance or support pricing; and so on. Fortunately, exit strategies exist through the open nature of these products. Link resolvers, proxy servers, ILS systems, and discovery services from different vendors can often be mixed and matched into a precise solution fitting the libraries' needs and workflow.

The Interface

Some services profiled in this report are more open than others. While all offer some level of customization allowing libraries to make the discovery service their own, the level of openness and flexibility vary. At a minimum, all offer a basic template for libraries wishing to make some choices but perhaps not deeply tinker. At the other end, some offer extreme flexibility, enabled and augmented by capable toolkits, flexible established APIs, use of modern open Web technologies, and a user group consisting of established customers who have already shared or are willing to share developed code and ideas. For those libraries with sufficient staffing and skill sets, such flexibility can be attractive.

As stated earlier, the purpose of Web scale discovery services is to connect users, as seamlessly and easily as possible, to content. Assuming the content is there to be discovered (and this is becoming less of a differentiator as all vendors ink more agreements with publishers and aggregators), and assuming the vendors have quality metadata and finely tuned relevancy algorithms (that's for the library to investigate), then a final question revolves around the interface. All vendors indicate they have conducted (and continue to conduct) extensive usability studies in designing their interface; some discovery services are becoming established to the degree that early-adopter libraries have conducted their own independent usability studies. Some vendors provide usability information directly on their websites, and others may be willing to share reports if asked. Assuming one is using the default template, the interfaces for the discovery services look quite similar—a search box at top, results presented in the middle of the screen, and facets and other search refinements in a pane along the left. That said, there are some differences, and how significant those differences are can be determined only by the prospective library customer for its environment. At the time of this writing, some differences exist. Some, but not all, alert the user that an item is a full-text item. Some allow you to limit to full text only, and some designate peer-review status. Some present additional information

when the user clicks on a tab or hyperlink; others rely more on mouseovers. Some accommodate the addition of widgets to provide additional services. Some offer boosting or highlighting of items, collections, or databases. Some have an index more open to search by unauthenticated users. One has established relationships with Google to help drive users to library-available content. All products offer rich export options for items of interest; some offer more than others. Some offer different facet refinement categories that may be of interest; one allows the library to define its own facet categories. All have advanced search modes with often similar capabilities, yet subtle differences exist. All discovery service search boxes can be embedded in different webpages or portals. Regarding resolution to the full text, some offer more streamlined access, at least for some resources from some content providers. Some offer a user account where researchers can save and later retrieve items of interest. Some offer rich social community tools, such as tagging and reviews; others don't (and suggest that tagging, ratings, and reviews benefit primarily the smaller sea of ILS and digital collection materials more than the ocean of articles and newspaper content).

A Final Note

The majority of vendors profiled in this report provided some details, included in the respective chapters, about potential near-future enhancements for their services, all geared toward refining these services to best meet the needs of today's generation. While library Web scale discovery has tremendous potential, there are several things to keep in mind.

First, such services do not cover everything of interest as pertains to a library's collections. This fact is due to a variety of factors, such as some publishers having yet to come on board and open up their content for indexing by these third-party discovery services. In some cases, contributing factors may also include current technical or compatibility issues.

Second, specialized databases may have search or presentation capabilities not easily integrated into the discovery service interface, at least at this early stage of development; as a result, database recommendations are starting to be integrated within discovery service search results. To a degree, silos of information—various repositories of information and their associated interfaces—will remain for the foreseeable future. For the present at least, library staff will continue performing cataloging and metadata work within their local ILS systems, digital content management systems, and institutional repositories.

Third, current discovery services can't read the researcher's mind and know precisely what he or she is searching for. However, apart from continued

refinement of relevancy algorithms, various recommender feature components are advancing the goal of returning relevant information to a given search.

Fourth, existing resources to which students have flocked for research needs are not going away. Google and Wikipedia are two of the most popular websites in the world, with good reason. Purchase and implementation of a library-focused Web scale discovery service is a first step; libraries will still need to studiously work to steer users to these services.

Libraries purchasing a Web scale discovery service obviously have implementation and marketing decisions to consider. Many libraries that have implemented a Web scale discovery service place the search box on the library's homepage, recognizing the importance of such services. Often, libraries provide a tabbed search box approach, allowing the user to choose which resource they want to search—be it the discovery service index, the local catalog (whether a traditional ILS or a “next generation” ILS discovery layer), a list of databases, an A–Z list of journals, etc.). Whether libraries choose to make the discovery service index the default search or not is a local library decision. Indeed, adoption of a Web scale discovery service can impact design decisions throughout a library's website—as mentioned, a search box for the service can be placed in multiple areas of not only the library website (and external websites for which the library has an account—such as Facebook), but other (often university controlled) sites such as course management systems. The level of marketing and bibliographic instruction can range from minimal to extensive. Web scale discovery vendors often suggest no instruction is needed, given the ease of use of such tools. Finally, “pleasing” all user groups is always challenging for libraries. Established faculty instructors (and librarians) may be used to the existing ILS, have their favorite topical databases, and enjoy browsing the table of contents of favorite journals. This competes with freshman undergraduates, who, as research shows, want quick, relevant information from the first tool they search. They perhaps (or likely) have no previous exposure to the university's ILS, or have favorite scholarly databases or journals. For the library, perhaps it's all about striking a happy medium. At the present stage of development, one new resource—no matter how promising—can (or should) immediately supplant a host of other, established resources. It is possible for several discovery systems to continue to coexist, and can be a fascinating (or frustrating) exercise for libraries to best choose how to design their webpages, market the strengths of the different systems, and provide appropriate instruction where needed.

Acknowledging some of the above challenges—or considerations—it will be fascinating to watch as these infant services mature. Eric Lease Morgan suggests lots

of interesting possibilities, noting that opportunities for future library catalogs (and, the author suggests by extension, Web scale discovery services) can be found in services—services that help researchers use the information they’ve found and better sense who the researcher is (such as a student or an instructor).¹ He offers examples of potential services, such as compare-and-contrast functionality, the ability to create different versions of a document, services to plot on a map, and services to translate. His extensive possibilities list includes many items that are beginning to appear in next-generation library catalogs and Web scale discovery services alike.

This issue of *Library Technology Reports* concludes where it began—with an acknowledgement that library-focused Web scale discovery services hold great potential and are evolving rapidly. This report has provided a snapshot of several Web scale discovery services developed and marketed by major

established vendors. The marketplace and development environment are still young for next-generation library catalogs, and younger still when Web scale discovery is added to the mix. Features, functionality, level of integration with other systems, scope of content, and soundness of metadata are all evolving, and, it’s hoped, will continue to evolve, better meeting the needs and expectations of today’s researchers. Things not offered by a service today may be offered tomorrow; things not quite envisioned are ripe to be imagined.

Note

1. Eric Lease Morgan, “Next Generation Data Format,” May 2008, Infomotions website, Infomotions’ Musings on Information and Librarianship section, <http://infomotions.com/musings/ngc4mla>.