

2013

Discovery Systems: The Promise and the Reality. A Report of the RUSA MARS Local Systems & Services Program, American Library Association Annual Conference, Anaheim, June 2012

Amy Fry
Bowling Green State University, afry@bgsu.edu

Follow this and additional works at: https://scholarworks.bgsu.edu/ul_pub



Part of the [Library and Information Science Commons](#)

Repository Citation

Fry, Amy, "Discovery Systems: The Promise and the Reality. A Report of the RUSA MARS Local Systems & Services Program, American Library Association Annual Conference, Anaheim, June 2012" (2013). *University Libraries Faculty Publications*. 11.
https://scholarworks.bgsu.edu/ul_pub/11

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@BGSU. It has been accepted for inclusion in University Libraries Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

Discovery Systems: The Promise and the Reality

At the 2012 ALA Annual Conference, an audience of over 300 gathered for the RUSA MARS Local Systems & Services Program “Discovery Systems: The Promise and the Reality.” Five speakers shared ideas and experiences that were informed by their roles, institutions, and the specific discovery services their institutions are using. Their comments highlighted the goals these very different libraries have in common, as well as the diversity of the decisions they have made and the things they have emphasized in order to achieve them.

Library automation expert and founder of *Library Technology Guides* Marshall Breeding first provided an overview of discovery technology: its evolution and current issues. Breeding described the technology as giving users access to library collections in a way that breaks down content silos and works towards comprehensive searching. One of the failings of the traditional online catalog, he said, is that it “misses fundamental things in libraries,” because an ever greater part of libraries’ collections and services – in particular, the articles in our journals – are not represented in the integrated library system. In 2002, “next-generation” catalogs came on the market which were “cosmetically and functionally” better than the old web OPACs (they included “do you mean” prompts to correct spelling, more visual elements like book covers, and more social and self-service features) but didn’t expand the scope of the catalog. The addition of federated search technology provided some movement in this direction, but discovery products, which he described as having evolved from next-generation catalogs with discovery interfaces to index-based discovery services, give users access to library collections in a way that is “fundamentally different” from both traditional and “next-gen” catalogs and federated searching.

Index-based discovery services (Breeding listed Ex Libris' PrimoCentral, Serials Solutions' Summon, OCLC's WorldCat Local, EBSCO Discovery Service, and, though it works differently from the others, Innovative's EncoreSynergy as examples) are often described as being "web-scale" – meaning they search library collections the way Google searches the web: by searching the entire breadth of content available in the library's collection. The services work by gathering and maintaining an index of the library's collection via a knowledgebase and ingests of local catalog records, metadata from local digital repositories, etc. Breeding stressed that discovery indexes encompass much larger numbers of items than were ever represented in any library's OPAC – about a billion. Indeed, Andrew Nagy announced, earlier that morning, that as of June 2012 the Summon index included 950 million items (Nagy, 2012).

While Breeding believes discovery technology is definitely moving libraries forward (in discussing the adoption rate of index-based discovery services, he asked the audience, "What are you waiting for?"), he also outlined several areas of concern. The first is the issue of participation – not all subscriptions and content, whether because of technology or the inability of vendors and publishers to come to an agreement, are available in every discovery service. Unrepresented items need to be spotlighted or otherwise specially promoted or they don't get used. Breeding mentioned data from the University of Huddersfield to illustrate this phenomenon (Pattern, 2012); at Grand Valley State University in Michigan (Way, 2010) and at Bowling Green State University in Ohio (Fry, 2012), libraries have also noticed that usage statistics for unrepresented collections drop precipitously after the adoption of a discovery service – whereas usage statistics for represented collections (as Michael Kucsak demonstrated later in the panel) can rise dramatically. The development of competing products by major content providers that either are themselves publishers or retain exclusive rights to publisher content has likely been a disincentive for cooperation in some instances; the rate and reach of the adoption of discovery

technology (which Breeding showed lags far behind that of the traditional integrated library system) is perhaps another.

The presence of full text in the index also adds a challenge for search and relevancy ranking. Items with full text, Breeding said, are “in a different search universe” than those without it – the full text of an entire book provides a lot more information about the book’s content than its MARC record, and discovery systems must search both types of content representations side by side. Therefore, use, citation and other measures are being brought into relevancy ranking: Breeding mentioned Ex Libris’s ScholarRank, used in Primo, as an example (Mark Dehmlow later described it in more detail), but Summon also incorporates elements such as length, times cited, and language rules into its relevancy ranking, and is working on incorporating recommendations (pulled from a library’s LibGuides) as well (Nagy, 2012).

Evaluating discovery products to know how if and how collections are represented is another problem facing libraries that adopt this technology. To address this, Breeding is co-chairing a NISO (National Information Standards Organization) workgroup with Ex Libris’s Jenny Walker called the Open Discovery Initiative which is working on issues of transparency in discovery services, including establishing best practices for content representation and facilitating interactions between publishers and service providers (NISO, 2012).

A fourth challenge especially concerns public libraries, Breeding said: the challenge of blending ebook content and services (which can have different checkout and simultaneous user models as well as complicated digital rights management requirements) with other collections within the discovery interface. Breeding mentioned several projects addressing this challenge: BiblioCommons; a partnership

between 3M, Penguin, and the New York Public Library; and another between SirsiDynix and EBSCOhost (Jefferson, 2011; Schwartz, 2012; SirsiDynix, 2012).

In his presentation, Breeding discussed what he feels is “increasing differentiation” between the needs and wants of public and academic libraries in their adoption of discovery technology: public libraries want more engagement and social features, while academic libraries are focused on delivering academic content, particularly journal articles, directly to end-users. These differences have led to differences in system design and adoption in these two library types. The other panelists, though they all came from academic libraries, really illustrated how different types of libraries – with their unique collections, blends of patron bases, and traditions of user instruction – value different things in discovery services. While all wanted to break down content silos and provide, as Breeding said, a “single unified layer of user experience” for searching their collections, each library’s vision for achieving this was somewhat different in viewpoint and emphasis.

Following Breeding’s presentation, Barbara DeFelice, Director of Digital Resources & Scholarly Communications Programs at the Dartmouth College Libraries, discussed insights from surveys and focus groups that were recently conducted with Dartmouth users. The library plans to use this information to generate ideas for future development of library services and tools.

The focus groups took place in 2012 as part of a needs assessment aimed at identifying an institutional approach to the next generation of library management systems. DeFelice and her colleagues explored user needs “across all library processes,” including, but not limited to, discovery (Dartmouth implemented the discovery service Summon during 2009). They developed questions that, though they asked users to discuss the tools they used, were themselves deliberately “tool-agnostic.”

Undergraduate participants mainly discussed “discovery and access,” naming tools to help them find and evaluate content. Graduate students, on the other hand, primarily discussed tools to “manage content” and work with it – analyze text, identify networks of papers, annotate. Faculty, not surprisingly, were “coming from a much more specialized research viewpoint” than students, particularly undergraduates: they listed many subject specific resources and discussed their need for “tools to help make sense of the world of scholarship” – distinguishing between authors’ names (authority control) and discovering citation rankings – the kinds of activities that are necessary for experts seeking specific kinds of research in narrow sub-disciplines. DeFelice hopes this focus group data can spur thinking on how discovery services might enhance their features.

In her opening remarks, DeFelice talked about how she felt it was important for her library to provide a comprehensive search experience for Dartmouth library users (one place to search the breadth of the library’s collections and services, even its people and guides). In closing, she discussed how important she felt it also is to create a comprehensive *research* experience, where the diverse experiences of users at all levels can find accommodation and even be brought closer together. “This is the most amazing collection of metadata, full text and other kinds of content that we have,” she said, speaking about Dartmouth’s Summon index. She would like to see the tool provide more ways to work with and manage this content as well as find it, providing comprehensiveness of both resources and activity in order to provide “common ground” for faculty and students engaging in research together. While she sees ongoing technological, legal and social challenges for discovery services like Summon, she also sees their potential to bring users together.

Mark Dehmlow, Head of Library Web Development at the University of Notre Dame’s Hesburgh Libraries, began by explaining his institution’s selection of Primo (with its cloud-based index,

PrimoCentral), which has been being soft-launched at Notre Dame over the last two years. The university evaluated a number of tools in twelve different areas, focusing on search functionality, user experience and APIs (application programming interfaces). Notre Dame, an institution that has a history of building sophisticated tools to link and manage its library collections (such as the open-source ERM platform CORAL), wanted a discovery product that was also a “platform for innovation” - something they could customize and build on and which was extensible and modular. Primo, they felt, best provided this option.

For Dehmlow and his colleagues, system quality goes beyond numbers: while many vendors talk about quantity (numbers of scholarly titles, numbers of full text pages), they feel that a smaller amount of information, better organized, can make a better service. “Something well-organized is going to work very well,” Dehmlow said. “Something less well-organized makes you a candidate for a reality TV show.”

Like Breeding, Dehmlow stressed that relevance ranking has to go beyond word matching – it needs to combine information about the user (user type, purpose, discipline, and entry point, as well as query) with information from the network (examining what other users have chosen to do and then feeding that information back into the relevancy). Ex Libris’s ScholarRank relevancy engine for Primo does this, giving results based on the user, the document and the query itself.

While Dehmlow extolled Primo’s ability to bump up the visibility of local resources in the index (such as catalog records) while presenting them alongside global information resources (like WorldCat and HathiTrust results) (“the blend and the bump” he called this), he was also wary of over-reliance on electronic content. “A lot of these systems focus very heavily on the electronic materials. Our collection and our services are much broader than that,” he said. While direct-linking to electronic content is

valuable, he also wants to make sure that the system will show the availability of print when something is available in that format but not electronically, as well as push people to interlibrary loan when neither is available. “It’s not enough to just get people to some records,” he said, “we want to get the material to them that they’re actually looking for.”

Dehmlow feels the biggest deficit in Primo now is its service logic and integration. Unlike Summon, Primo integrates OPAC functionality into the discovery interface. He wants the system to be able to display appropriate user options based on each context – for example, print when there is not electronic; recall links when that is an option.

Dehmlow closed by discussing how the adoption of a discovery service is really a paradigm shift for both librarians and patrons. “Users have been using these somewhat complex systems for a long time and there’s a feeling that if you spend a long time learning something, it’s probably really valuable,” he said. As librarians, “we need to start recontextualizing the notion of complexity and of robustness.” At Notre Dame, the implementers have both been explaining how the service works to librarians so they can better teach it to patrons and visiting departments around campus to demonstrate and explain it to faculty.

Though Dehmlow talked about a paradigm shift for users and public service librarians, Michael Kucsak (Director of Library Systems & Technology at the University of North Florida) really showed how implementing a web-scale discovery service can impact work across all areas of the library, from acquisitions to cataloging and interlibrary loan as well as reference and instruction.

North Florida implemented EBSCO EDS in October 2011 not for flexibility and wide integration with diverse homegrown tools, but to minimize disruption between systems and promote uniformity of service. When looking at systems, they were interested in breadth of coverage, relevance of results, and reliability of full-text linking. They asked, “Which webscale system and database combination will give us the best product?” Having already implemented the EBSCO link resolver, LinkSource, and knowing they could migrate many databases to the EBSCO platform, they felt they would achieve the best overall result implementing EBSCO EDS. So far, full text linking has been very reliable, Kucsak reported.

Kucsak favors discovery technology because it supports interdisciplinarity by bringing together results from a variety of databases – resources students might not think to search because they aren’t core to a discipline, but which may still contain valuable and pertinent information. In addition, having a discovery service provides the university a cushion if they lose a database. Unlike Notre Dame, the University of North Florida’s collections are almost entirely focused on electronic materials: a deep last-minute budget cut the previous summer was largely absorbed by the library’s print budget, which was cut 86%, mostly in response to declining circulation but also because of increased spending on e-resources (which went up 70% over the previous four years). With the print budget already cut so precipitously, Kucsak said, there is “nothing left” to absorb double-digit percentage price increases to databases and e-journals.

Kucsak compared full text downloads from four e-journal providers before and after North Florida’s EDS implementation. Though downloads rose 11% in 2010-2011 (a statistic that makes sense in that institution’s environment of growing reliance on electronic content), they rose seven times that after EDS was implemented – an average of 76% between October of 2011 and April of 2012 over the same

period the previous year. “The numbers are beyond our expectations since implementing webscale,” he said. The library saw an even more surprising rise in ebrary use – though it had purchased ebrary’s Academic Complete in 2011 and promoted it on the library website, blog, LibGuides, and in instruction sessions, use was sluggish (no MARC records for these titles had been loaded into the catalog). After activating the collection in EDS, however, ebrary use went up 2,351%. As a result, the library has stopped all MARC record loads for its e-products. Interlibrary loan use has also been flat, and ILL costs are down 43%. Kucsak speculated that the reason is that EDS, pulling diverse content from many sources, is a more efficient way to leverage the library’s collection and deliver content to users, who have now found that their need for ILL is diminishing.

While Kucsak did not also explore other potential effects on full-text downloads or ILL use other than the implementation of EBSCO EDS, there is little reason to doubt that implementation of EDS has had a profound effect on the University of North Florida library – both because of how it is getting students to content and because of administrative and procedural decisions the library is making as a result. Like Breeding, Kucsak addressed the problem of participation in discovery, but more bluntly: “You’ve got to be in,” he said. “You’re in – you win. You’re out – you’re not long for the world.” The first thing they now ask about new products is whether they are discoverable in EBSCO EDS. Not only is forging agreements among vendors for discoverability of content crucial for the use of resources, it soon may be so also for the continued viability of those resources as subscription products.

The final speaker, Cody Hanson, joked that he could have called his presentation “What We Were Up To at the University of Minnesota Libraries While Our Peer Institutions Were Implementing Cool Webscale Discovery Systems.” After implementing Primo in 2007-2008 (before the release of PrimoCentral), Hanson (Web Architect and User Experience Analyst at the University of Minnesota Libraries) and his

colleagues spent several years investigating discovery under the project name “Discoverability” (Hanson et al., 2009; Hanson et al., 2011). Not unlike the research DeFelice discussed at Dartmouth (though done before, and not after, product selection and implementation), the group very deliberately ignored the marketplace, focusing instead on user needs and collections to ultimately define a vision for discovery at the University of Minnesota that would be “very holistic,” encompassing not just one system but how the library could overall “best provide access to our collections in the broadest way possible.”

In Phase I, the group looked closely at data and trends, bringing together statistics from a wide range of resources and systems. They found that discovery started and happened in many places and with many tools, which validated a broad definition of discovery. In Phase II, they did a detailed investigation of user needs and finally codified their “high level vision” for discovery at Minnesota, which involved not just bringing content silos together but providing “a common set of tools” (such as filters and de-duplication) to help users work with the content as well as scopes to drop users directly into relevant subsets of the collection depending on the context of their query and how each scope is best defined. The resulting image of their vision is something that looks “very common now,” Hanson joked, “but we came up with it ourselves.”

At the time of his talk, the project was currently in Phase III: a year-long procurement process resulting from an RFP “which our university’s purchasing office described as the most complex RFP that they had ever seen.” Looking, now, at commercial products, Hanson said he’s learned that, though it is convenient to think of these products as “interchangeable widgets,” really they are “as different as they are similar:” unique not just because of their coverage or features but because of “the business model and the motivation of the vendor.”

While Hanson admits Minnesota is not on the cutting edge of discovery implementation, he also doesn't believe he and his colleagues have done their users a disservice by waiting, and he echoed Dehmlow's sentiments about the implementation of discovery being a cultural shift. Disruptive change exacts the biggest cost to the trust of people "for whom," Hanson said, "the library plays a very small role in their work and their research life." "Introducing interface change after interface change exacts a serious toll and serious cost to our institution" where many users are concerned, he said, and he is grateful that his library's slower adoption pace has perhaps had the effect of building up some residual good will in them.

Index-based, web-scale discovery systems have helped libraries move closer to Breeding's "single unified layer of user experience," but have also had the effect of precipitating cultural change in libraries – in our administrative and technical processes, our users' expectations, and how we as librarians and our patrons interact with and access collections. They also have put into stark relief some of the divisions and differences we face – differences between types of patrons (undergraduates and faculty), types of libraries (public and academic) and types of vendors (service providers, content providers, and publishers). The technology, while not as ubiquitous as the ILS, is definitely growing in adoption, and the experiences and decisions that Michael Kucsak related (re-vending databases, re-focusing budgets, reviewing collection development decisions, and stopping certain processes like MARC record loads) will be visited by many more libraries. While Dehmlow and Hanson both emphasized the importance of maintaining their catalogs as collection inventories for now, the fluid nature of what Dehmlow called the "world of knowledge" and the adoption of global knowledgebases will probably make that activity redundant in the not too distant future. As the panelists made clear, discovery technology comes with a paradigm shift – but of which this technology can hopefully be seen as an effect, not a cause. As DeFelice's and Hanson's work makes clear, the emphasis for libraries continues to be on collections and

patrons more than technology. As Hanson said, “Our stuff, no matter who the user; our user, no matter what the stuff.”

After the presentations, one attendee from a small library asked how big a library’s collection needed to be before a discovery system would be useful for its patrons. Hanson replied, “If your collection is big enough that the search box on your library home page doesn’t work the way users expect, you probably need a discovery system, and if you have a catalog and a database, it probably doesn’t work the way its users expect.”

Works Cited

Fry, A. (2012, April). The Biggest Winner: An ‘Urgent, Social, Blissful and Epic’ Competition to Promote Underused Databases. Lecture conducted from the Electronic Resources & Libraries Annual Conference, Austin, TX.

Hanson, C., Hessel, H., Barneson, J., Boudewyns, D., Fransen, J., Friedman-Shedlov, L., ...Traill, S. (2009). *Discoverability Phase 1 Final Report*. Retrieved from <http://purl.umn.edu/48258>.

Hanson, C., Hessel, H., Boudewyns, D., Fransen, J., Friedman-Shedlov, L., Hearn, S., ...West, A. (2011). *Discoverability Phase 2 Final Report*. Retrieved from <http://purl.umn.edu/99734>.

Jefferson, B. (2011). BiblioCommons. Retrieved from <http://www.bibliocommons.com>.

Nagy, A. (2012, June). Summon Update. *Summon Boot Camp*. Lecture conducted from the American Library Association Annual Conference, Anaheim, CA.

NISO (National Information Standards Organization). (2012). Open Discovery Initiative: Background. <http://www.niso.org/workrooms/odi/>.

Pattern, D. (2012, June). Summon Use at the University of Huddersfield. *Three Years of Innovation: How the Summon Service Meets User Expectations and Research Needs*. Lecture conducted from the American Library Association Annual Conference, Anaheim, CA.

Schwartz, M. (2012, June 21). Penguin, 3M Test Ebook Pilot at NYPL, BPL. *Library Journal*. Retrieved from <http://lj.libraryjournal.com/2012/06/shows-events/ala/penguin-3m-test-ebook-pilot-at-nypl-bpl/>.

SirsiDynix. (2012, June 22). EBSCO Publishing and SirsiDynix Sign Partnership Agreement to Enhance Discovery for Mutual Customers. *SirsiDynix Newsroom*. Retrieved from <http://www.sirsidynix.com/press/ebSCO-publishing-and-sirsidynix-sign-partnership-agreement-to-enhance-discovery-for-mutual>.

Way, D. (2010). The Impact of Web-Scale Discovery on the Use of a Library Collection. *Serials Review*, 36 (4), 214-220.

Special thanks to the speakers on our panel, members of the RUSA MARS Local Systems & Services Committee, particularly Chanitra Bishop and Geoff Morse, and Andrew Whitis.