2013

Web-scale discovery in an academic health sciences library: Development and implementation of the EBSCO discovery service

JoLinda Thompson  
George Washington University

Kathe Obrig  
George Washington University

Laura E. Abate  
George Washington University

Follow this and additional works at: http://hsrchimmelfarb.gwu.edu/libfacpubs

Part of the Library and Information Science Commons, and the Medicine and Health Sciences Commons

Recommended Citation


This Journal Article is brought to you for free and open access by the Himmelfarb Health Sciences Library at Health Sciences Research Commons. It has been accepted for inclusion in Himmelfarb Library Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.
Web-Scale Discovery in an Academic Health Sciences Library: Development and Implementation of the \textit{EBSCO Discovery Service}

JoLinda L. Thompson  
Kathe S. Obrig  
Laura E. Abate

\textbf{ABSTRACT.} Funds made available at the close of the 2010/2011 fiscal year allowed purchase of the \textit{EBSCO Discovery Service (EDS)} for a year-long trial. The appeal of this web-scale discovery product that offers a Google-like interface to library resources was counter-balanced by concerns about quality of search results in an academic health science setting and the challenge of configuring an interface that serves the needs of a diverse group of library users. After initial configuration, usability testing with library users revealed the need for further work before general release. Of greatest concern were continuing issues with the relevance of items retrieved, appropriateness of system-supplied facet terms, and user difficulties with navigating the interface. EBSCO has worked with the library to better understand and identify problems and solutions. External roll-out to users occurred in June 2012.

\textbf{KEYWORDS.} Discovery layers, discovery tools, \textit{EBSCO Discovery Service (EDS)}, federated search, Google, health sciences libraries, library discovery services, resource discovery tools, usability study, usability testing, user experience, web-scale discovery
Authors.

JoLinda L. Thompson, MLS, AHIP (jlt@gwu.edu) is Technical Systems Coordinator, Himmelfarb Health Sciences Library, The George Washington University, 2300 I Street NW, Washington, DC 20037.

Kathe S. Obrig, MLS (obrigk@gwu.edu) is Associate Director, Collections and Access Services, Himmelfarb Health Sciences Library, The George Washington University, 2300 I Street NW, Washington, DC 20037.

Laura E. Abate, MSLS (leabate@gwu.edu) is Electronic Resources & Instructional Librarian, Himmelfarb Health Sciences Library, The George Washington University, 2300 I Street NW, Washington, DC 20037.
Himmelfarb Health Sciences Library is an academic health sciences library serving 3,000 graduate and undergraduate students in the School of Medicine and Health Sciences, School of Public Health and Health Services, and School of Nursing at The George Washington University, and the faculty and staff who support those programs. It operates independently from the other two libraries on campus, with a separate administrative and fiscal structure, and separate integrated library system (ILS) and electronic resource management systems (ERMS).

The Library introduced a federated search service in 2007, which was featured in a search widget on the library home page. Though federated search provided users with a one-stop search solution for many online resources, it had drawbacks. Retrieval time was slow as each native database was searched sequentially and results were ranked based on their speed in delivering results, which skewed relevance rankings. As users frequently opted to “search all databases” in the federated search system, resources for which the Library offered limited simultaneous users were often unavailable to other users in their native interface because the seats were being occupied by the federated search. Librarians, however, did appreciate the ability to build multiple profiles to serve different groups and the capacity to conduct keyword searches across the full-text contents of several e-text providers. Despite the availability of federated search and efforts to update the library catalog with enhanced content, library users were increasingly using Google and Google Scholar as their primary search tools.

Simultaneously, a new category of search system was coming onto the market. Web-scale discovery was touted as a giant leap forward from federated search and possibly the answer to Google for libraries. Instead of the federated search model, which constructs a search that
translates across multiple database structures, the new model harvests the content to a central searchable index. This requires agreements with publishers providing the content, so early entries to this market from Serials Solutions with Summon\(^1\) and EBSCO with its EBSCO Discovery Service (EDS)\(^2\) product were not surprising. Other services, including Ex Libris’ Primo and OCLC’s WorldCat Local which started out as next generation catalogs, also made agreements with publishers to add electronic and database content, putting them in the new discovery model.

Web-scale discovery promises the ease and speed of Google’s keyword search, sorting massive retrieval sets with complex relevancy-ranking algorithms to bring the most relevant resources to the top of the results lists. The faceting features give users ways to drill down to the specific results desired by sorting source types, locations, publishers, author, etc. In addition, libraries can add content from their catalogs and digital repositories, making all of their local content accessible through the same search interface. A more in-depth discussion of web-scale discovery can be found in Matthew Hoy’s “An Introduction to Web-Scale Discovery Systems”\(^3\) and Athena Hoeppner defines and summarizes the vocabulary surrounding these systems in her recent Computers in Libraries article.\(^4\)

Himmelfarb identified web-scale discovery as a potential successor to federated search and a way to make the library’s print resources more visible to library users. The Library was concerned about how a single search interface could serve the needs of the diverse populations that use the Library. Another major concern was the large pool of content included in services like Summon and EDS, much of which does not have a health sciences focus, and the effectiveness of relevancy ranking, which if not finely tuned, could return many results not fully on target for the search. Because these systems were launched in general academic and public
libraries, it was not known if they could be tailored to work for a more specialized collection and audience.

OCLC’s WorldCat Local, EBSCO’s EDS, and Serials Solution’s Summon were all investigated. Himmelfarb decided not to pursue Summon, mostly due to pricing constraints. Six week long simultaneous trials of WorldCat Local and EDS in May of 2011 allowed Himmelfarb to evaluate both products side by side. Librarians liked the search interface and presentation of results on WorldCat Local and some felt that relevancy ranking was more effective on it than EDS, but EBSCO offered a better content match to Himmelfarb’s electronic collections and had superior citation management and post-search tools than WorldCat Local at the time of evaluation. Funds were available at the close of the 2010/2011 fiscal year to purchase EDS. It was anticipated that a one-year commitment would provide adequate time to configure and customize this complex product and expose it to librarians and users to reveal if it was an effective one-stop search solution for an academic health sciences library environment. (For a detailed profile of EDS, see the chapter on it in Library Technology Reports’ 2011 issue on web-scale discovery services.5)

**CONFIGURATION**

A rough timeline for implementation was determined shortly after purchase in spring 2011. Initial configuration, including loading of all catalog data and establishing Z39.50 links to live status of items, selection of content sources, and branding were targeted for completion by the end of summer. Release to staff for internal testing was to occur in fall 2011, with final changes and release to library users projected by Spring Semester 2012. It was hoped this would provide
adequate experience with the system to make an informed decision about subscription renewal by May 2012.

One of the first configuration decisions was identifying the target audience for the tool and determining if there was a need for multiple profiles to address the needs of divergent groups. The main audience was defined as graduate-level students including public health, nursing, and first- and second-year medical students. It was recognized that the search tool might also be useful at times to other user groups who would want a more generalized tool for initial searching about unfamiliar topics, or would be searching for a topic that was multidisciplinary in nature, the types of searches frequently being done in Google and Google Scholar. After these discussions, the development team decided that the best strategy was to build a general profile that would provide a comprehensive, but health sciences-focused tool.

EBSCO builds a unique index for each EDS customer including the base index (known as the Foundation Index), content from EBSCO-hosted databases subscribed to by the customer, and local content sources (catalog data and digital repository data) that the library may prefer to include. Much of Himmelfarb Library’s critical health sciences content was covered either through direct subscription or presence in the Foundation Index. The Library’s initial index included a number of university-wide subscriptions and open access sources. Some subscription content sources were obviously desirable, including key health sciences resources and databases like MEDLINE, CINAHL, PsycINFO, SPORTDiscus, and Health Policy Reference Center. The development team had more difficulty deciding and agreeing on whether to include sources of peripheral interest, but possible usefulness in selected situations: news sources, economic- and business-focused databases, general academic databases like Academic Search Premier, and broad-based archival and open access collections. The inclusion of these sources presented the
possibility of increasing “noise” in a search and possibly moving more relevant resources lower in the results ranking, but their exclusion presented the risk of limiting users’ access to possibly useful information. In the end, only some reputable business, economic, news, and general academic sources that seemed most compatible to Himmelfarb library user needs were retained. Appendix A contains a list of the sources selected for inclusion, but as content sources are added frequently, this list is subject to change.

The Foundation Index includes content from over 20,000 publishers and content providers who agree to make their content available to EBSCO for searching by all EDS customers. Though this provides metadata and full text for searching for many resources, access to the actual full-text content is determined by whether the library subscribes to that specific content. The team discussed possibly eliminating this data source due to a large quantity of non-health science specific content but also realized that important subscription content that is not included in EBSCO-subscribed databases is only available for searching in EDS by virtue of its availability in the Foundation Index. Himmelfarb elected to retain the Foundation Index to maintain access to these critical resources with the knowledge that it provides indexing and access to portions of the Library’s collection while also including resources for which full text would not be available.

The ease of getting to full text in EDS was one of the selling points of the system. Content that comes directly from an EBSCO full-text source that the library subscribes to, and content from open access sources, have PDF full-text links prominently displayed in the records on the results page. For content that the library provides full-text access to from other sources to which Himmelfarb subscribes, a Full-Text@ Himmelfarb button to the link resolver appears in the records on the results page. A Find It @ Himmelfarb button is provided for items that have
no full-text access available. A click of this button directs users to other document delivery options.

One of the most difficult parts of the configuration process was the branding and interface design. EDS allows customers to insert logos and customize colors and fonts through cascading style sheets (css) to conform with the institution’s existing web presence. Due to University restrictions regarding logos and strict policies governing the use of the University’s branding, it was difficult to come up with visually interesting graphics that would tie the website directly to Himmelfarb Health Sciences Library. Compounding the problem was the simultaneous internal vacancy in the Electronic Services Librarian position who serves as the Library’s web developer. EBSCO technical staff assisted with initial set-up by pulling in images and content from the Library’s home page. Subsequent design decisions moved away from mirroring the web design on the Library website and in favor of an EDS search page, which is cleaner and focused on the search box as shown in Figure 1.

[PLACE FIGURE 1 HERE]

Legend: FIGURE 1. Screenshot of main EDS page

In designing the default presentation for the EDS search box, Himmelfarb desired to keep it as simple as search boxes users encounter on other search sites. Basic search for keywords is the default setting, though users can click through to an “Advanced” search page if desired. Libraries have the option to present the Advanced search as the default. EDS is delivered with two toolbar menus on the upper left and right of all pages that can be customized. Himmelfarb populated the left-hand toolbar with links to the E-Databases, E-Journals, and E-Texts web pages
to make it easy for users to navigate directly to A-Z list access to these sources. This toolbar originally included additional options for the Library’s implementation of LibGuides, the catalog, and other tools, but wrapping on some screen resolutions made it look jumbled and those that were not considered as essential for quick access were subsequently removed. Links to an “FAQ” document and “Feedback” form were added later. The menu on the right includes access to user preferences and login options and also provides a quick link to Ask A Librarian/Chat help.

TESTING AND REFINEMENT

After initial configuration was completed in early October 2011, the service was released to library staff for internal testing and comment. Staff was queried regularly for input on their search experience, but thoughtful and detailed feedback was slow to come in with the demands of many other projects during the busy fall semester. Though some improvements were made based on limited staff feedback, it became clear that a more targeted effort was needed to assess satisfaction and usability. In early December, a focus group of all librarians was convened to determine what changes were required before a more general release to users.

The currency of citations retrieved was identified as an issue early on, and EBSCO was able to do some fine-tuning to make improvements. Another concern that was frequently shared was ineffectiveness of relevancy ranking. Librarians were finding that sources which did not seem on target or desirable were often filtering to the top. For example, poster abstracts would come up before more substantive review articles, or titles with just one or two of the matching search terms would come up before others with all the matching terms. Items with more
metadata (e.g., full text) seemed to outweigh items with less metadata and displayed first, sometimes skewing the relevance of results.

Limiting retrieval with facets, which appear on the left side of the retrieval display screen, was a potential way of getting to more targeted or desired sources. As delivered, EDS had many facets available but not all were useful in a health sciences context. The administrative module allowed facet sets that were obviously of no use to a health sciences audience (e.g., NAICS/Industry) to be turned off. But there were issues with those that remained. For example, the system supplied three separate subject facets, but it was unclear which should be used when and they could not be consolidated to one. Under Source Types, “reviews” meant literary/critical reviews, not review articles as expected or understood in a medical context. This continues to be a problem as selections within a facet set cannot be edited or removed, although a facet can be suppressed in its entirety. One of the biggest disappointments with the facets as delivered was the lack of a language facet. To limit to English language retrieval, users had to locate the option on the “Advanced Search” screen. Development of a “review” facet that would work as Himmelfarb users might expect and a language facet were requested as enhancements.

Not surprisingly, the areas identified as needing improvement in the focus group involved relevancy-ranking and confusion/frustration with facets. Librarians also did not understand the difference between limiting results to “Full-text” and limiting to “Available in Library Collection” as the results were often similar. Following is a list of all the issues identified for further investigation and improvement:

1) Presence of three different subject facets
   - [ ] What exactly does each represent?
   - [ ] Is there overlap?
Why are they separated?

Sometimes Subject-Thesaurus facet drops out when sort is changed from relevancy to date-based.

2) English language limit:

   Must be more prominent; should appear in facet list near top.

   Does not work for some resources (removes all newspaper content, for example).

3) Source types:

   Academic journal vs. periodical—“Periodical” is the word the system supplies, but in the health sciences library environment, it should more appropriately be titled “Academic journal”

   System-supplied label “Review” does not mean the same as “review articles” as our users think of it and is problematic.

4) “Full-text” and “Available in Library Collection”:

   “Full-text” choice does not represent all full-text resources available to our users.

   “Available in Library Collection” seems to map directly to the full text.

   Can we build a “Library Collection” based on our online holdings?

5) Relevancy Ranking:

   Results are inconsistent—seems to work well on some searches and not others.

   How can we boost the relevance of our catalog items and collections?

6) Related Images:

   This system-supplied facet returns results less relevant to research (e.g., pictures of people) and needs to be refined.
Although some librarians were enthusiastic about EDS and ready for its release to users, others had strong concerns about releasing something that would not live up to user expectations. A recommendation was made to conduct usability testing with Himmelfarb users to assess how they approached searching EDS and their degree of satisfaction with search results.

A meeting was held with Himmelfarb’s EBSCO medical sales representative and implementation specialist to discuss focus group results. This resulted in changes made by EBSCO to relevancy ranking to boost catalog records and the creation of a local collection to improve coverage for the “Available in Library Collection” limiter and the ability to mark resources with full-text availability. Discussion of how the facets worked revealed that differences in the fields available in different database records often could not be overcome, resulting in inconsistency with how faceting limits performed. Applying a facet could cause all the results from a database without a matching field to disappear. Facets whose content was not consistently available across all the databases and content sources were targeted for removal. Two of the three subject facets were removed based on this criterion. EBSCO has since collapsed the three subject facet choices to one for all EDS sites.

Usability testing was targeted for very early Spring Semester 2012. The development team opted to keep usability testing as simple as possible, not only due to time and resource limitation but also due to the strong desire to avoid formal IRB involvement. A literature review revealed recent articles on usability testing of discovery layers including Endeca\textsuperscript{6}, Primo\textsuperscript{7}, Aquabrowser\textsuperscript{8}, and VuFind\textsuperscript{9}. Most employed a staff administrator with a script who would direct the test subject to find known items and/or a set of topical searching tasks and at least one staff observer who recorded the subject’s actions and comments. Employing this technique, test subjects were encouraged to verbalize what they were thinking as they made decisions. Some
studies used recording technologies like Adobe Captiva to capture a testing session for analysis. Recording methods like this were quickly rejected because they were more likely to require IRB review.

The other discovery layer usability studies tested 5 to 12 participants. The development team decided that four student and four staff/faculty volunteers reflecting as closely as possible the diverse user population would be ideal. Volunteers were recruited via a call to student and faculty listservs with an appeal to help the library and offering either a $10 copy card or $5 coffee card incentive. Response was overwhelming from students and surprisingly strong from staff and faculty. It was relatively easy to select a cross-section of student and faculty/staff volunteers representing all three schools and varying types of information needs.

The script for the usability testing sessions is in Appendix B. It included a series of pre-search questions to assess the users’ familiarity with search systems and prior search experience; post-search questions regarding the users assessment of the new search system; two standard searches for students and two for faculty/staff; and time for each volunteer searcher to try a search of his or her own. Each session was scheduled for an hour, with most running about 45 minutes in length. The usability study searches for students were:

- Are there any scholarly, peer-reviewed studies on arsenic in children’s juices?
- Does the Library have the electronic version of the journal *Nature*?

The usability study searches for faculty/staff were:

- Find one of your recent journal publications.
- Does the library have resources on mentoring new faculty?

The pre-search questions revealed that the volunteer searchers were split between using Google (5) and using PubMed/MEDLINE (5) as their default resource when starting a health
sciences-focused search. Only one reported being satisfied with the federated search tool currently available on the library home page. Three (less than half) still used the library catalog on a regular basis.

Analysis of the notes taken in the eight usability testing sessions revealed the following problems with using EDS as configured for Himmelfarb at the time:

1. **Difficulty finding “Advanced Search”** – A surprising number of volunteer searchers went looking for “Advanced Search” immediately, several because they did not understand the implied Boolean AND of the basic search box and wanted to “and” search terms.

2. **Search results lost under search history** – “Advanced search” was configured to automatically display search history at the top of the retrieval page. If search history was lengthy, volunteers couldn’t find actual retrieval below the search history.

3. **Find It @ Himmelfarb?** – The button with the words “Find It @ Himmelfarb” which linked to the link resolver was too small to be noticed or not understood by some volunteer searchers. They couldn’t find where to go to get full text.

4. **Facet display** – Volunteer searchers did not scroll down far enough on the results page to see the full list of facets available.

5. **Limits/Facet selections carrying forward** – In several cases, previously applied limits or facet selections were carried forward and skewed results of newer searches.

Based on post-search question responses, the limiter and facet options were what the volunteer searchers liked best about EDS, despite problems encountered with them during the search sessions. All but one said they would use EDS if the library made it available to users. Six of the eight said that their least favorite thing about EDS was the inability to get to a good set
of results, or to get results that were consistently on target. Relevancy ranking clearly needed to be improved for EDS to be successful.

A call was scheduled with the implementation specialist at EBSCO to discuss usability results. Some of the searches that retrieved very off-topic results were shared in advance of the meeting. Several strategies were suggested for improving relevancy based on these findings, including turning full-text searching off. A search of “arsenic in juice” could retrieve an article with both terms, but in unrelated contexts. Though proximity and phrase searching were available and could improve the quality of retrieval, many users would not have the knowledge or desire to use these more advanced search techniques. Also suggested was limiting retrieval to “Available in Library Collections” by default. Both of these changes were made, and relevancy of retrieval improved markedly. Defaulting to “Available in Library Collections” also removed retrieval from the Foundation Index for content to which Himmelfarb did not have a full-text subscription. Users have the option to disable this default if they wish to broaden their search.

Other changes made before general release included removing the automatic display of search history from “Advanced Search” and increasing the size of the “Find It @ Himmelfarb” and “Full-text @ Himmelfarb” link resolver buttons. EBSCO delivered a language selection facet that was placed toward the top of the facet list. More facets judged to be extraneous were removed (Age, Geography). The facet set is constantly evolving and some of the facets eliminated during configuration are no longer available. Below are the currently available facets with those enabled by Himmelfarb in italics.

- **Content Provider**
- **Language**
- **Publication**
- **Geography**
- **Location**
- **Publisher**
RELEASE TO USERS

The time had come to decide between continuation of the federated search box or replacement with an EDS search box. Substantial and systematically obtained feedback or other evaluative measures from library users beyond the usability tests were not available to support a decision one way or another. However, a fair number of users had commented individually about dissatisfaction with federated searching. After discussion with the administrative team, Himmelfarb decided to cancel the federated search service and proceed with release of EDS to users in June 2012. Widespread dissatisfaction with federated search among library staff meant that most supported this decision and were of the opinion that EDS would offer a better search experience to users.

One of the final challenges was to find a suitable name for the service. A succinct name that communicated the health sciences focus of the service, the affiliation with Himmelfarb Library, and the “one search box” concept was desired. The name also had to be distinctive from the one being used by the main-campus library for its Summon application. Currently, it is called “Health Information @ Himmelfarb,” and the text “Start searching Health Information @ Himmelfarb” appears under the search box.

EBSCO technical staff assisted greatly with configuration of a widget for the Library’s home page that replaced federated search with EDS (see Figure 2). This tabbed search box widget also provides access to the catalog and a website search feature. The tab for EDS is labeled “Articles + Books” to better communicate to users what they are searching. Users can
change the search index from keyword to titles or authors with a pull-down menu to the right. A prominent link to “Advanced Search” also appears to the right which takes users directly to the Advanced Search page in EDS.

[PLACE FIGURE 2 HERE]

Legend: FIGURE 2. Image of home page with search box widget.

An FAQ document was developed to describe Health Information @ Himmelfarb and provide guidance on how to search, view, and manage retrieval; it is accessible from the menu that runs along the upper left of all EDS screens. A link to a feedback form that elicits information about the user and invites comment or suggestions appears next to the FAQ link. Instant Messaging (IM) chat assistance is available from a link at the top right of all web pages.

The new search widget went up on the Library’s home page at the end of June 2012, launching a one-year pilot service for library users. Marketing and evaluative efforts, including surveys and focus groups are planned for the upcoming academic year.

DISCUSSION

Himmelfarb Health Sciences Library was in the position to be an early academic health sciences adopter of web-scale discovery. EBSCO recognized the opportunity to work with a library, in this case an academic-health sciences library, with more specialized collections to see if EDS could be adapted to serve this market. Himmelfarb also recognized the opportunity to experience a lengthier and more detailed exposure to a product under serious consideration with
the Library’s actual resources, to better assess the product in the actual setting in which it would be used.

Although Himmelfarb librarians were aware that this was a complex and highly configurable service when the initial purchase decision was made, the amount of time and effort required to make it an effective search tool for an academic health sciences audience was greatly underestimated. A year was required to adequately assess the database possibilities and for the ongoing interaction with EBSCO necessary to modify and maximize the product’s capabilities.

The version of EDS that was rolled out in June 2012 to Himmelfarb’s users is quite different from the “out of the box” implementation the Library started with in summer 2011. It is a far more focused search tool, both in terms of resources included and options for narrowing retrieval. The Himmelfarb development team concentrated on cutting out the noise of less relevant resources that can create confusion and frustration for users, using a philosophy of discovery focused on recent and highly relevant return. Whether this effort was successful or not remains to be determined as the service is used by students, faculty, and staff for their diverse searching needs over the course of the academic year. As EDS is constantly changing, with new features and new resources becoming available on a regular basis, the work to continue to maintain and fine-tune this service will be ongoing. Formal evaluation of Health Information @ Himmelfarb is planned in the coming months, and many of the decisions made in configuration will be targeted for user feedback.

EBSCO is to be commended for all of the efforts made to assist and optimize EDS for Himmelfarb. Their openness to enhancements and suggestions, and willingness to find creative solutions to the challenges encountered during configuration, were key to the success of the project. A list of enhancements and improvements remains to be developed, including redesign
of the "Review" facet, more effective faceting to better account for differences in data structure between content sources, improved spell-checking and “Did you mean?” features, and enhanced relevancy-ranking algorithms that will bring the best quality sources more consistently to the top of hit lists.

CONCLUSION

The pace of technological development in search and discovery has been rapid, making it difficult for libraries to identify the right time to purchase and implement new products. Though web-scale discovery was new and relatively untried in an academic health sciences environment, it had become apparent that federated search no longer met Himmelfarb’s users’ needs. Configuring and implementing EDS was time- and labor-intensive, but paid off in an improved understanding of library users’ needs and the potential of delivering a search system that will meet them.

Received: September 18, 2012
Revised: October 26, 2012
Accepted: October 28, 2012
REFERENCES


