Comparison of three web-scale discovery services for health sciences research*

Rosie Hanneke, MLS; Kelly K. O'Brien, MLIS

See end of article for authors' affiliations.

Objective: The purpose of this study was to investigate the relative effectiveness of three web-scale discovery (WSD) tools in answering health sciences search queries.

Methods: Simple keyword searches, based on topics from six health sciences disciplines, were run at multiple real-world implementations of EBSCO Discovery Service (EDS), Ex Libris's Primo, and ProQuest's Summon. Each WSD tool was evaluated in its ability to retrieve relevant results and in its coverage of MEDLINE content.

Results: All WSD tools returned between 50%–60% relevant results. Primo returned a higher number of duplicate results than the other 2 WSD products. Summon results were more relevant when search terms were automatically mapped to controlled vocabulary. EDS indexed the largest number of MEDLINE citations, followed closely by Summon. Additionally, keyword searches in all 3 WSD tools retrieved relevant material that was not found with precision (Medical Subject Headings) searches in MEDLINE.

Conclusions: None of the 3 WSD products studied was overwhelmingly more effective in returning relevant results. While difficult to place the figure of 50%–60% relevance in context, it implies a strong likelihood that the average user would be able to find satisfactory sources on the first page of search results using a rudimentary keyword search. The discovery of additional relevant material beyond that retrieved from MEDLINE indicates WSD tools' value as a supplement to traditional resources for health sciences researchers.

Keywords: Information Storage and Retrieval; Databases as Topic; Information Seeking Behavior; Databases, Bibliographic; Libraries, Medical; Web-Scale Discovery Services

Web-scale discovery (WSD) services offer user-friendly search interfaces, relevance ranking, and large, centralized indexes, allowing rapid, simultaneous searching of multiple library resources. These features pose attractive advantages to health sciences library users, including undergraduate and graduate students as well as health sciences faculty. In addition to ample literature demonstrating undergraduates' strong tendency to opt for the

convenience and familiarity of the web over library databases [1–5], research has demonstrated that graduate students, including those in the health sciences, prefer Google as a search tool due to its speed and ease of use [6–8], and health sciences faculty are likewise increasingly using Google and Google Scholar for research [9]. Academic researchers frequently choose these tools over traditional bibliographic databases because of their efficiency and relevance ranking [10]. WSD tools, which allow users to search the majority of a library's resources conveniently through a single search box, merit further investigation from health sciences librarians due to users' demonstrated preference for this easy-to-use, relevance-ranked format.

However, to this point, literature on the applicability of WSD tools for health sciences users

has been slim. Implementation case reports speak to the time- and labor-intensive process of customizing a WSD product and conducting usability testing, but also confirm the product's potential to better meet library users' needs [11] and deliver a satisfying user experience [12]. Crook investigated the effectiveness of OCLC's WSD tool, WorldCat Local, as a potential research tool for busy clinicians, using simple keyword searches to test known-item searching [13]. The study found that WorldCat Local could successfully be used to locate material by both title and topic, and emphasized the need for further research on relevance ranking. Most recently, Ketterman and Inman conducted side-by-side search comparisons of PubMed and ProQuest's Summon service [14]. The authors did not evaluate the relevance of search results, but concluded through comparison of journal coverage and currency, that Summon is a valid research tool for the health sciences. While Ketterman and Inman did not recommend it as a replacement for PubMed, they argued that Summon was one of many resources that the library should promote to its users, when applicable to their needs.

A logical next step in this nascent area of study is to investigate the effectiveness of WSD search tools in returning relevant results on topics of interest to health sciences researchers. The purpose of this study is to compare multiple WSD tools, evaluating their respective abilities to return scholarly articles related to health sciences search queries and abilities to assess duplicate search results. It will also provide a glimpse into each product's overlap with MEDLINE content. The primary focus of this study is not to compare the search functionality of WSD tools with that of PubMed/MEDLINE, nor will it investigate WSD tools as a potential substitute for PubMed/MEDLINE (or any other traditional biomedical database). Instead, it is primarily intended to serve as a resource for libraries that serve large populations of health sciences users, presenting a side-by-side comparison of existing discovery products in order to provide quantitative, evaluative information in the most evenhanded manner possible.

METHODS

We evaluated WSD services from three major commercial vendors: Ex Libris's Primo, ProQuest's Summon, and EBSCO Discovery Service (EDS). Each of these three services were evaluated at two separate institutions that had implemented the tool,

resulting in six data collection sites. Research sites were chosen based on the following criteria: libraries were selected only from academic institutions with a Carnegie classification of "Research University High" or "Very High," and only universities with health sciences graduate programs, including a college of medicine, were considered. From this relatively small set of universities, we contacted those with WSD tools in the authors' state and the five surrounding states until we received permission to collect data at six libraries: two with EDS, two with Primo, and two with Summon. Our own institution was included. Two implementations were tested for each product to control for different institutions enabling different features. By collecting data from more than one institution, it would be easier to determine which data were outliers and which were generalizable to all implementations of the products.

Relevance of search results

With the aim of determining the relevance of search results retrieved by each WSD tool, we created questions for 6 major health sciences disciplines: applied health sciences, dentistry, medicine, nursing, public health, and pharmacy. We created 3 reference questions per discipline. We ran searches on the 18 health sciences reference questions at the 6 research sites and thus collected data from a total of 108 search queries.

The reference questions originated from (a) real student and faculty questions posed to liaison librarians at the reference desk or over email, (b) recent University of Illinois at Chicago faculty publication topics, or (c) liaison librarian expertise. All searches were completed between January 14 and February 13, 2015.

We translated each reference question into a simple, short keyword search string (Appendix A, online only), representative of typical student search techniques [15–19]. Commonly used medical abbreviations taken from student and faculty reference questions, such as *CHF* and *EKG*, were entered as is to test each WSD tool's ability to interpret keywords and search alternate terms, similar to PubMed's Automatic Term Mapping feature. We entered each search query in the WSD tool and captured the first page of results, as studies have shown this is typically where students stop reading [15, 16, 20–22]. Since each search tool displayed a different number of records on the first

results page, and Summon's recently upgraded 2.0 platform inserts no page breaks at all, instead displaying all results on a single screen (a.k.a., "infinite scrolling"), we captured the first twenty results to represent the first "page" of results from each search.

We independently rated each captured reference and marked it relevant, irrelevant, duplicate, or non-scholarly; each of these four classifications was mutually exclusive. Any article considered at least possibly relevant, based on title or abstract, was marked as such. Relevance was judged based on the simple criterion of whether the article appeared to address the proposed research question. Full text was accessed when necessary to determine relevance.

Duplicate results were preserved in the data set because their presence on a results screen negatively impacts user experience. Counting them among search results was, therefore, of interest in determining each WSD tool's ability to deduplicate results and retrieve unique, relevant citations in response to a given user query.

Any disagreements in our ratings were discussed and reconciled. We compared and discussed ratings after we had evaluated the first 40 references, then after the first 10% (220 references). Cohen's kappa (κ) was 0.543 and 0.553, respectively, at these 2 points, indicating moderate observer agreement, where κ=1.000 indicates perfect observer agreement and κ=0 indicates actual observer agreement that does not differ from that which would occur by chance [23]. In total, 2,086 citations were evaluated, κ=0.725 (good observer agreement). The noticeable improvement in agreement indicates that our discussion helped to resolve inconsistencies in our respective evaluation methods.

MEDLINE retrieval and coverage

To compare each WSD tool's retrieval and coverage of MEDLINE content, we searched 1 research topic from each of the 6 health sciences disciplines in PubMed, using a combination of keywords and Medical Subject Headings (MeSH) (Appendix B, online only). These search strings favored precision in order to produce a highly relevant set of articles—a "gold standard"—against which to compare the WSD search results. In addition to our own assessment of the 2,086 citations collected during WSD searching, this gold standard set from

MEDLINE would serve as an external indicator of relevance for a smaller subset of the WSD results. That is, we used it to judge how many relevant MEDLINE results the WSD tool could retrieve with a simple keyword search.

For each of the six MEDLINE search queries, we cross-checked results with the first page of results from the corresponding WSD query; for example, we compared the search "tamsulosin ureteric stones expulsion" (RXQ1) in the WSD tool to "tamsulosin"[Supplementary Concept] AND "Ureteral Calculi/drug therapy"[MeSH] (PMQ6) in PubMed. We did not do this to compare the search effectiveness of the WSD tools with that of PubMed, as our search of the latter was a more precise search of a narrower index of sources and a direct comparison of the two would not be valid. Instead, the intended goal was to investigate whether one WSD tool was more successful than the others in retrieving relevant MEDLINE material.

To test each product's overall coverage of MEDLINE content in its central index, we captured the first 50 references from each of these 6 results sets (300 total citations) into an Excel spreadsheet. When collecting data at each research site, we searched for these 300 citations as single, known-item searches to determine whether or not they were indexed by the WSD tool. These MEDLINE citations were searched first by article title. Those not located by title were searched by multiple combinations of author names, journal titles, and keywords, until either we retrieved the citation or were confident that all search options had been exhausted and the citation could be marked as "not indexed." These searches were conducted with no other limits placed on the query, so as to access the entirety of the WSD tool's index.

Process

One notable feature of WSD tools that had to be controlled for was each institution's ability to customize its implementation to accommodate its collections and preferences. To control for such customizations, we took the following steps: first, we ran identical search queries on two distinct implementations of the same service, allowing us to compare and contrast results and note any differences that might be attributed to institutional customizations. Second, we determined that the inclusion of local collections and integrated library system (ILS) holdings in results would cause the widest variance in searches among implementations.

	EBSCO Discovery Service (EDS) 1	EDS2	PRIMO1	PRIMO2	SUMMON
Relevant/Included	195	216	167	169	192
Excluded†	165	144	161	156	168
Total	360	360	328	325	360

* "First page" defined as first 20 search results or entire results set when <20.
† Total excluded results comprise irrelevant, duplicate, and non-scholarly results.

Table 1
Relevance of first page* of results from 18 search queries

By limiting our queries to scholarly and peer-reviewed journal articles, we excluded these records, avoiding the issue of local holdings altogether. In cases where institutions offered their WSD search through both a general or multidisciplinary interface as well as through a custom search interface for medical disciplines (an option available from all three vendors), we selected the general or multidisciplinary profile in order to maintain consistency across research sites.

RESULTS

Only five of the six research sites are reported on here. At the time of data collection, Summon's 2.0 platform had been recently made available, but its original platform was still in place at many institutions that had not yet chosen to upgrade. One of the two Summon implementations tested for this study was running on what is now considered its legacy platform. The pending retirement of this version renders the corresponding data obsolete, and the legacy platform data have, therefore, been excluded from analysis. The second implementation of Summon tested had been upgraded to the newer 2.0 platform, so the corresponding data set remains. A final total of five sites were included for analysis: two EDS, two Primo, and one Summon. All references to "Summon" that follow below refer to its 2.0 platform.

Relevance of search results

All 3 WSD tools at the 5 locations returned between 50.0% and 60.0% relevant results (Table 1). The 2 implementations of EDS returned the highest number of relevant articles on the first results page (i.e., the first 20 retrieved items), followed closely by Summon. Primo returned a noticeably lower number of relevant results compared with the other 2 products, largely impacted by the frequency of duplicates among Primo results. While EDS and Summon demonstrated similar success rates in deduplication (between 96.0% and 97.0% unique citations on the first page of search results), both instances of Primo returned considerably fewer unique results: 84.9% and 82.9%, respectively. The remaining results were reappearances of previously displayed citations.

MEDLINE retrieval and coverage

All three WSD tools retrieved relevant literature that was not indexed in MEDLINE, as well as relevant literature that was indexed in MEDLINE but not retrieved by our precision searches (Table 2). This could have been caused by several factors, among them the use of different keywords between PubMed and the WSD tools or the use of MeSH in precision searches eliminating recent literature that was not yet indexed. EDS and Summon retrieved a larger number of relevant citations from non-MEDLINE source titles than did either implementation of Primo. One EDS implementation's keyword search results had the most results in common with our precision searches in MEDLINE (31), followed by Summon and the second EDS implementation (25).

The largest overlap between the WSD tools' central indexes and the set of 300 MEDLINE citations was observed with EDS (Table 3). We successfully retrieved all 300 citations in one EDS implementation, while in the second implementation of EDS, we retrieved 299 of the 300 citations. Summon and the Primo services returned a lower percentage of the references, neglecting to index between 8.0% (Summon) and 23.3% (PRIMO2) of the MEDLINE citations sought. The disparity in results between Primo implementations may be attributable to

	EDS1	EDS2	PRIMO1	PRIMO2	SUMMON
Total relevant	82	93	67	65	81
MEDLINE	70	78	64	62	66
Non-MEDLINE	12	15	3	3	15
Also retrieved in MEDLINE with authors' precision search	25	31	23	22	25

* PRIMO1 returned 105 results for the 6 evaluated search queries; PRIMO2, 101 results.

Table 2

Of 120* results, source of relevant items (MEDLINE vs. non-MEDLINE)

decisions made to activate or deactivate individual publications in the Primo Central index, based on institutional subscriptions.

Total number of search results

The total number of search results was noted for each search query. Primo consistently returned the fewest search results, with a median results set of 59 references in both implementations and an average of 1,668 (PRIMO1) and 1,379 (PRIMO2). Both Summon and EDS returned notably higher numbers: median result totals ranged from 1,356 (Summon) to 3,997 (EDS1), and averages ranged from 2,945 (Summon) to 9,141 (EDS1). In the case of EDS, deduplication occurs as the user scrolls through results, so the final number of unique results is fewer than the initial figure noted. However, it is safe to remark that, regardless of this feature, EDS consistently returned higher numbers of unique references than did Primo. Its initial results numbers, and possibly total unique results, were also higher than those of Summon.

Mapping of less frequent terms

The search "mandatory EKG preparticipation"—taken from a reference question on administering EKGs to athletes before participation in sporting events—returned the fewest results of any query

	EDS1	EDS2	PRIMO1	PRIMO2	SUMMON
Indexed	299	300	267	230	276
Not indexed	1	0	33	70	24
Total	300	300	300	300	300

Table 3

Web-scale discovery coverage of 300 MEDLINE citations

across all 6 tools. In EDS and Primo, the WSD tool searched the abbreviation EKG without substituting alternatives. The low number of results can presumably be attributed to the infrequent use of the term "EKG" in the literature, as compared with "ECG," "electrocardiogram," or "electrocardiography." When this query was run in Summon 2.0, the platform's Automated Query Expansion feature mapped the abbreviation to "electrocardiography," automatically searching this term in addition to "EKG." This produced 97 results, compared with 12 total results in the search on Summon's legacy platform, which did not employ the term mapping feature. In addition, this was the only query out of 108 total searches that returned an entirely relevant "first page" of results, that is, all 20 of the first 20 results were unique, relevant scholarly journal articles.

DISCUSSION

In all areas of evaluation, EDS results appear slightly superior to those of Primo and Summon. EDS returned the highest overall number of relevant results; indexed the highest total number of MEDLINE citations; and in six search queries, EDS both found the most relevant results from MEDLINE source titles and tied with Summon for the most relevant citations found in less-common, non-MEDLINE-indexed sources. However, these results reflect a small sample and were not exceptional to a degree that would definitively recommend EDS over the other two WSD products.

Comparison of results from isolated implementations of each product (i.e., EDS1 versus EDS2, PRIMO1 versus PRIMO2) reflect similar search results, despite the fact that librarians at every research site confirmed having customized their implementation of the product. This suggests that

results retrieved from the selected research sites would not likely differ from other product implementations. Numbers differed slightly between Primo MEDLINE indexing rates, as noted above. This divergence aside, the authors did not notice any institutional customizations that they believe to have significantly impacted the results of the study.

We had anticipated that each library's decision to include or exclude certain collections from the searchable index might substantially impact search results. However, these customizations either did not have a noticeable impact on search results or the same changes had been made at both research sites for each product, making institutional modifications of the WSD search index indiscernible from the "out-of-the-box" or standard version in any of the implementations examined. Given that searches were run at only two libraries per WSD tool, it is not possible to know with certainty that all implementations would produce similar results.

While the WSD products varied widely in the total number of search results retrieved, this detail is irrelevant to the majority of searchers, who do not look past the first or second page of results [15, 16, 20–22]. It is, therefore, more pertinent to consider the relevance and quality of the first page of search results when evaluating these products. In these aspects, a few notable observations emerge.

Relevance was observed multiple times to be a function of the topic itself, rather than of the WSD tool. For example, queries with extremely specific terms, such as drug names, were across the board more successful than those using more ambiguous terms. A topic's coverage in the literature also seemed to have a direct effect on the relevance of search results. When a topic was not as widely covered, first-page citations resulted from the WSD tool searching beyond abstract and metadata for search terms. On several occasions, an irrelevant article was retrieved because one of the search terms simply appeared in the full text, generating a "false positive" result.

The success of the search in Summon 2.0 using its Automated Query Expansion feature indicates a promising direction for all WSD vendors to investigate. Maximizing the controlled vocabularies of the databases that contribute content to WSD tools' indexes could help "interpret" simple keyword searches, in the same way that PubMed uses Automatic Term Mapping to enhance user queries. At the time the searches were conducted, Summon

was the only product of the three tested offering this feature, although EDS has since added controlled vocabulary enhancements to its search features.

As noted, Primo returned a significantly higher instance of duplicate citations, which negatively impacted the number of unique relevant citations counted for that tool. This deviation notwithstanding, none of the three WSD tools demonstrated overwhelmingly better or worse success than the others in returning relevant results. It would be helpful to have a benchmark to which the relevance numbers could be compared in order to place the success rate of 50%–60% relevant results in context. Unfortunately, due to fundamental differences between WSD search tools and traditional databases, it was not possible to use existing research on the relevance of traditional database search results to compare the two. A product designed to be searched differently by users must necessarily be evaluated differently. For example, in the process of designing the study, we observed multiple times that a short keyword search that delivers multiple relevant results in a WSD search might not retrieve a single record in PubMed. To judge one resource within the context of the other would be unfair.

In the same vein, relevance numbers would almost certainly have been higher if the study design had incorporated the use of product features such as Boolean operators, search limiters and facets, and field searching, all of which were available in all three WSD tools. However, to mimic typical student searching in the present study, we deliberately avoided advanced search constructions requiring a high level of skill and often intensive prior instruction, and is, therefore, irrelevant to the question at hand. The appropriate question to pose when evaluating a resource is not how well it functions under ideal conditions when operated by an expert user, but rather what can be expected when operated in the most likely of circumstances. Under this examining lens, achieving an average of 50%–60% relevance among the first 20 results of every search implies a strong chance that the typical, inexpert user would be able to find 1 or more satisfactory sources on the first page of results. Note that this is accomplished with a simple search of 3 to 6 keywords. While librarians always hope for the opportunity to provide instruction on advanced searching, the data resulting from this study indicate that students and faculty could successfully use WSD tools to start or supplement the research

process, in the event that searching instruction were either not available or not desired.

Perhaps the greatest evidence of WSD's value to health sciences researchers was demonstrated when all three tools unexpectedly retrieved, by simple keyword search, a considerable amount of relevant MEDLINE literature that had not appeared in our precision searches in PubMed. In this way, it could be argued that WSD tools may be more effective than PubMed/MEDLINE or other traditional databases in certain cases: chiefly for inexpert users, but also as a supplementary source for those conducting comprehensive literature reviews. The ease with which all three products uncovered literature that was missed with a highly skilled search of MEDLINE, as well as relevant literature from less-common sources that the average user might not think to search (or know how to access), illustrates the appeal of this search format and encourages the recommendation of WSD tools to library patrons. Results suggest that health sciences students with information needs ranging from simple to complex will be able to locate relevant content—from both MEDLINE and less-common sources—quickly, with a simple keyword search from the library home page.

It would have been ideal to include OCLC's WSD tool, WorldCat Local, in this study; however, it was excluded for two reasons. First, no libraries fitting the profile for inclusion in the study were subscribers to the product at the time the study took place, making real-world implementations unavailable for testing. Second, in light of the fact that future customers will adopt the new WorldCat Discovery service, as WorldCat Local is to be retired [24], we decided not to pursue evaluation of a product that would not be available to new customers. Instead, we hope to update the current study in the near future, after WorldCat Discovery has been implemented at major libraries in our region.

Limitations

The study's methodology was limited in that we evaluated only 18 search queries for relevance and searched only 300 MEDLINE citations in each WSD tool's index. The preference for a larger data set was balanced against the amount of data that we felt we could collect at a research site within a limited amount of time and analyze within a reasonable timeframe. Additionally, search queries were constructed by librarians, not taken directly from

actual search query logs, though every effort was made to imitate real-world student and faculty search technique and syntax, as we have observed in the literature and in practice.

It should be noted that we are not subject experts in the respective disciplines of each query that we evaluated. However, the high inter-rater agreement measured after all 2,086 citations had been evaluated indicates that our inclusion and exclusion criteria, along with reference to full text when necessary, are a reliable gauge of relevance.

While results from precision searches of MEDLINE are referred to here as a "gold standard" set of articles, they are but one example of a highly specific topical MEDLINE search. These searches were constructed by one person and were not vetted by any external authority. Only six topics were searched in MEDLINE; these data therefore may not be generalizable to other scenarios. Because the primary goal in constructing these searches was that of precision—to produce mostly, if not exclusively, relevant search results—the WSD tools' ability to retrieve MEDLINE literature not found in a "comparable" PubMed search may be overrepresented. A more valid comparison would be to construct PubMed search strings that are better balanced in precision and recall; comparing these results to those of keyword searches in WSD tools might reflect different numbers and would make an interesting subject for future research. Similarly, while excluding local holdings from WSD search results allowed for generalizability, including them may retrieve more uncommon (non-MEDLINE) sources in response to the average search query.

Future directions

For libraries considering adopting a WSD service and for early adopters now contemplating a change in products, there are points to consider beyond the quantitative factors measured here. Chief among these is usability, which has a substantial impact on the desirability of any resource. We observed several usability issues during the collection of data for our study, including a wide variance in compatibility with citation management tools, problems with "freezing" and error messages, and basic differences in interface and user options. These would have an obvious influence on the decision-making process when selecting a WSD tool, but we did not describe them here because

they were not the aim of our study. In addition to usability, other factors such as compatibility with a library's existing electronic resource management tools [25], cost, or manpower required for setup and ongoing maintenance may also ultimately influence this choice.

ACKNOWLEDGMENTS

REFERENCES

1. Griffiths JR, Brophy P. Student searching behavior and the web: use of academic resources and Google. Libr Trends. 2005 Spring;53(4):539–54.

2. Rowlands I, Nicholas D, Williams P, Huntington P, Fieldhouse M, Gunter B, Withey R, Jamali HR, Dobrowolski T, Tenopir C. The Google generation: the information behaviour of the researcher of the future. Aslib Proc. 2008 Jul 6;60(4):290–310. DOI: http://dx.doi.org/10.1108/00012530810887953.

3. Martin J. The information seeking behavior of undergraduate education majors: does library instruction play a role? Evid Based Libr Inf Pract. 2008 Dec 13;3(4):4–17.

4. He D, Wu D, Yue Z, Fu A, Vo KT. Undergraduate students' interaction with online information resources in their academic tasks. Aslib Proc. 2012;64(6):615–40. DOI: http://dx.doi.org/10.1108/00012531211281715.

5. Colon-Aguirre M, Fleming-May RA. "You just type in what you are looking for": undergraduates' use of library resources vs. Wikipedia. J Acad Libr. 2012 Nov;38(6):391–9. DOI: http://dx.doi.org/10.1016/j.acalib.2012.09.013.

6. Judd T, Kennedy G. Expediency-based practice? medical students' reliance on Google and Wikipedia for biomedical inquiries. Br J Educ Tech. 2011 Mar;42(2):351–60. DOI: http://dx.doi.org/10.1111/j.1467-8535.2009.01019.x.

7. Connaway LS, Dickey TJ, Radford ML. "If it is too inconvenient I'm not going after it": convenience as a critical factor in information-seeking behaviors. Libr Inf Sci Res. 2011 Jul;33(3):179–90. DOI: http://dx.doi.org/10.1016/j.lisr.2010.12.002.

8. Cothran T. Google Scholar acceptance and use among graduate students: a quantitative study. Libr Inf Sci Res. 2011 Oct;33(4):293–301. DOI: http://dx.doi.org/10.1016/j.lisr.2011.02.001.

9. De Groote SL, Shultz M, Blecic DD. Information-seeking behavior and the use of online resources: a snapshot of current health sciences faculty. J Med Libr Assoc. 2014 Jul;102(3):169–76. DOI: http://dx.doi.org/10.3163/1536-5050.102.3.006.

10. Olle C, Borrego A. A qualitative study of the impact of electronic journals on scholarly information behavior. Libr Inf Sci Res. 2010 Jul;32(3):221–8. DOI: http://dx.doi.org/10.1016/j.lisr.2010.02.002.

11. Thompson JL, Obrig KS, Abate LE. Web-scale discovery in an academic health sciences library: development and implementation of the EBSCO Discovery Service. Med Ref Serv Q. 2013 Feb 8;32(1):26–41. DOI: http://dx.doi.org/10.1080/02763869.2013.749111.

12. Pinkas MM, Del Baglivo M, Klein IR, Brown E, Harris R, Gerhart B. Selecting and implementing a discovery tool: the University of Maryland Health Sciences and Human Services Library experience. J Elec Resour Med Libr. 2014 Jan;11(1):1–12. DOI: http://dx.doi.org/10.1080/15424065.2013.876574.

13. Crook L. WorldCat Local and the clinician: is discovery improved? Endnotes. 2011 Nov;2(1):1–11.

14. Ketterman E, Inman ME. Discovery tool vs. PubMed: a health sciences literature comparison analysis. J Elec Resour Med Libr. 2014 Jul;11(3):115–23. DOI: http://dx.doi.org/10.1080/15424065.2014.938999.

15. Asher AD, Duke LM, Wilson S. Paths of discovery: comparing the search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and conventional library resources. Coll Res Libr. 2013 Sep;74(5):464–88. DOI: http://dx.doi.org/10.5860/crl-374.

16. Asher AD, Duke LM. Searching for answers: student research behavior at Illinois Wesleyan University. In: Duke LM, Asher AD, eds. College libraries and student culture: what we now know. Chicago, IL: ALA Editions; 2011. p. 71–86.

17. Holman L. Millennial students' mental models of search: implications for academic librarians and database developers. J Acad Libr. 2011 Jan;37(1):19–27. DOI: http://dx.doi.org/10.1016/j.acalib.2010.10.003.

18. Mischo WH, Schlembach MC, Bishoff J, German EM. User search activities within an academic library gateway: implications for web-scale discovery systems. In: Popp MP, Dallis D, eds. Planning and implementing resource discovery tools in academic libraries. Hershey, PA: IGI Global; 2012. p. 153–73.

19. Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. J Am Med Inform Assoc. 2007 Mar 1;14(2):212–20. DOI: http://dx.doi.org/10.1197/jamia.M2191.

20. Majors R. Comparative user experiences of next-generation catalogue interfaces. Libr Trends. 2012 Summer;61(1):186–207. DOI: http://dx.doi.org/10.1353/lib.2012.0029.

21. Bar-Ilan J, Keenoy K, Levene M, Yaari E. Presentation bias is significant in determining user preference for search results—a user study. J Am Soc Inf Sci. 2009 Jan; 60(1):135–49. DOI: http://dx.doi.org/10.1002/asi.20941.

22. Jansen BJ, Spink A, Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the web. Inf Process Manag. 2000 Mar 1;36(2):207–27. DOI: http://dx.doi.org/10.1016/S0306-4573(99)00056-4.

23. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.

24. WorldCat Discovery migration [Internet]. Dublin, OH: OCLC Online Computer Library Center; 2015 [cited 3 Jun 2015]. <http://www.oclc.org/worldcat-discovery/migration.en.html>.

25. Djenno M, Insua G, Gregory GM, Brantley JS. Discovery usability: comparing two discovery systems at one academic library. J Web Libr. 2014 Jul 16;8(3):263–85. DOI: http://dx.doi.org/10.1080/19322909.2014.933690.

AUTHORS' AFFILIATIONS

Rosie Hanneke, MLS, rhanneke@uic.edu, Assistant Professor and Assistant Information Services Librarian, Library of the Health Sciences–Chicago, University of Illinois at Chicago, 1750 West Polk Street, Chicago, IL 60612; **Kelly K. O'Brien, MLIS**†, kkobrien90@gmail.com, Assistant Professor and Regional Assistant Librarian, Crawford Library of the Health Sciences–Rockford, University of Illinois at Chicago, 1601 Parkview Avenue, Rockford, IL 61107

† Current contact information: 18120 Wildwood Springs Parkway, Spring Lake, MI 49456.