

Discovery Tool vs. PubMed: A Health Sciences Literature Comparison Analysis

The XX University libraries, though administratively separate, jointly subscribe to and collaborate on enhancements for a shared instance of the Summon Discovery Service. Based on usage, enhancements to the discovery tool over the past few years, and the perceived ease of searching in Summon, health sciences librarians have questioned whether Summon could now be considered a legitimate competitor to PubMed. This paper includes results of a citation comparison between the two databases and the conclusion that Summon produces quality results, but should not be considered an adequate replacement information source for a subject-specific database like PubMed.

KEYWORDS discovery, Summon, PubMed, citation analysis, comparison analysis, health sciences libraries

INTRODUCTION

Discovery services are becoming a staple resource in academic libraries. A likely reason is that many patrons desire a “Google-like” search experience which is easy to navigate and can search multiple formats, material types, and locations at once. XX University librarians recognized this need and conducted a thorough analysis of available discovery services in 2010¹. The libraries decided to begin a subscription to ProQuest’s Summon Discovery System and, since implementation in August of that year, patron use of Summon has grown immensely (40,102

searches conducted in 9,037 visits in the typically heavy use month of October 2010, compared to 142,684 searches conducted in 28,003 visits in October 2013). We have also seen a steady decrease in overall database searches per year at the XX Libraries. From 2008-09 through 2012-13, searches decreased from 4.09 million to 3.36 million. We can draw an indirect correlation between the uptick in Summon searches and the decrease in overall searches. These usage statistics have informed decisions by XX Librarians to promote Summon in areas of prominence on our websites and increase its visibility in our course management system.

As the Summon discovery service's popularity has grown at our campus, and as discovery services in general continue to improve², the authors have questioned whether Summon may be considered a legitimate competitor to specialized health sciences information sources. In this paper, we attempt to answer the question by presenting findings of a citation comparison study between PubMed and Summon.

LITERATURE SCAN AND BACKGROUND

Marshall Breeding states, "It's imperative that... discovery services deliver search results consistent with library values, which differ considerably from what applies in the commercial Web".² As such, discovery services in a health sciences setting should become competitive with vetted article and indexing services. A literature search was conducted to see if there is published evidence comparing discovery tools to major abstracting and indexing databases, particularly in the health sciences, but little to none was found. An excellent article on implementation of the EBSCO Discovery Service at Himmelfarb Health Sciences Library

indicated that test searchers, “said their least favorite thing about EDS was the inability to get a good set of results, or to get a set of results that were consistently on target”, and, “...relevancy ranking needed to be improved for EDS to be successful”.³ Another excellent article by Baykoucheva compared drug information retrieval in well-known databases frequently subscribed to by health sciences libraries; unfortunately, none of those databases was categorized as a discovery service.⁴

Finally, Way mentions in his 2010 article that after implementation of Summon at Grand Valley State, “An examination of usage statistics showed a dramatic decrease in the use of traditional abstracting and indexing databases and an equally dramatic increase in the use of full-text resources from full-text database and online journal collections. The author concludes that the increase in full-text use is linked to the implementation of a web-scale discovery tool”.⁵ XX University has tracked trends in our usage post-Summon implementation and also seen an increase across the board. Way’s evidence suggests that there is a correlation between discovery implementation and patron use of library resources, which further encourages the need to explore whether a discovery service can provide comparable results to that of a database like PubMed.

METHODS

The first step in the research process was to choose which database should be compared to Summon. Multiple databases were considered, including CINAHL and OVID Medline, but ultimately PubMed was selected. PubMed seemed the logical choice because of the accessibility of the database to any researcher with internet access, and also because of the

breadth and depth of indexing on multiple health sciences topics. Summon's discovery service utilizes its own unified index

(<http://www.serialssolutions.com/en/services/summon/content#singleunified>) and PubMed

relies on the MeSH vocabulary

(<http://www.nlm.nih.gov/bsd/disted/meshtutorial/searchingpubmedusingmeshtags/>).

Next was to decide what topics to compare in each database. The Allied Health liaison librarian was familiar with typical searches requested of her by faculty and students, so we limited to three topics most frequently requested for search help: medical informatics, health information management, and electronic health records. We also did a preliminary search for these topics in both Summon and PubMed and knew we'd have a large sample of search results for our comparison process. The searches were conducted without using quotations around the keyword terms.

Following this came the decision of what data to pull from the databases for purposes of comparison. After reading other comparison studies^{3,4} and reviewing search result capabilities of Summon and PubMed, we determined the following points of comparison: total records retrieved, total records for the last three years, and top twenty journals by number of records retrieved in each database for each term. The total number of records retrieved would indicate the volume of data in each database, the total records for the last three years would indicate the amount of current data, and the top twenty journals would indicate common sources where data is being pulled and the relative quality of the source.

Collecting and organizing the data came next. The authors ran the search and recorded the date and number of total retrieved search records for each search topic in an Excel file. The search results were then limited to journal article and to the date range of 2010-2012, to contain three full years' worth of results. These results were then recorded in the same Excel file.

RefWorks bibliographic management software was utilized to determine the top journals for each search. The authors created a new RefWorks account and then created a folder for PubMed and a folder for Summon for each of the three search topics. In Summon, twenty-five citations were saved to the Save Items folder at a time due to technical limitations, and then exported to the RefWorks folder. This step was repeated until a total of 500 citations were moved into RefWorks.

PubMed proved to be slightly easier. The 500 total citations were added to the Clipboard feature and then exported 200 at a time to RefWorks via the special instructions required by RefWorks for the PubMed database. Once all the citations for all searches were added, each folder was saved as a Tab Delimited file and saved in that format. That file was then saved as plain text, then opened as an Excel file. At that point every citation was available to organize by the well-known sorting features in Excel. Using Excel, the citations for each of the three search terms were organized by total number of journal records, in both PubMed and Summon. The top journals with the highest number of records were then recorded.

RESULTS

The total number of retrieved records for Summon greatly outnumbered the results from PubMed. For the phrase health information management, Summon had over 4 million results and PubMed had 38,505 results. The term electronic health record retrieved a little over 574,000 hits and returned 13,396 in PubMed. Medical Informatics produced more results in PubMed with 304,431 than in Summon which contained 173,403.

When narrowing the results to the last three years, the phrase health information management contained close to 836,000 results within Summon and 8,000 results in PubMed. Electronic health record had a large difference, with Summon at 103,850 and PubMed 6,304. PubMed produced more results for medical informatics with 69,020 than Summon which had 52,093.

Based on the number of records in Summon the top three journals for the term Health Information Management were *Health Information Management Journal*, *Journal of AHIMA*, and *Perspectives in Health Information Management* (see Table 1a). In PubMed the top three journals were *Studies in Health Technology and Informatics*, *PloS One*, and *BMC Health Services Research* (See Table 1b).

INSERT TABLE 1a HERE

INSERT TABLE 1b HERE

The phrase electronic health record produced *The New England Journal of Medicine*, *Health Affairs*, and *Journal of the American Medical Informatics Association* as the top journals within Summon (see Table 2a). Within PubMed, *Studies of Health Technology and Informatics*,

Clinical Medicine & Research, and *Journal of the American Medical Informatics Association* were the top three journals (see Table 2b).

INSERT TABLE 2a HERE

INSERT TABLE 2b HERE

In Summon, the phrase Medical Informatics contained the following as the top three journals: *Information Technology Newsweekly*, *Health & Medicine Week*, and *Journal of the American Medical Informatics Association* (see Table 3a). PubMed produced *Conference proceedings: ...Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *Health Devices*, and *Computational and Mathematical Methods in Medicine* for its' top journals (see Table 3b).

INSERT TABLE 3a HERE

INSERT TABLE 3b HERE

CONCLUSION

Not surprisingly, Summon produced a much larger return of search hits than PubMed for two of the search topics in both initial total records retrieved and when limiting the searches to the last three years. This is probably due to vast amount of literature that is indexed within Summon and its full text searching ability. Interestingly, the term medical

informatics generated greater initial results and a higher number of records for the last three years.

The journal *Studies in Health Technology and Informatics* populated a large number of records across the board with the specified search terms in both databases. Another frequently occurring title included *Journal of the Medical Informatics Association*.

As expected, there are limitations with this study. The content in each database varies. PubMed searches mostly MEDLINE content, which consists of roughly 5,000 journals that meet the criteria set forth by the National Library of Medicine.⁶ Summon indexes journal, book, and other, "...content at the record level (full text and/or metadata) from information provided directly by publishers and content providers and not at a database level. Therefore, coverage of databases and packages is determined differently based on whether a publisher or provider of a database participates in the Summon service".⁷ (This difference in indexing practices may reflect the differences between non-profit and for-profit vendors, and is also why the citation analysis required the extra limit to journal articles for searches run in Summon.) Also, data comparison between total number of journal records for search terms is dependent on indexing coverage with the Summon and PubMed.

The data from this study can be utilized when trying to determine appropriate journal coverage for the growing field of health information management and medical informatics. The large amount of results produced in PubMed and Summon from basic keyword searching can serve as a reminder to encourage users of library databases in the use of proper searching techniques beyond basic keyword searching to aid in producing relevant search results.

A discovery tool, such as Summon, is useful for library patrons and staff when used properly. The X Health Sciences Library will continue to promote it as a place to start a search to find general information, for students looking to find quick yet reliable information for class papers, and for researchers who have exhausted known or familiar resources and would like to try one last source to make sure no information has been overlooked. There will also always be a place for subject specific databases like PubMed. The quality and specificity of information in a trusted source like PubMed continues to be worth the time investment in learning how to conduct a quality search in this database. Both Summon and PubMed should be considered valid tools for conducting research.

REFERENCES

1. Ketterman, E., Megan E. Besaw, and Michael Tucker. "Rethinking How Patrons Discover Information: Implementing a Discovery Tool." Presentation at the Medical Library Association Annual Meeting, Minneapolis, MN, May 17, 2011. Available from <http://thescholarship.ecu.edu/handle/10342/3506>. Accessed December 2, 2013.
2. Breeding, Marshall. "Smarter Libraries Through Technology: Discovery Services Improving." *Smart Libraries* 33, no. 5 (2013): 1-2.
3. Thompson, JoLinda L., Kathe S. Obrig, and Laura E. Abate. "Web-Scale Discovery in an Academic Health Sciences Library: Development and Implementation of the EBSCO Discovery Service." *Medical Reference Services Quarterly*, 32, no. 1 (2013): 26-41.

4. Baykoucheva, Svetla. "Selecting a Database for Drug Literature Retrieval: A Comparison of MEDLINE, Scopus, and Web of Science." *Science & Technology Libraries* 29, (2010): 276-288.
5. Way, Doug. "The Impact of Web-Scale Discovery on the Use of the Library Collection." *Serials Review* 36, no. 4 (2010): 214-220.
6. "Serials Solutions| Resources Solutions."
Available: <http://www.serialssolutions.com/en/resources/detail/summon-key-databases-and-packages-full-text>. Accessed: February 20, 2014.
7. "NIH: US National Library of Medicine. Fact Sheet: MEDLINE Journal Selection."
Available: <http://www.nlm.nih.gov/pubs/factsheets/jsel.html> Accessed: February 20, 2014.