# Imperfect tools: Google Scholar vs. traditional commercial library databases

Julie Arendt

Submitted for publication in *Against the Grain*

27 February 2008

Revised

17 March 2008

Like every other resource that a library might offer, Google Scholar has strengths and limitations. Instead of rejecting Google Scholar because it does not do everything that the library or librarians do, Google Scholar should be accepted or rejected based on how well it assists in a particular step in information seeking. That step traditionally has been assisted by indexing and abstracting resources. In some circumstances Google Scholar is a better tool than the indexing and abstracting resources; in other circumstances it is not. This article examines the strengths and weaknesses of Google Scholar compared to subscription indexing and abstracting databases. It critiques college and university libraries' continued use of subscription databases that fail to provide a clear advantage over Google Scholar.

When Google Scholar was introduced, it initially met with some praise and a fair amount of criticism from the library world. Both the praise and criticism generally were deserved. Unfortunately, early responses sometimes compared Google Scholar to the library as a whole[1] or to an idealized vision of library databases[2] rather than to the real, imperfect indexing and abstracting databases offered through the library. Some of the faults that early commentators found in Google Scholar included lack of a controlled vocabulary, lack of authority control, incomplete or uneven coverage depending on discipline, and time lags between publication and appearance in the database. These same faults could be pointed out for Web of Science, a venerable subscription database. Another criticism of Google Scholar was that its definition of “scholarly” includes materials that have not undergone peer review, so it may lead users to this unvetted material. Again, this criticism also could be leveled against a subscription database. Book reviews, editorials and commentaries regularly appear in search results from Academic Search Premier, even when the search is limited to scholarly (peer reviewed) journals. Instead of comparing Google Scholar to the ideal resource, a fairer comparison would be to actual subscription databases.

Some evaluations have explored whether a subscription database produces better results than Google Scholar. When librarians conduct test searches using advanced search features in library databases, they get somewhat better results with the database than with Google Scholar.[3-5] When college students, rather than

librarians, conduct the searches, the advantage for the subscription database evaporates. The sources students find from Google Scholar are as good as or better than those found through the library's databases.[6,7] For these novice users, often subscription databases do not provide a clear advantage over Google Scholar.

Librarians may be able to use controlled vocabularies to produce more precise results from a database than from Google Scholar or to find special materials that could not be found through Google Scholar, but library patrons are not librarians. Simply having a controlled vocabulary or special materials is not good enough for a novice user. If users cannot figure out the controlled vocabulary or find the special materials, they cannot experience these supposed advantages. For there to be a clear advantage of a subscription database over Google Scholar, novice users should be able complete their work more easily with the subscription database than they can with Google Scholar. Many subscription databases provide a clear advantage by simplifying access to special materials or by leveraging their controlled vocabularies. The interface designs that highlight subject terms next to results sets, such as those in EBSCOhost and Engineering Village, should be commended for their effort to guide novices to controlled vocabularies without interrupting users' searches. Some databases and interfaces simplify users' work in other ways. For example, Web of Knowledge provides citation assistance through EndNote Web, and full-text resources like JSTOR provide easy access to complete documents.

The databases and interfaces that do not deliver such clear advantages deserve scrutiny. In the past, scrutiny may have led librarians to complain to the vendor about a poor interface. When the alternatives were not too expensive, libraries would switch databases or interfaces. Now that a free alternative is available, this scrutiny can go one step further to switching to Google Scholar instead.

It has been argued that the subscription database is better than Google Scholar after a user learns how to use it.[8] For most students, especially undergraduates, this amount of database knowledge is unrealistic. Often the end users of the library's databases have not had any formal training.[9] If they receive training, it often is a one-time guest lecture by a librarian or informal instruction at the reference desk. Because one-shot

instruction sessions and brief instruction during reference encounters are the norm, there rarely is time for most users to thoroughly learn how to use a database. Typically there is only enough time to transform a complete novice database user into a slightly-less-than-complete novice. Even when there is more time, the time spent teaching a database reduces the time available to teach other important skills.

The Information Literacy Competency Standards for Higher Education from the Association of College and Research Libraries lists five capacities of an information literate person.[10] Only one of those capacities deals directly with searching techniques. It is not best use of time to teach a subscription database when the database has no clear advantages over Google Scholar. That limited time could be spent on other important knowledge such as information literacy.

As suggested by Diane Zabel, perhaps it would be better to for librarians to have regular, ongoing collaboration with faculty to integrate information literacy throughout students' disciplinary studies.[11] Perhaps it would be better to teach the broader information literacy concepts in a separate, mandatory course and to use one-shot instruction sessions for discipline-specific bibliographic and database instruction.[12] In colleges and universities that manage to successfully implement either model, librarians would have the luxury of approaching reference and one-shot instruction sessions with the knowledge that students will cover the other important ideas somewhere else. I do not have that luxury, and many of my colleagues at other institutions also work without that luxury. I go to classes where students' exposure to information literacy is as varied as the courses and instructors they have experienced up to that point. I am not the first to suggest that in a world with Google Scholar, it is time to move away from teaching the mechanics of searching databases to teaching more of the whole of information seeking.[13] I try to approach these teaching opportunities with two questions, "What are the most important things for them to learn from my presentation?" and, "What can I teach them that will help them the most on their work for this course?" My answers to these questions are always more than I can fit into a fifty-minute session. I have to jettison the material that is less essential.

If the best reason I can find for teaching a particular subscription database to undergraduates is simply to expose them to the database of a particular discipline, it is a topic that I consider less essential. After they graduate, most students will no longer be affiliated with a university and will no longer have easy access to university-level subscription databases. Although it may be possible for graduates to travel to the nearest public university library or to find a way to purchase short-term access to a database, the time, effort, and expense involved are substantial barriers that should not be ignored. Doesn't it make sense to introduce students to appropriate free resources rather than expecting them find a way to get access to subscription resources? In many cases, the appropriate free resource is Google Scholar, although it could be ERIC, PubMed, AGRICOLA, or another conventional library resource that does not require payment. Even for graduate students, where familiarity with the most important databases in their field should be a part of students' education, Google Scholar has value. Many graduate students will go on to be faculty at teaching colleges. Even though they will still conduct research, they will not have the same library resources. At colleges with small budgets, the premier database for a discipline may be too expensive. As Yvonne Jones described, alternatives for faculty in this situation can be to search multiple subscription databases to get about half the coverage of the premier database or to search Google Scholar to get about half the coverage of the premier database.[14] With those options, searching Google Scholar is a reasonable choice.

Another reason to teach a subscription database is to present general tactics for database searching. Students can apply skills, such as selecting keywords, leveraging controlled vocabularies, using Boolean logic, and broadening or narrowing a search, to other situations. Some of these skills are possible to teach within Google Scholar, but some of these tools are unavailable in Google Scholar. The trouble is that the skills are taught at the same time as the arbitrary mechanics of where to click to get a particular database to work. Even for databases with the best interfaces, it takes several steps of navigation through the library Web site just to get to the database. When the database requires several additional clicks, I wonder if the core message will get buried in the procedures. Every minute spent teaching these mechanics is a minute less

spent on teaching general concepts in database searching.

Sometimes those extra minutes on database navigation are worthwhile. In some subjects, the appropriate disciplinary database may produce better results with less effort for students despite the extra navigation. Google Scholar is weaker in the social sciences and humanities than it is in the sciences.[15] Some disciplinary databases have useful search features that are unavailable in Google Scholar. When students know how to use these features, they appreciate them.[16] On the other hand, when the interface is hard to use and the advantages over Google Scholar are small, those extra minutes spent on navigation pale in comparison to the other things that could be taught.

Navigational details will quickly become obsolete, and other things which can serve students for much longer could be taught instead. Although an hour is too short to build "an intellectual framework for understanding, finding, evaluating, and using information,"[17] it is enough time to encourage students to think critically about the information they find and to think about the legal and social issues involved. Knowing why it matters that there are differences between a white paper, newspaper, magazine, or scholarly journal article, or some other type of source will serve students a lot longer than knowing where to click in a particular database interface to find its advanced search tools. Knowing why it is important to cite sources should be useful after graduation, unlike knowing where to click on the college library's Web site.

One objection that may be raised to teaching Google Scholar is that it will direct students away from subscription databases that the library spends so much to have. In writing this article, I felt apprehensive that I would be accused of disloyalty to the library and to the profession for directing students to a non-library resource like Google Scholar. I believe that presenting arcane or confusing databases with no clear advantage over Google Scholar will do more to drive users away than directing them to Google Scholar will. As part of the library profession, my goal is guide patrons toward what I believe are the best resources for their research. Sometimes those resources are within the library, and sometimes they are not. From the student's perspective, the value of the database is not in the dollars that the library paid for it but in the usefulness of

the information it provides. For them, the database that can lead to the best resources for the task with the least effort is the one that is worth the most. It does patrons a disservice to direct them to library-paid resources out of tradition or because they are expensive.

From the library's perspective, it has to pay for a variety of resources because the members of the community it serves have a variety of needs. The best resources for a freshman composition usually are different from the best resources for an honors thesis, and the best resources for an honors thesis may be different from the best resources for a faculty member's research. Even with these competing needs, there are limits to what every library can afford. Libraries perennially have had the problem that more information exists than any one library can afford to possess. At one time, a library's indexing and abstracting databases were vital for patrons to discover information. Libraries willingly sacrificed the ability to possess some materials to pay for indexes and abstracts. Librarians knew that the information hidden in journals and books would stay hidden if their contents were too hard to find. Today libraries still deal with the problem that there is more information than any library can afford. Because Google Scholar offers an alternative, the subscription indexing and abstracting database is no longer the vital tool for discovery it once was. Money not spent on a hard-to-use indexing and abstracting database can instead be spent to supply the information itself. For some indexing and abstracting databases, it is time to reexamine their value.

I am not arguing that subscription indexing and abstracting databases should all be abandoned, but they should be compared with the alternatives. Two basic questions worth considering when evaluating subscription and instruction choices: 1. How is this database better than Google Scholar? 2. Assuming the subscription product is better, is the advantage worth the money and resources that would have to be devoted to it? These questions remain valid, but the answers will depend on the library patrons, budget and philosophy.

Endnotes

[1] Cathcart, Rachael, and Amanda Roberts. "Evaluating Google Scholar as a Tool for Information

Literacy." *Internet Reference Services Quarterly* 10, no. 3/4 (2005): 167-76.

[2] York, Maurice C. "Calling the Scholars Home: Google Scholar as a Tool for Rediscovering the Academic Library." *Internet Reference Services Quarterly* 10, no. 3/4 (2005): 117-33.

[3] Shultz, Mary. "Comparing Test Searches in PubMed and Google Scholar." *Journal of the Medical Library Association* 95, no. 4 (2007): 442-45.

[4] Levine-Clark, Michael, and Joseph Kraus. "Finding Chemistry Information Using Google Scholar: A Comparison with Chemical Abstracts Service." *Science & Technology Libraries* 27, no. 4 (2007): 3-17.

[5] Callicott, Burton, and Debbie Vaughn. "Google Scholar vs. Library Scholar: Testing the Performance of Schoogle." *Internet Reference Services Quarterly* 10, no. 3/4 (2005).

[6] Helms-Park, Rena, Pavlina Radia, and Paul Stapleton. "A Preliminary Assessment of Google Scholar as a Source of EAP Students' Research Materials." *Internet and Higher Education* 10, no. 1 (2007): 65-76.

[7] Jung, Seikyung, Jonathan L. Herlocker, Janet Webster, Margaret Mellinger, and Jeremy Frumkin. "Libraryfind: System Design and Usability Testing of Academic Metasearch System." *Journal of the American Society for Information Science and Technology* 59, no. 3 (2008): 375-89.

[8] Callicott, & Vaughn.

[9] Elsevier. "Usability Drives Value of Bibliographic Databases." (2003): http://web.archive.org/web/20040716112223/http://www.info.sciencedirect.com/content_coverage/databases/sd_bdwhitepaper.pdf.

[10] Association of College & Research Libraries. "Information Literacy Competency Standards for Higher Education." (2007): http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.cfm.

[11] Zabel, Diane. "A Reaction to ‘Information Literacy and Higher Education’." *Journal of Academic Librarianship* 30, no. 1 (2004): 17-21.

[12] Owusu-Ansah, Edward K. "Information Literacy and Higher Education: Placing the Academic Library in the Center of a Comprehensive Solution." *Journal of Academic Librarianship* 30, no. 1 (2004): 3-16.

[13] Williams, Genevieve. "Unclear on the Context: Refocusing on Information Literacy's Evaluative Component in the Age of Google." *Library Philosophy and Practice* 2007, June (2007).

[14] Jones, D. Yvonne. "Biology Article Retrieval from Various Databases: Making Good Choices with Limited Resources." *Issues in Science and Technology Librarianship* 44 (2005).

[15] Neuhaus, Chris, Ellen Neuhaus, Alan Asher, and Clint Wrede. "The Depth and Breadth of Google Scholar: An Empirical Study." *portal: Libraries and the Academy* 6, no. 2 (2006): 127-41.

[16] Jung, Herlocker, Webster, Mellinger, & Frumkin.

[17] Association of College & Research Libraries.