A COMPARATIVE ANALYSIS OF PRIMARY LITERATURE DATABASES FOR FRESHWATER BIOLOGY

Barry N. Brown
Associate Professor, Head, Access & Collection Services
Mansfield Library, The University of Montana
32 Campus Drive #9936
Missoula, MT 59812-9936.
Barry.Brown@umontana.edu

Abstract:
There are a bewildering array of databases currently available for literature searches. Major, traditional indexes to the primary literature of freshwater biology include: Aquatic Sciences and Fisheries Abstracts, Biological Abstracts, Biological and Agricultural Index, CAB Abstracts, CSA Biological Sciences, Fish and Fisheries Worldwide, Web of Science, and Zoological Record. New indexes and search engines have recently appeared; notably Scirus, Scopus, and Google Scholar. Large electronic journal packages that can function as indexes and provide significant access to the primary literature of freshwater biology include: BioOne, Blackwell Synergy, JSTOR, SpringerLink, and Elsevier ScienceDirect. All of these electronic databases (along with an interdisciplinary, undergraduate oriented, full text database, EBSCO Academic Search Premier) were compared and ranked using quantitative and qualitative criteria and search results based on the various chapter topics and content within *"Current and Selected Bibliographies on Benthic Biology"*; which is published annually by the North American Benthological Society. Overall the top five databases for freshwater biology, based on the bibliography content and criteria examined in this analysis, were: Google Scholar, Web of Science, Scopus, Zoological Record, and Biological Abstracts.

Keywords:
freshwater biology, databases, journals, primary literature, indexes, nabs, benthological society, bibliography

The University of Montana is a medium sized (almost 12,000 FTE), research oriented, public university. The University of Montana, Maureen and Mike Mansfield Library is a medium sized (1.4 million volumes), academic library. The Mansfield Library, like many similar sized academic libraries, has greatly increased access to databases (electronic indexes and ejournal packages) over the last few years. Access to journals has changed over the last 7 years from about 4,500 print subscriptions available in the library building

to over 20,000 ejournals available 24 hours a day, anywhere, to University of Montana student, staff, and faculty members. This is the result, to a large extent, of consortial, long term, contracts. Library users appear to be in a golden age of access.

The Mansfield Library, like many other medium sized university libraries, has hundreds of databases and arranges them on web pages alphabetically, and in department/ subject categories subdivided with two different listings of "try these resources first" and then "related resources" with titles and brief descriptions. Someone looking for an appropriate database for searching freshwater biology literature is faced with a bewildering array of choices. Major, traditional indexes to the primary literature of freshwater biology include: Aquatic Sciences and Fisheries Abstracts (ASFA), Biological Abstracts, Biological and Agricultural Index, CAB Abstracts, CSA Biological Sciences (which includes Aquatic Sciences and Fisheries Abstracts), Fish and Fisheries Worldwide, Web of Science, and Zoological Record. Some newer indexes and search engines that have recently appeared and are available include: Scirus, Scopus, and Google Scholar. Large electronic journal packages, listed amongst the databases, that can function as indexes (and some of which provide keyword searching of article full text) and provide significant access to the primary/ journal literature of freshwater biology include: BioOne, Blackwell Synergy, JSTOR, SpringerLink, and Elsevier ScienceDirect. Finally, there are number of interdisciplinary, undergraduate oriented, full text databases such as EBSCO Academic Search Premier which are presented as good places to start any literature search.

With so many database choices the obvious question arises of whether some databases are better than others for the general subject of freshwater biology or is one as good as another? Given some search term flexibility they certainly all will find something on most freshwater biology topics. One might suspect that many library users, and even librarians, select databases for literature searches based on past experience, familiarity, habit, availability, and/or the name and brief description of the database, or its web page listing order. Few users are likely to rigorously compare databases using standard criteria and evaluate results to determine the best information resource for a particular topic or discipline. And new databases may have a harder time getting used or reviewed adequately.

Surprisingly, there are few published studies in the library literature on database comparisons for particular disciplines using objective, content criteria. Several studies have looked at the overlap of coverage between selected, traditional science indexes (Bearman and Kunsberger 1977; Poyer 1984; Chisman 1989; Hughes 2001). Fewer studies have used specific criteria and evaluated indexes (Jatkevicius 2000; Parker 2005). The overall conclusion of most science index comparison studies is that there is a maximum of 60 to 70% overlap between indexes, and researchers should use multiple indexes for literature searches. Parker (2005) notes that Web of Science "remains a perennial favorite of scientists" and was included in her study "solely to prove that it should not serve as an ultimate resource for marine scientists". There are a number of reviews comparing Web of Science and Scopus (e.g. Deis and Goodman 2005; Dess 2006) which often can be boiled down to a recommendation of "keep Web of Science

and buy Scopus if you can afford to". And there are now many reviews of Google Scholar (e.g. Jacso 2005) typically pointing out its limitations (unknown and incomplete content); DeGuire (2006) states that "…Google Scholar will never be able to replace abstract databases…". In contrast a report by OCLC (2005) indicated that, based on an extensive survey, the vast majority of information consumers begin their information searches with search engines not library web sites, and that "….search engines deliver better quality and quantity of information than librarian-assisted searching and at greater speed….".

A survey of the top 5 databases for freshwater biology, was administered at the International Association of Aquatic and Marine Science Libraries and Information Centers (IAMSLIC) Conference, October 2006, Portland (Appendix 1). A survey sheet was distributed with a list of databases and librarians were asked to rank 1 through 5 (with 1 being the best). Based on 33 individual responses the top 5 databases for freshwater biology, as perceived by IAMSLIC attendees are listed below.

Top Databases for Freshwater Biology as Identified by IAMSLIC Conference Attendees:
1st) ASFA
2nd) Biological Abstracts
3rd) Web of Science
4th) Scopus
5th) Google Scholar

What content source should be used for an objective comparison and ranking of freshwater biology databases? The North American Benthological Society (NABS), founded in 1953 and international in membership, is arguably the premier society for scientists engaged in freshwater ecosystem science research. There are several publications put out by NABS including an annual bibliography (Current and Selected Bibliographies on Benthic Biology). The 2004 NABS Bibliography (published in 2005) was selected for this analysis and every citation in it was reviewed. This annual bibliography has individual chapters (18 chapters in the 2004 bibliography), each prepared by different authors who are typically expert in the field, with organismal coverage (e.g. chapters on periphyton, plecoptera, etc.) and environmental coverage (e.g. chapters on general aquatic ecology, macroinvertebrate toxicology, etc.). The chapters vary greatly in length (e.g. 1 page versus 19 pages in the 2004 bibliography). The content is primarily journal articles (the 2004 bibliography contained 3,990 journal articles out of 4,333 total citations). And the content is very diverse (citations were identified from over 850 different journal titles).

Given that the NABS bibliography is a good benchmark for analyzing freshwater biology databases what criteria should be used as measurements? Criteria, based on the NABS bibliography, that were identified for this analysis are as follows: Are the top journals indexed?; Is the most recent issue indexed for the top journals (and if not how long is the lag time)?; Are the specific citations indexed?; What is the amount and relevance of literature indexed matching topics found in the bibliography?

To identify the top journals in the 2004 NABS bibliography a straight count could be used (Appendix 2). However, given how greatly the chapters vary in length using a straight count might bias the identification of the top journals based on an individual chapter topic (e.g. Odonatologica, International Journal of Odonatology). Instead, to calculate the top journals overall for the 2004 NABS bibliography each chapter was examined, and the top 10 journals (based on number of citations) for each chapter were identified, and then those journals listed in the top 10 for 3 or more chapters were selected. There were 12 journal titles appearing in the top 10 for 3 or more chapters (out of 108 different titles from the combined top 10 lists of all chapters) and those are listed below (in ranked order with ties generating the same number order). Hydrobiologia was ranked number 1 for almost half the chapters of the bibliography (9 out of 19). The list of top journals based on a straight count (Appendix 2) was similar to the list below with 8 identical journal titles (out of the top 12 journals). The top journals listed below do not correlate well with the journals identified in the marine and freshwater biology category of the 2005 Journal Citation Reports as ranked by highest impact factor (e.g. Hydrobiologia is ranked 42nd out of 77 journals in that category by impact factor).

1) Hydrobiologia
2) Archiv fur Hydrobiologie
3) Freshwater Biology
3) Journal of the North American Benthological Society (JNABS)
5) Environmental Toxicology and Chemistry
6) Aquatic Insects
7) Journal of Freshwater Ecology
8) International Review of Hydrobiology
8) Journal of Great
8) Zootaxa
11) Canadian Journal of Fisheries and Aquatic Sciences
12) Entomological News

The databases were evaluated to determine coverage of the top 12 journals and results (i.e. the number of the top 12 journals not indexed) are listed and ranked below:

Database	# of Top 12 Journals Not Indexed
1) Google Scholar	0
1) Scopus	0
1) Web of Science	0
1) Zoological Record	0
2) ASFA	2
2) Biological Abstracts	2
2) CAB Abstracts	2
2) CSA Biological Sciences	2
3) EBSCO Academic Search Premier	6
3) Fish & Fisheries Worldwide	6
4) Biological & Agricultural Index	11
4) BioOne	11
4) Blackwell Synergy	11
4) SpringerLink	11
5) Elsevier Science Direct	12
5) Scirus	12

The databases were evaluated to determine how current the indexing of the top 12 journals is and results are listed (i.e. average # months behind) and ranked below.

Database	Average # Months Behind for Journals Indexed
1) BioOne	0
1) Blackwell Synergy	0
1) Science Direct	0
1) Springerlink	0
1) Scirus (NA)	0
2) Biological & Agricultural Index	1
3) Web of Science	1.16
4) EBSCO Academic Search Premier	1.33
5) Scopus	1.75
6) Google Scholar	4.16
7) Zoological Record	4.58
8) CAB Abstracts	5.1
9) Biological Abstracts	5.8
10) CSA Biological Sciences	6.4
11) ASFA	8.3
12) Fish & Fisheries Worldwide	11.33

A random number generator was used to select 20 numbers between 1 and 4,333. Each number was then used to find a correspondingly listed citation within the bibliography. The databases were evaluated to determine the number of the 20 citations indexed and the results are listed (i.e. # of citations not indexed) and ranked below.

Database	# of Citations (NABS Biblio) Not Indexed
1) Google Scholar	2
2) Scopus	4
2) Zoological Record	4
3) Biological Abstracts	5
3) Web of Science	5
4) CSA Biological Sciences	9
5) ASFA	11
6) Fish & Fisheries Worldwide	14
7) EBSCO Academic Search Premier	15
7) Scirus	15
8) CAB Abstracts	16
9) Elsevier Science Direct	17
10) BioOne	18
11) Biological & Agricultural Index	19
11) Springerlink	19
12) Blackwell Synergy	20

Ten keyword searches were crafted to capture NABS bibliography chapter topics. Six of the ten topics were organism oriented and four were concept oriented. Results were limited to the year 2004. An example of an organism oriented search was:
Keyword: plecoptera* or stonefl*
Limits: 2004-2004.

Searches were adapted to database interfaces. The databases were evaluated to determine number of results and relevance of results. Relevance was determined by examining the first 10 citations and noting the number of citations deemed likely to appear in the NABS bibliography. The procedure was admittedly subjective. The database ranking based on the number and relevance of results is listed below.
1) Google Scholar
2) Scirus
3) Biological Abstracts
4) Web of Science
5) CSA Biological Sciences
6) Fish & Fisheries Worldwide
6) Zoological Record
7) ASFA
8) Scopus
9) EBSCO Academic Search Premier

10) CAB Abstracts
11) Elsevier Science Direct
12) BioOne
13) Biological & Agricultural Index
14) Blackwell Synergy
SpringerLink – NA (couldn't apply searches)

To determine a final database ranking each of the four categories of criteria examined were equally weighted with 12 points each for a total of 48 points possible. Results were assigned point values within each category. The category for number of top journal titles not indexed had 0 points assigned for 0 journals not indexed and 12 points assigned for 12 journal titles not indexed. The category for number of months behind current journal issues indexed had an average for all journal titles with 0 points assigned for 0 time lag and 12 points assigned for 12 months or greater time lag. The category for number of the 20 citations not included had 0.6 points assigned for one citation not included and 12 points assigned for 20 citations not included. The category for number and relevance of keyword searches had an evaluation that resulted in rankings of 1 through 16 for each database and then assignments of 0.75 points per ranking. The final database ranking is listed below. If the fourth category of number and relevance of keyword searches is removed the same top ten databases remain with the only difference being that Scopus and Google Scholar switch rankings.

Top Databases for Freshwater Biology (based on the results of this study):

1) Google Scholar
2) Web of Science
3) Scopus
4) Zoological Record
5) Biological Abstracts
6) CSA Biological Sciences
7) ASFA
8) Scirus
9) EBSCO Academic Search Premier
10) CAB Abstracts
11) Fish and Fisheries Worldwide
12) BioOne
13) Elsevier Science Direct
14) Biological & Agricultural Index
15) Blackwell Synergy
16) Springerlink

Conclusions from this study include: top databases identified for freshwater biology are in general agreement with the collective judgment of IAMSLIC conference attendees on 4 out of 5 databases; Google Scholar performed better than expected but had a surprising lag time of several months for indexing current journal issues; some traditional indexes

performed better than expected and others worse than expected (depending on individual expectations!); not surprisingly publisher ejournal packages did not perform well. Google Scholar is still only listed as being in "beta" version; undoubtedly many traditional, specialized, commercial indexes may face increasing competition in the next few years with users preferentially selecting Google and Google Scholar for literature searches. It is worth periodically testing and reexamining assumptions about databases.

An obvious critique of this analysis would be the subjective nature of assessing the amount and relevance of citations indexed in databases matching topics found in the NABS bibliography. No rebuttal is offered; that is why the total ranking was assessed both with and without the literature amount and relevance measurements included (and the top ten databases stayed the same in both scenarios with only slightly different ranking results). Critiques of this analysis might also include a concern about what databases were originally used by the authors of the NABS bibliography (i.e. if Web of Science, as a perennial favorite of scientists, was primarily used to generate most of the citations in the NABS bibliography and analyzed in this study, then of course Web of Science would be identified as one of the top databases). That is an important concern however, using a bibliography with 18 authors from 18 different institutions makes it extremely unlikely that the same tool was used by all authors for finding literature. Furthermore, scientists often find literature not through indexes or databases but rather via personal networks, known authors, known journals, and the bibliographies in journal articles. One additional database that ideally should have been evaluated as part of this study is Aquatic Biology, Aquaculture and Fisheries Resources. A final concern might be the "macroinvertebrate-centric" nature of the NABS bibliography. An argument could be made that while no bibliography is comprehensive, the NABS bibliography has a very wide scope and given the fundamental position of macroinvertebrates in ecosystem studies most aspects of freshwater biology are covered each year in it. Nevertheless, an additional comparative analysis ideally should be performed using a freshwater "fish-centric" bibliography and the results compared with this study.

REFERENCES:

Bearman, T.C. & Kunberger, W.A. 1977. *A Study of Coverage Overlap Among Fourteen Major Science and Technology Abstracting and Indexing Services.* National Federation of Abstracting and Indexing Services, Philadelphia.

Chisman, J.K. 1989. Zoological Record, Biological Abstracts and Biological Abstracts/ RRM: a comparison of overlap. *RQ* 29 (Winter): 242- 247.

DeGuire, E.J. 2006. Publish or perish: afterlife of a published article. *CSA Discovery Guides* [http://www.csa.com/discoveryguides/publish/review.php].

Deis, L. & Goodman, D. 2005. Web of Science (2004 version) and Scopus. *The Charleston Advisor* 6(3): [http://www.charlestonco.com/comp.cfm?id=43].

Dess, H.M. 2006. Database reviews and reports: Scopus. *Issues in Science and Technology Librarianship* 45 (Winter): [http://www.istl.org].

Hughes, J. 2001. Characterization of unique serials indexed in the Zoological Record. *Issues in Science and Technology Librarianship* 30 (Spring): [http://www.istl.org].

Jacso, P. 2005. Google Scholar (redux). *Peter's Digital Reference Shelf, Thomson Gale* (June): [http://www2.hawaii.edu/~jacso/]

Jatkevicius, J. 2000. Biological sciences databases in academic libraries: comparing Cambridge Scientific Abstracts with Biosis. *Econtent* 23 (February/March): 55-59.

OCLC. 2005. *Perceptions of Libraries and Information Resources (2005).* OCLC [http://www.oclc.org/reports/2005perceptions.htm].

Parker, J. 2005. Evaluating bibliographic database overlap for marine science literature using an ecological concept. *Issues in Science and Technology Librarianship* 42 (Spring): [http://www.istl.org].

Poyer, B.K. 1984. Journal article overlap among Index Medicus, Science Citation Index, Biological Abstracts, and Chemical Abstracts. *Bulletin of the Medical Library Association* 72 (October): 353- 357.

Appendix 1: Librarian Survey of Top 5 Databases for Freshwater Biology (administered by Barry Brown at IAMSLIC Conference, October 2006, Portland during his presentation). Survey sheet was distributed with list of databases shown below. Librarians were asked to rank (1 through 5) the top five databases (with 1 being the best). Number of responses and numeric responses for each database are listed below; X indicates a circled nonranked response. There was a total number of 33 submitted surveys.

ASFA: 2, 2, 1, 3, 3, 1, 3, 1, 2, 1, 3, 1, 2, 2, 2, 2, 2, 1, 1, 3, 2, 1, X, X, X, X, X, X, X
Biological Abstracts: 1, 1, 1, 1, 5, 4, 1, 2, 2, 4, 1, 3, 2, 3, 1, X, X, X, X, X
Biological & Agricultural Index: 2, 2, 3
BioOne:
Blackwell Synergy: 4, 3
CAB Abstracts: 1, 2, 1, 5
CSA Biological Sciences: 2, 2, 1, 3, X, X, X, X
Fish and Fisheries Worldwide: 2, 1, 4, 4, X
JSTOR: 5
Google Scholar: 4, 3, 3, 4, 3, 3, 3, 4, 2, 2, X
ScienceDirect: 3, 4, 1, X
Scirus: 1, 5, 3, X
Scopus: 1, 3, 2, 2, 1, 4, 2, 5, X, X
SpringerLink: 5, 5
Web of Science: 3, 2, 4, 2, 3, 3, 1, 2, 3, 1, 1, 3, 2, X, X, X
Zoological Record: 2, 1, 4, 4, X, X

Appendix 2: Top 20 Journals in the 2004 NABS Bibliography as identified by total citation count of all chapters.
Hydrobiologia (295)
Freshwater Biology (131)
JNABS (110)
Archiv fur Hydrobiologie (90)
Environmental Toxicology & Chemistry (63)
Journal of Freshwater Ecology (56)
Journal of Great Lakes Research (43)
Zootaxa (40)
Odonatologica (38)
Ecology (35)
International Journal of Odonatology (35)
Marine & Freshwater Research (34)
Canadian Journal of Fisheries & Aquatic Sciences (33)
Environmental Pollution (32)
International Review of Hydrobiology (32)
Transactions of the American Fisheries Society (32)
Aquatic Insects (31)
Aquatic Ecosystem Health & Management (30)
Ecological Applications (30)
Archives of Environmental Contamination & Toxicology (28)