Article ID: WMC003159 ISSN 2046-1690

WebmedCentral.com
building science together

Is Your Journal Indexed in PubMed? Relevance of PubMed in Biomedical Scientific Literature Today

Corresponding Author:
Mr. Kamal K Mahawar,
Consultant Surgeon, City Hospitals Sunderland NHS Trust, Sunderland - United Kingdom

Submitting Author:
Dr. Deepak Kejariwal,
Consultant Gastroenterologist, University Hospital of North Durham - United Kingdom

Article ID: WMC003159

Article Type: Review articles

Submitted on:17-Mar-2012, 05:12:39 AM GMT **Published on:** 17-Mar-2012, 07:16:40 AM GMT

Article URL: http://www.webmedcentral.com/article_view/3159

Subject Categories:MISCELLANEOUS

Keywords: Literature, PubMed, Bibliographic Database, Google Scholar, Embase

How to cite the article: Kejariwal D, Mahawar KK. Is Your Journal Indexed in PubMed? Relevance of PubMed in Biomedical Scientific Literature Today . WebmedCentral MISCELLANEOUS 2012;3(3):WMC003159

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Source(s) of Funding:

None. Authors have used their personal time to write this article.

Competing Interests:

Webmed Limited, UK owns the portal WebmedCentral. DK is its chairman and KM is its CEO. Both are shareholders and directors in the company.

Is Your Journal Indexed in PubMed? Relevance of PubMed in Biomedical Scientific Literature Today

Author(s): Kejariwal D, Mahawar KK

Introduction

The arrival and spread of internet has enabled development of better bibliographic scientific databases with significantly improved capacity for storage and retrieval. In recent years, electronic database searching has become the default mode of bio-medical information retrieval. The National Library of Medicine (NLM) in the United States introduced the first interactive searchable database (Medline) in 1971. Subsequently, in 1996, it added the “Old Medline” database with coverage of publications between 1950 and 1965. In 1997, NLM launched PubMed (a combination of both Old Medline and Medline) (1). Several other databases have emerged over the last two decades, each tackling the issue from a slightly different angle. Here, we examine these databases against our criteria for an ideal database and suggest areas for further improvement.

Characteristics of an Ideal Bibliographic Database

Since these databases are so important for information storage and retrieval, it is crucial that we have scientific databases fit for purpose. In our opinion, a scientific database should ideally have following characteristics.

a. **Inclusive:** It should cover all scientific research and must not exclude any piece of research. Ideally, it should also include non peer reviewed (for example scientific reports, conference papers, webpage/blogs etc) information, the so called “gray literature”.
b. **Specific Refined Search**: At the same time, it should bring out the most relevant information to a user.
c. **Advanced Filters**: It should have a variety of filters for specialty, dates, keywords, peer reviewed/non peer reviewed, blogs, geographical location etc for an ideal user experience.
d. Link to Full text articles: It should provide link to full text articles. Print only journals need to be encouraged to look for ways for electronically archiving their work.
e. **Advanced Citation Analysis**: An ideal database should have sophisticated features to track and analyse citations
f. **Free:** Ideally, such a resource should be completely free for users along with free access to full text articles.

We discuss here briefly the major scientific databases available currently, their background, and their important characteristics. We also examine how they fair against the criteria above.

Important Databases

PubMed:

Advent of PubMed was a major landmark in history of electronic archiving of biomedical scientific literature. It is a free resource that is developed and maintained by NCBI (National Center for Biotechnology Information) at the NLM (National Library of Medicine), at US National Institute of Health (NIH) (2). It comprises over 21 million citations from biomedical literature from MEDLINE (a huge database of over 19 million references to articles published in approximately 5,600 current biomedical journals from the United States and over 80 foreign countries), some additional life science journals(not indexed in MEDLINE), and online books (2). As of November 2011, 5,582 journals were indexed with MEDLINE (3).

PubMed citations and abstracts include the fields of biomedicine and health, covering portions of the life sciences, behavioral sciences, chemical sciences, and bioengineering. PubMed also provides access to additional relevant web sites and links to the other NCBI molecular biology resource (2). Publishers of journals can submit their citations to NCBI. If the publisher has a web site that offers full-text of its journals, PubMed provides links to that site as well as biological resources, consumer health information, research tools, and more.

Criteria for inclusion of a journal into MEDLINE (4) (and hence into PubMed) can seem arbitrary. There is no single criterion which is crucial and they seem to be using a number of them to determine the suitability of journal. Even though it is widely held that PubMed will only consider Peer Reviewed literature (5), they explicitly state on their website that this is not the case, “Most journals in PubMed are peer-reviewed or

refereed. Non-editorial journal-staff review original articles before the articles are accepted for publication. Criteria for peer review and the qualifications of peers or referees vary among publishers. We have no list of peer-reviewed/refereed journals in PubMed; and you cannot limit your search to peer-reviewed journals using PubMed" (6). Overall 20-25% of titles reviewed are selected for indexing with MEDLINE.

PubMed stands out as a bibliographic database for biomedical scientists. It has much to recommend itself. It has been successful in encouraging a large number of publishers to work with it and its user base is constantly growing. It is totally free to use and provides free abstracts. Its filters are reasonably good. It links out to full text articles and, with the help of PubMed Central, is building a great resource of free full text articles. However, it excludes a large body of both peer reviewed and also the so called "gray" literature (non peer reviewed). Both are significant sources of knowledge, we should make better use of. Peer Reviewed literature (that has undergone pre publication peer review) has its own limitations (7), that gray literature is free from. Gray literature does not suffer from any form of publication bias, is instant, and guarantees true freedom of expression to scientists with contrarian views. Conn et al (8) noted that meta-analyses that exclude gray literature likely over-represent studies with statistically significant findings. As early as 2004, Banks had noted that the bigger challenge would be to develop bibliographic resources for gray literature (9). Others (10) have tried to develop models to facilitate this. However, by and large, this challenge persists for all of us in 2012.

Literature published in a climate of post publication peer review, as on WebmedCentral (11), is a further subset of scientific literature that aims to combine best of both the worlds of peer reviewed literature and gray literature. This model of publishing has some support (12) and other commercial players are keen to venture into the foray (13). PubMed will need to find ways to include these too. If it does not, it risks losing some of its relevance in future.

By virtue of its size, PubMed has a dominant position in the field. Because of its dominant position, journals not considered suitable by PubMed are at an obvious disadvantage compared to their PubMed included counterparts. One could argue that by excluding a large section of academic literature, PubMed is also putting science and scientists at a disadvantage.

Google Scholar:

Launched in beta version in 2004, it is a Google® product, company with world's best internet search engine. Even before the advent of Google Scholar and also today, many years after it, a large number of academicians rely on simple Google search to look for scientific content. However, the search on Google, looking for academic content, can be very frustrating with popular websites, promotional content and blogs competing with scholarly content written by scientists for a scientific audience. Users have no way of knowing how to interpret relative merit of search results. Google Scholar would seem like an attempt from Google to make this search more specific and relevant for academic community.

It obtains its information directly from publishers and by crawling the web for scholarly content. It searches for scholarly materials such as peer-reviewed papers, theses, books, preprints, abstracts and technical reports from broad areas of research. Google Scholar searches a variety of undisclosed academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles. Its search results will hence include a variety of peer reviewed scholarly content but also include non peer reviewed scholarly content. There is no clear way of identifying only peer reviewed content. There has been some criticism that all the content it shows may not be scholarly either. Moreover there seems to be some secrecy about what it covers. Google Scholar does not publish a list of scientific journals crawled, and the frequency of its updates is unknown. It is therefore impossible to know how current or exhaustive searches are in Google Scholar.

The exact ranking algorithm used by Google Scholar is undisclosed. According to them, "It ranks documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature" (14). It is suspected that citation count is probably the most important of several factors it takes into account (15). Google Scholar allows citation analysis. It may be vulnerable to spam (16). Use of Google Scholar is free and many articles are freely available full text. In other cases, it will take you to full text article on publishers' website. It provides single interface to access library catalogues indexes and websites and is relatively easy to use and navigate.

To some extent, it suffers from the same drawbacks that Google search engine does. Users cannot clearly make out what is peer reviewed and what is not; what is scholarly and what is not. The search results are less cluttered than Google search engine as it excludes the large body of supposedly non scholarly information out there. But that also makes it less

sensitive than Google search engine and perhaps less attractive, if one is looking for a relatively rare topic. At the same time, coverage is unclear and incomplete. Moreover, it is not always scholarly either. Because of the secrecy involved with regards to what it includes and how it ranks information, it is not widely regarded by biomedical scientists. Google says that MEDLINE citations are a part of Google Scholar but it does not always stand to independent verification. It does however complement a researcher's needs by covering resources not covered by other citation indexing bodies. Citation tracking is an advantage and it is free to users. It would seem as if Google Scholar is currently struggling to find its niche but does have a significant role to play in free search for scholarly content. It is bound to improve with time and could threaten commercial citation tracking databases.

Scopus:

It is a bibliographic database (17) containing abstracts and citations for academic journal articles. It covers nearly 18,000 titles from over 5,000 international publishers, including coverage of 16,500 peer-reviewed journals in the scientific, technical, medical, and social sciences (including arts and humanities). The only criteria it mentions for inclusion in the database are ISSN number and 'scholarly/academic' publication. It is owned by Elsevier and is available online by subscription. It allows citation analysis.

Web of science:

It is provided by Thomson Reuters (18). It covers 11,261 journals (as of September 5, 2009) spanning multiple academic disciplines including the sciences, social sciences, arts, and humanities, and across disciplines. Web of Science does not cover all journals, and its coverage in some fields is less complete than in others. Web of science mention multiple criteria for inclusion in its database. However peer review as a minimum criterion is not mentioned on its website. It includes non peer reviewed literature like seminars, symposia, conference proceedings etc. Web of Science does not provide any data regarding open access articles that it includes (if any). It is fee based and allows citation analysis.

Embase:

It is a commercial abstracts and index database maintained by Elsevier (19). It includes all MEDLINE records produced by the National Library of Medicine (NLM), as well as over 5 million records not covered on MEDLINE (over 2,000 currently indexed Embase journals are unique). Journals indexed in Embase are mostly peer reviewed. Use of Embase is subscription based.

Discussion

There are many bibliographic databases out there but none of them perform well against all of the criteria laid out by us above. PubMed and Embase focus mainly on medicine and biomedical sciences, whereas Scopus, Web of Science, and Google Scholar cover most scientific fields. PubMed and Google Scholar are free to use whereas Embase, Scopus and Web of Science are databases that belong to commercial providers and require an access fee. PubMed and Google Scholar are the only free ones and hence have been discussed in some depth above. However, the selection criteria and the journals indexed for all databases are unclear. All of these databases discussed above include non peer reviewed literature despite the widely held view to the contrary. Further. None of the databases are all inclusive.

Some databases often use Bradford's law (20) to justify their non inclusion of all scientific journals. Bradford's law states that relatively small numbers of journals publish the bulk of significant scientific results. Hence, in pre electronic era, the law was used to economically plan information systems and library services (21). This is clearly not relevant with the arrival of electronic databases where the costs of including all scientific literature would not be prohibitive and good search functionality would sort out the literature in a variety of filters. Bradford's analysis of choosing journals for databases has also been criticised for not being objective and neutral (22). Further, Bradford law or selectivity in databases clearly tends to favour dominant theories and views while suppressing views other than the mainstream. As a result of this selectivity, scientists have felt under pressure to publish in the best journals, and universities to ensure access to that core set of journals. There is clearly a danger of only representing majority views if journals are selected in this fashion. In this day and age, with the available technology, it should be possible for databases to include everything in their fold and identify them by giving them appropriate labels so that users can differentiate between sources of information.

PubMed is the most used biomedical bibliographic database. However it excludes a significant body of literature that diminishes its usefulness. Academicians are increasingly using multiple databases to provide a more inclusive information search. Google scholar is the only free creditable option at present apart from PubMed. Google is easier to use compared to

PubMed, which many users find tedious even with all its help tutorials (23). Internet search is improving everyday and also pulls out the relevant PubMed results. However, it may be too sensitive (24-25). In 2007, Shultz (26) examined PubMed and Google Scholar and found the two complementary.

It should be possible for PubMed to include all scholarly content and identify them with appropriate tags. For example PubMed content that has not undergone any peer review could be clearly identified. Bodies like PubMed can reduce the dominance of selected players and themselves become more useful for scientists in the process. Not surprisingly, scientists are looking elsewhere to find literature not covered by PubMed (27). This has led to development of multiple commercial databases listed above. PubMed however remains an important resource for clinicians and researchers (28).

In summary, none of the above databases include all scholarly content and their selection criteria are not completely obvious. They all include non peer reviewed literature and many are subscription based.

Conclusion

There is a need to develop better scientific databases for improved storage and retrieval of biomedical scientific content. It will lead to an improved user search experience and quality. PubMed currently maintains a dominant position in the field but risks losing it if it does not keep pace with time. It will need to evolve and become more inclusive in future to stay ahead in the game. Biomedical scientists need a bibliographic database that will archive all scientific content, identify them with appropriate tags, and is free to use.

Abbreviations

NCBI: National Center for Biotechnology Information
NLM: National Library of Medicine
NIH: National Institute of Health

References

1. PubMed Overview. http://www.ncbi.nlm.nih.gov/corehtml/query/static/overview.html Last accessed on 11/03/2012
2. http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.PubMed_Quick_Start Last accessed on 11/03/2012
3. http://www.nlm.nih.gov/bsd/num_titles.html Last accessed on 11/03/2012
4. http://www.nlm.nih.gov/pubs/factsheets/jsel.html Last accessed on 11/03/2012
5. http://www.fmhs.uaeu.ac.ae/FMHSWEBUserFiles/file/Scholarly%20Peer%20Reviewed%20Journals(1).pptx Last accessed on 11/03/2012
6. http://www.nlm.nih.gov/services/peerrev.html Last accessed on 11/03/2012
7. Mahawar KK. Role of Peer Review in Biomedical Publishing. WebmedCentral MISCELLANEOUS 2011; 2(4): WMC001863. Last accessed on 11/03/2012
8. Conn VS, Valentine JC, Cooper HM, Rantz MJ. Gray literature in meta-analyses. Nurs Res 2003; 52(4): 256-61.
9. Banks M. Connections between open access publishing and access to gray literature. J Med Libr Assoc 2004; 92(2): 164–166.
10. Turner AM, Liddy Ed, Bradley J, Wheatley JA . Modelling public health interventions for improved access to the gray literature. J Med Libr Assoc 2005; 93(4): 487–494.
11. http://www.webmedcentral.com/ Last accessed on 11/03/2012
12. Smith R. Classical peer review: an empty gun. Breast Cancer Res. 2010; 12(Suppl 4): S13.
13. http://f1000research.com/ Last accessed on 11/03/2012
14. http://scholar.google.com/intl/en/scholar/about.html Last accessed on 11/03/2012
15. Beel J, Gipp B. Google Scholar's Ranking Algorithm: An Introductory Overview. In Birger Larsen and Jacqueline Leta, editors, Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. Preprint downloaded from http://www.sciplore.org/publications/2009-Google_Scholar%27s_Ranking_Algorithm_--_An_Introductory_Overview_--_preprint.pdf. Last accessed on 11/03/2012
16. Joeran Beel and Bela Gipp. On the Robustness of Google Scholar Against Spam. In Proceedings of the 21th ACM Conference on Hyptertext and Hypermedia. ACM, June 2010 Pre print pdf downloaded from http://www.sciplore.org/publications/2010-On_the_Robustness_of_Google_Scholar_against_Spam--preprint.pdf Last accessed on 11/03/2012
17. http://www.scopus.com/home.url Last accessed on 11/03/2012
18. http://thomsonreuters.com/products_services/science/

science_products/a-z/web_of_science/ Last accessed on 11/03/2012
19. http://www.embase.com/info/what-embase Last accessed on 11/03/2012
20. http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/ Last accessed on 11/03/2012
21. Brookes BC. Bradford's law and the bibliography of science. Nature 1969; 224: 953-6.
22. Nicolaisen J, Hjørland B. Practical potentials of Bradford's law: a critical examination of the received view. Journal of Documentation 2007; 63 (3): 359 – 377
23. http://laikaspoetnik.wordpress.com/2008/06/11/pubmed-past-present-and-future-part-i/ Last accessed on 11/03/2012
24. Freeman MK, Lauderdale SA, Kendrach MG, Woolley TW. Google Scholar versus PubMed in locating primary literature to answer drug-related questions. Ann Pharmacother 2009; 43(3): 478-84.
25. Anders ME, Evans DP. Comparison of PubMed and Google Scholar literature searches. Respir Care 2010 May; 55(5): 578-83.
26. Shultz M. Comparing test searches in PubMed and Google Scholar. J Med Libr Assoc. 2007 October; 95(4): 442–445.
27. Hendersen J. Google Scholar: A source for clinicians? CMAJ. 2005 June 7; 172(12): 1549–1550.
28. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb; 22(2): 338-42.

Disclaimer

This article has been downloaded from WebmedCentral. With our unique author driven post publication peer review, contents posted on this web portal do not undergo any prepublication peer or editorial review. It is completely the responsibility of the authors to ensure not only scientific and ethical standards of the manuscript but also its grammatical accuracy. Authors must ensure that they obtain all the necessary permissions before submitting any information that requires obtaining a consent or approval from a third party. Authors should also ensure not to submit any information which they do not have the copyright of or of which they have transferred the copyrights to a third party.

Contents on WebmedCentral are purely for biomedical researchers and scientists. They are not meant to cater to the needs of an individual patient. The web portal or any content(s) therein is neither designed to support, nor replace, the relationship that exists between a patient/site visitor and his/her physician. Your use of the WebmedCentral site and its contents is entirely at your own risk. We do not take any responsibility for any harm that you may suffer or inflict on a third person by following the contents of this website.