



Review Paper

Overview of Journal Bibliographic Databases

Priyadarshini M. Deodurg, Srikanth*, Shireen Rana and Banti Brahma

Department of Pharmacology, Khaja Banda Nawaz institute of medical sciences, Gulbarga-585104, INDIA

Available online at: www.isca.in, www.isca.me

Received 22nd May 2013, revised 28th June 2013, accepted 25th July 2013

Abstract

A bibliographic database is a database of bibliographic records with organized collection of references to published literature. There are several types of medical and general bibliographic databases. Indexation of a journal in a specific database considered a reflection of its quality. Indexed journals receive more submissions when compared to non-indexed journals and many authors prefer to submit their work to indexed journals than non-indexed journals. The oldest index is the Index Medicus published since 1879 is considered to be the most comprehensive index of medical journals. Over the years, due to internet revolution, many other popular indexation services have developed such as, PubMed, MedLine, SCOPUS, EMBASE, EBSCO Publishing's Electronic Databases etc.. There are various regional and national versions of Index Medicus such as Index Medicus for South East Asia region, Index Medicus for the Eastern Mediterranean Region, African Index Medicus, etc. Knowing the basics of these databases will help research scholars and authors to enhance their writing skills and help in publication their research work.

Keywords: Indexing, journal, bibliographic database, journal quality.

Introduction

Indexation is considered a reflection of quality of a journal. Indexed journals are considered to be of higher scientific quality as compared to non-indexed journals. The spread of internet revolution has enabled development of several bibliographic (medical and other scientific) databases with significantly improved capacity for storage and retrieval. In recent years, for bio-medical information retrieval, electronic databases have become the preferred mode.

Some promotions or grants for research give more preference to journals indexed with PubMed or Science Citation Index (SCI). Many job applications in research institutions or in a medical college, and for grant applications specifically mention to list the publications based on the PubMed indexation and/or Impact Factor (IF) of the journal. It is considered that papers published in journals with high IF or indexed with PubMed /SCI are likely to have a higher research impact compared to papers published in non-indexed or low-IF¹ journals. Hence, a author gives preference to journals indexed with PubMed /SCI.

There are several types of medical bibliographic databases which cover various fields such as biomedical and scientific literature, therapeutic regimens, medical records, images and reviews of evidence-based medicine etc.

A bibliographic database is a database of bibliographic records, an organized collection of references to published literature, which includes journals specific to that particular field, newspaper articles, conference proceedings, books, government

and legal publications, patents etc. In contrast to library catalogue entries, a large proportion of the bibliographic records in bibliographic databases describe analytics (articles, conference papers, etc.) rather than complete monographs, and they generally contain very rich subject descriptions in the form of keywords, subject classification terms, or abstracts².

In medicine, bibliographic databases are one of the most essential search tools for both research and clinical practice, because they provide the latest scientific insights for evidence-based medicine³. There are thousands of biomedical journals are published worldwide, but most bibliographic databases index only those that meet their specific requirements. Selection criteria are usually based on various factors, such as whether the journal is peer reviewed, impact factor, longevity of the journal, language of publication etc. Most databases contain the article citations and selected abstracts, but some contain the full text or links to the full-text sources³.

Indexation services are meant to give journals wider coverage for easy accessibility to their published articles. For a long-time Index Medicus has been the most comprehensive index of medical scientific journal articles. Over the years, many other popular indexation services have developed. This article gives an overview of the various journal bibliographic databases, describing their salient features.

General Medical Databases

Most commonly used databases within this discipline are PubMed, EMBASE and MEDLINE.

PubMed: 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)⁴. PubMed comprises over 22 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed citations and abstracts include the fields of biomedicine and health, covering portions of the life sciences, behavioral sciences, chemical sciences, and bioengineering⁴.

Criteria for inclusion of a journal into MEDLINE⁵ (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. PubMed (which encompasses MEDLINE) is the database of choice for both researchers and clinicians to locate relevant articles, and in many cases, it links directly to a publisher's site for the full text. Most of the authors prefer their work to be published in PubMed indexed journals. Advantages of PubMed are it is totally free, provides free abstracts and has filters which 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 articles⁶.

In conclusion, PubMed citations come from 1) MEDLINE indexed journals, 2) journals/manuscripts deposited in PMC, and 3) NCBI Bookshelf. Both MEDLINE and other PubMed citations may have links to full-text articles or manuscripts in PMC, NCBI Bookshelf, and publishers' Web sites. If you limit your PubMed search to MeSH controlled vocabulary or the MEDLINE subset, you will see only MEDLINE citations in your results.

However, PubMed is not without limitations, like 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, and should be made available. Peer Reviewed literature has its own limitations⁷ 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 al⁸ noted that meta-analysis that exclude gray literature likely over-represent studies with statistically significant findings. Literature published in a climate of post publication peer review, as on WebmedCentral⁹, 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 support¹⁰ and other commercial players are keen to venture into the foray¹¹. PubMed will need to find ways to include these too, to avoid losing some of its relevance in future.

Medline: Medline is the U.S. National Library of Medicine's (NLM) premier bibliographic database that contains over 19 million references to journal articles in life sciences with a

concentration on biomedicine. A distinctive feature of MEDLINE is that the records are indexed with NLM Medical Subject Headings¹². The great majority of journals are selected for MEDLINE based on the recommendation of the Literature Selection Technical Review Committee (LSTRC), an NIH-chartered advisory committee of external experts analogous to the committees that review NIH grant applications. MEDLINE is the primary component of PubMed, part of the Entrez series of databases provided by the NLM National Center for Biotechnology Information (NCBI)¹².

In addition to PubMed, MEDLINE can also be accessed through other resources such as EBSCO interface and an Institute for Scientific Information (ISI) interface, all of which access the MEDLINE, but each has different capabilities and functions i.e. PubMed is one of several ways to access MEDLINE.

EMBASE: It is maintained by Elsevier. Embase's comprehensive journal and conference coverage, together with in-depth drug indexing and daily updates, supports tracking and precise retrieval of drug and disease information. From preclinical study to the search for important toxicological information, Embase offers the confidence and tools you need to capture the most relevant and up-to-date biomedical study research¹³. If you're searching MEDLINE, you're only seeing part of the picture. Embase holds millions of indexed records from thousands of active, authoritative journals, including all of MEDLINE as well as 5 million+ records and 2,000 biomedical journals not currently covered by MEDLINE¹⁴.

About 30% of journals that may be searched through EMBASE also appear in MEDLINE; however, EMBASE has a more European emphasis than MEDLINE and includes more non-English language biomedical journals than MEDLINE. Thus, it is useful for identifying citations in non-English language journals. EMBASE is a subscription-only database that is available mostly through university or medical libraries¹⁵.

Specialized Biomedical Databases¹⁵

These are databases covering specialized biomedical areas, ranging from complementary medicine to allied health sciences. Some of the important databases are as follows.

The Cochrane Library: The Cochrane library was developed by The Cochrane Collaboration (www.cochrane.org/). The Cochrane Library is a collection of six databases that contain different types of high-quality, independent evidence to inform healthcare decision-making, and a seventh database that provides information about groups in The Cochrane Collaboration¹⁶. i. Cochrane Database of Systematic Reviews, ii. Cochrane Central Register of Controlled Trials, iii. Cochrane Methodology Register, iv. Database of Abstracts of Reviews of Effects, v. Health Technology Assessment Database, vi. NHS Economic Evaluation Database, vii. About The Cochrane Collaboration.

The most important is randomized controlled trials with earliest record dating back to 1898 and the total number of trials is 692036¹⁶.

Clinical Evidence: It published by the BMJ Group, comprises an international database of high-quality, rigorously developed systematic overviews assessing the benefits and harms of treatments, and a suite of Evidence Based Medicine resources and training materials¹⁷. It is a subscription service but is available to developing countries free of charge or at a reduced rate through HINARI.

Global Health: CAB Direct is a source of references for the applied life sciences. It incorporates two bibliographic databases: CAB Abstracts and Global Health. CAB Direct is CABI's electronic platform delivering access to over 9 million abstracts across the applied life sciences. Coverage includes our leading bibliographic databases CAB Abstracts and Global Health, and databases from our Internet Resources and Abstract Journals¹⁸. Users have access to: i. over 9 million bibliographic records, ii. over 130,000 full text articles hosted by CABI, iii. many other authoritative reviews, news articles and reports.

Global Health has excellent coverage of health in the developing countries, and its sister database Global Health Archive (www.cabdirect.org/) covers the history of public health and research from the end of the 19th century to 1972.

Other specialized databases are listed below: i. BioMed Central Databases: databases.biomedcentral.com/browsecatalog, ii. Cumulative Index to Nursing and Allied Health Literature (CINAHL) Databases: www.ebscohost.com/cinahl/, iii. EBSCO Medical Databases: www.ebscohost.com/, iv. POPLINE (POPulation information onLINE): www.popline.org, v. TOXLINE (Toxicology Literature Online, a bibliographic database for toxicology): toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE, vi. DynaMed: www.ebscohost.com/dynamed/

General Databases

Google Scholar: Launched in beta version in 2004, it is a Google product, company with world's most widely used internet search engine. Google Scholar provides a simple way to broadly search for scholarly literature. The search options vary across many disciplines and sources like articles, dissertation, theses, books, abstracts, online repositories, universities and other web sites. Google Scholar helps you find relevant work across the world of scholarly research¹⁹.

Features of Google Scholar¹⁹: i. Search all scholarly literature from one convenient place, ii. Explore related works, citations, authors, and publications, iii. Locate the complete document through your library or on the web, iv. Keep up with recent developments in any area of research, v. Check who's citing your publications, create a public author profile.

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"¹⁹. It is suspected that citation count is probably the most important of several factors it takes into account²⁰.

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 scientists⁶.

Scopus: Scopus is a new abstract and citation database, as well as a web-based research tool provided by Elsevier. It contains over 20,500 titles from 5,000 publishers worldwide, 49 million records, 78% with abstracts, includes over 5.3 million conference papers and it provides 100% Medline coverage²¹.

Scopus tracks the same journals covered by MEDLINE and EMBASE, as well as many additional journals from a broad range of disciplines such as journal articles in chemistry, physics, mathematics, engineering, social sciences, psychology, economics, and general, biological, agricultural and environmental sciences. It is available online by subscription. It allows citation analysis.

A study from 2008²² compares PubMed, Scopus, Web of Science, and Google Scholar and concludes "PubMed and Google Scholar are accessed for free. Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy. PubMed remains an optimal tool in biomedical electronic research. Scopus covers a wider journal range but it is currently limited to recent articles (published after 1995) compared with Web of Science.

Web of Knowledge: Web of Knowledge (formerly known as ISI Web of Knowledge) is an academic citation indexing and search service, which is combined with web linking and is provided by Thomson Reuters. Its multidisciplinary databases, including the Web of Science (WOS), can be searched simultaneously. Web of Knowledge covers the sciences, social sciences, arts and humanities.

Features of Web of Knowledge²³: i. 23,000 journals (including over 12,000 unique titles in Web of ScienceSM), ii. 23,000,000 patents, iii. 148,000 proceedings, iv. 250 + product categories,

v. More than 100 years of back files, vi. Over 40 million source items and 760 million cited references.

The Web of Knowledge encompasses the following databases²³: i. Arts & Humanities Citation Index is accessed via the Web of Science, ii. Biological Abstracts, iii. Biosis Citation Index on the Web of Knowledge platform, iv. CAB Abstracts from CAB Direct, v. Current Chemical Reactions Web of Science access, vi. Current Contents Connect, vii. Derwent Innovations Index, viii. Essential Science Indicators, ix. Web of Science, x. Zoological Record, xi. Global Health from CAB Direct, xii. Index Chemicus, xiii. Journal Citation Reports, xiv. MEDLINE bibliographic database for the U.S. National Library of Medicine, xv. Science Citation Index Expanded, xvi. Social Sciences Citation Index.

Web of Science: It covers various academic disciplines including the sciences, social sciences, arts, and humanities, and across disciplines. However, Web of Science doesn't cover all journals, and its coverage in some fields is not complete.

Features of Web of Science²⁴: i. Coverage of over 12,000 journals from 256 categories, ii. Coverage of over 148,000 proceedings from the most significant conferences, symposia, seminars, colloquia, workshops and conventions worldwide, iii. Journal back files to 1900, iv. Cover-to-cover indexing, v. Cited reference and chemical structure searches, vi. Author identification tools, vii. Analysis capabilities and visualization tools, viii. Direct links to full text and library holdings information.

Web of science consists of six databases: Science Citation Index Expanded (covers over 7,100 major journals across 150 disciplines); Social Science Citation Index; Arts and Humanities Citation Index; Conference Proceedings Citation Index; Index Chemicus; and Current Chemical Reactions.

Web of science mentions multiple criteria for inclusion in its database. but an interesting point is, 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 analysis⁶.

Web of science is different from other "abstracts and indexing" databases because of the strict journal selection process. Journals selected for Web of science will have impact factor calculated by the Journal Citation Report (JCR). JCR offers a systematic and objective means to critically evaluate all internationally peer-reviewed influential journals, across all publishers with quantifiable, statistical information based on citation data, although the application impact factor is controversial.

Directory of Open Access Journals: The Directory of Open Access Journals (DOAJ) aims to be comprehensive and to cover all open access scientific and scholarly journals. It is maintained by Infrastructure Services for Open Access²⁵. As of April 2013, the database contains 9054 journals from over 1,000 publishers. The aim of the DOAJ is to increase the visibility and ease of use of open access scientific and scholarly journals, thereby promoting their increased usage and impact. The DOAJ aims to be comprehensive and cover all open access scientific and scholarly journals that use a quality control system to guarantee the content. In short, the DOAJ aims to be THE one stop shop for users of open access journals²⁶.

Selection Criteria²⁶

Coverage: i. Subject: all scientific and scholarly subjects are covered, ii. Types of resource: scientific and scholarly periodicals that publish research or review papers in full text. iii. Acceptable sources: academic, government, commercial, non-profit private sources are all acceptable. iv. Level: the target group for included journals should primarily be researchers. v. Content: a substantive part of the journal should consist of research papers. All content should be available in full text. vi. All languages.

Access: i. All content freely available. ii. Registration: Free user registration online is acceptable. iii. Open Access without delay (e.g. no embargo period).

Quality: For a journal to be included it should exercise quality control on submitted papers through an editor, editorial board and/or a peer-review system.

Periodical: The journal should have an ISSN (International Standard Serial Number, for information see <http://www.issn.org>). Online journals should have an eISSN.

Index Copernicus: 'Index Copernicus' (IC) is a web-based research infrastructure since 1999, which provides essential tools for scientists, research administrators and government agencies. Index Copernicus also provides an interactive and reliable scientists' evaluation system.

IC Journal Master List (JML) is a journal indexing, ranking and abstracting system. This service helps a journal to grow from a local level to a global one as well as providing complete web-based solution for small editorial teams. This service accepts peer-reviews and tracks manuscripts online and builds up relations with authors, reviewers and readers as well as increasing the international awareness of a journal. Currently JML includes over 8,000 journals from all over the world including medical and non medical journals²⁷.

WHO Global Health Library: The Global Health Library increases access to information and scientific evidence on health, particularly in developing regions. The Global Health

Library narrows the knowledge divide between developed and developing countries and enhances sharing and transfer of knowledge. It provides an international platform and virtual space for collection, organization, dissemination and access to reliable health sciences information sources and for sharing and networking between communities in different languages and formats²⁸.

Some of the regional offices of the World Health Organization (WHO) have produced their own Index.

Medicus, which includes regional medical journals that are not indexed in the western bibliographic databases. Some examples are as follows: i. Western Pacific Region Index Medicus (WPRIM) (www.wprim.org/), ii. Index Medicus for the WHO Eastern Mediterranean Region (IMEMR) (www.who.int/library/databases/emro/en/), iii. African Index Medicus (AIM) (indexmedicus.afro.who.int/); iv. Index Medicus for South-East Asia Region (IMSEAR) (imsear.hellis.org/);

HINARI: The HINARI (Health InterNetwork Access to Research Initiative) programme, established by WHO together with major publishers, enables developing countries to gain access to one of the world's largest collections of biomedical and health literature. More than 8500 journals and 7000 e-books (in 30 different languages) are available to health institutions in more than 100 countries, areas and territories benefiting many thousands of health workers and researchers, and in turn, contributing to improving world health²⁹.

Discussion: Bibliographic databases offer a centralized medium to search and access the journal contents and are important for increasing the visibility of a journal. Since these databases are so important for spread of knowledge and information, it is crucial that we have specific scientific databases fit for a particular purpose.

Some of the ideal characteristics that a bibliographic database should have are as follows⁶

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".

Specific Refined Search: At the same time, it should bring out the most relevant information to a user.

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.

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.

Advanced Citation Analysis: An ideal database should have sophisticated features to track and analyze citations.

Free: Ideally, such a resource should be completely free for users along with free access to full text articles

None of the databases perform well against above criteria. PubMed and Embase focus mainly on medicine and biomedical sciences, whereas Scopus, Web of Science, and Google Scholar cover most other scientific fields. PubMed and Google Scholar are free to use whereas Embase, Scopus and Web of Science are subscription based.

Indexed journals, particularly those indexed with PubMed or Thomson Reuters are considered to be of higher scientific quality as compared to non-indexed journals. But due to internet revolution indexation of medical journals has become a debatable issue.

Amongst the bibliographic databases in medical science, the Index Medicus and its online version MEDLINE are probably the most commonly used. But now preference is also given to impact factor of Thomson Reuters.

PubMed is the most used biomedical bibliographic database. However it has some limitations such as it excludes a significant body of gray 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.

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 PubMed³⁰. However PubMed still remains an important resource for clinicians and researchers²².

Conclusion

Having a basic knowledge of these databases and using them effectively will help researchers and authors to improve their writing and publishing skills. There is a scope and need to develop better bibliographic databases for which will have features like improved storage, improved user search, experience and quality. PubMed is one of the better tools to seek in the biomedical literature, with the exception of certain very specific documents and gray literature. PubMed need to evolve along the changing trends in medicine and become more inclusive in future. Biomedical researchers are in need of a

better scientific database which will archive all scientific content, more easy and free to use.

References

1. Sahu D.K., Small journal vs. high impact journal: Dilemma of a new author, *Indian J Pharmacol*, **41**(3), 95-96, (2009)
2. Feather John and Sturges Paul, eds. International Encyclopedia of Information and Library Science, 2nd edition, London: Routledge, 127 (2003)
3. Holzer L. and Eysers J., Medical Bibliographic Database, *Student BMJ*, **16**, 366-7 (2008)
4. Available from <http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp> (2012)
5. Available from <http://www.nlm.nih.gov/pubs/factsheets/jse1.html> (2013)
6. Kejariwal D. and Mahawar K.K., Is Your Journal Indexed in PubMed? Relevance of PubMed in Biomedical Scientific Literature Today, *WebmedCentral MISCELLANEOUS*, **3**(3), WMC003159 (2012)
7. Mahawar K.K., Role of Peer Review in Biomedical Publishing, *WebmedCentral MISCELLANEOUS*, **2**(4), WMC001863 (2011)
8. Conn V.S., Valentine J.C., Cooper H.M. and Rantz M.J., Gray literature in meta-analyses, *Nurs Res*, **52**(4), 256-61 (2003)
9. Available from <http://www.webmedcentral.com> (2013)
10. Smith R., Classical peer review: an empty gun, *Breast Cancer Res* **12**(Suppl 4), S13, (2010)
11. Available from <http://f1000research.com/>, (2013)
12. Available from <http://www.nlm.nih.gov/pubs/factsheets/medline.html>. (2013)
13. Available from <http://www.elsevier.com/online-tools/embase/about?q=what-embase> (2013)
14. Available from <http://www.elsevier.com/online-tools/embase/about> (2013)
15. Ng K.H., Peh W.C.G., Getting to know journal bibliographic databases, *Singapore Med J*, **51**(10), 757-761 (2010)
16. Available from <http://www.thecochranelibrary.com/view/0/AboutTheCochraneLibrary.html>. (2013)
17. Available from <http://clinicalevidence.bmj.com/x/index.html>. (2013)
18. Available from <http://www.cabdirect.org/web/about.html>. (2013)
19. Available from <http://www.google.com/intl/en/scholar/about.html>. (2013)
20. Beel J. and 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 (2009)
21. Available from <http://www.info.sciverse.com/scopus/>. (2013)
22. Falagas M.E., Pitsouni E.I., Malietzis G.A., Pappas G., Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses, *FASEB J.*, **22**(2), 338-42 (2008)
23. Available from http://thomsonreuters.com/content/science/scripts/elqnow/elqRedir.htm?ref=http://thomsonreuters.com/content/science/pdf/Web_of_Knowledge_factsheet.pdf (2013)
24. Available from http://thomsonreuters.com/content/science/scripts/elqnow/elqRedir.htm?ref=http://thomsonreuters.com/content/science/pdf/Web_of_Science_factsheet.pdf. (2013)
25. Available from <http://is4oa.org/doi/>. (2013)
26. Available from <http://www.doaj.org/doaj?func=loadTemplate&template=about&uiLanguage=en>. (2013)
27. Available from <http://journals.indexcopernicus.com/>. (2013)
28. Available from <http://www.emro.who.int/information-resources/health-knowledge-initiatives-ghl/global-health-library.html>. (2013)
29. Available from <http://www.emro.who.int/information-resources/hinari/hinari.html>. (2013)
30. Hendersen J. Google Scholar: A source for clinicians? *CMAJ* **172**(12), 1549–50, (2005)