

A Database Analysis Kit for Your Information Toolbox

JAY MACLEAN
ICLARM

After five years of using aquatic-science related computer databases via our online connection at ICLARM, I can confirm that they are a most important information tool in this field. I know this not only because of our own expertise but also through feedback from users of our Selective Information Service, a free service for developing-country researchers funded by the International Development Research Centre (IDRC) of Canada, which provides a package of references and addresses of workers in the same field and, usually, copies of important articles. A search of the relevant computer database, especially of the Aquatic Science and Fisheries Abstracts (ASFA) database, is the prime retrieval method. Our users like what they get, responding with comments like "very useful service", "very efficient", "excellent", and "the information you send us is irreplaceable".

While continuing to advocate database use I feel now that some caution is needed. Perhaps our users accept computer printouts as "gospel" because they are from a specialized service and after all, they do emanate from the infallible computer. Alas, the databases were, probably without exception, designed by committees and are compromises in many ways—in subject coverage, in depth of indexing and in document type coverage to name only the more obvious. Add to their dubious parentage the probability that the database searcher is not an expert in your field and you have a situation where you are "buying a pig in a poke", that is, taking a chance on the result.

A good response from the computer would be "WYSIWYG", a reborn software acronym that says "what you see is what you get". This article is intended to help you see what you get. That is very important in research, since full knowledge of previous work may save you repeating someone else's work, provide an essential contact address or save the embarrassment of having your results rejected by a journal for not citing important relevant articles.

A database should be equivalent to a bibliography over the years the former has been in operation. Unless research in the particular subject is dying, which is rare in the life sciences, the literature (number of articles published) should show a more or less logarithmic increase over time. Some illustrations are given in Fig. 1. If this pattern is not followed, then the reasons must be sought.

Database Analysis

Database analysis here simply means listing various parts of the references and extracting useful information, such as the identities of major authors, institutions and research topics or gaps in the database. No computer or calculator is needed.

The procedure I use is to make a table in which the columns are: authors (senior or all); major institution; broad subject; published year; and sometimes other columns, depending on the purpose. From such a table one can derive the temporal pattern of the references; the major authors and their affiliations; and the major research themes. One could add journal titles to see the most popular outlet, number of pages, document type, etc. If you captured your database search on a computer, some of these factors can be quickly derived by sorting.

Some Examples

An example of database use to find major authors, institutions and themes was given in the January issue of *Naga* (p. 17), in which *Eucheuma*, the red seaweed, was the subject and ASFA was the database. Seven authors and institutions and three major research themes were identified.

A second example is given in this issue (p. 22) on artificial reefs. The analysis is interesting in showing the heavy preponderance of USA authors and institutions in artificial reef research and the rapid growth of literature on this

subject. However, the computer database (ASFA again) was shown, by reference to a recent bibliography, to be deficient in its coverage of the subject. Compilers of the bibliography had clearly used sources not accessed by the ASFA compilers.

Another example, dealing with the well-studied common carp, is given by Bryan Pierce on p. 5. He concludes that "at best, more than half of the pertinent literature has been missed by any search".

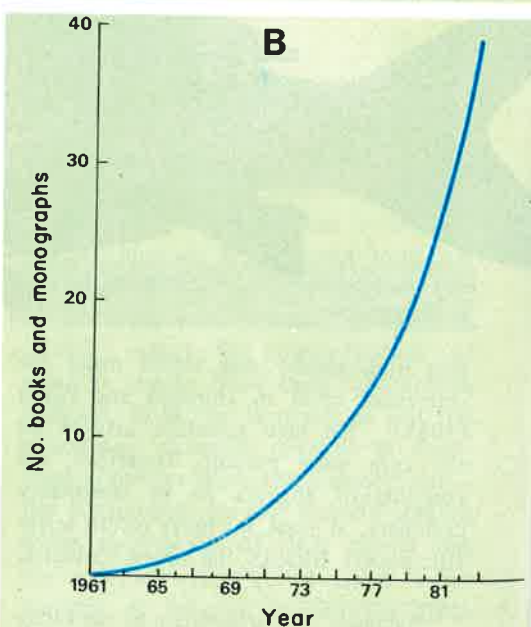
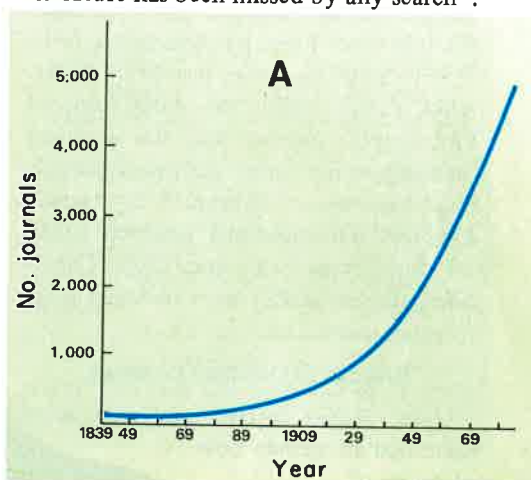


Fig. 1A: The logarithmic increase in number of scientific and technical journals in the USA, doubling every 15-20 years. B: Number of books and monographs on aquaculture published each year since 1961, doubling every five years.

Inadequacy of coverage by computer databases is not confined to fisheries. Fig. 2 shows the growth of rice research literature according to an annual bibliography published by the International Rice Research Institute (IRRI) and the performance of various databases which include rice. The disparity is alarming.

A different kind of inadequacy is shown in the computer-generated bibliography in Fig. 3 of Southeast Asian fish-

eries. The printout was published as a bibliography covering the years 1976-1982. Fig. 3 shows that 1979 was probably the last year properly covered. The following years should have shown numbers at least similar to the late 1970s.

The Old-Fashioned Bibliography

In the course of investigating the usefulness of databases, I had despaired

about the future of the old-fashioned manually prepared bibliography. It seemed that the computer could do in seconds what the bibliographer took weeks to accomplish. The bibliographer patiently seeks out all sources of publications about a subject, going from source to source until they are all collected. Database compilers work from a set of journals and whatever other material they happen across, such as conference proceedings and monographs.

For example, ICLARM sends copies of its publications to ASFA (c/o Fisheries Department, FAO, Rome) as a matter of course. If the mail goes astray, publications are not recorded in the database. And many institutions do not send their publications to ASFA at all.

Databases may hold the future but not the present; the gaps in their coverage are still too big. The old-fashioned bibliography still has a long life. Here is what Ms. Clare Cuerden, librarian of the FAO Fisheries Branch Library, has to say: "Online facilities are by no means available to everyone and in any case most databases store data only from the last 10-12 years. A printed bibliography can and does form a base from which to start looking for information and therefore published bibliographies with annotations, appropriate indexes and citations back as far as the turn of the century (as for example the ICLARM bibliographies on tilapia), still have a valid place in research work and scientific libraries will continue to buy them". Considering that her library is the nerve center for developing-country fisheries documentation, these are encouraging words for bibliographers.

Researchers should always treat a database printout with caution; remember WYSIWYG. You can expect enough coverage to identify major authors, institutions and research areas. For complete coverage of a subject, you also need to scan all available journals and bibliographies, contact other authors and where possible enlist the help of librarians and information services. In ICLARM's Selective Information Service, we combine database use with the expertise and collections of the scientific staff as well as the resources of the ICLARM library.

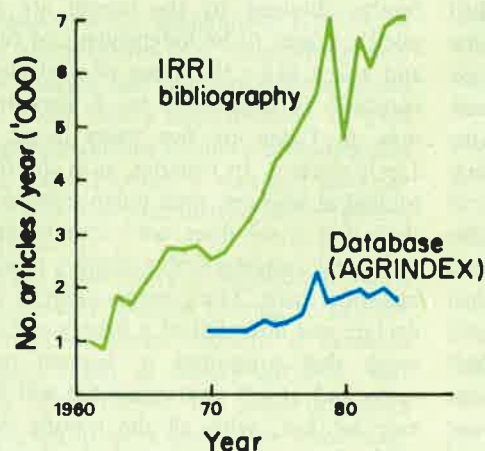


Fig. 2. (left) Comparison of rice research articles since 1961 in a rice bibliography by IRRI and an agricultural database, AGRINDEX. Note: the database began only in 1970. Fig. 3. (below) Number of articles by year in a computer-based bibliography of Southeast Asian fisheries, 1976-1982.

