

Regular columns

Savvy searching

Free government databases

Savvy searchers always keep an eye on government databases. They have known for years that some of the databases of the US government are available free of charge on the Web for anyone, anywhere without showing proof of having paid US federal taxes. Not all of these databases have the software trimmings that the best online and CD-ROM versions like Ovid and KnowledgeFinder have been offering, but they have been more than satisfactory for most of the users most of the time. Many of the free government databases are put on the Web by the agency that is in charge of producing the data, although there are also third parties outside the government that offer government databases free to increase traffic to their Web sites. The most notable are the many versions of MEDLINE beyond the one offered directly by the government.

The venerable free databases

The ERIC database was the first to go free courtesy of the Information Technology Clearinghouse of ERIC at Syracuse University <<http://askeric.org/Eric/>>. It has excellent free text searching capabilities, a perfect solution for those who want to find a few good articles on a subject without first learning the intricacies of the ERIC Thesaurus, like undergraduate students. Laudably, another ERIC Clearinghouse, on Assessment and Evaluation, introduced another version, serving the other camp by bringing out the best of the ERIC Thesaurus (<http://www.ericae.net>). Both versions run circles of remarkably poor implementation of ERIC – like the one by CARL that many libraries subscribe to for thousands of dollars annually. Even the DIALOG version leaves much to be desired, especially in the handling of the thesaurus with chopped off scope notes. Others are paying dearly for some ERIC records when they search the *Information Science Abstracts* database that has at least 20,000 ERIC records; at ISA's prices this makes the pure ERIC versions look like a good deal. There is hardly any reason to pay

for ERIC searches with such two excellent versions available for free and now covering the entire timespan of this database.

MEDLINE is probably the best known and most used government database. There are about a dozen free versions now, such as the commercial health-related Web sites Medscape (<http://www.medscape.com>), HealthGate (<http://www.healthgate.com>) and IntelliHealth (<http://www.intelihealth.com>). Arguably, the one that has the best search features is PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>). Grateful Med (<http://igm.nlm.nih.gov/>) stands out by virtue of offering a variety of specialty health-related government databases like AIDSLINE, AIDSDRUGS, BIOETHICSLINE, HEALTHSTAR, HISTLINE, and TOXLINE.

Patent databases have been notoriously expensive both for searching and full-record printouts. IBM changed this overnight when a couple of years ago it released a top-notch and free version of the US Patents database (<http://www.patents.ibm.com/>). This has awesome search capabilities that only Questel can compete with, a masterfully designed interface, and several print formats, including PDF version of patents. The USPTO (<http://www.uspto.gov/patft/>) itself launched its own version recently, but once you have used the splendid IBM version you will not want to look elsewhere for free patent information.

The relative newcomers

Beyond the educational, medical and patent databases of the government, its other agencies have also been busy making available their indexing and abstracting databases with links to the increasing number of full-text documents in digital format.

The National Agricultural Library offers the entire AGRICOLA database (<http://www.nal.usda.gov/ag98/>) that has the same content that you would find in many commercial services like OCLC and DIALOG for a fee. The software is nothing to write home about but does an adequate job in finding relevant articles. Importantly, it has hot-linked URLs in many records that DIALOG still has not implemented, depriving users of instant access to thousands of full-text HTML and PDF documents. It is also remarkable that you may display the records in MARC format.

The Department of Justice, more specifically the Justice Information Services,

offers for free the National Criminal Justice Reference Services (NCJRS) Abstracts (<http://www.ncjrs.org/database.htm>). Its software has power search features like proximity searching, but its real stunner is a pattern searching feature not available in the CD-ROM and DIALOG online versions. It recognizes misspelled words and finds the records with the correct spelling. If you misspell “cannabis” as “canabis”, it still finds all the 410 records that include cannabis and 40 other records with variant and incorrect spellings. It is a significant bonus that most of the documents published by the sponsoring agencies of NCJRS (like the National Institute of Justice, the Office of Juvenile Justice) are available in one or more full text format, including ASCII, HTML and/or PDF formats.

The Department of Transportation has a beta version of its Transportation Information System (TRIS) database (<http://tris.amti.com/sundev/about.cfm>) of well over 400,000 records. It does not have the records from the International Road Research Centre that the DIALOG version has, but even without this 70,000 record subset TRIS is a pretty comprehensive database for most subjects. The records already have links to full-text government reports and statistics, but these were cold links as I went to press with this column. I hope that by the time you read this the links will be hot. It is to be noted that the DIALOG version also has only cold links as in all of its databases.

The Department of Energy launched last year the PubSci service that includes, among other features, a subset of the Energy Science and Technology database for free. This is not the same full database carried by a number of commercial vendors, because the free version excludes non-US records. I am sure many of the users can live with this limitation and restrict themselves to the millions of US records in the knowledge that they save a lot of money by not paying, for example, the \$5.10 DialUnit charge and the \$1.55 display/print fees.

Statistics galore

The US Census Bureau makes available free of charge a plethora of excellent statistical

collections that cost an arm and a leg through commercial vendors. Beyond such essential works as Statistical Abstracts (<http://www.census.gov/statab/www/>), you will find various factographic databases with highly intuitive graphical interfaces as is the case with American FactFinder (<http://factfinder.census.gov>). The Census Bureau’s hub is the perfect starting point at <http://www.census.gov> for exploring the sources. The 1990 census data are far more accurate through the above government site than the versions through commercial vendors. In DIALOG at least the savvy searcher can exclude, for example, the 6,000 records where the population reported is zero and hence all the other statistical tables are filled with zeroes – not particularly relevant statistical information. Though you do pay the browsing and searching charges at \$7:35/DialUnit to find and exclude the useless records, it is still cheaper than the \$13.40 charge for the full records that are filled with zeroes and other useless values. If this sounds incredible, have a look at some of the weird examples from the pricey Population Demographics database at <http://www2.hawaii.edu/~jacso/extra/>.

One of the most widely used government databases, NTIS is not available free of charge on the Web, but there is a version that provides free bibliographic citations from NTIS and charges just a dollar for the abstract. Although the Web-borne equivalent is not really free, it almost feels like it compared with the commercial alternatives that charge much more per record. The provider is no surprise for savvy searchers – Northern Lights that has revolutionized the commercial online information services by free abstracts and very reasonably priced ASCII full-text format of the source documents at (<http://usgovsearch.northernlight.com/>)

Even sweeter are the free versions of proprietary databases or their subsets that can replace the fee-based information services or reduce their use and associated costs significantly. I will discuss some of these in a later issue.

Péter Jacsó

University of Hawaii at Manoa