

Database Subsets

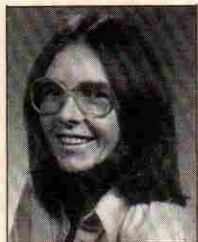
By Carol Tenopir

LOCALLY held databases may be one answer, for librarians who would like to allow patrons to conduct database searches themselves, but are hesitant to commit the time and resources necessary for direct end user access to on-line systems. Some database publishers are now selling portions of their databases on floppy disks or other storage media. Microcomputers and advances in computer storage technology make such products feasible.

Subset characteristics

Database subsets are created in two ways. Either the subscriber submits a customized subject profile to the publisher or the publisher offers databases on predetermined topics. If the publisher offers a subscription service, disks or tapes with new citations on the chosen topics are sent at regular intervals, usually monthly or quarterly. Prices vary from a flat yearly fee to a charge for each citation. Most publishers either provide the needed search and retrieval software as a part of the initial subscription or offer it for an additional charge.

With database subsets, librarians return to a familiar purchasing model. There is a one-time subscription charge for the information, regardless of how much of it is actually used, instead of the access charges for information used in the newer online model. Librarians are comfortable with this method because it facilitates budgeting and existing ordering and receiving procedures. The cost of locally-held materials (whether books, magazines, or databases) is part of the library's budget. The cost of access to the information is equal for all users. Access to information again becomes centralized in the institution that collects materials from many different sources.



Carol Tenopir is Assistant Professor at the Graduate School of Library Studies, University of Hawaii, Manoa

A small but growing number of publishers are beginning to offer database subsets in addition to their online and print products. The scope of the subsets offered varies from floppy disks with 200 records each to laserdisk subsets with over one million records each. Costs vary accordingly.

Floppy disk subsets

ERIC MICROsearch is produced by the ERIC Clearinghouse on Information Resources. Subscribers may choose between two subject profiles that reflect the subject areas of this clearinghouse: Educational Technology or Library/Information Science. Each quarter subscribers receive 200 to 600 new records per subject. Since each floppy disk holds 200-300 bibliographic records without abstracts, two disks per subject are usually needed.

Librarians can also opt for retrospective subject bibliographies on Computer Literacy, Selection and Evaluation of Computer Software and Hardware, Basic Skills, or Resources for Library Instruction. These disks contain references on the subjects taken from a three- to six-year portion of the ERIC database which were closed from March 1983 to October 1984. Because they are out-of-date, updated versions of some may be forthcoming.

Subscribers to *ERIC MICROsearch* pay an initial subscription fee of \$2, which includes the *MICROsearch* retrieval programs, one data disk on a subject of your choice, and a users manual. A quarterly subscription or a nonsubscription disk costs \$7.50.

ERIC MICROsearch runs only on the Apple II Plus and IIe computers. A simple step-by-step search and retrieval program written in *Apple BASIC* allows limited Boolean searching of subjects and the ability to print a bibliography from each update disk. Future plans include an IBM PC version written in *PASCAL* with more search capabilities and disk capacity. Contact: ERIC

Clearinghouse on Information Resources, 130 Huntington Hall, Syracuse University, Syracuse, NY 13210.

Microcomputer Index (MI) indexes about 65 popular microcomputer journals. Its new producer, Database Services, P.O. Box 50545, Palo Alto, CA 94303 (415) 948-8339, now offers *MIND (Microcomputer Index on Disk)*, a one-time subset on one double-sided, double-density data disk and a search and retrieval program on an additional disk, that provides over 2500 citations to reviews of software, hardware, and books from *MI*'s 1984 entries for \$125. *MIND* and the retrieval program *MicroAccess* (a product of Knowledge Access Inc.) are available only for the IBM PC. Database Services is now considering offering additional subject profiles on a subscription basis. They dropped plans to offer user profile subsets when a market survey showed most users prefer more general standard profiles.

MicroAccess is also available for other databases listed in the catalog available from Knowledge Access (445 E. Charleston Road, Palo Alto, CA 94306). The *MicroAccess* program is menu driven and allows Boolean logic searching.

For libraries interested in offering access to biological information, *BIO-SIS B-I-T-S* (BIOSIS Information Transfer System) has provided subsets since 1983. Subscribers receive floppy disks (*MICRO/B-I-T-S*) or magnetic tapes (*MACRO/B-I-T-S*) each month with new BIOSIS records. Options available include formulating a customized search profile, paying BIOSIS a fee to help develop a customized profile, or choosing from a list of over a dozen "pre-profiled" subjects.

Subscription prices for the customized search profile are based on an estimate of the number of citations that will be retrieved each year. *MICRO/B-I-T-S* charges 40¢ per citation with abstract or 20¢ per citation without for up to 10,000

records. Quantities over 10,000 are 30¢ each with abstracts and 15¢ each without. Typical costs vary from a low of \$100 for up to 500 records without abstracts to \$3200 for up to 10,000 records with abstracts. There is an additional charge of just under \$3 per disk. (Each floppy disk holds from 150-400 references without abstracts.)

Perhaps of more use to most libraries is the pre-profile option. BIOSIS offers *B-I-T-S* subscriptions on such broad topics as aging; industrial health and toxicology; pesticides; addiction; anticancer agents; and population, fertility, and birth control. The estimated number of citations per year is different for each topic, varying from 1000 to 2800 records, so yearly subscription rates vary from \$300 to \$840.

MICRO/B-I-T-S is available for most major microcomputers. *BioSuperfile*, a search and retrieval software package especially designed to search *B-I-T-S* citations, may be purchased from BIOSIS for \$100. *B-I-T-S* is also compatible with other retrieval programs. Contact: BIOSIS, User Services Department, 2100 Arch Street, Philadelphia, PA 19103, (215) 587-4800 or (800) 523-4806.

Magnetic tape subsets

Before online access was widespread, many databases were available only via magnetic tape subscriptions for mounting on your large in-house computer. Some publishers have continued to offer tapes of their entire database, but this requires large computer capabilities and gives you all data in the database whether or not it is of likely interest. Subsets on magnetic tape are a newer and more customized option (as mentioned, *BIOSIS B-I-T-S* provides the option of magnetic tape subsets (*MACRO/B-I-T-S*) in addition to their floppy disk option.

In late 1984, the National Library of Medicine (NLM) announced plans to make *MEDLINE* subsets available on magnetic tape and floppy disk. So far only the magnetic tape option is available.

By this fall, subscribers may choose either a "predetermined" subset where NLM has designed the subject profiles or design a customized profile ("individualized subsets"). Currently only the monthly updated individualized subsets are available. Subscribers must have their own search software.

NLM offers two plans. The "personal use only" plan agreement restricts access to one individual who pays a flat yearly fee of \$1000 for predetermined subsets or \$1500 for individualized subsets. Of more interest to libraries are multiple use agreements which allow use of the subsets by users in one institution. A \$10,000 fee covers

annual subscriptions for current year magnetic tape subsets for a single institution, with a \$1000 charge for another institution sharing the subset.

Another option is an annual subscription fee based on a connect hour charge and citation printing charge (the basic minimum charge is determined by what percent of the *MEDLINE* database is included in your subset). Annual minimum charges for 1985 data vary from \$1250 for up to 25 percent to \$5000 for from 76 percent to the full file. (Approximately 22,000 records are added each month to the full *MEDLINE* file.) Backfile subsets are priced separately. Use charges of \$4 per local primetime hour or \$3 per nonprimetime hour plus a penny per printed citation offset annual minimums. Institutions must be able to monitor the amount and time of use and report it to NLM to be eligible for this option. Contact MEDLARS Management Section, National Library of Medicine, Building 38A, Room 4N421, Bethesda, MD 20209, Attn: Subsets.

Laserdisk systems

International Thompson/Carrollton Press and Information Access Company are each using laserdisk storage devices (the latest storage technology) to make subsets of their databases available on a subscription basis to libraries for direct patron access. Each laserdisk holds from 800 megabytes (800 million characters) to a gigabyte (one billion characters) of information (most floppy disks hold only a million characters).

Subscription prices for these databases include the laserdisk hardware and microcomputers in addition to the search software and data disks. This means a substantial investment by the subscribing library, but these systems open up new possibilities for patron access to large general interest databases.

Libraries know about the Carrollton Press *REMARC* and *MARC* locally-held databases for retrospective conversion. The company is now offering a laserdisk system for patron access. *MARVLS* (the *MARC* and *REMAC* Videodisk Library System) provides a million record subset of the *MARC* and *REMARC* files stored on a single five-inch laserdisk. Contact: Carrollton Press Inc., 1611 N. Kent St., Arlington, VA 22209, (800) 368-3008.

Simplified searching can be done on the search software provided or users can purchase powerful Cuadra Associates, 2001 Wilshire Boulevard, Suite 305, Santa Monica, CA 90403, (213) 829-9972 *STAR* search software.

InfoTrac, from Information Access Company, 11 Davis Drive, Belmont, CA 94002, (800) 227-8431; (415) 591-2333, has been available since March 1985. Initially it contains approximately a half million references

from over 1000 business, technical, legal, and general interest publications reflecting "the periodical collections of major research libraries." Citations are included from IAC's *Business Index*, *Legal Resources Index*, *Trade and Industry Index*, *Area Business Collection*, *National Newspaper Index*, and *Magazine Index*. Journal citations go back to January 1982, newspapers only 90 days.

InfoTrac, updated and cumulated monthly, carries an annual subscription price depending on the hardware configuration chosen. The most complete system includes four workstations (four micros and four printers), the laserdisk player and controller, a security enclosure, hardware maintenance, menu-driven search software, and monthly cumulated updates at \$16,000 for the first five years. At the end of five years, the hardware becomes the property of the institution and the yearly price is reduced to \$9500 for the data disks and maintenance of the disk player and controller. Yearly subscription price with three workstations is \$15,000 or with two workstations, \$14,000. If you have your own IBM PC with 256K and one disk drive, you may get only the player/controller and monthly disks for \$12,000 per year.

IAC's primary market for *InfoTrac* is academic, research, and large public libraries. More specialized subsets may be developed in the future for smaller libraries or special libraries. IAC may develop different disks for different subjects or provide databases from other publishers through the *InfoTrac* system. This would allow more publishers to provide database subsets without incurring the development and start-up costs and provide a single source for library users to access many different kinds of locally-held databases.

Clearly new developments in hardware and storage technology have made database subsets an increasingly attractive option for publishers and libraries. Subsets provide an alternative to downloading and can allow unrestricted patron access to database information.

The biggest drawback (other than the sometimes hefty prices) is the variety of types of microcomputers and retrieval software needed to access the databases. Nearly every subset requires a different software package or different kinds of hardware. The easiest way to cope with this confusing diversity is to select the one or two database subsets that are of most interest to your patrons. A better solution may emerge soon when a few subset providers such as IAC and Knowledge Access offer many subsets from a variety of publishers. Ironically, it is this "vendor" service that made online database searching feasible in the first place.

