LIBRARY JOURNAL digital
Visit Our Web Site
www.libraryjournal.com

BY CAROL TENOPIR

Time To Redefine *Database*

IT USED TO BE EASY to define a database. It was "a continuously updated computer file of related information, abstracts, or references on a particular subject, arranged for ease and speed of search and retrieval" (ODLIS: Online Dictionary of Library and Information Science, www.wcsu.ctstateu.edu/library/odlis.html). A database publisher such as Psychological Abstracts or Engineering Information was responsible for creating the content (and perhaps distributing the printed indexes), but the vendor, such as Dialog or SilverPlatter, transformed the content to make it searchable and then provided access.

This model has served us since the early 1970s. Sometimes large database producers such as H.W. Wilson or Information Access Company/Gale Group also serve as aggregators and vendors for their own and other companies' content, but they kept the original database model. When SilverPlatter, OCLC, or Dialog say they offer hundreds of databases, we know that means hundreds of choices of separate containers of information, each provided by a content producer.

As described by the *Gale Directory of Online, Portable, and Internet Databases* (2001 ed.), databases may be bibliographic (indexing/abstracting), directory, statistical/numeric, full text, or of other types. No matter the type,we in the library profession (at least in North America) we call each of these individual files a *database*.

There were always exceptions to this model. Lexis-Nexis preferred a more fluid model from the start. Most Lexis-Nexis sources come directly from the primary publisher (or from the courts) rather than from a second-party indexing company, so the size of the Nexis system is better described by the number of full-text titles included rather than the number of databases. The company prefers the term *files* to *databases* and uses "libraries" to describe large groups of files. A file can be a single journal title or a topic-related or date-related group of sources.



Carol Tenopir (ctenopir@utk.edu) is Professor at the School of Library and Information Science, University of Tennessee at Knoxville

A troubling term

I have long been troubled by the term *database* for several reasons. First, only information professionals use it to mean a file of discrete content, and even we use it to mean something different when we are describing "database management system" software. The distinction sometimes made (especially in Europe) between individual databases and databanks (systems) never really caught on. Moreover, the term is used to describe a huge variety of types of information sources. An index is quite different from an encyclopedia, yet the digital versions of each are called databases. The term *full-text databases* is so broad as to be meaningless. Full text includes everything from a collection of journal articles to newspapers to almanacs to even fiction.

Finally, counting databases when comparing systems has never been easy, as each system might segment the same content differently. For example, Lexis-Nexis separates Medline by article characteristics, such as articles that report human studies only or those written in English, while Dialog segments this huge database just by date.

Megadatabases

It may be time to put the term *database* to rest, as we enter an era that is dominated by full-text articles linked to bibliographic records, hybrid web sites that contain all types of information in one site, and fluid segmenting of content, depending on subscriptions or the individual user's needs. Digital indexes, journals, encyclopedias, or newspapers need no special designation. Even the lines between databases within a single system are blurring.

For example, in the fourth quarter of 2000 Gale Group announced that the InfoTrac family of databases would be offered as a single megadatabase option, called OneFile. It provides access to bibliographic records from over 6200 magazines, journals, newswires, and newspapers, some of which have files that go back 20 years. Nearly 2700 of the periodicals also include full-text articles.

All told, OneFile includes over 16 million records (either bibliographic or full text) with the popular InfoTrac interface and search engine. Libraries that now subscribe to one or more InfoTrac databases (such as Academic ASAP, LegalTrac, or National Newspaper Index) know the databases provide a mix of scholarly journals, general magazines, specialized literature, and general interest news sources.

Avoiding overlap

The sum of records offered in each of the separate InfoTrac databases would appear to be much greater than 16 million, because, although InfoTrac offers over 40 separate databases, there is much overlap among them. Academic ASAP, for example, includes sources that are also a part of Expanded Academic ASAP. InfoTrac Kids Edition includes a subset of InfoTrac Student Edition. OneFile removes the considerable duplication.

Since it combines diverse individual databases, OneFile includes many subjects. Almost one-third of the titles are general interest magazines, while a quarter are academic journals, and another quarter are business or technical publications. The remaining titles are law journals, healthcare periodicals, and newspapers or newswires.

Large libraries or consortia may want to subscribe to OneFile, but unless a library now subscribes to most of the InfoTrac databases separately, the cost of OneFile may be prohibitive. Individual libraries can still subscribe to specific InfoTrac databases.

I asked Bonnie Hawkwood, product manager for InfoTrac OneFile, why a library should choose OneFile rather than individual InfoTrac databases. She pointed out that "for some interdisciplinary topics, accessing a megafile allows

the patron to get comprehensive answers quickly without having to guess which combination of databases to use."

Megafiles may better match the way today's library users think of online resources. According to Hawkwood, "Users look at the Internet as a vast resource of information. Libraries are competing for the user's attention and frequently need to remind patrons that subscription databases are the most effective way to conduct research. The 16 million articles in OneFile may have overwhelmed students ten years ago, but users are now more savvy."

OneFile is part of an emerging pattern of giant general interest databases, such as those sold by Lexis-Nexis Universe and ProQuest, two of Gale's main competitors in the academic library market. So, are individual databases becoming obsolete? Hawkwood speculates that "the traditional publisher/aggregator model was developed when transactional pricing was the norm and became further defined during the CD-ROM era." Now that the market has become dominated by online subscription services, it's no longer technically or financially necessary to break up databases.

Hybrid resources

The web model and mindset have spawned another trend: combining a variety of types of databases into a single, subject-related resource center. Health and Wellness Resource Center, Gale's latest hybrid database, integrates indexing and full texts from 400 medical journals and 2100 general interest periodicals, with 800 pamphlets and dozens of healthcare reference books.

Health and Wellness Resource Center will replace the Health Reference Center databases. It joins other Resource Centers introduced recently, including those for Business & Companies, History, Literature, and various Resource Centers for students from college to junior high. All mix the periodical indexes and full-text databases created by Information Access Company/InfoTrac with the many reference books published by Gale Group. This allows a user to find summary information from an encyclopedia or directory and then get more details from periodical articles.

ProQuest's dual track

ProQuest (formerly UMI/Bell & Howell Information and Learning) also sells its content in a megafile or as separate database pieces. ProQuest 5000 combines the most popular ProQuest content, including bibliographic and full-text records for nearly 5000 newspapers, magazines, and journals for both general interest sources and more specialized sources (such as the business database ABI/INFORM).

Most ProQuest subscribers, however, still select individual databases. Tina Creguer, director of communications/public relations for ProQuest, tells me that "the large integrated set is typically purchased by libraries that so value the individual elements like ABI/INFORM and Research Library that they decide to purchase nearly every database we offer."

Traditional descriptive terms—such as dictionaries or indexes—are more useful

This suggests that the concept of separate databases is not yet obsolete. Creguer explains that:

> [L]ibraries still use a subject basis to evaluate content. There is no one-size-fits-all solution. Business libraries need powerful and deep business products, engineering libraries need deep and powerful scientific databases, law libraries need law databases. The megadatabase implies that there is a place in libraries for a huge generic database. While we do have a suite of all our major products, we do not subscribe to the notion that it should contain generic content.

She believes that "bigger is not always better. We do not add content to ProQuest simply to build its bulk." ProQuest databases are indexed with subject-specific terminology or in a way that matches the main target audience. ABI/INFORM, for example, uses indexing aimed at business researchers.

H.W. Wilson's new mix

The H.W. Wilson Company has also started grouping its considerable content into megadatabases and resources that mix diverse content. Wilson OmniFile Full Text Mega includes all indexing (1982 to present), abstracting (1984 to present), and full texts (1994 to present) from six Wilson databases (Education, General Science, Humanities, Readers' Guide, Social Sciences, and Business Full Text files), plus full-text articles from an additional five databases. OmniFile Full Text Select includes full-text records only from 1994 to the present, selected from many Wilson Full Text databases. Wilson's meticulous editors have reconciled indexing differences among the databases in the OmniFile products, improving searching. Wilson's name and company authority files also carry across the OmniFiles.

Even consistent indexing doesn't mean megadatabases are best for everyone. Wilson's web site explains that this "one-stop searching" approach makes "OmniFile especially convenient for users with inquiries in areas that cross disciplines—ethnic studies, women's studies, history," for example. Although they may be good for users who don't know where to begin, OmniFiles may prove overwhelming for some users.

For biographical information needs, Wilson is also moving into the hybrid model of collected information with its new Wilson Biographies Plus Illustrated. All people-related information resources created or licensed by Wilson are mixed in this attractive resource, including over 100 books, 72,000 biographical profiles, 26,000 images, and links to articles from 4000 periodicals. A user can read a short profile about someone, view a gallery of pictures, and then find more detailed information in journal or magazine articles.

Retiring the term *database*?

Is it time to retire the term *database*? New hybrid products like Wilson Biographies Plus Illustrated or Gale Group's Resource Centers go far beyond the original meaning of the term. Megadatabases blur the lines between separate files. Perhaps it is time to recognize that reference sources come in many forms, whether they are print or digital, and that *database* is no longer descriptive enough. Traditional descriptive terms—such as dictionaries, encyclopedias, journals, and indexes—are more useful, with the digital versions allowing users to link among the types to get whatever amount of information is needed on a topic.