

LJ INFOTECH □ ONLINE DATABASES □

BY CAROL TENOPIR

Getting What You Pay For?

THE OLD ADAGE "you get what you pay for" doesn't always apply to the World Wide Web. Certainly, you frequently encounter worthless information (or wrong information, which is worse), but there's a wealth of very useful and valuable free information. Many government agencies, professional societies, medical support groups, museums, and libraries create content-rich free web sites.

Still, although the number of authoritative sites grows daily, not every topic is successfully or best searched on the web. Library funders, however, may not understand why libraries also continue to pay for a variety of electronic information sources. Isn't everything needed, they ask, available for free on the web?

Commercial online services and fee-based web sites, of course, have a vested interest in the answers. Recently I asked representatives from Lexis-Nexis and Factiva, a Dow Jones & Reuters Company (formerly Dow Jones/Reuters Business Interactive) what they suggest librarians tell questioning funders. Although their answers often relate specifically to their own online systems, the arguments are broadly applicable.

Lexis-Nexis vs. the web

Jill Konieczko, a regional representative for Lexis-Nexis, points out that L-N surpasses the web in several ways. I've grouped her points into six themes:

Size. A new web page is put up every four seconds, but L-N puts up 15.7 documents per second; L-N includes approximately one billion documents, the web just 40 million.

Organization. L-N information is grouped into easily accessible and logical libraries or group files, while infor-

mation is scattered on the web. Similar information resources are grouped together on L-N for easier access (such as caselaw and filings or Investext and Dun & Bradstreet information), but web sites from different providers are maintained separately.

Archives. L-N maintains deep archives, usually from the time a resource was added to the system, while there is little or no archive on much of the web.

Searching. L-N includes powerful search tools in both its Boolean logic search engine and its relevance ranking/statistical search engine (FreeStyle). These tools include the capacity to improve precision with features such as focus, segments (fields), and proximity connectors. Advanced search features on many web search engines are often limited to Boolean and/or and adjacency.

Response time. L-N claims 99.84 percent reliability and less than six second response times on searches, while response time on the web is highly variable, not to mention dependent on local conditions.

Customer service. L-N (and other online services) offer 24-hour customer service for technical and search assistance. On the web, well, you're on your own.

While people may quibble about statistics used for size, and response time can be improved locally, organization of resources and efficiency of searching almost always clearly favor fee-based online aggregator services.

Factiva vs. the web

Perhaps the best way to compare the relative merits of web resources with fee-based services is to conduct head-to-head tests, so Factiva commissioned a "White Paper" by writer and information entrepreneur Mary Ellen Bates: "Selecting Business Intelligence Sources: The Public Web vs. Value-Added Online Services." (Dow Jones Interactive subscribers can find the paper in PDF format.)

Bates tested three common ways to access information: the free web, fee-

based web sites (such as Northern Light and the Electric Library), and value-added business information services (like Dow Jones Interactive).

Bates's six business queries ranged from the straightforward ("we need a copy of the article that appeared in yesterday's *Washington Post* about the EPA's rules..." and "a company profile of Genentech") to the more complex ("market analysis of nicotine patches" and "product development of in-seat power supplies for airline passengers"). The paper includes in-depth analysis for every search, including cost, completeness of answers, time needed, and depth of archives.

When comprehensive or complete answers to complex information are required, searchers must access a variety of sources. Bates found that "at least part of every business question could be answered using web resources alone or in conjunction with fee-based sources. However, in almost every case, complete answers were only available through the value-added service."

The web compared well on very simple searches, such as finding a recent newspaper article, but, as the questions became more complex, the web took longer to search and retrieved less useful information. Because most commercial online services aggregate different types of sources—including indexes and abstracts, full texts of journals, and statistical sources—and because they have deeper archives, users find complete answers faster.

Research time, or cost-efficiency, is another important area where commercial online services compare well (at least for an experienced online searcher). In Bates's study, every question (except retrieving a current *Washington Post* article) took longer on the web, while yielding a less complete answer.

She also calculated "true research costs" by adding in personnel costs. In many cases, these actual costs came out lower for Dow Jones Interactive than for the web or fee-based web sites because of time saved. In one question, the searcher spent 33 minutes on the



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

ONLINE DATABASES

web (calculated at a personnel cost of \$25.74) and found no useful material. Although the cost was \$76.47 on Dow Jones Interactive, it took the experienced searcher only 14 minutes to get a complete answer. For some but not all the other questions the "true research costs" were actually higher on the web.

If asked to assemble a presentation listing relative strengths and weaknesses of the free web vs. commercial online services, you might include the following points that are detailed in Bates's White Paper.

Strengths and limitations

For the free web, the strengths: many expert sites, maintained by experts on a particular topic; links to other resources; government web sites offer reliable information; information available at little cost for good searchers; cheap access to newspapers and other current information.

The limitations: no capacity to aggregate or integrate searches; limited archives; instability of sites; limited power searching features; questionable reliability and accuracy of sites. Another limitation of web search engines is the growing practice of adjusting relevance rankings if a site pays a premium to the search engine (or web sites may just manipulate the relevance ranking of search engines by "spamindexing").

For fee-based web sites (such as Northern Light or Electric Library), the strengths include: usually provide an archive as well as current information; allow power searching; provide copyrighted materials such as journal and magazine articles for a relatively low fee; provide collections by topic.

The limitations: usually require credit cards for purchase; encourage impulse buying of articles when the same source may already be available (from the library) to the person doing the ordering; usually limited to one type of information.

For commercial online services, Bates highlights five main strengths of services such as Dow Jones Interactive, Lexis-Nexis, and Dialog. They provide aggregation of hundreds of sources that can be searched simultaneously; extensive archives; a wide variety of types of sources; powerful search and output features; and automatic alert or current awareness search services.

Bates does not summarize limitations of value-added online services, but

some should be mentioned. High out-of-pocket costs are the most obvious drawback. Another is the need to have someone trained to get the most out of the many power-searching features. Reliability is always cited as a strength of fee-based publications, but sometimes resources such as directories are inaccurate because they are not kept up-to-date by the publishers. Online aggregators make

Most often, users must search both the web and traditional online sources for complete information

no claims as to the accuracy or reliability of the sources they provide, although, unlike with the web, there is always a publisher for these resources with a reputation to uphold. Also, web sources usually include graphics and links.

Other comparisons

Bates was not the first expert searcher to run tests to compare the web with commercial online services. In a 1998 column ("Online Meetings of Minds," *LJ* 10/1/98, p. 38,40), I mentioned Susan Feldman's "Internet Search-Off," results of which were presented at the 1998 National Online Meeting (a preliminary description was published in *Searcher*, 2/98 [p. 28-35]). Her findings remain valid today.

Feldman challenged experienced searchers to run the same searches in Dialog, Dow Jones Interactive, and on the free web. Like Bates, Feldman found that web searching takes much more time to get the same or less complete results for many topics. She found the web to be better for current information about specific, small companies; locating pictures and illustrations; product information directly from a company; and current medical statistics.

Most often, users must search both the web and traditional online sources for complete information. Feldman reported they complement each other, especially when searches seek a range of items: standards; general interest articles; popular subjects;

organizations or directory information; reviews, evaluations, and how-to information; government regulations; competitive intelligence; and information about "obscure topics."

Traditional online services alone were found to give better results for more specific items: archival or longitudinal information; chemical or electrical engineering; history; market or industry reports; current drug studies; industry newsletters and journals; financial industry coverage; scholarly journal articles; high-quality information; and quick searches when the user knows the information is likely to be there and time is a factor.

Like Bates, Feldman reminds fellow-searchers to "always remember—time is money. Free information that takes too long to find and format is expensive information."

In 1998 *Online* magazine began a feature called "Head-to-Head," which compared the results of searching on traditional online services and the web for different questions. The results varied with the type of information needed. The expert searchers who participated show remarkable variations in strategies. Almost all came to the conclusion that both resources are needed.

Nancy Garman of *Online* suggests instead of viewing the question as the Internet vs. traditional online, "why not think in terms of ORing the Internet with the traditional online services for many of our searches."

As studies repeatedly show, comparisons should not be used to choose one type of online resource over another. To offer complete information services, every library needs to offer web access and a variety of fee-based online services, not to mention CD-ROM and print.

CLARIFICATION: Helen Atkins, director of database development at the Institute for Scientific Information, sent a correction and a clarification to my *LJ* 11/1/99 column (p. 38) about ISI's enhanced indexing (KeyWords Plus). KeyWords Plus (KW+) was introduced in 1991, not 1995, and, Atkins emphasizes, "KW+ only supplements, never replaces, author-assigned Keywords." Also, she clarifies how the words are derived: "The cited reference field does not contain article titles. Access to the titles of cited articles for use in creating KW+ is by matching the reference information as it is captured to previously captured source articles. Article titles used to create KW+ are those for which we have a source record in previous data years."