

LJ INFOTECH
ONLINEDATABASES
BY CAROL TENOPIR

Database and Online System Usage

A DECADE AGO I wrote a column called "The Most Popular Databases" (*LJ* 4/1/91, p. 96,98). I used data from a company called Information Market Indicators (IMI) that tracked usage in a wide sample of libraries, plus CD-ROM best sellers lists from popular computing magazines. It is harder now to track database and system usage over all types of libraries because IMI no longer collects the data, CD-ROM isn't as important in libraries, and there are now many more choices of systems and databases.

The founder of IMI, Martha E. Williams of the University of Illinois, still tracks the database industry but on a much broader scale. Each year she writes an introduction to the *Gale Directory of Databases* (which she founded), which traces the growth of the database industry.

Her introduction, "The State of Databases Today," doesn't give figures on usage of individual files or systems and after 1998 stopped collecting data on overall online use. (In 1998 there were over 90 million database searches conducted in the online systems most used in libraries.) Williams instead focuses on tracking the number of online systems (approaching 3000), databases (over 12,000), database producers (over 4000), and records in these databases (over 15 billion). "Word-oriented databases" (full text and bibliographic) continue to be the main focus of the *Gale Directory*.

Most popular university systems

These numbers show that libraries have many choices for their online services, but they don't show which systems libraries use most. I asked heads of reference at over 60 university libraries what online systems they use for their

mediated online search services and which they provide for their end users. Not surprisingly, almost all libraries provide a wide range of online systems, each of which may have dozens or hundreds of individual databases.

For intermediary searching, Dialog remains the most popular, but libraries are cutting down on their intermediary search services and the number of systems they support for the few searches they still do.

The opposite trend seems to be happening for end user online searching, as

ing standards for how database use statistics should be collected or reported.

Nearly three years ago, the International Coalition of Library Consortia (ICOLC) published "Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full Text Resources," based on the guidelines developed by the JSTOR Web Statistics Task Force (www.library.yale.edu/consortia). The ICOLC guidelines call for usage statistics to be reported broken down into several subdivisions, e.g.,

Systems Used by University Libraries for Intermediary Searching

COMPANY	1991	1994	1998	2001
Dialog	98%	100%	98%	95%
STN	61%	60%	35%	44%
LexisNexis	44%	50%	26%	18%
Ovid/BRS	95%*	79%	16%	11%
Westlaw	29%	29%	5%	7%
DataStar	N/A	17%	9%	7%
Orbit	29%	26%	2%	0%

*Then known as BRS

In 1991, also appearing were Wilsonline 67%, EPIC 59%, NLM 56%, VuText 44%, DJNR 38%

libraries keep offering a wider array of systems. The most popular end user online service in university libraries is still OCLC's FirstSearch, but SilverPlatter, LexisNexis, and ProQuest are strong players as well (see p. 45).

Data from online systems

Collecting usage data for databases and systems has become, if not easier, at least a focus of attention in individual libraries. Tracking which databases and which online systems are most used helps libraries make important collection management decisions.

Unfortunately, not all online systems report database use in the same ways or employ the same units of measurement. It may be difficult to compare usage data for a system that reports total number of logins daily per database, for example, with one that reports hourly usage by duration of system sessions. About a dozen organizations or initiatives around the world are now develop-

each specific database; each set of IP addresses; account or ID number; time period (ideally by hour of the day, minimally by month); and a total for consortia. Within these breakdowns, each system should provide (as appropriate) the number of searches performed; number of menu selections; number of logins for simultaneous use; number of turn-aways (to show if attempted use exceeds a simultaneous use contract); number of items viewed, marked, downloaded, etc.; and the specific items (citations or full text) selected.

Responding to customers

A session at the 2001 annual Special Libraries Association meeting sponsored by the Physics-Astronomy-Mathematics Division featured several representatives from scientific, technical, and medical (STM) database and e-journal publishers. The publishers all reported efforts to give customers the types of usage reports they want, but there are still problems.



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

ONLINE DATABASES

For example, in late May 2001 the American Institute of Physics and the American Physical Society began providing an ambitious user statistics service in response to consumer demand. Still, they are concerned about how these numbers will be used because they see "no comparability between publishers." The service, they note, is "all cost, no revenue." Less than a month after it debuted, 266 institutions

had registered, and the flood continues.

A SIAM (Society for Industrial and Applied Mathematics) representative described successes and problems with usage statistics for full-text articles. SIAM has sent out usage statistics since 1997 by providing web logs four times per year to customers. Overall the SIAM online system recorded 800,000 downloads from 1997 through April 2001, with 200,000 of those in the first four

months of 2001. (Half of the downloads to free SIAM journals are not from subscribers.) The statistics show journal, volume, issue, starting page, and file type, and users aren't logged until they go into full text. Results, however, are skewed by robots that download everything and by papers that are huge files, such as those with color graphics, which require connecting and reconnecting. These reconnects to finish a single download request are counted by the SIAM computer as multiple downloads.

IEEE "is planning big things with user statistics," while IEE and Inspec now send usage statistics to users via e-mail if requested. Usage is reported by IP address and tells the number of searches, number of views of journal Table of Contents, number of abstracts viewed, and number of downloads of PDF articles.

Statistics wanted

Additional statistics would be useful. For example, now that so many full-text journal articles are accessed through a link from a bibliographic database, publishers like IEE (and librarians) would like to know how many downloads of full articles are initiated by a search of an index and from what index.

Usage statistics become quite complex when systems get large. ScienceDirect, for example, offers 1,285 journals, 1.4 million full-text articles, and 35 million abstracts. The system is growing at a rate of 250,000 records per year, and there are now 28 million page requests per year. To count usage in a meaningful way, ScienceDirect's usage count removes hits on images, all unsuccessful logins, all double-clicks within ten seconds, and multiple identical requests for full-text articles from the same IP address within 65 seconds. Customers receive 12 standard reports monthly through a password-protected web site. The reports include data on users, IP addresses, session duration, articles by type (PDF or HTML), and use of bibliographic databases. The reports differentiate between articles downloaded from subscribed journals and those from nonsubscribed journals.

ScienceDirect promises in the future to offer more flexibility in reports issued for customers, more frequent reports, more comparative measures, and information on the value (cost) per article. A ScienceDirect representative admitted that the company was "terrified" of this last item at first, but they now think informing customers is important.

**For 20 years, we've been helping
libraries improve their catalogs.**



**We're MARCIVE ... with you then ...
now ... and always!**

Here's how we can help:

- **Cataloging (MarciveWeb SELECT)** customized, current, and economical for books, audiovisuals, and more.
- **Catalog cards from your MARC records.**
- **Government document conversion** for new and retrospective holdings.
- **Documents Without Shelves** puts e-documents just a click away. Cataloging containing URL's for US government websites can be loaded directly into your OPAC.
- **Retrospective conversion** from your shelflist or brief records. Reclassification from Dewey to LC also available.
- **Item record build & smart barcodes.**
- **Overnight Authorities** for daily cataloging. Overnight we flip obsolete headings and find matching authority records.
- **MARC Record Enrichment** adds Table of Contents (TOC) data, Fiction and Biography access points to facilitate retrieval and help patrons determine relevancy.

MARCIVE, Inc.
Contact us today!

1-800-531-7678 • fax 210-646-0167
info@marcive.com • www.marcive.com
P.O. Box 47508, San Antonio, TX 78265-7508

ONLINE DATABASES

Systems Used by University Libraries for End User Searching

COMPANY	1991	1994	1998	2001
FirstSearch	N/A	35%	100%	89%
SilverPlatter	N/A	N/A	61%	87%
LexisNexis	19%	33%	90%	83%
ProQuest	N/A	N/A	46%	76%
InfoTrac	N/A	N/A	49%	56%
EBSCOhost	N/A	N/A	22%	44%
DJNR	12%	18%	41%	43%
STN	14%	14%	10%	21%
Westlaw	8%	12%	17%	19%

In 1991, also appearing were BRS/AfterDark 23% and Dialog Knowledge Index 22%, which no longer exist

ISI has provided usage statistics since the late 1990s. The monthly statistics are based on IP address and include number of sessions and number of queries for each day of the month. One drawback for Web of Science subscribers is the statistics are only reported for the entire Web of Science system and are not broken down by individual indexes (e.g., Science Citation Index, Social Science Citation Index).

Collect your own usage data

Although more companies provide a greater amount of usage data to their customers, inconsistencies in reporting make it difficult to compare systems. One solution is to collect your own data through the library's web front end. Gayle Baker, electronic services librarian at the University of Tennessee (UT), Knoxville, does just this. (For more information contact her at gwbaker@utk.edu.)

From the library's web page, each request for an online service is automatically recorded as a script file. The data can then be imported into a simple database or spreadsheet program and can then be sorted by frequency of use, time of use, or other factors. This analysis shows which bibliographic or full-text databases are requested the most often, when they are searched, and whether the requests come from within the library or from offices, campus labs, outside the university, or dorms. The library system cannot record which full-text articles are requested from full-text databases. Still, says Baker, these data give "a good picture looking across all of the databases." Because vendor-provided data are in so many different formats, the library's own data are more consistent.

The UT library uses these data "to

cut down simultaneous user contracts and for decisions to not renew and to assist in continuation decisions." Baker also can calculate the cost per use of all online systems—which helps justify purchasing decisions. A high-priced system like Web of Science, for example, gets used a lot, making its cost per use reasonable.

Full-text systems that appeal to a wide range of undergraduates are the most popular systems at UT. In 1999 and 2000, the most heavily used systems were ProQuest Research Library, LexisNexis Academic Universe, FirstSearch (half of use was through a subscription and half was per search), Dow Jones, Web of Science, and the Tennessee Electronic Library (Gale's InfoTrac databases). Databases from SilverPlatter are counted separately; when added together, SilverPlatter is in this list, too.

Some special-interest bibliographic databases are heavily used, including, in rank order, PsycInfo, Medline, ERIC, MLA International, Humanities Abstracts, Agricola, Biological Abstracts, CINAHL, Social Sciences Abstracts, Books in Print, ABI/INFORM, CAB Abstracts, and General Science Index.

Good data, good decisions

Good data can help librarians make decisions on renewing or canceling subscriptions, determine optimal numbers of simultaneous users, set staffing levels for traditional or virtual reference service, and demonstrate the cost-effectiveness of products. Academic libraries can use these numbers to help determine which academic disciplines use online systems the most and help measure the impact of user instruction. Effort is required to make data comparable across systems and the most meaningful to each library.

KAPCO LIBRARY PRODUCTS
PRESENTS...

FOUR WAYS TO MAKE YOUR LIFE EASIER:

1 Buy direct from Kapco

Call for our brand new catalog then place your order by phone or by fax. Or shop online 24 hours a day at kapcolibrary.com to take advantage of our Internet-only specials.

2 Order books fitted with Easy Cover® II

Most major book wholesalers and jobbers now offer paperbacks pre-processed with Kapco's Easy Cover II self-adhesive book covers. Check our Web site for contact information.

3 Major catalogs carry Kapco products

Order Easy Cover II, Easy Bind® Repair Tape and other Kapco products from well-known library supply catalogs. Call your favorite supplier for availability.

4 Join the TechTeam!

The TechTeam is a partnership between libraries and Kapco for the purpose of extending the shelf life of circulating materials through preventive maintenance and repair. Be sure to visit kapcolibrary.com to learn about our NEW Training Kit and to sign up for our e-mail series of TechTeam™ book repair and preservation tips.



Make Kapco part of your preservation team. Call our customer service representatives at 800-791-8965 for additional information, samples or application tips.

Kapco®
LIBRARY PRODUCTS



Toll Free: **800.791.8965**
Local: **330.678.1626**
Fax: **800.451.3724**