

Factors That Influence Online Database Use

Carol Tenopir

Introduction

What makes a student or faculty member select a particular online database to search? We would like to think they go through a systematic process of determining which of the databases offered by their library is best for their needs each time they search or that their knowledge of reliable sources on a topic always leads them to search the best possible resource.

Realistically, however, the choice of database is more likely made by a complex mix of factors, centering around such things as convenience, recommendations by librarians, placement of a database icon on a library welcome screen, or availability for remote access. The library and library staff influence this process by their control over many of these factors.

This paper reports on a two-phase study of academic libraries to identify patterns of database use and what subtle factors might influence this use. Online data from 96 academic libraries reveal how often and when selected databases are used. Usage data do not show what each library is doing to encourage (or discourage) use of these databases, however, so usage data were supplemented with a survey. The survey questionnaire asked

each library about their specific environment for online access and gathered information about what factors influence online use.

Review of the Literature

The information reported here is part of a larger study that examines patterns and factors of use for both academic and public libraries. (Tenopir, Green.) All sizes of academic and public libraries have similar usage patterns, although the amount of use varies considerably. The American Library Association's Office for Research and Statistics participates in studies regularly to gauge amounts of library use. In 1997, in cooperation with the Association of College and Research Libraries, ALA published a survey that showed how all types of academic libraries have embraced electronic services (ALA).

Early usage studies for automated resources often were done to help libraries determine how many terminals were required when they first brought up an online catalog (just as much earlier studies tried to predict an appropriate number of chairs to provide in academic libraries.) A 1983 report incorporated queuing models to recommend appropriate numbers of terminals (Tolle,

Carol Tenopir is professor, School of Information Sciences, University of Tennessee.

Table 1. Carnegie Categories For Academic Institutions
(Source: www.carnegiefoundation.org)

Category	Carnegie Class	Number in Sample	Surveys Received
B1	Baccalaureate (Liberal Arts) Colleges I	7	3
B2	Baccalaureate Colleges II	9	5
M1	Master's (Comprehensive) I Universities and Colleges	32	21
M2	Master's (Comprehensive) II Universities and Colleges	3	1
D1	Doctoral Universities I	3	2
D2	Doctoral Universities II	6	5
R1	Research Universities I	15	6
R2	Research Universities II	6	5
AA	Associate of Arts Colleges	12	7
Bus	Schools of Business and Management	3	3
Total		96	58

et al.). Turnstile counts were used to optimize reference department staffing or pickup schedules for shelving. Turnstile counts show that peak usage periods in academic libraries correspond to the academic calendar and daily class schedules (Murfin, McGrath.)

Turnstile counts are not enough to see why a particular database is used. Librarians' attitudes might be expected to influence the amount of use an online database receives. Throughout the 1990s, I measured reference librarians' attitudes toward electronic reference products and changes in university reference services (Tenopir, Ennis). Reference rooms in libraries grew busier throughout the decade. Many more workstations were added in libraries at the same time remote

access was made available. Users prefer electronic reference products, both from within the library and through dial-up access, while library instruction classes now most often focus on electronic resources.

Methodology: Phase 1—Usage Data

Measuring online usage is more complex than turnstile counts, sampling workstation queues, or even measuring access to a single library's online catalog. To get both remote and in-house

online activity for selected commercial databases from many libraries, a major database producer and aggregator provided me with usage data for all of its online databases. This database aggregator provides online access to many bibliographic, full text, and directory databases, 38 of which were used by one or more academic libraries in this study.

This database aggregator provides over 100 database titles, many of which contain overlapping information that is aimed at different audiences. For example, the same journals and magazines may be available in an indexing only version, a full text version, or a combination version. Versions indexing thousands of titles may be sold to university libraries; small colleges may prefer

Table 2. Simultaneous Users, All Databases, Academic Libraries by Carnegie Class

No. of Users	Cumulative Percent Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
1	97.8	96.8	94.5	91.5	96.4	95.2	91.8	89.9	96.1	90.5
2	99.2	98.7	96.9	96.1	99.0	97.8	95.0	93.3	98.2	94.9
5	99.9	99.9	99.1	99.7	99.8	99.7	98.3	97.8	99.8	99.3
10	100.0	100.0	99.7	100.0	100.0	100.0	99.4	99.7	100.0	100.0
15	100.0	100.0	99.8	—	100.0	100.0	99.6	99.9	100.0	100.0
25	—	—	99.9	—	—	—	99.8	100.0	—	—
66	—	—	—	—	—	—	100.0	—	—	—

versions with fewer, selected titles. Rarely does the same library purchase overlapping titles, although a library may purchase separate current and backfile versions or choose a combined version. Some databases are subject specific (business journals, for example); others are aimed at a general interest academic audience. For the purposes of analysis, the 38 databases used by the libraries were examined in three different groupings: 1) all 38 databases together; 2) eight databases that together cover general magazines and journals for academic libraries (including full text, indexing only, backfiles, and current files); 3) the single most heavily used database among these academic libraries (a current general magazine and journal title.)

From a customer list of over 1200 libraries, a random sample of 100 academic libraries in the U.S. and Canada was taken. From these, usable online usage data was available for 96 libraries. The 96 libraries cover every Carnegie Class of parent academic institution, with the largest number from MI: Master's Universities and Colleges I (32 libraries). Table 1 shows how Carnegie Class was distributed in the study sample.

Online usage for every library is captured automatically by the database provider in five-minute intervals 24 hours a day. Even with a random sample of only 96 libraries, a year's worth of data would thus yield over 10 million data points for every database. For this study, usage data were sampled once per hour (on each half-hour), from the hours of 8 a.m. to midnight, for a period of 6 months (July to December.)¹ Still, over 281,000 data points per database are included in this sample of academic libraries.

Usage data reveal how many simultaneous users are logged on to any one database at any of the sampled times. Time stamps on the data allow us to draw patterns that show average numbers of users by time of day, day of the week, and day of the month for each class of library and each database or database group.

Methodology: Phase 2—Questionnaires

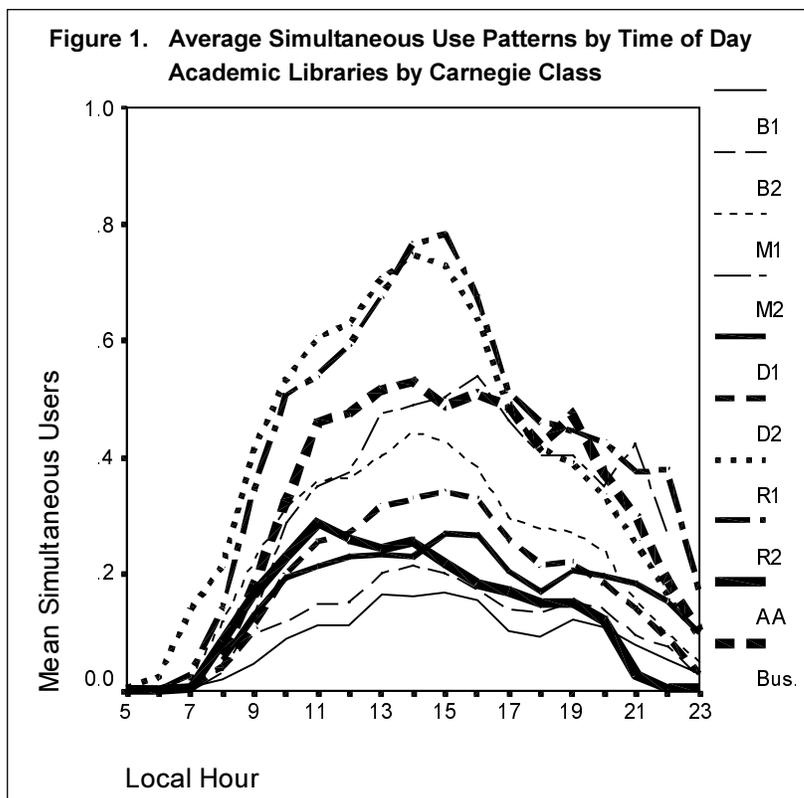
Usage data reveal typical patterns of use within classes of academic libraries, but

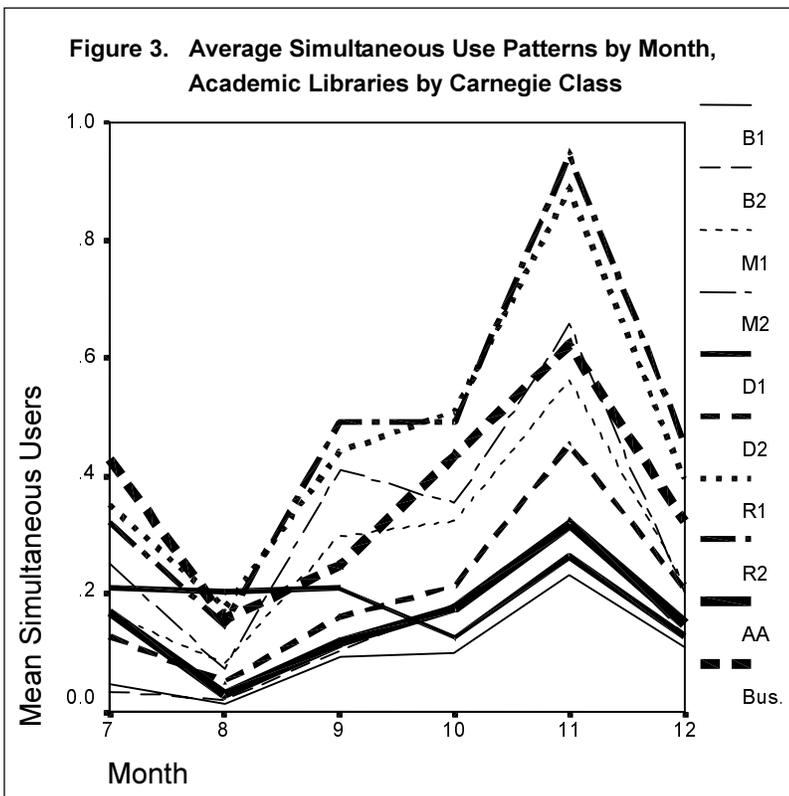
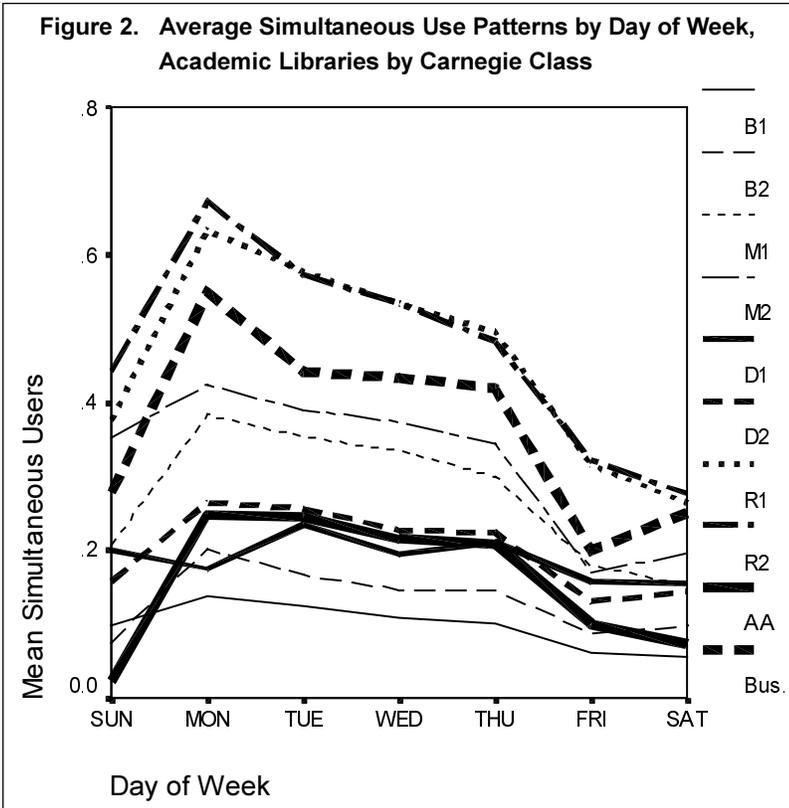
do not identify why or how the specific environment might influence online use. To begin to answer the questions of how and why, information beyond usage data is needed. Information about unique environmental factors that may influence online use in individual libraries was sought by sending a questionnaire to each of the academic libraries for which we had usage data. Fifty-eight libraries responded, for a return rate of 60%. Their distribution by Carnegie Class is shown in table 1.

Analyzed alone, the survey questions reveal much about the different environments in academic libraries and the role librarians play in influencing online use. A final step will be to analyze the survey answers in conjunction with the usage pattern data.

Analysis of Usage Patterns

The number of simultaneous users for all 38 databases in all 96 libraries ranges from zero users (the mode and median) to 66 users (the mean is .28 and the standard deviation is 1.36). Periods of high use follow clearly defined patterns. Figures 1, 2 and 3 show the patterns of use aggregated for all databases, but separated by class of library. Although the range and exact number of us-





ers varies by class of library, the patterns of use, including peaks and valleys, are quite similar. Not surprisingly, the busiest time for online research in or from academic libraries is between 10 a.m. and 5 p.m., Mondays and Tuesdays, in November. (It is likely that April and May would also be high use months.)

Patterns of use for eight general databases together show similar patterns of use (figures 4, 5, and 6.) (The mode and median are still zero, but the mean is .56 and standard deviation is 2.15.) The patterns of use for the one most heavily used general database are very similar, so are not presented here. Amounts of use are higher and vary more, however, with a mean of .84 simultaneous users and a standard deviation of 2.77 (the mode and median are still zero.)

Averages do not show the true impact of multiple users on workstations, online ports, and staff. Patterns do a better job, by helping to identify times of heavy demand. Another way to show impact is to measure how often multiple users are logged on. Tables 2 and 3 show how many simultaneous users are logged on at any one time on any single database, aggregated first for all databases and secondly for the eight general interest databases. Unlike usage of a library's catalog, simultaneous usage of any one reference databases is relatively uncommon. Providing access to only one user for a general research database, for example, would be satisfactory 80.8% of the time in research libraries (R2) and over 95% of the time in the smaller baccalaureate colleges (B1). If five simultaneous users could be accommodated for a database, users would be accommodated 99.8% of the time in B1 libraries and 94.8% of the time in R2 institutions (table 3).

The single most used database in these libraries was also analyzed alone to mitigate any effects on the data from seldom-used files such as backfiles. Table 4 shows the simultaneous use figures for the general magazine/journal database used the most

in these libraries. Clearly databases that are expected to be used by a wide variety of students and faculty and have general current interest appeal will attract more simultaneous users.

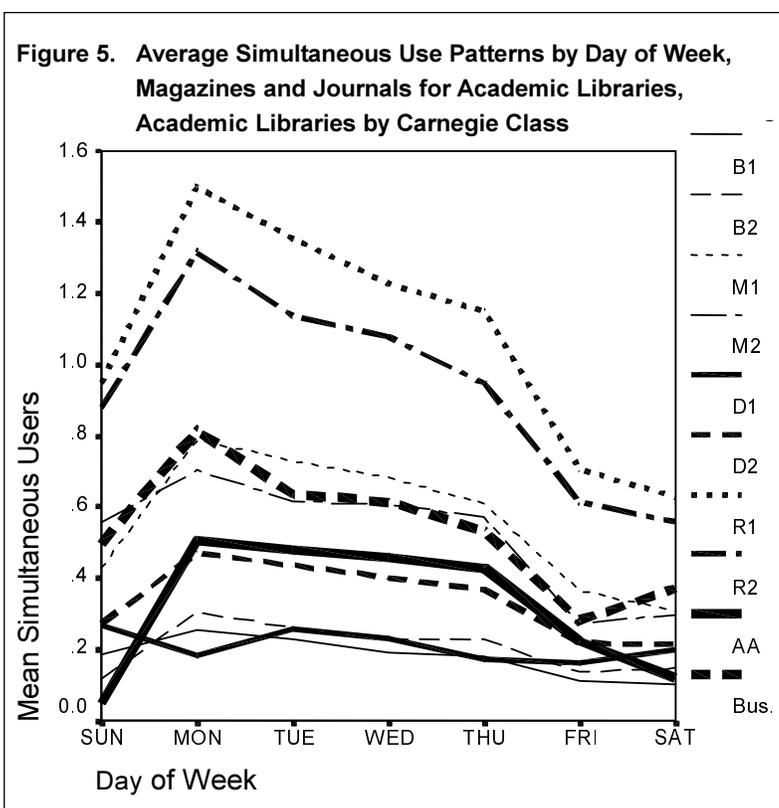
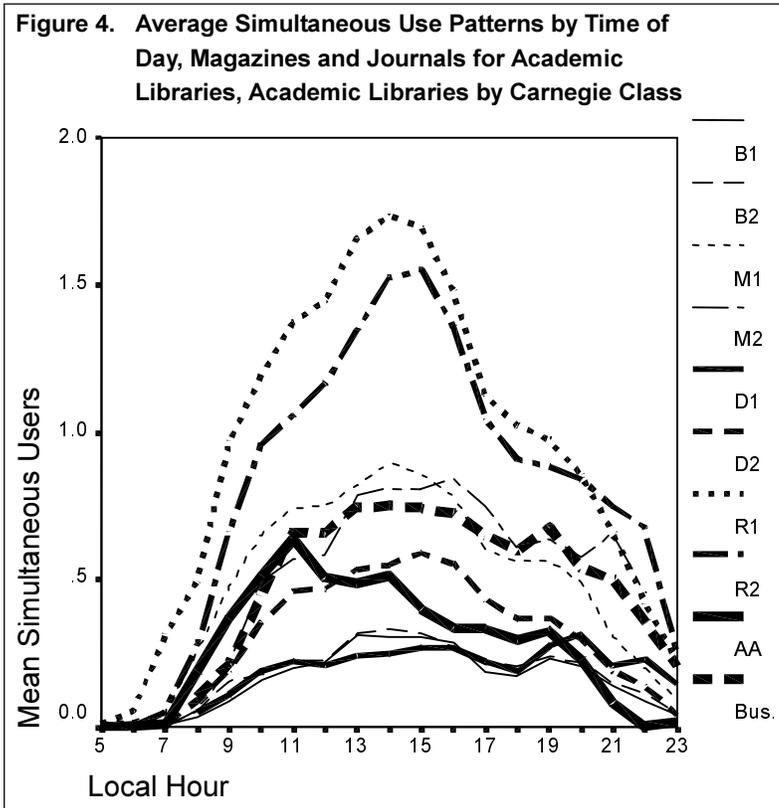
Analysis of Questionnaires

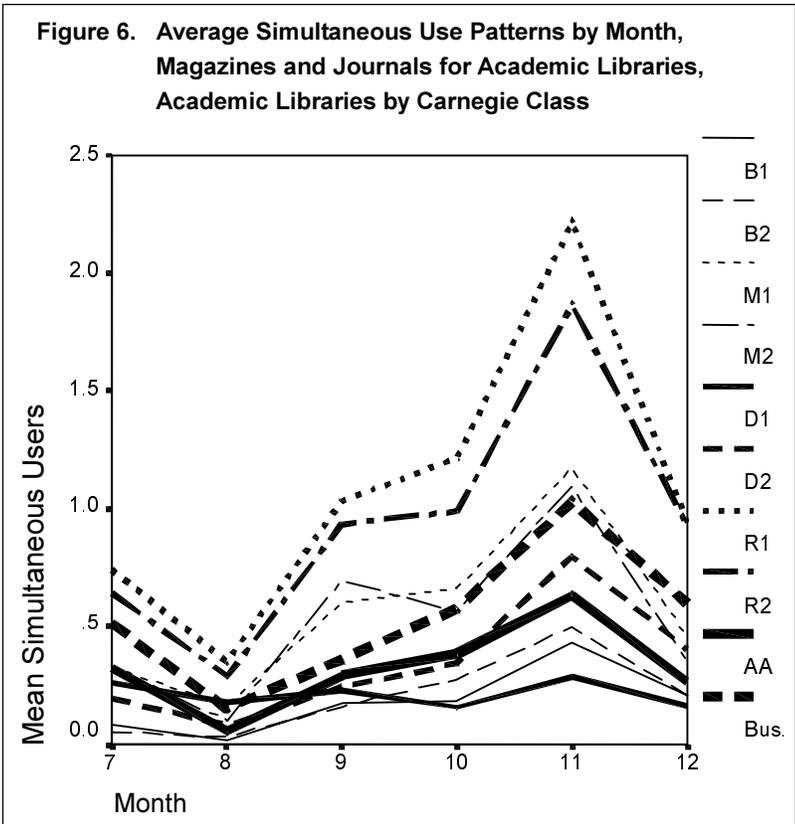
The 58 libraries that responded to our questionnaire offer a variety of electronic media for end user searching. Over 90% offer CD-ROM, commercial online, and World Wide Web access. Many provide access to several commercial online services. Seventy-two percent provide both in-house and remote access (table 5).

Approximately 95% of the libraries that responded to my survey offer access on ten or more workstations to the databases provided by the company for which I have usage data, but only 7% have workstations dedicated to these databases. Multipurpose workstations that allow access to the library's online catalog and other online databases are the rule.

Librarians believe that the databases analyzed in this study are some of the most popular among library users. The largest number of libraries (44.8%) reported these are the most popular databases. An additional 20.7% believe these databases rank second in popularity. Some libraries actively promote one or more of these databases or, at least, make those databases easier to find. Of the libraries that responded to the survey, 62% note these specific databases on the library system's main menu, 21% post signs that promote them, 57% provide handouts that describe them, and over 79% offer training that specifically mentions this company's databases. Once a user logs in or sits down at a workstation, it requires just 1-3 steps to reach the databases in 83% of the libraries (table 6.)

Close ties with academic classes should also influence how much a database is used. Over 72% of the libraries said all of the databases from this company have subject matter related to academic classes. Although





the week, mid-day, in the month when term papers are due. A majority of academic users are accessing databases at the time they typically use the library. For all types of academic libraries there are clear valleys and peaks for online use, following the rhythm of academic life.

Even for general interest magazine databases that are available both in-house and through remote access, at many times no one is online. Peak usage can be quite high in some libraries, but average usage for any one database or group of similar databases is low.

A combination of subtle factors may influence use. Librarians influence use in a variety of subtle and obvious ways, including mentioning specific products in a user instruction class, advocating use of a specific database in specific class assignments, referring to a database on a library's welcome screen, or otherwise reminding users about a specific database and making it easy to get to it. The next step will be to seek statistical correlation between deviations in amounts or patterns of use and environmental factors in the libraries.

many librarians don't know for sure if the databases are mentioned in academic classes, 64% report that they are mentioned specifically and 55% said there are specific class assignments that require use of the databases.

Conclusions

It will come as no surprise to academic reference librarians that peak online usage follows clearly defined patterns. The greatest number of users is online early in

Acknowledgments

The author would like to thank Danielle M. Green for her able assistance with bringing this project to a conclusion. George Banks and Leslie Preston were involved in the project at its early stages, including the overwhelming task of figuring out how to deal with tens of millions of online data points and how to present the data

Table 3. Simultaneous Users, Magazines And Journals For Academic Libraries Database

No. of Users	Cumulative Percent Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
1	95.6	94.8	89.3	85.7	95.1	91.7	83.5	80.8	91.7	85.6
2	98.2	97.7	93.0	92.8	98.9	95.5	88.3	86.0	95.6	92.0
5	99.8	99.8	97.8	99.4	100.0	99.2	95.0	94.8	99.4	99.1
15	100.0	100.0	99.6	—	—	100.0	98.8	99.9	100.0	100.0
25	—	—	99.9	—	—	—	99.4	—	100.0	—
66	—	—	—	—	—	—	100.0	—	—	—

(“Factors that Influence Online Database Use” Database Use,” Carol Tenopir, p. 73)

Table 4. Simultaneous Users, Most Used Database (General Magazine/Journal)

No. of Users	Cumulative Percent Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
1	95.7	92.1	84.2	85.7	91.3	86.9	78.9	68.9	86.7	85.3
2	98.5	96.4	89.4	92.8	98.2	92.4	84.0	76.3	92.6	92.2
5	100.0	99.5	96.4	99.4	100.0	98.4	91.9	90.6	99.1	99.1
15	—	100.0	99.2	—	—	100.0	97.7	99.7	99.9	100.0
25	—	—	99.7	—	—	—	98.9	—	100.0	—
66	—	—	—	—	—	—	100.0	—	—	—

Table 5. Electronic Media for End Users

Media Types Used By Respondents	Percent Carnegie Class										
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.	
CD-ROM	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	85.7	100.0
Locally loaded	33.3	0.0	33.3	0.0	0.0	40.0	83.3	40.0	57.1	33.3	33.3
Loaded on another library's computer	66.7	80.0	33.3	0.0	50.0	40.0	33.3	40.0	28.6	33.3	33.3
Commercial online from a vendor	100.0	100.0	95.2	100.0	100.0	100.0	100.0	100.0	71.4	100.0	100.0
Commercial online from an OPAC company	33.3	20.0	23.8	100.0	0.0	20.0	50.0	0.0	14.3	33.3	33.3
World Wide Web	66.7	100.0	100.0	100.0	100.0	80.0	100.0	80.0	100.0	100.0	33.3

in a meaningful way. Shawn Collins, technology coordinator at the School of Information Sciences, University of Tennessee, has offered expert support throughout.

Note

1. We have converted the hourly data into the local time zone for each library, but, since the database company's computer was set to eastern time, data from libraries in western time zones begin and end earlier (7 a.m.–11 p.m. central time; 6 a.m.–10 p.m. mountain time; 5 a.m.–9 p.m. pacific time; 4 a.m.–8 p.m. Alaska time; and 3 a.m.–7 p.m. Hawaii time).

References

- American Library Association, *Electronic Services in Academic Libraries*, Chicago: ALA, 1997. [Executive summary available at www.ala.org/alaorg/ors/electsvcs.html].
- McGrath, W. E., "Periodicity in Academic Library Circulation: A Spectral Analysis," *Journal of the American Society for Information Science* 47 (February 1996): 136–45.
- Murfin, Marjorie E., "National Reference Measurement: What Can it Tell Us About Staffing," *College and Research Libraries* 44 (September 1983): 321–33.
- Tenopir, Carol and Ennis, Lisa, "The Digital Reference World of Academic Librarians," *Online* 19 (July/August 1998): 22–28.
- Tenopir, Carol and Green, Danielle M., "Simultaneous Usage of Online Databases in Academic and Public Libraries," *Proceedings of the National Online Meeting, New York, May 1999*.
- Tolle, John E., et al., "Determining the Required Number of Online Catalog Terminals," *Information Technology & Libraries* 2 (September 1983): 261–65.

Table 6. Ways Libraries Promote The Specific Databases Analyzed Here

Methods of Promotion Used	Percent Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
Remote access	33.3	80.0	85.0	100.0	100.0	100.0	83.3	60.0	28.6	66.7
Workstations dedicated to these databases	0.0	20.0	4.8	0.0	0.0	0.0	0.0	20.0	14.3	33.3
Databases noted on library system main menu	66.7	100.0	61.9	—	50.0	60.0	50.0	40.0	85.7	33.3
Signs	33.3	20.0	33.3	0.0	0.0	0.0	0.0	20.0	28.6	0.0
Handouts	0.0	60.0	66.7	100.0	50.0	40.0	50.0	40.0	71.4	66.7
Databases mentioned specifically in training	100.0	60.0	85.7	100.0	100.0	80.0	66.7	80.0	57.1	100.0
Databases specifically mentioned in academic classes	66.7	80.0	71.4	0.0	100.0	60.0	33.3	20.0	71.4	100.0
Specific class assignments require use of these databases	66.7	40.0	61.9	0.0	50.0	20.0	66.7	40.0	57.1	100.0