Data Needs and Use of Electronic Resources and Services at Academic Research Libraries

Wonsik Shim and Charles R. McClure

abstract: The ARL (Association of Research Libraries) E-Metrics project is a concerted effort by the research library community to investigate various problems related to collecting and using data for electronic materials and services. A survey was conducted involving twenty-four libraries during the first phase of the project to learn about current data collection activities. Twenty-two libraries responded to the survey. The results show that a number of libraries are not well prepared to collect, to analyze, and to report data related to networked resources and services. Libraries need to determine their data needs and the amount of resource that they can commit for data collection and use activities. The results also identified the lack of consistent and comparable statistics from database vendors as a main challenge and an area in which libraries and vendors must work together for solutions.

Research libraries have been increasing their acquisition of electronic information resources for some time now. Recent statistics estimate that in 1999–2000 research libraries spent on average 12.9 percent of their materials budget on electronic resources, a sharp increase from a mere 3.6 percent in 1992–93.[1] The trend is largely in response to the availability of electronic products and services as well as the increasing demand from the users for such products.

The working definition of networked services is those electronic information resources and/or services that users access electronically via a computing network, (1) from on-site in the library (2) remote to the library, but from a campus facility, or (3) remote from the library and campus. Examples of networked resources include local, regional, and statewide library hosted or authored websites and library-licensed databases (e.g., ScienceDirect, EBSCOHost, JSTOR, Project Muse). Examples of networked services include:

- Text and numerical databases, electronic journals and books;
- E-mail, listservs, online reference/assistance;
- Training in the use of these resources and services;
- Request for services via online forms (i.e., interlibrary loans).

While electronic resources significantly enhance user access to library materials, they present several problems in terms of gathering data on materials use for various managerial and operational purposes. For example, libraries now need a range of data about the use of electronic resources to justify the growing expenditure. Libraries are able to keep track of the use of library materials (most notably scholarly journals) that were difficult to monitor in the print collection environment. Automated means of tracking materials use, such as transaction and web server logs provide vast amounts of information for libraries to process and to analyze. However, since libraries now depend heavily on outside information providers (namely, database vendors or content providers) for data collection of electronic resources, they must manage inconsistent and incompatible data provided by these vendors. Several authors have pointed out the inconsistencies of data collection methods resulting from the proliferation of electronic materials in libraries.[2]

Overall, it seems there is a general lack of preparation and understanding of the issues on the part of libraries and vendors concerning statistics and measures to describe the use and users of electronic resources. Changes need to be made in describing these uses and demonstrating the value of electronic resources to the users and various stakeholders.

Background

The ARL (Association of Research Libraries) E-Metrics project is an attempt by the research library community to investigate various problems related to collecting and using data on electronic materials and services. The project began in April 2000, is scheduled for completion in December 2001, and is funded by a group of twenty four libraries in ARL. Figure 1 identifies the participants of the ARL E-Metrics project.

University of Alberta	Arizona State University
Auburn University	University of Chicago
University of Connecticut	Cornell University
University of Illinois-Chicago	University of Manitoba
University of Maryland-College Park	University of Massachusetts
University of Nebraska-Lincoln	University of Notre Dame
University of Pennsylvania	Pennsylvania State University
University of Pittsburgh	Purdue University
University of Southern California	Texas A&M University
Virginia Polytechnic Institute and State University	University of Western Ontario
University of Wisconsin-Madison	Yale University
Library of Congress	New York Public Library, The Research Libraries

Figure 1. ARL E-Metrics Project Participants

The E-Metrics project aims to:

- Develop, test, and refine selected statistics and performance measures to describe electronic services and resources in ARL libraries;
- Engage in a collaborative effort with selected database vendors to establish an ongoing means to produce selected descriptive statistics on database use, users, and services;
- Develop a model to describe possible relationships between library activities and library/institutional outcomes;
- Develop a proposal for external funding to maintain the development and refinement of networked statistics and performance measures.

This paper reports on the results of a survey conducted during the initial phase of the E-Metrics project. Detailed project information and reports can be found at <http://www.arl.org/stats/newmeas/emetrics/index.html> [January 28, 2002].

The proliferation of networked electronic information resources and services prompted interest and research in developing statistics and measures to describe this emerging information provision environment. The following sources are useful introductory and background readings to the topic. Charles R. McClure and Cynthia Lopata offer a number of strategies and measures to assess the academic networked environment.[3] Young identifies issues that make the measurement of electronic resources and services challenging.[4] Carol Tenopir and Eleanor Read offer an excellent analysis of database use at fifty-seven academic institutions.[5] They conclude that for all types of academic libraries, user demands are concentrated on a fairly predictable span of time, "early in the week, at midday, in the month when term papers are due."[6] They also point out that electronic library databases are underutilized by students compared to other electronic media such as chat rooms and general Internet resources. Deborah Blecic, Joan Fiscella, and Stephen Wiberley, Jr. provide an example of how vendor supplied usage statistics are being used at an individual institution.[7] A recent compilation by Charles R. McClure and John Carlo Bertot provides an overview of a wide array of issues surrounding evaluation of networked information services.[8] Regarding usage statistics from vendors, the guidelines produced by the International Coalition of Library Consortia are widely recognized by both the library and vendor communities.[9]

> **The proliferation of networked electronic information resources and services prompted interest and research in developing statistics and measures to describe this emerging information provision environment.**

While the above-mentioned sources aid our understanding of the context and issues surrounding measuring and evaluating networked information sources and services, there has been no comprehensive investigation of the data needs and current data collection practices regarding electronic resources and services at research libraries. This paper begins to fill that void and to provide perspectives for library administrators as they begin planning and implementing strategies to collect and use data on electronic resources.

Specifically, the paper examines the results of a survey conducted to investigate:

1) What kind of data describing electronic resources and services currently are being collected at research libraries;
2) How are those data being used;
3) What kind of data needs exist in the libraries?

Based on the findings of the survey and other activities in the first phase of the project (May 2000–October 2000), we present recommendations and future directions to improve the collection and use of data regarding electronic resources and services.

Methodology

A survey questionnaire was developed in consultation with several project participants.[10] Of the twenty-four e-mail surveys mailed to project participants, we received twenty-two completed surveys.

The survey focused on the following categories of information:

• Currently collected measures of electronic resources and services;
• Data use in terms of specific decision making; and
• Issues and challenges facing libraries in the data collection efforts.

Note that the second and third items were asked in open-ended questions. We also asked the libraries to include sample reports related to electronic resources and services.

With regard to currently collected measures, we asked the respondents to put them in the following five categories:

• Measures related to patron accessible resources;
• Measures related to patron use;
• Measures related to users;
• Measures related to cost; and
• Other measures that cannot be classified above but are important to electronic resources and services.

In addition, we asked the respondents to tell us how long each measure has been collected and how frequently it is collected. We also ask for the name of the entity initiating the collection if the measure was requested—by someone or an entity either inside (e.g., library director, collection specialists) or outside the library (e.g., deans, provost, associations, accreditation agencies).

Survey responses were aggregated in a Microsoft Excel spreadsheet for analysis. For measures collected, two research assistants read the responses and constructed an authority list for each category where applicable. They then returned to each response to tally them according to the authority list. The analysis and summary of the frequency of data collection, the time data collection began, and the person or organization requesting the data proved problematic because of the variations in responses. For this information, we created simple frequency tables. Through an iterative process, we developed a data-derived taxonomy in an effort to put responses to open-ended questions (decision-making and issues) into some context. Responses were then assigned to specific classes in the taxonomy.

The study sample consists of a self-selected group of twenty-four libraries. There is substantial willingness on the part of the libraries to cooperate in the data collection activities because of vested interests. Also, the sample is a subset of the ARL member libraries. Therefore the results may not represent the ARL membership as a whole.

Findings

The survey findings are presented in the order of the questions asked on the survey. Note that the intention of the survey was to obtain the range of measures currently collected among participating ARL libraries rather than to create an exhaustive list. The survey did not ask the libraries to supply the definitions of the measures that they reported on the survey form.

Data Collection

In the data collection part of the survey, we asked the respondents to list the names of collected statistics, the frequency of data collection (from weekly to yearly), the inception of data collection, and whether the data are required, and if so, by whom. We provided five categories for reporting statistics, along with several specific examples under each category.

Category 1: Measures of Patron Accessible Resources (e.g. number of electronic database titles served, number of library web pages in service, number of e-books, number of full-text e-journals, number of librarians providing electronic reference).

Table 1

Collected Measures on Electronic Resources by Number of Mentions

Collected Data	Times Mentioned
Number of electronic database titles	17
Number of electronic (full-text) journals	16
Number of electronic books	6
Number of library web pages in service	4
Number of new electronic titles	3
Number of cancelled electronic titles	2
Number of electronic resources	2
Number of computer files	1
Number of networked CD-ROMs	1
Number of librarians providing electronic reference	1
Number of public workstations available	1
Number of classes on electronic resources	1
Other	3

As illustrated in table 1, most libraries seem to collect counts of various types of electronic materials. Electronic databases and electronic full-text journals represent the bulk of the measure related to electronic collection of licensed materials. Most libraries that collected a count of the number of electronic databases also said they kept separate counts for full-text journals. Three libraries gathered information about the number of databases added or deleted periodically.

Only six libraries reported that they collected information about electronic books (e-books). At this point, it is not clear whether e-books include only those e-books that are commercially available through companies, such as netLibrary, or scanned copies of books generated locally within the library.

Four libraries reported that they collected information about the number of web pages they provide to the users. However, a page in one library can be quite different from a page in another library. This information can provide not only a trend-line in terms of the amount of information in the local context but also some crude measure of staff productivity. Among the less frequently mentioned measures is the number of networked CD-ROMs (1). This may be interpreted as an indicator that CD-ROMs are gradually disappearing from the library and that their role as a major networked resource in research libraries is diminishing. Only one library reported that it counts the public workstation available in the library.

The measures reported in the "other" category include the number of digitized images, the number of print or CD-ROM indexes migrated to the Web, and the number of locally mounted database records. All three measures were reported from the same library.

In terms of the collection frequencies of these measures, most libraries provide these numbers annually. While this remains unclear, we suspect that many libraries used the information for reporting purposes, both internally (e.g., annual report) and externally (e.g., ARL statistics). Only about one-third of the libraries (8) indicated that they collect the key measures monthly. A few libraries said they collected the information quarterly.

There is a wide variation, in terms of when the data collection began for different measures in the same institution as well as for the same measure among reporting libraries. For instance, one library reported that it started collecting the number of electronic databases in 1991, the number of electronic journals in 1995, and the number of web pages in service in 1998. Among libraries that count the number of electronic databases, nine out of seventeen libraries indicated that they started collecting the data after 1997. The earliest collecting goes back to 1990 and 1991. Three libraries did not specify when they began data collection.

Seven libraries reported that they collect the data to fulfill requirements of external organizations such as ARL and CARL (Canadian Association of Research Libraries), including the ARL Supplementary Statistics on Electronic Resources (<http://www.arl.org/stats/arlstat/#sup>). Almost an equal number of libraries (6) said that library administration or staff required the data. At least one library was able to list a rather comprehensive group of consumers of the collected information: University and Library Administration, Library Committee responsible for networked electronic resources, Serials Department, Libraries Electronic Technologies and Services (LETS), bibliographers, and Library Unit Heads. Only one library reported that the information is required as part of strategic planning efforts.

Category 2: Patron Use Measures (e.g. number of logins or visits, number of library web pages accessed, number of documents downloaded, and number of electronic reference service transactions).

It is quite difficult to summarize the responses for this category given the wide range of measures reported that are likely to be confounded by the inconsistent use of terms. As far as the licensed vendor materials are concerned, nearly all libraries responded that they collect usage statistics provided by vendors. These include sessions, searches, downloaded records, rejected logins (i.e., turn-aways), and others. Some of these data are not available from all vendors. The investigators realize that it is not possible to collect these measures uniformly across the database vendors.

Several libraries seemed to have a formal report that summarizes various measures available either from the vendors or from internal sources. A sample report sent by one library compiles a quarterly *ER (Electronic Resources) Usage Report* that includes (by database) measures such as:

- sessions,
- connect time,
- searches,
- records and articles (downloaded or viewed),
- and lock-outs.

The investigators noticed that there are a number of empty rows in the report. Several other libraries also sent sample reports that are less formal but nonetheless attempt to include as much information as possible using a common set of core measures supplied by vendors.

A handful of libraries (4) seemed to have some form of a click-through mechanism where they captured "attempted" log-ons to electronic databases and full-text journals from their electronic resource pages.[11] One library went even further to map these click-throughs to schools and departments based on IP domains associated with campus units.

Most libraries capture information about library web page usage through programs such as *WebTrends* and *Analog*.[12] However, when we examined the sample reports from select libraries, we find they are mostly raw data, such as page access, generated by the software programs. Other than a few factoids such as top-ten most visited pages (or areas) and peak use period, we did not see good examples of succinct summary of the data.

Other use measures worth mentioning included the number of electronic reference transactions, the number of documents downloaded from electronic reserve, and the number of electronic document delivery requests (e.g., from Current Contents database). Approximately half the libraries (10) reported that they collected the number of electronic reference transactions.

Regarding frequency of data collection, an overwhelming number of respondent libraries reported that they collected statistics monthly. This coincides with the fact that many database vendors supply monthly usage statistics and that many library system log files (e.g., web statistics) are also captured on a monthly basis. The only exception is the number of electronic reference transactions. Four out of ten libraries that reported the measure said they collected it annually.

Again, it is only quite recently that libraries began collecting various electronic usage statistics such as web access statistics, licensed database use statistics, and the number of electronic reference questions. Libraries first collected input-oriented information such as database titles and gradually moved to collecting usage statistics as they received information from database vendors. Additionally, some needs (such as budget justification) required them to report these types of information.

While more than two-thirds of the libraries were required to report various electronic resource related measures, only about one-third of the libraries said they collected usage statistics because someone requested such data.

While more than two-thirds of the libraries were required to report various electronic resource related measures, only about one-third of the libraries said they collected usage statistics because someone requested such data. The person who is mostly likely to request data is either a collection development manager or individual collection development staff member. This suggests that usage statistics currently assist the collection development activities like renewal/cancellation of subscriptions.

Category 3: Measures of Users of Electronic Resources and Services (e.g. percent of undergraduates who have used the e-books, percent of grads who have used the electronic reference service, and number of users by type of services).

Respondents all agree that they do not have a way to distinguish individual users of electronic resources and services. This is a serious problem as libraries strive to collect such crucial information as user penetration as an indication of the library providing value to the user community.

With the introduction of OPACs, libraries were able to track who was borrowing what materials rather easily. Now the Web, on which most electronic services are offered, presents a substantial obstacle as far as libraries' ability to describe their users. However, obtaining this information is not impossible. There are techniques that libraries can use to track users and their activities.

One strategy for tracking is to validate user access to licensed electronic resources through a login. At the moment, however, libraries are reluctant to put obstacles, such as forcing users to enter a library card number and a few clicks to authenticate themselves, that may hinder the use of electronic materials. In addition, many licensed materials are essentially IP address validated to ease the burden on the users to authenticate themselves to each database and also to allow the database vendor to manage access without too much overhead.[13]

The exception, where a comprehensive set of data is available, is two libraries tracking the number of uses originating from various academic units including libraries, computer labs and other buildings identified with a set of IP addresses. In rare instances, locally mounted databases can keep track of users by type (faculty, undergraduate, graduate, and staff). There is only one mention of that kind of configuration.

Otherwise, libraries collect bits and pieces about users of electronic resources and services through various means. Three libraries reported that they ask the users to identify themselves (by type) when they submit electronic reference queries. One library conducted a user survey on electronic resource use where users were identified by type. Another library said they monitored user comments submitted electronically based on the type of users.

In two instances, respondents mentioned the use of a proxy server for obtaining user information. One library has information about users by remote vs. on-campus connections. The other library has not yet collected information but plans to use the proxy server log to obtain user demographics.

Category 4: Cost Measures (e.g. cost per electronic document delivered, cost of database subscription fee, and expenditures for electronic journals)

Table 2 shows the frequency of cost measures reported by respondents. Many libraries collect the overall cost for electronic database subscriptions. All libraries, except for one library with no response for this category, reported that they have this data. Some libraries (8) further differentiate the costs for electronic full-text journals.

Table 2

Measures of Cost

Collected Data	Times Mentioned
Total costs of database subscriptions	21
Expenditures for e-journals	8
Cost per electronic document delivered	4
Cost per search	3
Average cost per database subscription	2
Cost per login	2
Other	16
No response	1

Less than half of the libraries (9) reported that they linked the cost per use to some kind of usage measure such as the number of searches, logins, or documents delivered. In some instances, the information is available from the vendor as indicators of value or cost savings.

Some libraries commented that calculating cost per use for types of electronic resources and services is not valid. While this may be true, we also suspect that some libraries fear that producing cost per use data might be risky if the ratio turns out to be too expensive.

There are a variety of measures collected at individual libraries. One library reported an extensive range of cost measures that include:

- percentage of collections budget spent on electronic resources;
- dollar amounts by college;
- cost projections;
- dollar amounts invested in digital library and web development;
- electronic resource spending per student;
- ratio of paper vs. electronic journals costs.

All these measures are incorporated in the library's strategic planning document. There was at least one other library that reported projected cost. Some libraries differentiated one-time costs (such as JSTOR membership) and ongoing subscription costs.[14] One library said it kept the cost for electronic books separately. Another library reported that it collected information about internal and external sources of funding for electronic resources.

There are some similarities between the patron-accessible resource measures and cost measures with regard to how frequently libraries collect the information, when they began collecting the information, and who requested the information. Except for a few cases, many libraries collect cost information annually, perhaps for annual budget preparation. Where data are reported monthly or quarterly, it appears that the libraries have a special electronic resource cost report type of arrangement that is reviewed by a committee in charge of database renewal.

Collecting cost measures started about at the same time as libraries started counting the number of electronic resources. Collection of cost measures at six libraries was required as part of the 1997/98 ARL survey on the topic (http://www.arl.org/stats/arlstat/#sup). Nine other libraries said the information was requested by either library committee/management (7), outside government body (1), or for strategic planning requirement (1). One library said that it reported the total cost of database subscription to a magazine.

Category 5: Other Measures related to electronic resources and services (such as service quality, the effects of library use on research and instruction or the percent of library users satisfied with libraries services).

When discussing the quality and value of services, this category captured ongoing data collection efforts on user satisfaction. However, it appears that the responses were not limited to electronic services. Instead, what emerges from the responses is a list of tools, summarized in table 3, that libraries use to measure some aspect of the aforementioned aspects of library service.

Most of the instruments are used irregularly or on an as needed basis. Only one library reported that it conducts biennial user satisfaction surveys. Three libraries mentioned participation in the ARL LibQual+ project as an effort to collect information regarding service quality.[15] Focus groups, user panel feedback, and user forums represent important, but less frequently used ways of collecting data.

In addition to the instruments mentioned in table 3, there are other aspects regarding electronic resources that some libraries measure. One library collects dollar amounts saved (cost avoidance) as part of the consortium report. Not clear is whether the

Table 3

Sources Used to Obtain Information Regarding Service Quality and User Satisfaction.

Instrument Types	Times Mentioned
User satisfaction survey	7
LibQual+	3
Focus Group	2
Library Class Evaluation	2
User Panel Feedback	1
User Forum	1

information is the cost savings at that institution or at a consortium level. One library lists the number of students who take library online tutorials and courses.

Data Use

1) Level of communication

The survey requested that libraries indicate the people or entities to which they report the collected data on electronic resources. Table 4 summarizes the results along with the response categories.

It seems that information on electronic resources and services (to the extent it is available) is being communicated widely within the library and within the parent organization. But we do not know the specifics of the nature (e.g., expenditure, user activities) and manner (e.g., formal vs. informal, annual report vs. monthly report) of those communications. The "others" category included ARL (2), outside accreditation body (1), external funding agency, consortia groups (1), magazine (1) and public (1).

2) Use of Collected Data for Decision-Making

We also asked the respondents to elaborate on the decision-making in the library based on the collected measures and statistics. Table 5 summarizes the responses.

Table 4

Dissemination of Data

Among library managers	Among library staff	With supporting institutions	Others
100%	95%	82%	32%

One of the immediate uses of measures of electronic resources for library decision-making is in the area of external resource contracts. Fourteen libraries specified that they used electronic resource statistics to make electronic database subscription decisions. Among those libraries, eight reported that they use turn-away or lock-out (logins that exceed the simultaneous user limit) data to change the simultaneous user (S/U) licenses. It is not clear whether the information is used both to increase and to decrease the number of S/U licenses. Three libraries responded that usage statistics of electronic resources affect subscription decisions of the same or similar counterpart print materials.

In the reporting and communication category, seven libraries reported that data were being used for budget-related activities. It is interesting to note that only one library specifically mentioned the use of such data in their annual report. The library wrote:

> The university administration has little interest in frequent statistical reports, but we make sure that the overall extent of our e-resources and bottom-line summaries of the extent of their use feature prominently in our most significant reports [annual report] to them.

Three libraries used the data in their strategic planning process and related documents. There was only one library that mentioned specifically comparison with peer institutions as one of the uses of the data.

Table 5

Taxonomy for Decision-Making Instances Affected by Information on the Use of Electronic Collection.

Main Category	Specific Category	Frequency
External Resource Contract	Journal renewal/cancellation	3
	Database renewal/cancellation	14
	Changes in simultaneous/user Limit	8
Reporting and Communication	Budget request, justification, and presentation.	7
	Annual report and other similar summary report	1
	Strategic planning	3
	Institutional comparison	1
Service Assessment & Improvement	Redesign of web pages	9
	Marketing of collections and services	3
	Instruction and training	4
	Changes in staffing	4
	Assessment of existing collections and services	9
	Assessment of pilot (trial) collections and services	1

In the category of service assessment and improvement, nine libraries said they use the data to redesign library web pages. We suspect the high number of responses in this category might have been influenced by the example given in the question. The same number of libraries responded using a more generic phrase that can be summarized as assessment of existing collections and services. Only one library said that it used the data to evaluate trial databases.

There are hints in the responses that some of the decision-making activities were performed only on occasion (as opposed to regularly or systematically), and in an *ad hoc* manner. The following comment from one library illustrates this view:

> At this point in time we have neither sufficiently comprehensive statistics nor the necessary processes in place to use these statistics to systematically evaluate our collections and service.

Nonetheless, libraries reported that they use the data to make improvements in service marketing (3), instruction and training (4), and staffing (4).

Issues Important to Statistics and Measurement

The survey asked the respondents to list their three most important issues related to collection of statistics and measures to describe networked resources and services. Eighteen libraries responded to the question. An overwhelming number of responses were related to problems associated with current vendor reports.

Table 6 describes the summary of these responses. The main category is broken into two sections: one that has to do with data collection and the other with data processing and use. The data collection category is further divided into issues related to vendor reports and others.

The data clearly show that libraries want consistent, comparable data delivered in a standardized method. The responses read like a scripted answer when asked "what's wrong with statistics on electronic resources?" Respondents suggest that vendors are to blame. A high level of frustration with the contents of vendor supplied usage reports and the way they are delivered were common responses. One respondent summarized it by saying, "for the most part, vendors provide what they want, not what libraries want."

However, one library expressed some optimism amid frustration about the vendor reports. "Despite these frustrations, vendors are making progress in moving towards simpler and more accessible reports and we are pleased that for our most important resources port contention [turn away] is reported to us."

Four libraries expressed concern that it takes too much time and effort to collect and compile various vendor statistics. Some pointed out problems of not having useful data (3) or data without enough detail (3). Two libraries said some vendors provide no statistics at all and one library complained that vendor reports are not timely.

The data clearly show that libraries want consistent, comparable data delivered in a standardized method.

Table 6

Summary of Issues Related to Measures of Electronic Resources and Services

Main Category		Specific Category	Frequency
Data Collection	In regard to vendor statistics	Lack of consistent definitions	9
		Lack of comparable measures	12
		Lack of standardized reporting method	10
		Lack of detailed, granular data	3
		Lack of useful data	3
		Availability of data	2
		Timeliness of data	1
		Effort to collect data	4
	Other	Difficult to measure web access	1
		Difficult to aggregate data among branch libraries	1
		Lack of qualitative data	1
		Lack of information about users	3
Data Processing and Use		Lack of supporting (technological and human resources) to facilitate data processing	2
		Urgency to justify expenditure	1
		Difficult to summarize or interpret data	2
		Inability to relate use of electronic materials with their physical counterparts	1
		Comparison with peer institutions	1
		Inability to link data to decision making	1
		Inability to link data to quality of service	1
		Inability to link data to outcomes	2

Among data collection issues other than those related to vendor reports, the lack of information about users and user behaviors tops the list of concerns. One library reported that aggregating measures across branch libraries on campus is a difficult thing to do given the fact that some libraries operate either independently or under different reporting environments. Another library expressed concern that collecting even a seemingly simple thing such as how many times library web pages were accessed for a given period proved to be a difficult task because of technological and organizational complexities. Finally, one library said that lack of qualitative information such as ease of access, quality of sources, and availability is an important issue.

The responses suggested relatively low interest related to data processing and use. While this may have to do with the way the question was phrased, it does raise the question "are libraries doing enough to collect useful data describing use of electronic resources and services?" At least two libraries commented that lack of organizational support in terms of technical and human resources is a major hindrance to the measurement effort. Two other libraries said they were having difficulty interpreting and summarizing the collected data. These kinds of responses acknowledge the fact that even when libraries receive reasonably satisfactory reports from database vendors, much work remains for the library. Four libraries pointed out their inability to link data to other important issues such as decision-making (1), quality of service (1), and educational outcome (2).

Conclusions

While analysis of the E-Metrics survey responses reveals a range of data collection and use activities among project participants, to a large degree respondents have just begun investigating the issue of data collection of electronic resources and services seriously, and formal assessment activities in this area differ among respondents. A number of libraries are not well prepared to collect, analyze, and report data related to networked resources and services. Clearly, there are libraries that are working in this area and gaining experience with some specific data collection procedures such as cost analysis and collection development. But as a group, it is difficult to point to specific best practices among participating libraries and offer a coherent picture of issues, strategies, and specific techniques related to producing statistics and performance measures for networked resources and services at respondents.

One of the findings from the survey and other activities in the initial phase of the E-Metrics project was that there is a range of unique situational factors at the various participating libraries that affect the:

- Library's need for statistics and performance measures;
- Degree to which the library is willing to commit resources to produce such statistics and performance measures;
- Range and type of databases being used by the library;
- Library's staff relationships with the vendors of these databases;
- Specific type of statistics and performance measures that would be useful in a particular library setting.

In some libraries, there is less interest from university administration in the benefit, impact, and use of electronic resources, with the result that there is "less pressure" to produce descriptive data. Other libraries have significant cost and administrative pressure to produce such data. Thus, the nature of the university context in which a library operates has a significant influence on the development of statistics and performance measures to describe their networked services and resources.

This diverse context suggests that while there may be a set of "core" statistics and performance measures that would be of use to all libraries, there is also likely to be specific statistics and measures that will be of greater interest to some more than others.[16] A better understanding of local factors that affect the need for specific types of statistics and performance measures will require additional research. Models that can help describe these local situations may also be useful in the selection of statistics and performance measures.

The results reported here suggest a wide range of self-induced responsibilities and levels of effort on the part of respondents for data collection, analysis, and reporting. To some degree, the amount of responsibility and level of effort that the library will commit to data collection and analysis of networked services and resources depends on their context, the perceived importance of having such data, and how such data might be used effectively in the local setting. We have yet to understand what constitutes "reasonable" levels of effort on the part of the library to collect and produce such data *at a given library or at libraries in general*. Some libraries may need only to commit minimal resources to obtain the data that they need, whereas others may have to commit significant resources in order to obtain the data required for their decision-making processes. Again, models that describe and relate the local level of efforts to uses and applications of statistics and performance measures may be useful.

Regarding the most important issues related to measurement of networked resources and services, the majority of respondents cite the lack of consistent and comparable statistics from database vendors as the most serious problem. Relatively few respondents recognized or identified problems associated with the library's inability to process and to utilize collected data. While the database statistics supplied by vendors are essential, libraries seem to ignore the fact that there are other types of data that they could collect and analyze. For example, statistics related to counting and assessing the use of electronic reference (web-based or e-mail-based) are not dependent on obtaining data from external vendors. Still, there is some logic in focusing on the vendor-based statistics rather than those of other types of networked resources and services in the library. While all agree that progress should be made to standardize usage statistics from database vendors, libraries should also look at networked services and resources for which data can be gathered internally.

Based on the findings of the Phase I of ARL E-Metrics projects, the investigators developed and field-tested a set of recommended statistics and measures that provide indicators of library networked services and resources.[17] The report and the enclosed manual to collect and use the recommended statistics and measures provide a beginning approach for research libraries to describe better the use and users of their networked services. At the same time, the manual increases the visibility and importance of developing such statistics and measures.

There are a number of issues and challenges that will affect the library's ability to collect statistics and measures to describe its electronic resources and services. Here are some of the issues for libraries to consider:

- *Library culture of assessment:* In addition to having a systematic approach to network statistics and performance measure activities, libraries need to adopt an overall culture of assessment. Amos Lakos defines a culture of assessment as:

  > The attitudinal or institutional changes that have to occur in order for library staff to be able to work in an environment where decisions are based on facts, research and analysis, and services are planned and delivered in order to maximize positive outcomes and impacts for the library clients.[18]

  As such, libraries need to focus on a systematic approach to the assessment of library services, resources, and initiatives in order to understand better the impact of those services, resources, and initiatives as well as to undertake changes and or modifications to best meet the needs of library users. While several E-Metrics project participants undertake a number of assessment activities, they are not, in general, part of a systematic evaluation and assessment process that permeates the library. It is within such an assessment framework that network statistics and performance measurement activities need to reside.

- *Library data collection, analysis, and presentation management system:* The E-Metrics survey presented in this article and subsequent research activities throughout the project, demonstrate that libraries overall do not engage in a systematic and focused data collection system. Current bibliographic and management information systems, for the most part, reflect practices in the pre-web, print dominant environment. The lack of efficient information systems that pull together elementary data elements force many libraries to resort to labor intensive processes to collect data. Few, if any, have staff specifically responsible for library wide data collection, analysis, or reporting and presentation efforts. Moreover, library staff may not be trained appropriately in the various methodologies, data analysis techniques, and reporting procedures required to engage in statistical and performance measurement activities. The data collection situation is particularly problematic in the networked environment, as data collection efforts in this area require additional technical and research skills.

> **The data collection situation is particularly problematic in the networked environment, as data collection efforts in this area require additional technical and research skills.**

- *Library staff development and training*: Given the culture of assessment issue, combined with a general lack of systematic network statistics and performance measurement activities in participating libraries, it is clear that there is a need for staff development and training in both assessment and network statistics

activities. This training should incorporate an overview of the benefits and impacts of evaluation activities; the value of evaluation in decision-making and resource allocation processes; network statistics and performance measure definitions, collection activities, methodologies, and reporting systems; and the incorporation of findings regarding network statistics and measures into decision-making and resource allocation activities.

- *Network planning and evaluation activities as part of a larger context*: Research library network activities reside in a larger organizational context. Despite varying local factors, it is important to consider the planning and evaluation of library networked resources and services as part of larger organizational planning and evaluation activities. For example, the decision to subscribe to various online databases needs to be made in the larger context regarding library collection development efforts.

- *Development of multi-agency reporting systems*: It is clear that libraries do not control the use data for all networked services and resources. This is clearly demonstrated by online database vendor statistics. Libraries that receive networked services and resources from other entities, such as state library agencies or regional consortia, also need to work with those administrative entities for usage reports as well. This will help them to get a better sense of the overall use of database services from a research library perspective, for example. In such cases, usage reports will go from the vendor directly to the subscribing entity – not necessarily to all the participating members. Thus, there is a need to develop a reporting structure that goes beyond the research library in such cases. It is also necessary to construct agreements to encourage individual library statistics and to work with consortia groups to generate meaningful reports for members, etc.

- *Investment and/or modifications in infrastructure*: Network statistics and performance measures are dependent on the information technology (IT) architecture of a library, consortia, vendor, or other service provider from which the library derives service. If a library finds certain statistics of interest and/or importance, it may require investment in an IT infrastructure that enables the collection of such statistics. That investment may require IT and library staff training in order to understand the configuration as well as the statistics enabled through such a configuration.

- *Fluid nature of electronic resources and services*: The nature of electronic resources and services is fluid and makes it difficult to devise clear-cut data definitions and data collection procedures. In some cases, electronic access can trigger an entirely new conceptualization of a given information object. Libraries need to deal with the implications of this changing environment and be more flexible. Any guidelines and progress including the E-Metrics work that we described in this article are temporary and will have to be revised as we make progress.

The report suggests that there is much work yet to be done—both at the individual libraries and in the library community as a whole. Given the rapidly changing technology environment, the changing milieu of higher education, changing organizational structures within libraries, and the complexity of measuring such networked services,

it is going to be a continuing challenge for libraries to articulate, to collect and to utilize statistics that can capture their activities and demonstrate a benefit to their users and parent organizations.

Acknowledgement

Wonsik Shim is an Assistant Professor at Florida State University; he may be contacted via e-mail at: wshim@lis.fsu.edu.

Charles R. McClure is Director of the Information Institute and is the Francis Eppes Professor at Florida State University at the School of Information Studies; his website URL is: http://slis-two.lis.fsu.edu/~cmcclure/. He may be contacted via e-mail at cmmclure@lis.fsu.edu.

Notes

1. Association of Research Libraries, *ARL Supplementary Statistics 1999–2000* (ARL, Washington, D. C., 2001). Available: <http://www.arl.org/stats/sup/sup00.pdf> [January 25, 2002].
2. Lorrie A. Knight and Kimberly A. Lyons-Mitchell, "Measure for Measure: Statistics about Statistics," *Information Technology and Libraries* 20 (March 2001): 34–38; Linda S. Mercer, "Measuring the Use and Value of Electronic Journals and Books," *Issues in Science and Technology Librarianship* (Winter 2000). Available: <http://www.library.ucsb.edu/istl/00-winter/article1.html> [January 25, 2002]; and Peter R. Young, "Measurement of Electronic Services in Libraries: Statistics for the Digital Age," *IFLA Journal* 24 (1998): 157–160.
3. Charles R. McClure and Cynthia L. Lopata, *Assessing the Academic Networked Environment: Strategies and Options*, (Washington, D.C.: Coalition for Networked Information, 1996).
4. Peter R. Young, "Measurement of Electronic Services in Libraries: Statistics for the Digital Age," *IFLA Journal* 24 (1998): 157–160.
5. Carol Tenopir and Eleanor Read, "Patterns of Use and Usage Factors for Online Databases in Academic Libraries," *College & Research Libraries* 61 (May 2000): 234–446.
6. Ibid., 244.
7. Deborah D. Blecic, Joan B. Fiscella and Stephen E. Wiberley, Jr., "The Measurement of Use of Web-based Information Resources: An Early Look at Vendor-supplied Data," *College & Research Libraries* 62 (September 2001): 434–453.
8. Charles R. McClure and John Carlo Bertot, ed., *Evaluating Networked Information Services: Techniques, Policy, and Issues* (Medford, NJ: American Society for Information Science and Technology, 2001).
9. International Coalition of Library Consortia, *Guidelines for statistical measures of usage of web-based indexed, abstracted and full text resources*, (1998). Available: <http://www.library.yale.edu/consortia/webstats.html> [January 25, 2002].
10. A draft version of the survey was provided to the participants of the ARL Working Group on Database Vendor Statistics meeting in Chicago on July 7, 2000. A revised pre-test

version of the survey was sent to four libraries that agreed to take part in the testing. Based on the comments from the pre-test, the study team distributed the final version to the project participants on July 31, 2000 in an email attachment. The survey instrument can be found in the E-Metrics Phase I Report available: <http://www.arl.org/stats/newmeas/emetrics/> [January 25, 2002].

11. Using a click-through mechanism, libraries can force users who try to access licensed databases to go through an intermediate web page that serves as a counter for access to databases. Two kinds of information are captured: number of times the page is accessed and the destination (databases). The mechanism works only for database access through library web pages. If the user bypasses the library web site (e.g., typing the database vendor's website directly or through stored bookmark), that access cannot be captured. The big advantage of the click-through mechanism is that the uniform usage data can be collected by the library as opposed to obtaining inconsistent usage data from different vendors.

12. The two programs seem widely used on many different system platforms. WebTrends is a commercial product of NetIQ, available: <http://www.webtrends.com> [January 25, 2002]. Analog is a free software available: <http://www.analog.cx> [January 25, 2002]. More information and sample reports are available from the websites.

13. Currently most electronic databases validate legitimate user logins by examining the origination IP (Internet Protocol) addresses included in the request. If the address matches with the one of the IP addresses provided by the library, the user is granted access.

14. For an academic institution to participate in JSTOR, it needs to pay two fees: ACF (Archive Capital Fee) and AAF (Annual Access Fee). Detailed information about the JSTOR fee structure can be found at <http://www.jstor.org/about/na_fees.html> [January 25, 2002].

15. Information on the LibQual+ project can be found at <http://www.arl.org/libqual/> [January 25, 2002].

16. This was the primary focus of the second phase of the E-Metrics project and has resulted in a set of standardized statistics and measures and procedures to collect them. Detailed information can be found in the ARL E-Metrics Phase II Report available: <http://www.arl.org/stats/newmeas/emetrics/> [January 25, 2002].

17. The ARL Phase II Report includes the methodologies and findings. Available: <http://www.arl.org/stats/newmeas/emetrics/> [January 25, 2002].

18. Amos Lakos, "The Missing Ingredient–Culture of Assessment in Libraries," *Performance Measurement and Metrics: The International Journal for Library and Information Services* 1, 1 (1999): 5.