Networked Resources, Assessment and Collection Development

by Sue Samson, Sebastian Derry, and Holly Eggleston

Available online 18 September 2004

This project provides a critical evaluation of networked resources as they relate to the library's collection development policy, identifies areas of the curriculum not well represented, establishes a reliable method of assessing usage across all resources, and develops a framework of quantitative data for collection development decision making.

Academic library collections have changed dramatically during the past decade. Reference collections, the bastion of bound volumes of indexes of peer-reviewed research literature, have been transformed into computer banks that offer these same resources in the form of networked databases offered via the World Wide Web. The benefits are overwhelming: desktop access from any computer with Internet access; keyword searching; combined author and keyword searching; limiters that include date ranges and language preferences; and, in some instances, the inclusion of or linking to the full-text of the articles retrieved. The networked environment provides challenges as well: networked resources and services tend to be more expensive; measuring and evaluating these resources is problematic; and, in most libraries, networked resources operate within the context of traditional services. The project described in this article was undertaken to provide a critical evaluation of networked resources as they relate to the library's collection development policy, to identify areas of the curriculum not well represented, to establish a reliable method of assessing usage across all resources, and to develop a framework of quantitative data for collection development decision making. The project was undertaken in four stages: a compilation of networked resources by subject discipline; an overview of costs by both individual resource and discipline; an assessment of comparable usage statistics for each resource; and an assessment of comparable usage statistics for each resource; and a review of overlap in full-text journal coverage among resources.

Curriculum Coverage

To establish an overview of the electronic resources as they relate to the collection development policy and to identify areas of the curriculum not well represented, subject bibliographers were asked to review the databases and identify those that were considered primary tools within their subjects. This methodology built on the efforts...
of Croneis and Dahlbach who used the Classification of Instructional Programs as a base for their categories. The framework for this selection was based on the joint efforts of the Collection Development Team in the design of the dynamic database that hosts the networked resources. The Web page that provides access to the dynamic database of networked resources offers four points of access: a basic link to core multi-disciplinary sources with full-text; an alphabetical link by database name; a keyword search that searches title, contents note, and keyword fields; and subject categories that list primary and related sources specific to a curricular area of study. Within these subject categories, bibliographers identified the primary databases considered to be core indexes for the curriculum. The numbers of resources identified for each curricular area were then tabulated (Fig. 2).

Cost Comparisons

Annual costs for each networked resource were also analyzed. This was accomplished in two ways. First, each electronic resource was tabulated with its annual cost. This was interestingly complicated. Certain resources are purchased in packages, others on a per-search basis, many depend on consortial arrangements, and others reflect collaborative purchasing agreements within and among departments. Identifying all of these unique elements in a spreadsheet and tabulating the total annual costs on a bar chart provided another step in the analysis.

The second analysis of the annual costs combined elements from the curricular analysis described above. Costs for each networked resource identified as primary in a particular subject category were totaled. This provided an overview of the distribution of acquisition dollars across the disciplines. Costs for multi-disciplinary databases considered primary for multiple subjects were divided by the number of disciplines and distributed.

Usage Statistics

The most complicated aspect of this analysis was in the compilation of usage statistics. For this analysis, usage was defined as a single access to a specific database. The target was to obtain comparable data that would be a general indicator of the use of each resource. The System Administrator had this exact data for a portion of the databases. These data are obtained by using EZproxy and a click-through method of counting access to each resource. When a user clicks on a proxied link to one of the resources, a unique entry is recorded into the proxy server log file. This entry records the time of the request, the specific resource requested, and the requesting computer’s IP address along with some other non-relevant information. Statistics are generated by running a small script against the log files to format them into Common Log Format (CLF). The CLF file is then

These two analyses of costs provided another aspect to the overview.

“’The most complicated aspect of this analysis was in the compilation of usage statistics.’”

Usage Statistics

The most complicated aspect of this analysis was in the compilation of usage statistics. For this analysis, usage was defined as a single access to a specific database. The target was to obtain comparable data that would be a general indicator of the use of each resource. The System Administrator had this exact data for a portion of the databases. These data are obtained by using EZproxy and a click-through method of counting access to each resource. When a user clicks on a proxied link to one of the resources, a unique entry is recorded into the proxy server log file. This entry records the time of the request, the specific resource requested, and the requesting computer’s IP address along with some other non-relevant information. Statistics are generated by running a small script against the log files to format them into Common Log Format (CLF). The CLF file is then

These two analyses of costs provided another aspect to the overview.

“’The most complicated aspect of this analysis was in the compilation of usage statistics.’”

Usage Statistics

The most complicated aspect of this analysis was in the compilation of usage statistics. For this analysis, usage was defined as a single access to a specific database. The target was to obtain comparable data that would be a general indicator of the use of each resource. The System Administrator had this exact data for a portion of the databases. These data are obtained by using EZproxy and a click-through method of counting access to each resource. When a user clicks on a proxied link to one of the resources, a unique entry is recorded into the proxy server log file. This entry records the time of the request, the specific resource requested, and the requesting computer’s IP address along with some other non-relevant information. Statistics are generated by running a small script against the log files to format them into Common Log Format (CLF). The CLF file is then
processed using a program called The Webalizer to produce usage analysis numbers. Since the analysis project was initiated, these data are now gathered for all databases and will establish a consistent counter method for comparing database use and identifying trends and patterns. This method is not perfect because it does not count those users who bookmark a direct link to a specific database. However, this method is consistent with models used in other academic libraries and provides comparable data across all resources.3,4

In order to gather this data for the databases that were not yet proxied, vendor usage reports were consulted. It was possible to refine these usage reports and select data that represented the access level targeted for this analysis. This included reports from two primary vendors who defined their data differently from one another but did include the data needed for their suite of databases. These results were tabulated for each database and provide a significant element in the overview that reflects comparable use and trends (Fig. 3).

**Full-Text Overlap**

The final step in the analysis was a comparison of journal-level full-text overlap between databases. Data was gathered to identify the percent of title overlap between the approximately sixty-five subscription databases providing full-text access to journals (Table 1). This overlap comparison was further refined by removing those databases considered to provide unstable or incomplete access to full-text journals. Databases that remained in the comparison met the following criteria: dates of coverage include access to current issues; the search capability of the database provides title level and table-of-contents searching; the complete contents of a journal is included; and the rate that titles are dropped and added is relatively low. Serials Solutions holdings data was imported into an Access database for comprehensive querying. Those databases that were identified to have a 50% or greater overlap in full-text coverage within the collection of stable database were targeted for a more detailed analysis of specific journal coverage (Table 2). The detailed analysis was also based on the specific criteria listed above but was expanded to include the value of the journal as a core title for specific curricula, dates of coverage, and the relative costs to maintain the title in multiple electronic versions as well as...
comparative overlap between individual selected databases.

**DISCUSSION**

With the advent of networked environments, the range of resources and accompanying services in academic libraries has undergone significant change during the past decade and impacted collection development priorities and acquisition budgets. In tandem with these changing services, the need to provide quantitative data in the framework of learning outcomes and substantive assessment that transition from traditional library use measures has become increasingly important as academic institutions compete for students in new market models. As networked resources were implemented, libraries began to review usage statistics provided by database vendors and were immediately confronted with a vast array of incompatible statistical data. Efforts to standardize vendor-supplied data have been initiated. The International Coalition of Library Consortia developed guidelines that included five data elements with implications for collection management: variability of data over time; ratios of queries per session; hourly use; use of e-journal collections; and the ratio of uses per FTE in a discipline; and are further discussed by Bleicic, et al. The Association of Research Libraries (ARL) initiated an E-Metrics project to develop statistics and performance measures for vendors to provide. Phase I identified current practices among ARL libraries based on: survey questionnaires; site visits to selected libraries; sample vendor reports; sample library generated reports; and follow-up interviews. Phase II identified selected key statistics and measures and standardized procedures and definitions for data collection. Another initiative is Project COUNTER (Counting Online Usage of NeTworked Electronic Resources) that is developing an international Code of Practice to govern the recording and exchange of online usage data.

These efforts are aimed at collaborating with vendors to provide libraries with appropriate and compatible usage data. Breeding and Van Epps provide practical in-house applications for librarians to document usage of electronic content and services in the current networked environment. Breeding’s approach is comprehensive and focuses on measuring the use of electronic and digital services and content within the broader scope of the library’s services, giving the traditional services equal voice. Van Epps targets the specific collection of usage statistics with a quick re-direct Web-log that provide numbers that can be compared to each other to determine which resources are being used and how much. This method of data collection does not include users who bookmark databases nor does it address how users search once they connect. However, while identifying the limitations of their approaches, Breeding and Van Epps emphasize the advantages of establishing a consistent usage measurement that reflects trends and patterns and that provides consistent data for comparison and decision making.

Another example of a locally-developed approach to collect usage data is described by Duy and Vaughan who compared locally collected data with vendor-provided data. They offer that the local tracking of data provides the benefits of being collected in a uniform manner for all products, collected in a library-controlled way, and compiled in a central location.

These local models define data collection in terms of counting the number of users that access an electronic resource and utilize the statistics to identify trends, to compare database use across all resources, and to broadly determine use patterns. How patrons use the electronic resources once that have access, the efficiencies of their search strategies, and other detailed information will have to come from vendors. To address the broad scope of service assessment, Bertol provides an excellent overview of measuring service quality in the networked environment by incorporating the findings from research efforts into four models of determining library network statistics and performance measures.

**PROJECT OUTCOMES**

This project expanded on the models described above by not only developing basic usage statistics for networked resources but by adding three more layers of quantitative data that provided a comprehensive overview of the electronic resources collection. This project was designed to provide an overview of networked resources as they relate to the collection development policy: 1. quantifying networked resources in terms of curriculum representation, 2. identifying costs in terms of both individual databases and curriculum, 3. establishing a reliable method to determine usage across all resources, and 4. comparing full-text journal coverage between databases.

“The administration has used this data to leverage collaborative buying power from individual Schools, Colleges, and Research and Development areas directly needing access to these resources.”

Library administration was initially interested in the databases used most frequently (Fig. 3). These data provide justification for their acquisition, establish a direct link to the curriculum and research efforts of students and faculty.
and underscore the collection development policy that targets this primary academic audience. The administration has used this data to leverage collaborative buying power from individual Schools, Colleges, and Research and Development areas directly needing access to these resources. These campus-wide initiatives have been a source of public relations for the library and have increased knowledge of the value of information resources in the Academy. Subject bibliographers were able to identify those curricular areas under-represented by networked resources, compare costs and numbers across the disciplines, identify usage patterns of each resource, and use all of this quantitative data in their renewal and selection decisions. As liaisons to curricular departments, the bibliographers were able to share specific data in support of the library collection decision-making.

For the first time, the Collection Development Team was able to review the networked resources collection from an overview of quantitative data and relate it to the collection development policy. Were all curricular areas represented? Do the distribution of costs reflect usage patterns, student and faculty numbers, and the collection development policy? More specifically, how much overlap of full-text journal coverage exists between complementary databases and can efficiencies be realized while maintaining a high level of service?

The data collected that compared journal full-text overlap between databases provided another access point for decision-making relative to the networked resources. These data when compared with costs, curricular coverage, and usage patterns for each database establishes the foundation for effective decision-making.

CONCLUSIONS

Three primary conclusions evolved from the process to evaluate networked resources. First, each discipline was identified relative to its representation within the networked collection both in number of primary databases and costs for these databases. This provided an immediate overview that could be used in conjunction with numbers of students, faculty, and degree programs to target those areas of the curriculum that were not well represented.

Second, the usage statistics provided a basis for the comparison of the use of all resources. Although not a perfect method of counting usage, these comparable data were used to identify the most frequently used databases, to compare usage across all resources, and to identify trends and patterns that exist among and between these networked resources. In addition, in the process of reviewing usage statistics, the systems administrator refined the collection of these data so that future assessment will be even more consistent and more easily compiled.

Third, the detailed analysis of overlap in full-text journal coverage provided critical data for a review of potential efficiencies. When used in concert with the data provided through an analysis of curriculum coverage, cost comparisons, and usage statistics, these data were essential in decision-making that made it possible for the Collection Development Team to target potential databases for review while maintaining a high level of service.

 NOTES AND REFERENCES

15. Duy & Vaughan, “Usage Data for Electronic Resources.”