

## Chapter 14

### **Uses of Necessity or Uses of Convenience? What Usage Statistics Reveal and Conceal About Electronic Serials**

NORM MEDEIROS

#### *INTRODUCTION*

The standardization and assessment of processes related to managing electronic resources have necessarily lagged a few years behind the inception of such processes. As an example, electronic journals (e-journals) had been available for several years before Nancy Olson's *Cataloging Internet Resources* was published in 1995.<sup>1</sup> Until the publication of Olson's seminal text, pioneering catalogers developed uncoordinated local practices within the construct of AACR2 and MARC to provide bibliographic description for these elusive resources. Several years before the Digital Library Federation (DLF) commissioned Tim Jewell of the University of Washington to investigate management issues associated with electronic resources, many academic libraries had been recording administrative metadata about e-resources within spreadsheets and homegrown databases. As a result of Jewell's study, the DLF sponsored the Electronic Resource Management Initiative (ERMI) to investigate the needs of libraries with respect to electronic resources. The final report of the ERMI group, two years in the making, was released in August 2004.<sup>2</sup> It contained detailed, comprehensive specifications that identified the elements, values, and relationships that formed the foundation of e-resource

administration. Functional specifications, 40.1 through 40.7, are devoted to usage statistics, one of the next important challenges for libraries.

### ***A BRIEF HISTORY OF E-JOURNAL USAGE STATISTICS***

Measuring the use of serials is not a new phenomenon. One needs only to browse the dusty bound volumes of *Library Literature* to see how ubiquitous this activity was in past decades. Goehlert,<sup>3</sup> Maxin,<sup>4</sup> and Broadus<sup>5</sup> have described the value of usage statistics when making collection development decisions about print journals. Impetus for these studies of the late 1970s and early 1980s was the rising cost of serials, especially scientific/technological/medical (STM) journals. Sadly, the serials cost crisis has not only persisted, but also been exacerbated by the advent of the e-journal.

In 1997, the JSTOR Users Group created a Web Statistics Task Force to investigate desirable usage statistics reporting.<sup>6</sup> Early in their work, the Task Force discovered that reporting inconsistencies and other problems, more plentiful than had been anticipated, might be alleviated if the vendor community adopted a common set of standards. In response to the JSTOR investigation, the International Coalition of Library Consortia (ICOLC) developed guidelines in the following year that were adopted by dozens of consortia throughout North America and Europe, citing a “responsibility to their library members to ensure the provision of usage information of licensed electronic resources.”<sup>7</sup> These guidelines included elements such as number of queries, logins, turnaways, and item-type displays (for example, tables of contents; abstracts; full-text articles). These data were considered desirable for learning how and how often e-resources were being used. On the heels of the ICOLC guidelines, Project COUNTER (Counting Online Usage of NeTworked Electronic Resources), an international initiative

developed in 2002 in large part by publishers in the United Kingdom, sought to create “services to be measured in a credible, consistent, and compatible way using vendor-generated data.”<sup>8</sup> This code, which dictates how sessions and file types should be counted, recorded, and delivered to libraries, has been adopted by a growing number of e-resource suppliers. The gradual increase in “COUNTER-compliant” suppliers has been a significant step toward realizing a global set of standards for measuring usage of electronic resources. Although other means of usage reporting have been implemented, COUNTER has become the de facto standard for the scholarly publishing community.<sup>9</sup>

### ***Why Do We Track Use?***

Libraries have long been fond of measuring usage of the resources they provide. As Weintraub notes, “Librarians in the all-print library were happy with their relative understanding of how much their collection was used. They could get information by keeping records of circulation, journal browsing, and even a cursory look at how well worn their books were. In the electronic world, we do not have those cues.”<sup>10</sup> Circulation is among the most tracked statistic in libraries, but other statistics, both for collections and services, are also monitored. Commonly tracked statistics include reference interviews, Web site accesses, and gate counts. Libraries are often asked for these and other data from statistics-collecting agencies. Consequently, reporting to these bodies is a key reason for maintaining such statistics.

Although, recently, at least one statistics-collecting agency has asked about e-journal usage for a certain set of resources, libraries that track usage do so because they feel such data can help with current or future collection development. Usage statistics often contribute to, and at times dictate, retention and cancellation decisions. Tracking

usage, in a manner that incorporates financial review, can determine whether it makes sense fiscally to retain a subscription or, instead, to gain access to needed articles through a document supplier. With respect to print journals, this practice has never been exact. As Galbraith reminds us, “Usage studies of print collections assign one use per journal issue or volume, regardless of how many articles within the issue or volume were reviewed.”<sup>11</sup> The inexplicable nature of article-level usage with print journals creates a fundamental flaw with the data collected. As a result, comparing or integrating print journal usage with e-journal usage, which is recorded at the article level, is worthless. In light of this “apples to oranges” comparison, and of the continued migration of print holdings to electronic, the need for maintaining statistics on print journal usage must be called into question. Yet, the continued rise of serials subscription costs, the splintering of journal titles, and the multi-payment of content (for example, paying for current issues as well as back issues as well as aggregator content) are strong catalysts for maintaining e-resource usage statistics.

### *Spreadsheets Galore*

Despite the adoption of COUNTER-compliant usage statistics by a growing bloc of e-journal providers, the maintenance of usage statistics is a labor-intensive activity. COUNTER statistics by rule port to spreadsheets ably, as do many usage reports provided by non-COUNTER-compliant providers. Merely accumulating spreadsheets of e-journal usage, however, does not provide the kind of immediate and longitudinal analysis libraries need, in order to make both informed and on-demand collection management decisions. Still, the practice of storing countless spreadsheets of usage data

predominates among statistics maintainers, since no agreeable alternative presently exists.

Complicating matters is the desire to analyze usage against expenditures. Since fee-based document suppliers can provide immediate online access to a large body of current journal articles, determining whether a journal subscription or a pay-per-use approach to article provision is the more fiscally sensible route has become an important decision making factor. Typical usage analysis compared to cost data might look like the following (Figure 14.1):

Journal Title	Use	Cost	Cost-Per-Use
<i>XYZ Journal</i>	25	\$1000.00	\$40.00

If individual articles from the notional *XYZ Journal* could be purchased for \$20 through a document supplier, a library might consider surrendering its subscription to that journal, since the cost of maintaining a subscription (\$1000) is greater than the amount the library would spend to purchase the needed articles (\$500). This example presupposes that the cost-per-use measurement is accurately calculated (although, it is not); that the number of uses from *XYZ Journal* would remain constant if a subscription was not maintained by the library (although, highly unlikely); and that the staff costs of mediated or even unmediated article delivery are comparable to staff costs associated with maintaining a subscription (although difficult to determine, but probably not).

### ***Cost-per-use: Meaningless and Error-Laden Data***

The easiest of these three presuppositions to discredit is the cost-per-use calculation. As already noted, an important reason for tracking the usage statistics of e-journals is for collection development purposes. The disaggregating of “big deals” is

opportune, but it also means that libraries must understand the curriculum and the research needs of faculty and students in order to identify the specific e-journals within a publisher's stable that are most valuable to their academic community. Part of this identification process is determining whether the retention of a subscription versus reliance on document delivery is the more fiscally-sensible option. Cost-per-use, as illustrated in the previous section, is the means most commonly employed to make this determination. The figure is calculated by dividing the number of full-text uses reported in a year by the subscription cost of the e-journal ( $\text{cost}/\text{use}=\text{cost-per-use}$ ). This calculation ignores staff costs incurred in maintaining a subscription, as well as staff costs that would be incurred if an article-based document delivery medium was employed in favor of a subscription. Staff costs aside, the cost-per-use calculation as generated in this section, is flawed.

This easily recognizable flaw is somehow overlooked or purposely ignored by most libraries maintaining such cost-per-use statistics. Holmstrom illustrates this flaw, replacing the phrase cost-per-use with "return on investment."<sup>12</sup> Holmstrom argues that cost-per-use calculations do not consider uses of journals from years other than the subscription period under review. Using our *XYZ Journal* as an example, if the 2004 subscription cost was \$1000, and the e-journal was accessed twenty-five times during that year, the cost-per-use of *XYZ Journal* as typically calculated would equal \$40. This calculation is inherently flawed, however, since publisher-provided statistics do not reveal the percentage of these twenty-five uses that were of 2004 journal issue content only. Even COUNTER-compliant usage statistics do not yet require this level of specificity—the age of the material being accessed—probably because few libraries are

prepared to utilize such detail. Nonetheless, if collection development decisions are being made based on cost-per-use calculations, libraries ought to apply a more rigorous calculus.

Holmstrom suggests an algorithm drawn from studies performed by Carol Tenopir and Donald King in the 1990s, and documented in their seminal text, *Towards Electronic Journals*.<sup>13</sup> Tenopir and King conclude that issues from the current year of science journals constitute 58.5 percent of all uses of these journals in a given year, whereas, the remaining 41.5 percent of uses are from issues published in previous years.<sup>14</sup> In the *XYZ Journal* example, the 25 uses recorded in 2004 would equal 14.6 uses from calendar-year 2004 content, using the Tenopir–King formula. Instead of a cost-per-use equaling \$40, the cost-per-use would equal \$68.49 (\$1,000 divided by 14.6). Yet, this cost-per-use calculation is not quite accurate either, since it does not take into account the future uses of 2004 journal content. According to the Tenopir–King formula, articles from the 2004 issues of *XYZ Journal* would represent 12.3 percent of the uses recorded for the journal in 2005. If usage for *XYZ Journal* grew by 20 percent in 2005, meaning the journal was accessed 30 times, the Tenopir–King formula estimates that 3.7 of these 30 uses, or 12.3 percent, would be from 2004 journal issue content. When added to the 14.6 uses of 2004 journal content accessed during 2004, the new cost-per-use figure is \$54.64 (\$1,000 divided by (14.6 + 3.7)). Looking outward one additional year, during calendar year 2006, usage of 2004 journal content would represent 6.2 percent of all uses, according to the Tenopir–King model. If usage of *XYZ Journal* in 2006 grew, once again by 20 percent, bringing overall usage to 36, 2.2 of these uses would be from 2004 content. When factored into the 2004 equation, the new cost-per-use equals \$48.78

(\$1,000 divided by (14.6 + 3.7 + 2.2)). Given the limitation of current statistics-reporting capabilities, cost-per-use should be derived from a Tenopir–King-like formula, which factors in all uses of an e-journal over the lifetime of that e-journal. This model seems to be a more precise means of calculating cost-per-use.

Unfortunately, usage statistics measure only utility, not value. As Luther notes, “It is dangerous to assume that a popular title, which is used by many students, is worth more than a research title that is used by only a few faculty members working in a specific discipline. Other factors need to be considered.”<sup>15</sup> Indeed, caution must be exercised when considering how to weigh publisher-provided usage statistics within the decision-making framework. In the era of print-only serials, librarians recognized that usage studies in open-stacks environments underrepresented actual use. Consequently, the role of these studies in journal renewal/cancellation decisions was comparatively inferior to the role that usage reports for e-journals are tending to play in today’s libraries. The ability to count article-level uses of e-journals, a relative impossibility with print journals, contributes to the exalted role that online usage statistics play in collection development decisions. One fundamental and critically important difference between library-administered journal use studies and publisher-provided usage statistics is motive. Libraries want to know which journals are being used and not being used by patrons, in order to make informed journal renewal and cancellation decisions. Publishers’ interest in what usage statistics reveal about their e-journals may not be so impartial.

### ***Trusting Usage Statistics***

Some authors have alluded to the potential of e-serials providers manipulating usage statistics in order to inflate usage.<sup>16, 17</sup> It is not hard to imagine some publishers

padding actual use as a way of enticing renewal of certain middle-tier journals. Perhaps in response to this concern, COUNTER employed a rigorous auditing requirement to ensure the authenticity of statistics. The *COUNTER Code of Practice* details the measures used to enforce “credibility, consistency, and compatibility” of usage statistics.<sup>18</sup> The audit test for the most coveted statistic, “Journal Report 1: Number of Successful Full-Text Article Requests,” includes the following:<sup>19</sup>

1. The auditor performs 100 requests for full-text articles from a selection of available journals.
2. The auditor records the requests he made from each journal accessed.
3. The auditor compiles a report that lists the requests, separating them by journal.
4. The publisher passes the audit test, when the auditor’s report and the publisher’s report are within a -8 percent and +2 percent reliability window.

A number of additional audit tests occur for COUNTER reports. These include testing “double-clicks” to ensure that clicks for an HTML file that occur within a ten-second window, and clicks for a PDF file that occur within a thirty-second window, are counted only once; verifying that the contractual number of simultaneous accesses (user seats) are available; and ensuring proper session and turnaway counts. This third party auditing should give libraries a measure of confidence about the credibility of publisher-supplied COUNTER-compliant usage statistics, particularly comforting as pricing models based on use begin to appear. At the same time, perhaps this rigorous auditing

procedure is a reason why at least some e-resource providers remain outside the COUNTER initiative.

### *Uses of Necessity, Uses of Convenience*

As useful as they may be, COUNTER-compliant usage statistics cannot record motive for use. As Phil Davis states, “While usage statistics can tell us so much about *how much* a journal or resource is being used, it cannot tell us *why* it was used” (emphasis in original).<sup>20</sup> Although empirical data on uses of necessity and uses of convenience will never become an available COUNTER report, clearly there is some number of e-journal uses that merely meet a professorial requirement (for example, a certain number of journal articles necessary in a bibliography) versus satisfying an information need; that is, being an article that answers well—the question at hand. Many a reference librarian has watched helplessly, while a patron accepts a full-text article that can be accessed immediately and printed, in place of the more-pertinent article for which only a citation exists online. Granted, if the student’s paper is due in one hour, and the journal in which the more-pertinent article resides would need to be requested via interlibrary loan, accessing the less-relevant but full-text article is understandable. Time constraints, or more typically—procrastination, however, aren’t always the culprit. Today’s culture of convenience and immediate gratification, which encourages this last-minute behavior, is further evidence that e-resource usage statistics should only augment, not dictate, collection development practices. There are numerous reasons why an obscure e-journal might receive relatively high usage, none of which may have anything to do with ongoing curricular or research needs. E-journals with longitudinal data revealing strong usage should cause notice; one-year wonders should not.

### *Administering Usage Statistics*

Despite flawed cost-per-use calculations, uses of convenience, and the possibility of artificially inflated usage statistics provided by deceitful publishers, the maintenance and analysis of usage statistics for e-journals are important activities. To date, most libraries that have chosen to expend the effort to compile usage statistics have created spreadsheets into which they input or download data. This means of maintaining statistics make data sharing difficult. A more attractive model is a statistics repository where various staff and/or institutions can store and retrieve statistics. For example, the centralization of usage statistics would be especially useful in a consortium environment where resources are purchased collectively and used on separate campuses. If COUNTER-compliant usage data could be ingested into a repository, statistics could be available to appropriate members of a library or consortium community in a more timely, less staff-time-intensive manner.

Before library information systems vendors begin expending energy building such statistics repositories, the library community must determine what it hopes to achieve by way of usage statistics. Caryn Anderson of Simmons College surveyed members of the Electronic Resources in Libraries (ERIL) LISTSERV in summer 2004 with just this question in mind.<sup>21</sup> Respondents indicated that a system, which could integrate usage statistics with financial data, make comparisons with peer institutions, and provide subject analysis, would be a powerful tool. Anderson used the survey to help her develop the model for her ERUS (Electronic Resource Usage Statistics) database, a repository that can store COUNTER statistics and provide access to these by subject and resource type. Andrew Nagy, a library technology development specialist at Villanova University,

has developed a similar statistics-ingesting/reporting/managing system named LibSGR (Library Statistics Gathering and Reporting). Nagy's application imports comma-delimited files of usage statistics, and is capable of delivering various reports generated from these data. Anderson, Nagy, and their colleague, Tim McGeary, a senior systems specialist at Lehigh University, have agreed to combine their efforts into building a decision support system. The system will utilize open standards and provide libraries with an application to help relate usage statistics with other decision criteria, such as resource price, impact factor, and faculty interest. The group plans to share the results of their collaboration in the future.

### ***CONCLUSION***

Few would argue that standardized and credible usage statistics for electronic resources should not be used as a complement to other journal utility and value measures. Fewer still would argue that usage statistics maintained in an automated fashion and fused with other decision criteria would be worthless to them. Clearly, usage statistics for electronic journals will become only more important as collections shift from print to electronic. Although libraries have little choice but to rely on publisher-provided statistics for electronic journals, at least those based on the *COUNTER Code of Practice* are standardized and audited. As use-based pricing takes hold as a viable, and perhaps, popular pricing model, libraries will have an added incentive to maintain timely statistics, whether in traditional spreadsheets, or within the framework of a centralized database along the Anderson/Nagy/McGeary model. Lastly, despite everything that usage statistics cannot reveal, they are still the best utility measure available to libraries. We must accept their shortcomings and utilize their potential.

## NOTES

1. Nancy B. Olson, *Cataloging Internet Resources: A Manual and Practical Guide* (Dublin, OH: OCLC, 1995).
2. Timothy D. Jewell, et al. *Electronic Resource Management: The Report of the DLF Initiative*, August 2004, <http://www.dlf.org/pubs/dlfermi0408> (April 18, 2005).
3. Robert Goehrlert, "Journal Use per Monetary Unit: A Reanalysis of Use Data," *Library Acquisitions: Practice and Theory* 3 (1979): 91–98.
4. Jacqueline A. Maxin, "Periodical Use and Collection Development," *College & Research Libraries* 40 (1979): 248–253.
5. Robert N. Broadus, "The Measurement of Periodicals Use," *Serials Review* 11 (1985): 57–61.
6. Jim Mullins, "Statistical Measures of Usage of Web-Based Resources," *Serials Librarian* 36 (1999): 207–210.
7. International Coalition of Library Consortia, *Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full-Text Resources*, *Serials Librarian*, November 1998, <http://www.library.yale.edu/consortia/webstats.html> (April 19, 2005).
8. *COUNTER Code of Practice*, December 1, 2002, [http://www.projectcounter.org/code\\_practice.html](http://www.projectcounter.org/code_practice.html) (April 15, 2005).
9. Peter T. Shepherd, "Industry Initiatives," in *Online Usage Statistics: A Publisher's Guide*, ed. Bernard Rous (New York, Association of American Publishers, 2004): 47–56.

10. Jennifer Weintraub, "Usage Statistics at Yale University Library," *Against the Grain* 15, no. 6 (2003): 32–34.
11. Betty Galbraith, "Journal Retention Decisions Incorporating Use-Statistics as a Measure of Value," *Collection Management* 27, no. 1 (2002): 79-90.
12. Jonas Holmstrom, "The Return on Investment of Electronic Journals: It is a Matter of Time," *D-Lib Magazine* 10, no. 4 (2004).  
<http://www.dlib.org/dlib/april04/holmstrom/04holmstrom.html> (April 15, 2005).
13. Carol Tenopir and Donald W. King, *Towards Electronic Journals* (Washington, DC: Special Libraries Association, 2000).
14. Holmstrom.
15. Judy Luther, "White Paper on Electronic Journal Usage Statistics," *Serials Librarian* 41 (2001): 119-148.
16. Ibid.
17. Terry B. Hill, *Using Traditional Methodologies and Electronic Usage Statistics as Indicators to Assess Campus-wide Journal Needs: Contexts, Trade-offs, and Processes*, April 13, 2004, Master's thesis, University of North Carolina at Chapel Hill,  
<http://hdl.handle.net/1901/64> (April 15, 2005).
18. *COUNTER Code of Practice*.
19. Ibid.
20. Phil M. Davis, "Why Usage Statistics Cannot Tell Us Everything, and Why We Shouldn't Dare to Ask," *Against the Grain* 15, no. 6 (2003): 24–26.
21. Caryn Anderson, "Survey Results," February 2, 2005,  
<http://web.simmons.edu/~andersoc/erus/results.html> (April 15, 2005).

22. Andrew Nagy, "LibSGR: Library Statistics Gathering and Reporting," c2004,  
<http://liboffice.villanova.edu/libmgr/> (April 15, 2005).