November 2005

Maximizing the Value of Usage Data

Elise Anderson

Wake Forest University, andersem@wfu.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Anderson, Elise (2005) "Maximizing the Value of Usage Data," *Against the Grain*: Vol. 17: Iss. 5, Article 6.

DOI: <https://doi.org/10.7771/2380-176X.4548>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

UNCG concluded that automatically dropping subtitles from entered searches would do more good than harm, and made the appropriate programming changes in **Journal Finder**. A subsequent analysis showed that this change eliminated over a thousand failed searches per year.

Another pattern that emerged in reviewing the "no hits" log was that the data in the added title fields was insufficient. For example, one user searched for "News Week" when they no doubt wanted "Newsweek." Another used the American spelling to search for "Behavioral Neurology" when they really wanted the British journal "Behavioural Neurology." In both cases, the patron entered the "wrong" title and thus did not get a title match. By subsequently including these types of "incorrect" title searches in the added title field, **UNCG's Journal Finder** learns from actual patron mistakes and improves service for future users. Augmenting **Journal Finder's** already strong added title data with these patron-generated titles improves search accuracy and is responsible for 1.3% drop in the number of searches that produce no results.

Conclusion

Most libraries spend a considerable percentage of their budget on e-journals and databases. If libraries can increase usage of these electronic resources, then they will get more from their collection dollars and will be better able to justify to their administrations such large expenditures. The integration of a library's complex resource infrastructure through such technologies as link resolvers drives a significant amount of new usage into electronic collections by simplifying user access. Removing even small barriers between patrons and their desired information can have a large impact on usage levels. By analyzing actual usage data and patterns, libraries can build systems that increase usage by improving the success rate of searchers. All of these methods of adapting library systems to patron behavior not only increase user satisfaction, but they also make good business sense by allowing maximal usage of electronic collections that have predictably fixed costs.

Maximizing the Value of Usage Data

by **Elise Anderson** (Collection Management Specialist and Statistician, Z. Smith Reynolds Library, Wake Forest University) <andersem@wfu.edu>

(Adapted from a chapter in the forthcoming Haworth Press Book, Usage Statistics for E-Series.)

In the past decade electronic serials have come to dominate the serials collections of most academic libraries, both numerically and financially. A lucky few libraries may be able to provide their patrons all of the serials that they want, without constraints. In their case, usage statistics may be irrelevant and dispensable. However, most libraries must choose which electronic resources they provide, including e-series, and for them usage statistics are an important part of the decision process. Whether a library manages its e-series statistics in-house or outsources the work, developing useful statistics requires a significant investment in time, personnel, and money. The library's goal should be to maximize the return on its investment in a statistics program. One obvious area where e-series usage data can have an impact is the budget process. However, once a library has usage data, the real benefit comes from relating it to a wide range of library functions. In the future, e-series usage statistics may play a significant role in library activities as wide-ranging as instructional programs, collection management, and system administration.

Many libraries already rely on cost-per-use analyses to identify when to switch a serial subscription between print and electronic formats. An e-serial or entire e-serial package may be canceled if there is insufficient use. If an e-serial is available from more than one source, it may be possible to cancel the more expensive version and direct the savings toward other purchases (Hiott 2004). Tracking e-serial use and cost over multiple years may enable collection management librarians to anticipate and prepare for new budget demands from developing academic programs. Linking IP data to other types of usage data from vendors significantly enhances the utility of that usage data, particularly for budget questions. As electronic resources continue to increase in cost, libraries increasingly unite to make joint purchases of e-series (Watson 2003a). If the cost of a resource is shared among multiple libraries, use data linked to IP information could identify each library's proportionate use of the product, and possibly, the relative proportion of cost adjustments. Electronic resources are a boon for academic researchers but carry a heavy cost for the libraries that

provide those databases and e-journals. When library administrators can document the tremendous increase in use of e-series, as well as their even-more-rapidly-increasing costs (Jewell et al. 2004; Watson 2004), it becomes easier for university administrators

to accept the library's need for strong financial support in providing this medium to the university community.

Besides budget questions, e-serial usage statistics can help with other collection management tasks. If a particular e-serial has a consistently high number of turnaways, then users may benefit if the library can pay for an increase in the number of simultaneous users for the product. Spikes in turnaways only at certain times of a semester may indicate classroom demonstrations of a product and that the current user license is generally adequate.

Combining statistics provides expanded opportunities for interesting analyses. When an e-serial shows a high number of searches for each downloaded unit, it may indicate that users are having problems with the serial. They may be misinformed about the purpose of the product or they may have problems negotiating its search interface (Hiott, 2004). Either way, a librarian may be able to enhance the user's success with the serial. In contrast, high downloads/session ratios usually indicate a product that users are pleased with.

New uses for e-serial statistics are multiplying beyond collection management. While bibliographic instruction programs have been staples of library educational offerings for many years, recently they have been joined by semester-long classes in information literacy. Both instructional programs introduce students to the concept of electronic resources and educate them in the use of databases and ejournals.

Usage statistics aid in determining if the repeated use of specific e-series in academic programs results in the higher use of those same products by students. Because bibliographic instruction programs began well before the availability of usage statistics, their effects on the use of e-series are harder to track. However, if a library tracked e-series use before the advent of information literacy classes, and the instructors can identify which e-series they included in their teaching, then it should be possible to identify significant changes in students' use of those serials. If statistics can document such a change in e-serial use patterns of students, this would provide strong evidence to libraries and university administrators of the effectiveness of these programs and of the need to carefully evaluate the resources used for teaching.

Evidence of the teacher's ability to influence students' choice of electronic resources might indicate a way to promote under-utilized electronic resources as well as a need to avoid endorsement or inadvertent censorship. Many libraries have large collections of electronic serials. **Reynolds Library** offers users more than 200 databases and several thousand electronic journals. Students and library staff alike are faced with a daunting task of identifying which e-series are best suited to which purpose. Many **Wake Forest University** students resort to using a standard cluster of products, such as **ProQuest** and **EBSCOHost** databases, to answer most of their research questions and consult other e-series on the advice of peers, professors, and librarians. Librarians, faced with a student who has waited

continued on page 22

Maximizing the Value ... from page 20

until the last minute to begin a research paper, can rarely interest the student in the complete range of e-serials that cover the topic. As a result, every library has paid for electronic databases, ejournals, and other resources that do not get the use they deserve. Often these end up being cancelled because the cost per use is too high to be affordable (Moore-Jansen, Williams, and Dadashzadeh 2001; Tenopir 2001, Hiott 2004).

Bibliographers, collection management librarians, and library instructors could work together in a process aimed at increasing the student's knowledge of the entire range of e-resources for the best academic performance. For example, at regular intervals, bibliographers and library liaisons could review the usage statistics for e-serials and identify likely candidates for cancellation. Library instructors would then take the lead in demonstrating one or more of these under-utilized serials in their classes. With usage data from before and after the instructors' intervention, it should be possible to detect if there was a positive or negative effect on students' use of the targeted e-serials.

If there are generally positive results with this technique then libraries could regularly identify their lowest performers and give them wider publicity through instructional programs. If an e-serial's use fails to improve after this "probationary period", then collection managers may feel more confident that librarians are canceling products that are unsuitable for their users and not just unknown to them.

Students are noted for wanting to use only products providing full text (Tenopir 2001, Dilevko and Gottlieb 2002, Black 2005). Yet citation databases often provide the best resources for a particular research topic. Student preference for full-text resources have been attributed to multiple factors, including the convenience of electronic full text and to the common situation where photocopying print materials costs the student directly but printing out electronic text does not (Black 2005). In addition, it is also well documented that students are very confused and frustrated by the process of moving from identifying a reference in one source to retrieving the full text for that reference from another source. (Black 2005, Labelle and Nicholson 2005).

If BI and information literacy programs are effective in making students more comfortable with the library research process, then usage statistics for citation databases will rise from previous levels. If so, libraries could again target low-performing citation databases for a little extra attention in instructional programs and use usage data to monitor the results.

Even relatively mundane activities such as system administration can benefit from input from usage statistics. Wake Forest University recently added a wireless component to its hardwired network system. The wireless system includes connections for guest, faculty and staff, and students. Guest users have no access

to networked printing and file sharing nor to the library's small collection of locally-mounted CD databases. In the first months of the dual system, large numbers of students were using the guest wireless system rather than the system intended for them. As a result, they were unable to access local databases. The only way that Reynolds Library personnel were alerted to the problem was by the sudden tripling and quadrupling of the usage statistics for those databases. Databases that previously averaged 10-12 sessions per month were abruptly showing 50 sessions per month, usually with a single IP address, which represented one user, attempting 2 and 3 login attempts within a period of 60-90 seconds. Once librarians began advising students of the need to use the student wireless for these databases, the usage statistics dropped significantly though not back to previous levels. The pattern of multiple attempts in a short time is still apparent, indicating that not all students received the message.

The biggest challenge in maximizing the utility of e-serial usage data may be simply recognizing new opportunities for using this data. As new technologies appear they can exert both positive and negative effects on the use of e-serials. For example, as bibliographic management software such as **EndNote** or **Reference Manager** becomes more prevalent on campuses, users may respond by focusing on e-serials that work best with those systems. Libraries could monitor its e-serial usage data for such indications and develop appropriate responses. Developing useful statistics requires a significant investment in time, personnel, and money. Diligence and creativity in using e-serials statistics can enable a library to reap the full benefits of its investment in a statistics program.

Deep Linking — Beyond Journal Articles

by **Chuck Hamaker** (Associate University Librarian
Collections and Technical Services, Atkins Library, University
of North Carolina at Charlotte) <cahamake@email.uncc.edu>

Usability has become the single most critical function for e-resources. Usability must be improved dramatically for libraries and the electronic resources they support. Deep linking must become ubiquitous from catalogs, from Abstracting and Indexing services and from other identification tools for content beyond journal articles. Its lack is a major obstacle to the full use of networked electronic content.

For music CD's **Gracenote** formally known as **CDDb** working with the music industry worldwide has compiled metadata for consumer use on 3.6 million CDs and 46 million tracks. **Apple's iTunes** supports 30 second previews for everything and links at the artist, album, and track levels for commercial use. Some similar

metadata center or centers, some clearing house of stable syntax or permanent networked identifiers, much as **CrossRef** provides for journal articles, needs to exist for all networked e-content supported by libraries, vendors and publishers of e-content. **CDDb's** goal is to "enable consumers to better manage, enjoy and discover digital media." That sounds similar to the goals for the library, vendor and publisher community with networked digital content: discover, use, manage, and preserve.

Site and title level linking is currently the standard for library catalogs (**OPACS**). Such linking fails to support robust exploration and retrieval - how well systems answer queries and deliver content. Title level linking is inadequate compared with the deep linking available from **OpenURL** supported indexes and from famil-

continued on page 24

Endnotes

- Black, Steve. 2005. Impact of full text on print journal use at a liberal arts college. *Library Resources & Technical Services* 49 (1):19-26, 56.
- Dilevko, Juris and Lisa Gottlieb. 2002. Print sources in an electronic age: A vital part of the research process for undergraduate students. *Journal of Academic Librarianship* 28 (6):381-192.
- Hiott, Judith. 2004. Collecting and using networked statistics: Current status, future goals. *Library Quarterly* 74 (4):441-454.
- Jewell, Timothy D., Ivy Anderson, Adam Chandler, Sharon E. Farb, Kimberly Parker, Angela Riggio, and Nathan D. M. Robertson. *Electronic resource management: Report of the DLF Initiative*. Digital Library Federation 2004 [cited April 1, 2005]. Available from <http://www.diglib.org/pubs/dlfermi0408/> <http://purl.oclc.org/DFL/dlfermi0408>.
- Labelle, Patrick R. and Karen Nicholson. 2005. Student information research skills: Report on a Quebec study on information literacy. *Feliciter* (1):47-49.
- Moore-Jansen, Cathy, John H. Williams, and Mohammad Dadashzadeh. 2001. Is a decision support system enough? Tactical versus strategic solutions to the serials pricing... *Serials Review* 27 (3-4):48-61.
- Tenopir, Carol. 2001. Database and online system usage. *Library Journal* 126 (16):41-45.
- Watson, Paula D. 2003. Acquisitions of e-journals. *Library Technology Reports* 39 (2):28-41.
- 2004. Raised expectations/rising prices. *Library Technology Reports* 40 (6):40-41.