November 2005

Getting More From Your Electronic Collections Through Studies of User Behavior

Tim Bucknall

University of North Carolina at Greensboro, bucknall@uncg.edu

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

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

Recommended Citation

Bucknall, Tim (2005) "Getting More From Your Electronic Collections Through Studies of User Behavior," *Against the Grain*: Vol. 17: Iss. 5, Article 5.

DOI: <https://doi.org/10.7771/2380-176X.4547>

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.

Against the Grain

c/o Katina Strauch
209 Richardson Avenue
MSC 98, The Citadel
Charleston, SC 29409

CHARLESTON CONFERENCE ISSUE

VOLUME 17, NUMBER 5

NOVEMBER 2005

Against the Grain

"Linking Publishers, Vendors and Librarians"

TM

ISSN: 1043-2094

Getting More From Your Electronic Collections Through Studies of User Behavior

by **Tim Bucknall** (Assistant Director - Jackson Library, Head, Information Technologies and Electronic Resources, University of North Carolina at Greensboro) <bucknall@uncg.edu>

Libraries spend a lot of money on electronic resources and understandably want to get the best possible return — in other words, the most usage — on that investment. Publicity, bibliographic instruction, and prominent links on Web pages are all common ways we seek to enhance awareness of expensive commercial databases and e-journals. While these methods have achieved some measure of success in generating more initial Web hits, they do little to increase levels of use once a given resource is being utilized. By analyzing patron behavior and preferences within Web-based information sources and services, libraries can adjust their electronic collections environments to better meet user needs and generate significantly more traffic.

Integration Boosts Use

At most libraries several years ago, one hundred percent of database accesses were initiated when a user clicked on the name (such as "Infotrac OneFile") somewhere on the Library's Website. As libraries and technology have become more sophisticated, additional access points have been created. For example, the recent widespread introduction of link resolvers has produced another major pathway into commercial databases, and has generated a significant increase in the statistical picture of patron activity.

At the **University of North Carolina at Greensboro (UNCG)**, the library catalog and the majority of the subscribed databases link directly to journal titles or, in many cases, articles

in other databases through a locally developed knowledge base and link resolver, **Journal Finder**. By comparing the number of times patrons access a given database by its name against the number of times they gain entry through links in **Journal Finder**, we can determine how expanded title and article level access has increased database usage. For the period of July 2002 to June 2003, I selected a random sample of UNCG's **OpenURL**-enabled databases, all of which were also linked from the catalog and from various Web pages listing our database offerings.

continued on page 18

If Rumors Were Horses

Well, the saddest news is that **Lyman Newlin — Papa Lyman** — passed away on Tuesday, September 20, at the ripe old age of 95. **Papa Lyman** will be remembered by all of us for his charm, wit, anecdotes, raspy voice, bell ringing at the **Conference**, outspokenness, incredible memory and I could go on and on. We are going to try to institute a new column called **Remembering Papa Lyman** and **John Perry Smith** has agreed to be the editor of it. So, send remembrances of **Papa** or of any of the era that was his (the twentieth century) to either me <kstrauch@comcast.net> or to **John** <jps@totalinformation.com>.

We also have an obituary of **Lyman** by **Richard Abel** and **Fred Gullette** in this issue, p. 8. May he rest in peace.

Julie K. Miller has been promoted from Publicist to **Marketing Manager** at **CQ Press**. For the moment she is handling publicity projects until a new publicist is hired. **Julie** is also now using her married name instead of "Keisman." Her

email address is <jmiller@cqpress.com>.
www.cqpress.com

Just had an interesting conversation with the astute **Douglas Black** <dblack@copyright.com> of **Copyright Clearance Center**. The news is that **Copyright Clearance Center** has combined with **Blackboard** to dovetail copyright licensing within the **Blackboard** workflow. Nearly all colleges and universities use course management systems, and many users are faculty. This means that copyright responsibility has passed from librarians and information specialists to faculty, who usually do not have extensive copyright training. Meanwhile, college and university administrators have heightened sensitivity to all forms of copyright compliance. This integration is an important development within that trend. The product will typically be implemented so that the licensing trans-

continued on page 6

What To Look For In This Issue:

<i>Biz of Acq</i>	66
<i>Case Study One</i>	72
<i>The More Things Change, the More They Stay the Same</i>	75
<i>Group Therapy</i>	78
<i>International Dateline — Open Access Gathers Pace in Europe</i>	80
<i>The Wide World of Wikis</i>	83
<i>I Hear the Train A Comin'</i>	84
<i>Something to Think About?</i>	86
<i>Subcultures on the Internet?</i>	92
<i>Information Commons</i>	94
<i>Profiles Encouraged</i>	
<i>David Moore</i>	69
<i>Interviews</i>	
<i>Jane Burke</i>	42
<i>Howard Harris</i>	46

1043-2094(200511)17:5:1-0

Getting More ... from page 1

The number of times users entered a database by clicking on a specific journal title or article ranged from 25% to 99% of that database's total usage, while the remainder was generated by clicks on the name of the database.

Acesses through	journal/article title	database name	% accesses via journal/article
ACM Digital Library	271	4	99%
Proquest Research Library	34207	1179	97%
Science Direct	17693	2050	90%
Muse	3279	466	88%
Infotrac Onefile	35570	13842	72%
Business Source Elite	4716	2478	66%
Emerald Library	1919	1313	59%
Proquest ABI	10782	8415	56%
JSTOR	5021	4051	55%
Academic Search Premier	23342	19593	54%
LexisNexis	5529	9924	36%
Library Literature FT	1026	3159	25%

For ten of these twelve databases, there were more total uses by way of a journal or article link than through the library's list of databases. In aggregate across these twelve products, there were 2.2 database accesses by patron selection of specific journals or articles for every 1 access resulting from name selection of a database from the library's database pages or catalog.

While this data strongly suggests that adding article and title linking via a link resolver drove substantial new traffic to databases, there could be alternative explanations for these usage patterns. For example, it could be true that deep linking changed how people accessed a database without changing the total number of people who accessed it. In other words, the 1919 people who opted to use journal and title links to **Emerald** could have been the same people who would have clicked on an "Emerald Library" link if a link resolver had not been available. Other data indicates that is unlikely to be the case. The annual rate of usage increase for **OpenURL**-compliant databases exceeds that of non-compliant databases at **UNCG**, which seems to show that it is indeed the enhanced linking features that are responsible for increased use.

It is apparent that adding a link resolver increases database traffic. At **UNCG**, that increase approximates 200%, yet the database vendors do not charge us additional fees for that increased usage. By keeping database costs fairly static and by increasing usage levels through improved integration of e-resources, libraries can reduce their cost-per-use dramatically and thus make their electronic collections much

better deals for both the library and its users.

Reduce or Eliminate Barriers

Improved linking and additional access points are effective in driving up traffic because they enhance user convenience when navigating libraries' often complex lists of electronic resources. An alternative way to accomplish the same goal is to reduce patron inconvenience.

Any barriers to users will, of course, reduce overall traffic. But it is surprising to see how even a relatively small barrier can have a measurable dampening effect on patrons' willingness to pursue a given resource.

In January 2004, **UNCG** implemented pay-per-view access to all of **Wiley's** journals. Because the institution was being charged for each article viewed, the library added an authentication screen to ensure that use was limited solely to its own students, faculty, and staff. In the first four months of that year, there were 543 instances in which someone clicked on a **Wiley** journal title or article in **Journal Finder**, **UNCG's** knowledge base and link resolver. Unfortunately, those 543 clicks generated only 141 full text views. Almost 75% of the users did not follow through and obtain the available content.

In January 2005, **UNCG** joined the **Carolina Consortium**, which shares access to **Wiley's** titles among thirty-six libraries. Under the terms of the agreement, **UNCG** no longer had to pay on an article-by-article basis and thus was able to remove the authentication screen for its on campus users. Remote users were still required to prove they were affiliated with the institution. In the first four months of 2005, there were 1501 recorded accesses of **Wiley** journals at the article or title level through **Journal Finder**, and those 1501 clicks resulted in 1420 full text views.

In 2004 and 2005 **UNCG** offered the same **Wiley** titles through the same links in its **OpenURL** link resolver, **Journal Finder**. The only differences between the access mechanisms for the two years were a single authenti-

cation screen to prove affiliation before entry and a **Wiley** message stating the cost and saying that the institution would pay for it. Surprisingly, the removal of these seemingly minor impediments to users resulted in the improvement of our full text "fill rate" from only 25% to nearly 95%.

The relationship between inconvenience and turn-away rates is supported by other data as well. An examination across five of **UNCG's** pay-per-view services shows a very strong inverse correlation between the number of steps required to retrieve an article, and the percentage of people interested in a given title who are willing to follow all the way through and obtain the full text. For minimally intrusive pay-per-view services such as **AIP** and **FirstSearch**, the percentage of people who ultimately view a full-text article is about 50%. **EBSCO's** system asks for an additional piece of data, the user's email address, which has proven to be a strong deterrent to usage. Only about 13% complete **EBSCO's** entire full text retrieval process. Both **Ingenta** and **Science Direct** require multiple screens and lots of personal data prior to article fulfillment. These barriers between the text and the patron mean that fewer than 7% conclude the process and retrieve complete articles. Since **UNCG** fully funds the entire cost of each of these services, it seems apparent that it is indeed the vendors' comparative inconveniences that cause significant differences in users' willingness to complete the process of obtaining full text.

Matching System Behavior to Patron Behavior

Integration of resources and removal of obstacles to access are general strategies which can improve access to electronic products in any library. There may, however, be other enhancements more clearly targeted at local user populations. By logging actual data input and usage patterns, a library can identify weaknesses within its electronic resource environment and make needed corrections. In other words, rather than forcing user behavior to fit the library's system, it might be more effective to make library systems adapt to actual user behavior.

By capturing all failed searches in a "no hits" logfile, **UNCG's Journal Finder** provides valuable insight into how patrons respond to the University's primary journal access mechanism. The logfile allows librarians to see every search that retrieves no results, which makes it possible to identify common searching errors. The system can then be reprogrammed to handle these problems automatically. An analysis of a large random sample of failed searches at **UNCG** led to an unusual conclusion related to the use of subtitles. It turned out that the majority of users were not using truncation and were instead entering entire journal titles, frequently including subtitles. The subtitles rarely helped the user get to a particular title. In fact, our sample data set included no single instance in which the main title alone was insufficient to accurately access the desired journal. But the sample set did contain a number of spelling and other errors in the subtitles that caused false negatives that prevented a successful title match.

continued on page 20

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