Pattern of online library resource usage per user in a distributed graduate education environment

By Stefan Kramer (kramer@scn.org)

After receiving his Master's in Librarianship from the University of Washington in 1993, Stefan Kramer worked in Internet user services, first for the Northwest Academic Computing Consortium's *NorthWestNet*, and then for the University of Washington's Network Information Center. From 1999 to mid-January 2007, he was Director of Library Services at Fielding Graduate University, where the research described in this article was conducted.

Introduction

The gathering of usage data for online library resources has increased in importance as libraries and their users increasingly rely on such resources - to inform e-collection development decisions, justify and plan expenditures, and gauge the need for information literacy initiatives. Methodologies for collecting data on such usage, and discovering usage patterns, have included surveys of actual and potential end users (Bar-Ilan, Peritz, & Wolman, 2003; Kim, 2006) and retrieval and analysis of data supplied by online library resource vendors for one or multiple institutions (Black & Sisson, 2003; Tenopir & Read, 2000). Most web-based online library resources are accessed on the servers of the vendors to whose products a library initiates an institutional subscription, so the vendors can and do collect data on the utilization of those resources, as much for their own purposes as those of the subscribing libraries. The inconsistencies in format, meaning and delivery method of such data or aggregate statistics, as made available from vendors, have presented challenges to libraries for a number of years; as Shim & McClure put it, "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" (Shim & McClure, 2002). In 2002, the COUNTER (Counting Online Usage of Networked Electronic Resources - www.projectcounter.org) initiative was launched to standardize usage statistics across vendors, in order to make those more comparable for subscribing libraries, with its first "code of Practice for Journals and Databases" published in January 2003, superseded by Release 2 published in April 2005. To address the problems of *retrieval of* usage statistics, even if increasingly comparable, by libraries from dozens to hundreds of different vendors, the Standardized Usage Statistics Harvesting Initiative (SUSHI - www.niso.org/committees/SUSHI/SUSHI_comm.html) was launched in 2005. However, these initiatives are aimed at gathering statistics *by resource used*, not *by resource user*.

Focusing on the users, in a survey of undergraduate students at Florida State University in 2004, Jong-Ae Kim (Kim, 2006) examined the frequency of use of subject-specific databases (ERIC, PsycINFO and Library Literature & Information Science Full Text) in the study participants' respective fields. Of those who had any experience with these databases, almost 50% reported having used them less than 4 times during the prior academic year, 34% reported 5-10 times, and 16.5% more than 10 times[*].

[*] Exactly four times was not covered by any response choice in the survey.

The following presents the application of a home-grown online library resource access provision and statistics gathering system, and some findings of usage patterns by user.

Methodology

Fielding Graduate University (FGU) is a distributed education institution with doctoral, masters and graduate certificate programs in the applied social and behavioral sciences, and its faculty and ca. 1500 students residing mostly in North America, but also increasingly in other countries.

In 2000/2001, FGU's Library Services and Information Technology groups jointly developed and deployed an authentication and authorization scheme to provide end users (mostly students and faculty, but also staff and to some extent alumni) access to the online library resources that the institution subscribes to; these resources are listed at www.fielding.edu/about/library.asp. This scheme is based on user information exported from FGU's Datatel-based ERP system, logins to FGU's online learning environment „FELIX" (Fielding Education Link and Information eXchange, utilizing SiteScape Forum software), and EZproxy software to relay accesses to the externally hosted resources. It is also utilized in providing authorized users direct access from the online catalog of Fielding dissertations (www.fielding.edu/library/dissertations/) to the online full text of those dissertations at ProQuest/UMI (Kramer, 2003).

In late 2001, preceding the COUNTER initiative, a usage statistics gathering function was added to this scheme. It counts each click on a link for an online resource from the web page that relays authorized access requests from the FGU web site to the online resource host, after which the end user's searching session there can begin; in other words, it works by click-through data capturing. As Shim & McClure (Shim & McClure, 2002) put it, 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." However, it should be noted that the measured click-throughs register *one count*, on the FGU web site end, for *each attempt* to access an online library database or e-journal. That attempt may (and hopefully usually does) succeed, or – in case of technical problems brought about by firewall software on the end user's computer, network connectivity or vendor system problems, or other factors – fail. If it succeeds, the initiated session might last moments or hours, consist of a few or hundreds of browsing/querying operations, result in no, few, or many downloaded articles – those types of statistics can only be captured on the resource vendor's side. In short, the usage statistics gathered here measure initiated access requests to externally hosted online library resources, not what happened after those access requests. In the subsequent discussion of results, they will be identified as "accesses" for brevity.

The main purpose of this scheme was to capture comparative access statistics *by online resource*, to help inform future decisions for or against renewals of subscriptions to those resources when considered along with their costs. However, as the captured data also includes the FELIX logins of end users (www.fielding.edu/library/databases/privacynotice2.asp), it became possible to detect resource usage patterns *by user*, addressing an issue revealed in Shim & McClure's survey (Shim & McClure, 2002) of 22 ARL member libraries: "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."

Results and Discussion

Access statistics by resource

In the time period of July 1, 2005 to June 30, 2006, a total of 67,060 online library resource accesses from FGU users were recorded - 92.4% of those from students, and 6.5% from faculty/staff. Among the several dozen available online resource choices,[†] the two most frequently selected ones each accounted for 12.3% of the total; the next two for 6.7% and 6.3%, respectively. The top seven "most popular" resources accounted for just over 50% of all accesses, the top eighteen for just over 80%.

Access statistics by user

These 67,060 accesses between July 2005 and June 2006 came from a total of 1656 distinct users. The mean number of accesses per user was 39.4, with a standard deviation of 61.3 – a distribution with pronounced skewness, as evident in Figure 1.

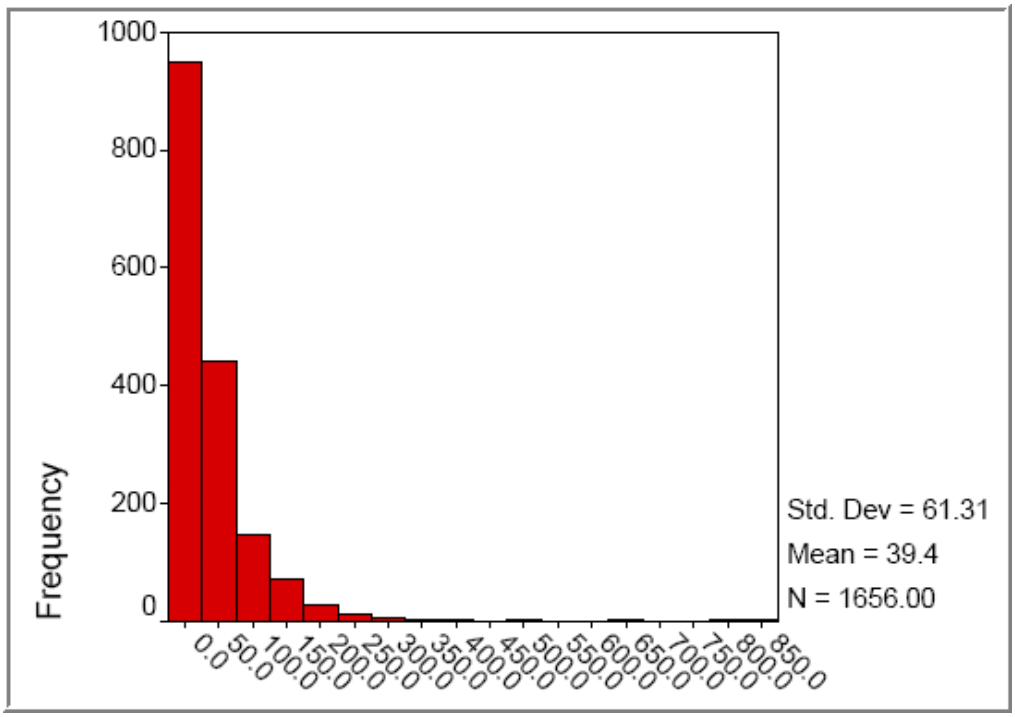

Figure 1: No. of online resource accesses per user, 7/2005-6/2006

The (here more meaningful) *median* number of accesses per user was 18, the maximum 855, and the minimum (also the mode, with 9.5%) was 1. It should be emphasized here these numbers are based on *recorded* accesses, so "zeroes" (no use whatsoever) could not be counted. The 25th

[†] These could be single databases, such as PsycARTICLES, or collections of databases, such as all available via the online service ProQuest.

and 75th percentiles were, respectively, 5 and 49 accesses. 11% of the users registered more 100 or more accesses.

The above analyses, which cover 2005-2006, had also been conducted for the time periods of July-June 2004-2005 and 2003-2004, and yielded substantially similar results. The overall number of accesses increased from year to year, but the distributions of frequency of usage per resource, and of number of accesses per user, were also quite skewed.

A question that poses itself from this distribution: aside from those faculty/students who just started their employment/enrollment near the end of one of those 12-month periods, or who left/graduated soon after its beginning, which would naturally not allow them to account for many accesses, what causes those many users represented by "the long tail" to apparently engage in only a few online library resource usage sessions? Following from that, what, if anything, could or should a library do to/for/about these "low-usage" users? Also, do they, who at *least once* became aware of the existence of these resources, differ from those students and faculty who *never* utilize them and conduct their literature research relying on other means (web search engines, print indexes/journals/books, etc.)? At FGU, a survey of "low-usage" users - students in the lowest 25%ile in terms of their online library databases access frequency from July 2003-June 2004 - conducted in mid-2005 was the first step in an ongoing effort to answer these questions. The most frequent responses to what might have caused them to utilize the databases more than they did ranged from their lack of ease of use, to having more time to explore and use them, to desiring a better orientation to them; this suggests the need for a multi-faceted approach to get "low-usage" users to increase their level of utilization of online library databases.

Summary and Conclusions

In this particular population of users of online library resources – mostly graduate (the majority of them doctoral) "mid-career" students, average age in the mid-40s; in other words, not a "traditional" university student group – the distribution of frequency of accesses is noticeably skewed. To ascertain to what degree this pattern is typical in academic settings, a comparison of published studies from various institutions, to the extent available, would be helpful. From studying such distributions, librarians may be able to learn more about the makeup of their online resource users, which may vary greatly in the degree to which they utilize these resources.

References

Bar-Ilan, J., Peritz, B. C., & Wolman, Y. (2003). A survey on the use of electronic databases and electronic journals accessed through the web by the academic staff of israeli universities. *Journal of Academic Librarianship, 29*(6), 346-361.

Black, S., & Sisson, A. (2003). Bradford's distribution, the 80/20 rule, and patterns of full-text database use. *Against the Grain*, *15*(6), 20-24.

Kim, J. (2006). Capturing metrics for undergraduate usage of subscription databases. *ONLINE, 30*(3), 32-37.

Kramer, S. (2003). Creation of an online catalog of dissertations using access & ASP. Paper presented at the *Proceedings of the Sixth International Symposium on Electronic Theses and Dissertations - ETD2003,* Berlin. from http://edoc.hu-berlin.de/conferences/etd2003/kramer-stefan/HTML/kramer.html

Shim, W., & McClure, C. R. (2002). Data needs and use of electronic resources and services at academic libraries. *Portal: Libraries and the Academy, 2*(2), 217-236.

Tenopir, C., & Read, E. (2000). Patterns of database use in academic libraries. *College and Research Libraries, 61*(3), 234-246.