December 2003

Evaluating Bibliographic Database Use: Beyond the Numbers

Steve Hiller

University of Washington Libraries, hiller@u.washington.edu

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

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

Recommended Citation

Hiller, Steve (2003) "Evaluating Bibliographic Database Use: Beyond the Numbers," *Against the Grain*: Vol. 15: Iss. 6, Article 9.
DOI: <https://doi.org/10.7771/2380-176X.4200>

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.

Second Degree

Second-degree price discrimination is where the customer is offered a choice of possible contracts that reveal information about themselves through their choice. Booking a flight months in advance to the Charleston Conference and including a Saturday stay-over distinguishes you from most business travelers who book very close to the travel date and often fly during the week. In this case, your choice reveals information about yourself and allows the producer to charge very different prices for the same service.

Third Degree

Third-degree price discrimination is where a producer can identify different types of customers and offer different contracts to each group based on their willingness to pay. Movie theatres offer one price to adults, and a discounted price to students or seniors. Many publishers distinguish different types of subscribers. For example, the *New England Journal of Medicine* sets different prices depending on whether you are a physician, student, resident, institution, or other. You are also required to disclose your country, and if subscribing as an individual, required to disclose your place of work or study. PNAS groups institutional subscribers by Carnegie classification and distinguishes academic from commercial institutions. The Ecological Society of America involves a more complex matrix for institutional subscribers involving country AND institution type. Many journals publishing online with **HighWire Press** offer free online access to developing countries. Other pricing classifications may include FTE counts (or subsets of faculty within a discipline). The practice of third-degree price discrimination is not limited to commercial publishers, and may not be considered "unfair." By charging different amounts to different subscribers a non-profit publisher may use commercial subscribers to help subsidize member subscriptions, student subscriptions, or to help run other society services.

Why knowing too much about our patrons is detrimental for library budgets (and really good for publishers).

As mentioned above, academic publishers have been practicing third-degree price discrimination for quite some time, offering differential pricing for classes of users. While more customer surplus can be extracted by practicing this type of discrimination over setting a uniform price, it is not nearly as efficient as first-degree price discrimination.

In order for publishers to move into first-degree price discrimination, they need to know more information about how their product is valued at each institution. Raw number of downloads would allow publishers to compare institutions and their use of the same product. Knowing *who* used their product and *how* or *why* they used it would allow greater leverage to price discriminate. At present, both librarians and publishers are only privy to anonymous down-

load statistics. We don't know whether it was a faculty member, a cataloger, or a student who downloaded an article, and certainly didn't know whether the download was for the purpose of research, verifying whether the resource works, or a random article pulled from a database search. If publishers knew this information, they could use it to move toward first-degree price discrimination and profit maximization. Charging libraries different amounts for downloads based on *who* and *why* they downloaded the article would yield more profits than setting a uniform 'cost per download' for all articles.

Defending Patron Privacy Can Have Same Results

Most librarians adamantly defend the right of patron privacy, and refuse in principle to sign licenses that require patrons to log on to publisher products. I say "most" since some of my colleagues do not think this is as important an issue as some believe. The unintended consequence of preserving patron confidentiality prevents personal information being gathered and analyzed by the publisher, who could then turn this knowledge into extracting consumer surpluses from each library — in other words, price maximize at each institution.

Other Ways to Gather Personal Information

There are other ways of gathering personal information about the user without requiring mandatory sign-in (and the subsequent wrath

of librarians). One way is to provide added services that go beyond the mere content of the journal. The *New England Journal of Medicine* offers a host of value-added services (like continuing medical education, table of contents services, personal customization services). In order to make these features available, the publisher must be able to clearly identify an individual, require email address, and may ask for additional personal information. Since the individual who uses these services does not pay the direct fees for the subscription, there is no incentive to withhold this information. The only caveat is that personal disclosure for services may not be accurately provided. An American publisher colleague of mine is registered for access to the NY Times online as a Chilean vintner — an illustration that many of us may not take personal disclosure very seriously.

Conclusion

Publishers currently practice third-degree price discrimination whereby different classes of subscribers are charged different prices based on their ability (or willingness) to pay. Distinguishing the type of user within the institution and determining why a resource was used would allow publishers to start practicing first-degree price discrimination — an economic environment where publishers can start maximizing profits from each institution. Protecting patron privacy also has the unintended consequence of protecting the library budget.

Evaluating Bibliographic Database Use: Beyond the Numbers

by **Steve Hiller** (Head, Science Libraries/Library Assessment Coordinator, University of Washington Libraries) <hiller@u.washington.edu>

Subscriptions to bibliographic databases comprise a substantial acquisitions investment for many libraries. At my institution, the **University of Washington**, we spend more than a million dollars annually to provide the academic community with online access to a wide range of bibliographic databases, including some full-text aggregated products. Librarians traditionally have seen indexes and bibliographic databases as essential to finding and locating the scholarly information that is at the core of academic teaching, learning, and research.

However, the times, and the information environment, they are a changing! A number of libraries, including my own, now find that use of many bibliographic databases is on the decline. So how do we measure use? How accurate and reliable are the data? What do the numbers mean? How can we use them? And why aren't our users searching these expensive databases like they used to? These are ques-

tions we need to ask and, hopefully, find answers to.

The Quantitative Approach: It's All in the Numbers

The two most common methods of determining use are number of login sessions and number of searches. Several acronymic groups have produced definitions of what constitutes a login session and a search and are working hard to standardize data definitions. NISO defines a session as:

A successful request of a database. It is one cycle of user activities that typically starts when a user connects to a database and ends by terminating activity in the database that is either explicit (by leaving the database through logout or exit) or implicit (timeout due to user inactivity). (NISO 2002)

A search is defined as representing a unique intellectual query. Typically a search is recorded every time a search request is submitted to the server.

In principle, a database pro-

continued on page 28

vider (or your local system) would provide a standardized number of login sessions and searches by database on a frequency cycle you define. You can then compare sessions and searches of that single database over time as well as usage with other databases. Given the amount spent on databases by many libraries, getting good, usable statistics would seem to be essential for evaluation and management purposes. ARL as part of its E-Metrics efforts is asking member libraries for the first time to collect and send statistics on database logins, searches, and items requested, for use in the ARL supplementary statistics. However, as **Duy and Vaughan** note in a recent article, "Perhaps one of the biggest reasons why it is difficult for libraries to use electronic resource data for decision making is the inconsistency across vendors" (Duy & Vaughan 2003). These inconsistencies (and other related factors) include:

- Vendors may define searches or sessions in different ways
- Statistics may not be provided by database, geographic site or consortium member
- Use data collectors may malfunction leaving gaps in the use record
- Time out thresholds can vary widely
- Sessions by different users can be counted as a single log-in on library computers
- Multiple sessions can occur due to system or interface problems
- Unintended user behavior may create "inadvertent" searches
- Simultaneous searching of multiple databases can be treated inconsistently
- Changes in platforms, database versions, interfaces, subscriptions, servers and data definitions can all affect data accuracy, reliability and comparability

There are other potential quantitative measures of bibliographic database use such as connect time, simultaneous users, the number of turn-aways, items retrieved or requested, and full-text links or views. However, these are generally secondary use measures and the data are even more problematic to collect, report, and analyze than login sessions and searches.

Although number of searches are considered by some to be a better measure of use and more precise than login sessions, I find number of searches a difficult figure to interpret. A large number of searches may indeed indicate a high level of activity but it can also result from interface problems or just poor search techniques. Conversely a small number of searches may reflect a higher proportion of "successful" searches, ease of interface use, or good linkages to full-text sources. Login sessions seem to have fewer variables, especially if used in an environment where much of the database searching occurs outside of library workstations. As the bound-

aries between bibliographic databases (especially with cross-database searching), full-text sources, general search engines, and the open Web further blurs, the meaning derived from some of these quantitative measures become even more difficult to ascertain.

What the Numbers Show at the University of Washington

The number of database login sessions has declined during the past several years at the **University of Washington**, especially for subject specific databases. According to our statistics, the number of sessions for the 30 databases provided through the WebSpirs interface fell from 1,121,597 in 2000 to 499,707 in 2002. That's a stunning decline of 55% - if the numbers are accurate and comparable. They're not. Server problems inflated the number of sessions during six months in 2000. During this period, some databases were dropped from WebSpirs and others were added. The most noticeable change was the switch from WebSpirs to open Internet access for Medline (PubMed). If we use 1999 as our starting point, drop 2000, and then work with the same set of databases, the 2002 figure is down about 25% from the 673,664 sessions in 1999.

While it's helpful to know that overall database use is declining, we need more detailed information for planning and management. The table below shows usage data each Spring for the past three years and includes the most heavily used WebSpirs databases. I have also provided usage numbers for three broad full-text aggregate databases.

Table 1. Bibliographic Database Usage (2001-2003).

	Spring 2001	Spring 2002	Spring 2003	Change
All WebSpirs	147,218	131,013	88,671	-39.8%
Aquatic Sci and Fish Ab	6,772	6,165	5,940	-12.3%
Biosis	10,590	9,297	6,800	-35.8%
CINAHL	5,432	5,303	4,574	-15.8%
ERIC	7,832	8,005	6,592	-15.8%
Inspec	7,862	7,385	5,102	-35.1%
Modern Language Asso	6,605	6,792	4,733	-28.3%
PsycINFO	29,570	21,605	18,774	-36.5%
Sociological Abstracts	5,236	4,335	4,178	-20.2%
3 Aggregate Databases	155,074	156,963	159,251	+2.7%
Expanded Academic	37,200	37,589	38,306	+3.0%
Lexis-Nexis	84,125	80,004	75,649	-10.1%
ProQuest Research	33,749	39,370	45,296	+34.2%

Now this is more useful data. We see the number of login sessions have declined for the WebSpirs databases, albeit at different rates. Meanwhile usage has remained relatively stable for the broad-based aggregated ones and they now account for a much larger proportion of use. Of course, this is only part of the electronic resources usage data. We have added a number of large journal packages (**Elsevier Science Direct, Wiley, Blackwell Synergy,**

PsycArticles) since 2001 and the number of full-text articles viewed and downloaded is large and growing rapidly. During our last major serials cancellation in 2001, we canceled a number of bibliographic databases that showed relatively low use or high subject overlap with other ones.

Qualitative Assessment: Another Way to Evaluate Use

Assuming the numbers are correct we know how many login sessions or searches occurred on a particular database or group of databases. That's it. We don't know who searches, why the database was selected, how it was searched, or what the results were. Not only are these outcomes important in themselves, but they also affect the number of sessions and searches. For example, if the user cannot easily find a database on the library Website or it's difficult to search, it confirms "Moors' Law" that "An information retrieval system will tend not to be used whenever it is more painful or troublesome for the customer to have information than for him not to have it" (Moors 1960).

Qualitative evaluation works directly with the target community to discover how they find and use information resources. Some common qualitative evaluation methods are:

- Guided observations and interviews
- Focus groups
- Usability studies
- Surveys

Guided observations and interviews provide the most direct experience of exam-

ining how the user finds information. There are a number of ways to set up the process and select participants, but it basically involves sitting down individually with someone and observing how they find databases and search them. We used this method in early 2003 as part of a review of our database vendors and interfaces. We visited faculty and graduate students in their offices and

continued on page 30

Evaluating Bibliographic ... from page 28

asked them to do some typical searches. While the number of observation sessions was relatively low (15), we found similarities in how our users searched, regardless of discipline or status. Some common observations included:

- Keyword searches are the most frequently used
- Little use of Boolean commands, indexes or thesauri
- Limits or format changes rarely employed
- Commands need to be on first page or won't be used
- Visible and quick links to full-text are critical
- Many use a bibliographic reference manager to import citations

These observations indicate that our subjects did not utilize many of the features that librarians think necessary in an interface and search engine, including Boolean searching, field indexes, thesauri, and advanced limiting features. We plan to do more observation studies this academic year.

Focus groups bring together a group for discussion on selected topics. A facilitator guides the discussion, but the real value is in hearing directly from participants and the interaction that occurs between them. We choose a different topic for our annual focus groups and in 2002 we chose the "libraries impact on research." One of our questions asked how they found the scholarly information needed for their work. Common observations made were:

- Library licensed databases are too complex to use
- Not enough time or frequency of use to learn how to use them well
- Need search engines that are more sympathetic to our needs
- **Google** is simple to use, intuitive, and comes up with items you can't find in databases
- Easier to work from key articles and authors through citation databases

As one faculty member bemoaned:

I'd like to use INSPEC more. I avoid it because I have problems with the search interface. And I know there are articles that should be coming up, but I'm not finding them. And I keep finding hundreds of garbage items. The librarian keeps saying, "Well sit down with me and I'll show you how to do it." But I can't remember how to do these complicated things from one day to the next.

Graduate students also mentioned that they found searching through electronic journal packages (e.g. Science Direct, ACM) more productive than using bibliographic databases. Our 2003 focus groups dealt with information literacy and library research skills from both the undergraduate and faculty perspectives. Again, the size and complexity of the information environment makes resource discovery and data-

base use challenging, especially for students, as this comment reveals:

There are so many [databases], there are so many different ways to search the databases I think that the biggest barrier is just working through and finding how to, what keywords to use, what connectors, things of that sort, trying to figure out which one is relevant to your research.

Usability (or user protocol) studies observe how a product, system or service is used. Often used for Website resource discovery and navigability issues, usability employs techniques such as computer log analysis and recorded observation. Usability studies allow providers to see if assumptions on how the product or service will be used are borne out in practice. Many usability studies, including our own, have shown that library jargon confuses users and that finding the appropriate resources and services is challenging.

Surveys can be used to get input from users by asking a series of questions. They can range from short, context appropriate ones on Websites to broader community surveys. The **University of Washington Libraries** has conducted large-scale surveys of faculty and students on a triennial basis beginning in 1992. One of our questions asks participants to rate the importance of a number of information resource types (books, print journals, electronic journals, bibliographic databases and so on). We can then compare results within the question, between groups, and between surveys. It's interesting to note that between 1998 and 2001, user reported importance of bibliographic databases decreased while importance of electronic journals increased:

Table 2. Importance of Bibliographic Databases and Electronic Journals to Work.

Importance of Resource Types (scale of 1 to 5)	Faculty 1998 (1503)	Faculty 2001 (1345)	Grad Students 1998 (457)	Grad Students 2001 (597)
Bibliographic Databases	4.09	3.76	4.12	3.55
Electronic Journals	3.50	3.94	3.74	4.05

Conclusion

Use statistics only tell part of the story in evaluating bibliographic database use. Assuming that your use data is accurate, complete, and reliable (and that's a BIG assumption), it's important to have the complementary qualitative information that provides critical context. By using multiple assessment methods libraries can gain an understanding of user behavior and how that impacts the use numbers. We know that our faculty and students have trouble finding the appropriate databases and how to use them. Consequently, we are improving the resource discovery features of our Website, looking at simpler search interfaces, and reviewing the value of our databases.

It's worth the time to look beyond the numbers and learn directly from your customers about how and why they search, barriers to their use, and how your library can better support their work.

References

1. NISO Z39.7-2002 Draft Standard for Trial Use. Information Services and Use: Metrics and statistics for libraries and information providers—Data Dictionary.
2. Duy, Joanna and Liwen Vaughan. "Usage Data for Electronic Resources: A Comparison between Locally Collected and Vendor-Provided Statistics." *Journal of Academic Librarianship*, 29 (1), 16.
3. Moors, Calvin. "Moors' Law or Why Some Information Retrieval Systems are Used and Others Are Not." *American Documentation*, 11, (July 1960); ii.

Rumors from page 16

demic and library communities. For the researcher community, now publishers allows copyright retention and self-archiving by the author. The library community benefits from a simple site license agreement with choice of purchasing options including campus-wide access for all students and faculty, and unlimited course pack use at subscribing institutions. www.nowpublishers.com

On Monday, January 12, 2004, from 9:30-11 AM in the San Diego Marriott/ New York/ Orlando Rooms. The **ALCTS PVL R (Publisher-Vendor-Library Relations) Interest**

Group will present an open forum at the ALA Midwinter Meeting: "Here Today, Gone Tomorrow: Why Online Content Disappears."

Unexpected withdrawal of online content from electronic databases is a major concern to library subscribers. This session will begin with a brief overview of the issue, describing high-profile incidents such as *Tasini v. New York Times*, the *Economist's* court-ordered withdrawal of content judged libelous, and *Sage Publications'* decision to remove its journal content from aggregator databases. Librarians' concerns will also be outlined briefly. Once the context has been established, panelists from Elsevier, EBSCO and LexisNexis will address the issue from their vantage points as publisher

continued on page 32