

Middle Ground: Reaching Agreement on Vendor-supplied database use statistics

Denise M. Davis
National Commission on Libraries and
Information Science (NCLIS)
Washington, DC

<http://www.nclis.gov/statsurv/statsurv.html>

Introductory remarks:

In the beginning there were databases. Online databases. They were numerous, expensive and use reports were plentiful. Plentiful because revenue was directly associated with the use data. In the second age of databases there were CD-ROMs. They, too, were numerous, not quite as expensive, and use reports were non-existent. In the third age of databases, there was a tangled Web of content, interspersed with subscription databases, and available worldwide. It was the time of the Internet and beings roamed the earth dragging enormous databases behind them, filled with mixed content of text, sound and images. The databases were loved and users consumed them with ever growing enthusiasm. The costs were high, the use data diverse and of mixed quality, and library administrators became weary. Database developers and publishers are weary, now, too. There is, however, far to go.

Historical Overview:

Work within information community regarding understanding use of electronic resources began in the mid-1990's. As far back as CD-ROM, libraries have wanted to capture use information. As we moved from signups to logins, clear patterns began to evolve about what sorts of information was needed to best understand "use". The labels applied to the activities, and the definitions associated with the labels, became the ground of considerable discussion and even controversy. The language used to describe login and search activities of users and the language used to describe computer sequences in response to the login and search actions became the place of considerable discussion. Those discussions continue today.

Simply saying "we want to capture database use information" was insufficient to describe the complexity of actions that make up a login, a "session", not to mention an output such as full text views, printing or downloading. And, just as people speak different languages, the language spoken by librarians and database developers also is different. To add even more complexity to the process the varying needs of users (real or perceived needs), the varying needs of libraries, the variety of methods for capturing and reporting use data by developers and publishers all feeds into a staggering scenario. Finding the middle ground has taken years of work, years of discussions, and

lots of “expectation modification” by all parties involved. You have surely heard the truism of something being like “herding cats”, or “nailing Jell-O to a tree”. This process has been more like taking a camel through the eye of a needle. With so much data to answer the myriad of questions, where does one begin?

We began looking at database use the way we look, and measure, use of traditional library materials (e.g., print, music, video, etc.). We took the round peg of circulation and crushed it into the square hole of full text views, printing and downloading. We took logins and sessions and ascribed them to in-library use. We took online tutorial use and merged it with bibliographic instruction figures. We remained in our print world grappling with making the new use figures align with the old methodologies for capturing use data. Working with database developers and publishers to derive the necessary use measures, have those figures consistently reported, and understand these figures in the context of a morphing collection continues to be the greatest challenge.

Easily 7 years into this process and several iterations of definitions, minimum use measures and uniform customer reporting later, we find ourselves at a crossroad. The gnashing of teeth over issues remains with us. These include

- What is a hit versus a search versus a view?
- Is a full text view the same as a download or a print if the inherent purpose is to retrieve the full text of the article or book?
- Are the number of full text journals available in a database in a given year valuable information to have and does it equate with anything we really need to know?

The key difference is that we have all matured around the databases and their use – librarians, publishers and database developers alike. We have worked toward the middle ground. The middle ground is that all of these machinations are irrelevant if we cannot interpret the data and use it in meaningful ways to describe how our libraries improve the lives of our users. What do we really need to know and how do we learn that?

Lead research and data collection initiatives:

A number of research initiatives have been underway to better understand our data gathering needs around electronic resources, network performance, and staff performance. With so many efforts behind us and so much yet to do, what are the next steps to move this all along? The most promising of the initiatives include:

Library surveys integrating network performance measures and other e-Metrics include:

- The State Library Agency Survey has adopted several of the e-Metrics from the work begun by ICOLC and refined by Bertot and McClure.
<http://www.nclis.gov/libraries/lsp/StLArevised2001.html>
- The Federal State Cooperative System for Public Library Data (FSCS) has adopted a few output measures, but those focusing on sessions, searches, full text views, etc. are being field tested under the Bertot grant from IMLS. The results of this field test will guide the adoption of these measures for the federal reporting. The FSCS survey site is
<http://nces.ed.gov/surveys/libraries/surveyPUB.html> and the field test information site at <http://www.ii.fsu.edu/Projects/IMLS/index.html>
- The PLA survey, Public Library Data Service, has adopted a few of the Bertot and McClure measures and included them on the FY2000 survey.
<http://www.pla.org/plds.html>
- The Association of Research Libraries (ARL) has integrated several e-Metrics into their annual member survey. <http://www.arl.org/stats/arlstat/index.html>

There are **standards activities** in the US and internationally that will consider or adopt the e-Metrics identified in the Bertot and McClure work. This includes

- The International Standards Organization (ISO) Library Statistics standard #2789. A final revision is in development and expected to be released for a final vote by member countries before the end of 2001.
- The revision of the National Information Standards Organization (NISO) Library Statistics Standard, Z39.7, which begins its formal review process in November 2001. I am chairing that committee and inclusion of e-Metrics will be a significant portion of the group's discussion.
<http://www.nclis.gov/statsurv/niso/z39.7/z39.7.html>
- Development of a "data dictionary" of e-Metrics terms, including network performance measures, inputs and outputs, etc. The preliminary meeting of a group of experts is planned for January 2002 in conjunction with ALA. EBSCO, the Florida State University School of Information Studies, Information Institute, NISO, and the US National Commission on Libraries and Information Science (NCLIS) make support for this effort available.

Guidelines development within the publishing community is most actively guided by Richard Gedye, Oxford University Press, in his role as chair of the PALS Working Group on Vendor-based Usage Statistics and the Joint Information Systems Committee (JISC) <http://www.jisc.ac.uk/>. This group is presently focusing on vendor-based usage statistics for online journals and databases, and the development of an agreed set of principles for the measurement and reporting of usage of online resources.

And, last but certainly not least, **ICOLC**, the US-based academic consortia focusing on vendor statistics, has just released revised guidelines. The 2001 guidelines were not yet posted, however the 1998 version is available at <http://www.library.yale.edu/consortia/>