December 2003

Electronic Use Statistics: A Panacea or Just A Problem?

John McDonald

California Institute of Technology, mcdonald@caltech.edu

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

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

Recommended Citation

McDonald, John (2003) "Electronic Use Statistics: A Panacea or Just A Problem?," *Against the Grain*: Vol. 15: Iss. 6, Article 5.
DOI: <https://doi.org/10.7771/2380-176X.4196>

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

ALA MIDWINTER ISSUE

VOLUME 15, NUMBER 6

DECEMBER 2003 - JANUARY 2004

Against the Grain

"Linking Publishers, Vendors and Librarians"

ISSN: 1043-2094

Electronic Use Statistics: A Panacea or Just A Problem?

by **John McDonald** (ATG, Guest Editor) (Acquisitions Librarian, California Institute of Technology;
Phone: 626-395-6427) <mcdonald@caltech.edu> <jmcdonald@caltech.edu>

The US GDP grew 7.2% in the third quarter of 2003. **Hideki Matsui** of the **New York Yankees** hit 2.17 groundballs for every flyball he hit in 2003. Fulltext articles from the *Journal of Organic Chemistry* were accessed 7,921 times by **Caltech** researchers in 2002. Very interesting, one might say, but so what? While statistics and empirical measures permeate our lives everyday, in every way, oftentimes we do not really understand the numbers that are thrown at us. Would it help your understanding of the preceding statements if you knew that the GDP growth was the highest rate of growth in 19 years, that Matsui was 7th in the Major Leagues in highest groundball to flyball ratio, and that the *Journal of Organic Chemistry* was the 12th most frequently used online journal at **Caltech**? Maybe or maybe not, but this example illustrates what libraries are now going through every day. We have more numbers that represent our user activities, but very little context, very little consensus, and very little agreement on what these numbers might mean or how to interpret them.

Although libraries have rarely relied solely on quantitative measures to inform their day-to-day actions, the increases in serials costs and the decreases in real budget dollars are forcing libraries to make tougher and tougher collections decisions each year. Luckily, we are starting to acquire real measures of use to supplement our qualitative judgments about resource quality and collection use.

The theme of this issue is electronic use statistics and we have some great articles by a wide variety of authors on a range of topics. First up is an excellent opinion piece about the issues surrounding use statistics by **Bob Molyneux**. Bob cautions that not every party is interested in the same objectives in usage gathering and reporting and that there has been little collective understanding of the processes involved in use of digital materials. Following Bob is an article by **Steve Black** and **Amy Sisson** who have contributed an excellent study of the use of fulltext aggregated databases and how they conform to the **Bradford** distribution and the 80/20 rule of journal use.

Phil Davis has written a very thought-provoking article on how increased monitoring of our patron's usage can possibly have very negative effects for our library acquisitions budgets. **Steve Hiller** follows with an interesting article on how to complement quantitative use measures with qualitative measures to better understand what users use and how they use it. **Jennifer Weintraub** describes **Yale's** use statistics collecting program and muses on the difficulties in comparing measures across resource types and vendors. I'm sure her description of their situation will resonate well with those of us who also have use statistics collecting responsibilities for our libraries.

continued on page 18

If Rumors Were Horses

Sadness brings us all together. **Dana Alessi** died on November 26. There is an obituary in this issue (p.8) by her cousin and mother. We will include some other special remembrances from many of her friends and colleagues in the next issues of **ATG**. If you wish to contact Dana's mother, the address is: **Louise Bennett**, 3636 Angela Robin Apt. 104, Las Vegas, NV 89129. **Tom Leonhardt** <leonhardt@libr.stedwards.edu> is coordinating the remembrances if you wish to submit one.

Phil Davis (AKA **Daddy**) <pmd8@cornell.edu> returned from the **Charleston Conference** early and was there for his wife's scheduled C-section. The newborn is **Anna Davis Cohen** (no hyphen... yes, she has his wife's last name...), was born on Monday Nov 10 at 2:24 PM at 6 lbs 10 oz with lots of dark hair. To satisfy all of us baby-lovers, Phil has scanned a few pictures. He hasn't created a Web page yet (he is sleep-deprived among other duties), so we have to go to these links. Phil says it's hard not to love children, even when

they keep you up most nights. <http://people.cornell.edu/pages/pmd8/anna/1.jpg> through <http://people.cornell.edu/pages/pmd8/anna/8.jpg>

In May, 2003, the awesome **Phil Dankert** <prd2@cornell.edu> was awarded the **SUNY Chancellor's Award for Excellence in Librarianship**. As the Award stipulates, his performance in the areas of skill in librarianship, service to the university and to the profession, and scholarship and professional growth and development have been truly extraordinary. No wonder! Did you know that Phil has been at the **Martin P. Catherwood Library** of the **ILR School** for 41 years, except for the first 5 years when he was at **Olin Library** in the **Acquisitions** Department? Anyway, Phil is retiring June 30, 2004. He says that the **Charleston Conferences** that he has at-

continued on page 6

What To Look For In This Issue:

Dana Alessi Obituary	8
To Thine Own Self Be Dispassionate ...	46
Adventures in Librarianship	60
Books That Matter	66
Books Are Us	78
New Programs and Accreditations	79
Lost in Austin	82
Talk of the Trade	83
Law Librarians Take a Positive Step	86
Profession Confusion	94
Profiles Encouraged	
Antje Mays	68
Interviews	
Robert Wedgeworth	50
Daniel Halloran	56
Bob Schatz	58

1043-2094 (200312/200401)15:6:1-4

Making Haste Slowly: E-metrics — Where Are We Now and What Can We Expect?

by **Bob Molyneux** (Director, Statistics and Surveys, US National Commission on Libraries and Information Science; Phone: 202-606-9181) <bmolyneux@NCLIS.gov>

Using digital materials as a method to disseminate information is an exciting development for information professionals and academic researchers. The transformation brought about by the digitization of the human record is proving revolutionary since it makes the transmission of information so much easier than earlier formats.

Part of this revolution is trying to get a handle on what is going on for purposes of planning, budgeting, and collection development. In fact, we have many people chasing fragmentary data, looking for insights into what is going on while often faced with budget decisions that have to be made by next week.

We largely have anecdotal evidence about the general situation but there is a budding community of people focused on systematically measuring the use of materials in digital formats. Anyone who wants answers to questions right now is going to be frustrated because measuring the impact of these materials has proven to be a difficult problem. Better data is coming but even though the data will be better, it will never be perfect. There are sure to be un-

available or unreliable data for some time to come that will not fully alleviate these areas of frustration.

COUNTER

I have argued in *ATG* before that **COUNTER** is the best shot we have for getting better, comparable, generally available, and believable data than any other initiative. The first **Code of Practice** is sound, systematic, and reflects serious attention to the problem. It is the *de facto* standard and once the XML and auditing infrastructures are in place, then we will begin to have basic information of use to both the practitioner and the researcher. We will be seeing **COUNTER** compliant data soon but XML formats, as I understand it, are not obligatory before 2005. Expect to see some of these data before then, and see the **COUNTER** Web site for some example reports <http://www.projectcounter.org/>.¹

XML formatted samples also exist. **COUNTER** is an organization where a process for development and improvement of electronic usage data will continue.

COUNTER is now soliciting memberships to fund its work and its work on data exchanges standards deserves support from the library community.² **COUNTER's** existence, though, will not end the quest for data for a set of reasons that are endemic to the problem at hand.

The first and easiest reason to understand is that there is a hard data problem since just about everything that makes data collection difficult exists: the processes being measured are complex, changing, and new. Data collection is easiest when you are collecting data on simple, stable, and old processes. So, even with the best intentions, there are technical difficulties that will require work to solve. Some, I fear, may prove intractable.

Cooperation and Contention

The second reason that **COUNTER** will not end the quest for data has to do with the different objectives the various parties bring to their quest for data. Data are not collected in a vacuum but to meet the objectives of those collecting them. These differing objectives create centrifugal forces that will bid to pull the **COUNTER** effort apart.

Librarians will not want the same things as publishers, so there will be areas where there will be agreement and other areas where there will be profound disagreement and we will all just have to get along if it is to succeed. For instance, librarians look at these collections of serials titles that they are paying for and look

for means to cut some of them or to rationalize the selection process so they can plan for the future and can tailor their collections to their users. From the librarian's perspective why have access to a given title from five different places? Why not get access to that title from the cheapest place? And with comparable use statistics from **COUNTER**, some titles could well be canceled.

Publishers and aggregators, of course, do not like the idea of cancellation so they have an incentive to prevent publication of some types of data or obscure others. We all may be better off in the long run if libraries can cancel some titles but each of us won't necessarily be better off tomorrow. Hence, it is arguably rational for those firms supplying data on use of electronic materials to supply non-comparable data because it makes comparisons that might lead to cancellations difficult. Similarly, it also seems rational to restrict sharing of contract terms and sharing of data between institutions: why inform the library that paid the most for a title that everyone else paid less? Why tell everyone which library got the best contract and have that become the floor for all future contract negotiations?

Neither community is monolithic, however. The library community has parties within it with different objectives that could also disagree with each other about **COUNTER**. I have had librarians say to me that they do not want data published because it would make them "look bad" so some may be ambivalent about various data items that are discussed in **COUNTER**.

Of course, those in the publishing and aggregator sector also have crosswinds blowing in their communities. Those who have read the **Morgan Stanley** report on **Reed Elsevier** know that clear-eyed appraisal of the situation libraries are in and how it can be exploited. It is tempting to think of this report as the blueprint for all publishers and aggregators. However, some publishers or aggregators are stronger than others and some are under more pressure than others to increase revenues, gain market share, or whatever. Some may take a long view while others a shorter one. These differing situations will lead to different attitudes about what should be reported.

Some publishers and aggregators must also be concerned about overreaching because if they go too far in increasing costs, the politicians get involved and then all bets—and business plans—are off. These facts in their environment will mean that they will disagree from time to time with each other both in which data to supply and in which data they can afford the programming to supply.

Now, none of this is either bad or good as long as everyone understands the different objectives. Collecting these data are difficult by

continued on page 20

Electronic Use Statistics: from page 1

Two publishers are represented among our authors. **Marthyn Borghuis** of **Elsevier ScienceDirect** describes the factors that influence usage of online journals and explains what publishers do with the usage data they collect. **John Sack**, of **Highwire Press**, explains the usage studies that Highwire has completed and the findings that their extensive analysis show about online journal use.

Both standards-based projects for collection and reporting of usage statistics are represented in the issue. **Denise Davis** brings us up to speed on the **NISO Committee AY** and the **Z39.7-2002** standard in her article, while **Peter Shepherd** describes the progress that **Project COUNTER** has made to date and its plans for the future.

I hope that all our readers enjoy this issue of *ATG* and find each article as thoroughly interesting and thought provoking as I did! And remember, we can look at statistics two ways: as **Paul Fisher** believes, "A passion for statistics is the earmark of a literate people," or as former major league pitcher **Jim Bouton** believes, "Statistics are about as interesting as first base coaches." 🐘