2003

Getting the Best Out of It! Usage Analysis from the Publishing Perspective

Marthyn G.M. Borghuis
ScienceDirect, M.Borghuis@elsevier.com

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation
DOI: http://dx.doi.org/10.7771/2380-176X.4202

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.
send them every quarter and do not. When I do receive them, the electronic journal statistics contain one number: downloads per journal per month. While this is the essential number I want to know, I conjecture how many tables of contents were viewed, whether the package as a whole was searched, much less any linking information. But judging from my cursory survey above, the situation is improving, though still time consuming. I also find most vendors to be receptive to complaints or questions about their statistics and have answered at least five surveys in the last calendar year about electronic resources that had some component devoted to usage statistics.

The inability of even some expensive vendors to provide us with the minimum level of acceptable statistics leads to holes in our analysis of how people are using electronic resources. Yale has been a member of the ARL E-Metrics project for the last few years, and as part of this we are asked to test various statistical measures of electronic usage and collections in libraries. The most troublesome question is “Total Number of Searches on Electronic Databases.” While it is possible (though incredibly time consuming) for me to give a number of “total searches we are told about,” it is unclear what this can really mean to us. It is difficult to extrapolate from our available statistics to what our actual number of database searches were. We also can’t tell how different groups are using our electronic resources, or how people are moving from one resource to another, or whether they use our linking tools or are unaware of them. This is important as we assist users in person and implement new services to exploit our full electronic potential.

My future goals for our usage statistics are to increase their relevance to librarians at Yale. First, I continue to strive to get information that can really help them make decisions about what kinds of resources are the best for our users. I also want to be able to demonstrate trends across subject areas, not just vendors. Setting up a “data farm” or database of the statistics is one way we can accomplish this. Finally, we support efforts to get better statistics from vendors and contribute to new research using this data, at Yale, within our consortia, and at the national level.

Our collection of usage statistics, while perhaps not as comprehensive as we’d ideally like it to be, is a step in the direction of being able to tell what our users are doing and what they want and need, even if they can’t verbalize it to us. Seating in our library, it can be hard to tell if the investment made in electronic books was worth the money. Once we created an eBooks database and put it prominently on our Web page, usage of electronic books increased immediately. Certainly it is hard to tell how satisfied people are with those books aside from the fact that usage keeps growing every month. But we do know that people are using them, and perhaps these statistics and our experiences in collecting them will lead to getting a more complete picture of our electronic resource use in the future.

Getting the Best Out of It! Usage Analysis from the Publishing Perspective
by Martyn G.M. Borghuis (ScienceDirect, Senior Manager, Usage Research, Amsterdam, The Netherlands) <M.Borghuis@elsevier.com>

What Factors Drive Usage?
The conversion from print to electronic publishing is proceeding rapidly as is the retrospective digitization of the archive of print articles. Availability of full text articles in electronic format does have a substantial impact on the volume of usage data generated, this is obvious. It seems, however, that for publishers the three crucial factors influencing usage levels are:

1. The number of universities, corporations, hospitals, and research organizations worldwide that have access to the online journal collection,
2. The number of users, within these institutions that have access to the collection, and finally,
3. The increase in the number of readings (and thus usage) by these users (usage intensity).

The Global Trend
Knowing these main usage factors, what then is the global usage trend across all publishers offering content online? Where is the sky and does it have limits? It is an opportunity to present to the readers of Against the Grain, a straightforward analysis of online journal usage developments worldwide to date and in the years to come. For that exercise the usage data of ScienceDirect have been used covering the period Jan 2001 to Sep 2003.

Analysing Elsevier’s ScienceDirect data, the following assumptions can be made: Extrapolating from ScienceDirect usage and the size of its journal collection relative to the total scientific journal literature available online, it is likely that across all primary publishers at least 90 million requests for full text articles were made on a global scale in September 2003. In addition it is essential to know, that full text usage is increasing by an average 15% in peak months (September to November & February through April) and about 8% in all other months. As a consequence, every year so far, usage has more than doubled and there is no sign whatsoever that this will decline in the next two to three years! It is thus expected that total full text article usage achieved in 12 months, across all libraries and publishers worldwide, will pass the 1 billion landmark in early 2004!

Looking at this trend it is very clear, that the transfer from library print collections to electronic full text collections has resulted in a tremendous increase in accessibility and thus usage of full text scholarly and scientific articles. In itself this is a true success story!
The Beginning of Value Assessment: Usage Information in the E-Journal Age

by John Sack (Director, HighWire Press, Stanford University; Phone: 650-723-0192) <sack@stanford.edu>

Right from the start of HighWire Press with the publication of the Journal of Biological Chemistry online in 1995, there was a strong interest in learning what we could from usage statistics, or “server logs.” We knew we should measure just about anything that moved or clicked, because the early days of e-journals were like the early days of space exploration: we didn’t know what observations would be useful because we didn’t know what phenomena in user behavior would occur. With over 350 online journals, from over 150 different societies and publishers, about all the HighWire-hosted publishers could agree on was that we needed to measure “hits.”

At first, each journal editor wanted to watch the hits climb - and they did climb: a number of journals in the 1990’s saw usage increase by factors anywhere from 2 to 10 from one year to the next as some whole user communities rapidly shifted from print to online use. But soon after, editors realized that they had data on usage patterns that they couldn't really obtain from counting print copies mailed, or even by reader surveys: what are readers reading?

This began our understanding that electronic journal usage information could be “mined” for more than just “hits”: it could tell those who have intellectual responsibility for collecting articles into journals - editors - and those with responsibility for collecting journals to support academic programs - librarians — what role a journal had in the intellectual life of a discipline or institutional community.

We also realized that the initial introduction of a new technology such as e-journals was the best and perhaps only time to instrument the shift in literature-research practices.