December 2003

The Beginning of Value Assessment: Usage Information in the E-Journal Age

John Sack

Stanford University, sack@stanford.edu

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

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

Recommended Citation

Sack, John (2003) "The Beginning of Value Assessment: Usage Information in the E-Journal Age," *Against the Grain*: Vol. 15: Iss. 6, Article 12.

DOI: <https://doi.org/10.7771/2380-176X.4203>

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.

Getting the Best Out of It! from page 34

usage market share which is 0.3×1.0 billion = 300M. The difference between the two totals can be explained by the assumption that not all articles retrieved electronically are read, due to crawler or robot activity among other reasons.

The authors further report that searching online bibliographic databases on one hand, and using enlarged electronic library collections on the other, are the two major factors that caused the increase in readings per academic reader and thus usage per user. In addition, they report a significant increase in readings from library collections and thus a diminishing share of readings from personal and member subscriptions. They also confirm more readings of electronic titles not previously subscribed to in print.

Collection Management: Libraries and Publishers Alike

Once publishers started to publish scholarly and scientific journals online and to collect and analyze usage of these journals, dramatic changes occurred in the way librarians and publishers communicate. The contact between the two parties is now direct, which immediately generates a need to better understand each other's business processes. The reporting of online usage of full text articles is a new phenomenon. At the time when only print collections existed, it was a cumbersome and often difficult exercise for either librarians or primary publishers to quantitatively establish the usage of journal collections. For librarians, available lending data (including interlibrary loan balances) per journal title was used to measure the attractiveness of a journal collection. For unbound issues, techniques like reshelving counts were applied; these techniques proved to be both time-consuming and often incomplete.

Publishers faced the same difficulties: measuring subscription figures, subscription income, or the number of articles or printed pages published, proved to be an inaccurate measure

of the attractiveness of published journals. Now, for the first time in history, both librarians and publishers are able to share the same usage information and, what is more, both parties should now be better equipped to have a better-informed opinion on the relevance and attractiveness of the e-journal collections they subscribe to or produce.

What Do Publishers Do with Usage Data?

First of all, I would advise publishers to make their own usage reports **COUNTER** compliant (for details on **COUNTER** see the contribution of Peter Shepherd in this issue of *ATG*) and not only the reports for their customers, but also those for in-house evaluation of their journal collections. The main drivers in primary publishing, however, are the editorial boards: they define the aims and scope of a journal and organize the refereeing process and peer review, thus guaranteeing the uniqueness of published research results and preventing duplication. It is a basic requirement to maintain a certain quality standard for articles to be published, which will enable librarians to earmark these publications.

A publisher's journal management staff, however, may add a new criterion to the ones they already use to evaluate the composition of the journal portfolio for which they are responsible. Circulation figures are supplemented by data about the number of (repeat) users and usage intensity of a journal. Usage analysis will also help in sustaining new and quickly developing areas of research and at a far earlier stage than other measures, such as citation data, can tell.

Comparing usage and citations from another perspective could reveal that there are substantial subject-specific differences in the volume of usage and citations. Especially in areas like health, clinical medicine, engineering, social sciences or psychology, usage outperforms the relative level of citations quite substantially. These areas do attract a lot of users, who seem to be active in more fundamental science areas;

they do want to combine their research with the discoveries in these other subject fields. Many users do also read articles in these areas out of personal interest, while working in other faculties of science. In this sense both groups are the "virtual" representation of so-called "walk-in" users as known in the print based libraries. There are also areas where reading has always been predominant and where writing and publishing articles is less common. Here we can localize the users active in the various engineering departments.

Finally, more detailed studies point out that for publishers like Elsevier, with a broad collection of journals, review book-series and major reference works in *electronic format*, it is obvious that not all science areas benefit from the strong increase in usage to the same degree. High usage areas are definitely found in the life sciences, chemistry, and environmental sciences.

To end with a controversial statement: where institutional print subscriptions see a rapid conversion to the electronic-only versions open to and accessible by the community of readers, it might occur that print based, personal and member subscriptions will only gradually disappear as they incorporate and bundle quite a few excellent reader needs at once: alerting (in the post-box or on the doormat), print-at-hand, and no need to log-in.

Endnotes

1. Tenopir, Carol and King, Donald W. *Towards Electronic Journals, Realities for Scientists, Librarians and Publishers*, Special Libraries Association, 2002, p.18-19.
2. See Tenopir and King (2002) p.19.
3. King, et. al., found a continuing increase in the number of readings per reader especially since the electronic availability of articles. See: King, Donald W., Tenopir, Carol, Hansen Montgomery, Carol Aerni, Sarah E. *Patterns of Journal Use by Faculty at Three Diverse Universities, D-Lib Magazine 2003 (Oct)* <http://www.dlib.org/dlib/october03/king/10king.html>.

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 new 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 use 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 much more than 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 pro-

continued on page 38

The Beginning of Value Assessment: from page 36

cesses. **Stanford University Libraries** — home to **HighWire Press** — received a grant from the **Mellon Foundation** to conduct a three-year e-Journal User Study (e-JUST) to document the shift in researchers' attitudes and work patterns with respect to their encounters with the literature. The study, completed at the end of 2002, is available at <http://ejust.stanford.edu>. The study did longitudinal surveys of tens of thousands of researchers in over a dozen societies, as well as studying Web server logs, and has findings and recommendations for each community that supports the research process through literature: societies, publishers, librarians, and system suppliers. One important insight to come out of the study is a model for how researchers interact with the literature during the entire research process. The best service providers will be the ones who provide support to all the stages of this circle of innovation:

Figure 1: Six Core Information Practices

See: Research Findings. <http://ejust.stanford.edu/SR-786.ejustfinal.html>

Because **HighWire** is located in the **Stanford University Libraries**, with a management team that has been working with libraries since the development of **BALLOTS** and **RLG**, we were naturally interested in how usage information might assist libraries in making selection and de-selection decisions — perhaps re-shaping the economics of journal publishing — as well as how the information might help librarians provide better public service. It seemed to us that the usage information contained in e-journal Web-server logs could provide a new source of information to librarians serving a more and more “remote” online community. That community could increasingly take advantage of library services without stepping into the physical library, and whose information needs could less and less be understood through personal interaction, much less through circulation and re-shelving statistics. This was especially true in the science-research community where **HighWire** got its start.

We began our analysis by asking librarians what type of information would be useful and/or interesting — the type of information that might provide its own insights! This was all being discussed before online publishers were regularly delivering usage information, before **ICOLC** developed general standards, and years before **COUNTER** began to codify uniform practice. At the same time, we asked journal editors and publishers about their interests in measurement. We thought that public service librarians and journal editors might have similar interests, and that selectors and publishers might have similar interests because each of these pairs serves a somewhat similar function but at different “ends of the pipeline.” And in all cases we felt that the principle of transparency should apply: information needed by either community should in principle be accessible to the other.

Were did this lead us? Of course, we saw the need for counts of various types of actions such as searches and article displays, just as all

cating articles by known-item searching by author/title, or by subject searching? We believe, from our usage statistics, that they are predominantly searching, by about 2:1.

4) How broadly distributed is usage across a patron population, and across a database of articles? Are there many “light” users, or just a few heavy users of a journal? Are many articles being read across many issues, or is reading primarily in the current issue, or in just a few articles? The “unique events” counts in the **HighWire** usage statistics tell the story here.

5) How does academic-period usage compare across time? We found there was interest in comparing academic-term (or a range of months’) usage in one year vs. another year, rather than comparing two subscription-years’ usage. **HighWire’s** “custom multi-month reporting” enables this function.

Of course, there were some “interesting” ideas that we couldn’t support, such as reporting on IP ranges within a subscription. But maybe someday!

We heard from readers that they were interested in knowing what others were reading, specifically in the categories of their studies. This gave rise to the “Most-frequently Read” reports, which some of the **HighWire**-hosted publishers post on their Websites for readers to peruse, for example see Gut Online’s most frequently read articles here: <http://gut.bmjournals.com/reports/mfr1.dtl>.

And, of course, we heard from authors that they wanted to know how often their papers were read and cited. This is a feature of the *New England Journal of Medicine* Website. We even heard from a tenure-review committee asking for data about how often the papers of a particular researcher were read!

More recently, we began work on a tool to help libraries study their patrons’ collective use of journals to which the library does not subscribe, and, reciprocally, a tool for publishers to identify non-subscribing institutions where there is high use of a journal. When we reviewed the idea for such a report with librarians, there was good agreement that this kind of information would be useful as long as the same information was available to both publisher and librarian. These new reports are just about ready for release to publishers and librarians. One might perhaps wonder where such use would be coming from if there were a lot of it, but the library didn’t hold a subscription. Because there is so much free full-text content available from **HighWire**-hosted publishers — there are almost 620,000 free full-text articles available from 165 journals participating in the “Free Back Issues” program, making the **HighWire** archive the largest free full-text science article repository that we know of — there is a good opportunity for librarians and publishers to judge demand for a journal’s content without ever having to implement and promote trial subscription programs.

Since the **HighWire** usage reports provide

continued on page 40

The Beginning of Value Assessment: from page 38

quite a bit more data than **COUNTER** compliance required, and because **COUNTER** specifies both the format as well as the content of the usage information, **HighWire** had a dilemma in becoming **COUNTER** compliant: either eliminate the "useful and/or interesting" information that fell outside the **COUNTER** specification, or not be **COUNTER** compliant. We decided the best course was to continue to have a venue in which to provide innovative usage information, but also to offer strictly-compliant usage reports; thus starting in November there will be two parallel usage reports for each month in the **HighWire** system, one **COUNTER**-compliant, and one that offers extra-**COUNTER** information.

We feel it is too early in the life of e-journals to limit everyone to a standard, yet a standard is obviously necessary to enable librarians to make apples-to-apples comparisons across titles. We need to support innovation and interpretation. As a simple but common example: most **HighWire**-hosted journals now make available "data supplements" - information that journal editors decide can't or won't appear in print, but will be online only; how much use do these data supplements get? Many **HighWire**-hosted journals post articles online ahead of their appearance in print; publishers definitely want to know how much use these "ahead of print" versions receive, and we wonder if librarians have any interest in this information as well. A less common example: the *BMJ* produces "abridged" versions of its articles in print, with the full-text online as well as the abridgement; these abridgements are very popular with busy clinicians. Would librarians want to differentiate these versions of articles from the unabridged versions?

What kind of innovation may be ahead in usage reporting? There are many possibilities, and I've learned from experience not to gaze into the crystal ball in anything that will be printed or posted online. But I'll speculate based on what we've heard from librarians, authors, readers and publishers.

As publishers deliver content through more avenues — a journal might be available via **HighWire**, via **Ovid**, in **PubMed Central**, and in abstract form in **PubMed** and **Medline** — it becomes harder and harder for a librarians to

understand how valuable the content is in general to an academic community, because the use is dispersed through multiple "channels." There is no place or report that pulls all the use together of the intellectual content in a title.

As librarians try to decide whether an "aged" version of a journal in **Ovid** (a version whose newest issue is three issues back from current), or an incomplete version of a journal at **Ingenta** (a version that does not include non-print material), are sufficient to support a research and instruction program relative to the complete "journal of record" available at a publisher's own site, we may see that usage information that is insensitive to the age of content being accessed will not suffice.

Librarians may need better or different statistics to calculate the increasingly popular "cost per use" metric. Such a metric gives librarians — and publishers — a tool for validating that value is commensurate with cost. A good set of metrics in this regard should help with librarian/publisher negotiations, especially in complex multi-institution cases.

Ultimately, science and publishing will need an additional metric going beyond the "impact factor" as a way of judging the importance of a journal. While no one seems to have come up with a replacement for impact factor - it is tempting to say that like democracy it is the worst possible approach except for all the others - with the availability of much more fine-grained and pervasive metrics on usage it should be possible to build alternatives that might, for example, more accurately indicate the value of clinical journals whose articles are heavily used by practitioners, but not widely cited by researchers. Impact factor emphasizes one aspect of value over other aspects, and some of the others may now be visible and useful in usage reporting. For example, in the Web world where the impact of a site is measured by its raw traffic, a more useful statistic could be the total-citations to a journal (i.e., the impact factor without the division by the number of articles in a journal), because that is the best predictor of the total amount of visitors a journal site will get; measured this way, the *Journal of Biological Chemistry* is the "highest impact" STM journal, because it is the most-frequently cited. And, it might be possible for libraries to calculate the impact factor of a journal relative to their own academic programs.

Librarians may want to find a way to use usage information to negotiate prices for products when portions of a product — such

as a database containing journal articles — contain content that has already been purchased through another channel. This would require fine-grained analysis of usage of portions of a database.

Librarians may want to assess the amount of use that publisher archives of 20, 50 or 100-year old articles have, so they know whether a continuing subscription to such an online archive is worthwhile.

Online journals provide services in addition to content. **HighWire**-hosted journals provide not only the usual email tables of contents, but also personalized (keyword and subject-specific) alerts when new content is published. Part of the value of a journal is in these services. Online journals are beginning to provide e-journal content in PDAs and via RSS (Really Simple Syndication) protocols to campus communities. Should the use of these services by a campus community be tracked?

Authors may want to know who cites their work — this information is available at many **HighWire**-hosted journals through a partnership with ISI on its **Web of Knowledge** product line — and even what institutions most often read their work and incorporate it into course packs. Certainly as publishers increasingly allow authors to retain copyright, we will see the authors' interests in the usage-end of publishing increase.

Usage statistics are becoming an important way for a library — and for a publisher — to assess whether value matches cost. Perhaps as these assessments evolve we'll see the statistics evolve as well, beyond "usage" in the traditional sense. For example, under the new American Physical Society "blended pricing" model, an institution's frequency-of-authorship is taken into account in establishing the subscription price for a journal. Librarians may even want to identify authorship to suggest alternative publications to faculty who publish articles in expensive titles. As additional delivery platforms become available to authors — such as institutional self-archiving programs starting at a number of institutions — these calculations of a journal's value will become more complicated.

Clearly, significant challenges and "interesting" times are ahead for us! Usage, content and value continue to be a dynamic equation depending on whose usage, whose content and whose value is at issue.

Rumors from page 32

Editor of the **American Academy of Pediatrics**, who led the **Shop for Journals** initiative. "The concept of the **Open Package** grew out of our conversations with serials librarians. They told us that they liked the idea that the **Big Deal** offers: holding down costs by acquiring a large number of titles at reduced prices. However, they also told us that the problem with **Big Deal** packages is that they may include a lot of titles that an institution does not need or want." The **Shop for Journals Open Package** is a way for institutions to pick and choose which titles they want, obtain the correct pricing for their institution,

and to purchase them through one simple mechanism. Along with the customized price chart, there are options for a librarian to send a custom package of titles and prices to their regular subscription agent, to contact the publishers directly to purchase, or to send a subscription request through the **HighWire** system to serials agent, **Otto Harrassowitz KG**. The advantage of this third option is that most new subscriptions will be available online immediately after confirmation of the order. "In the future, our intent is for the participating publishers to introduce a discounting system for new subscriptions," says **Mr. Clarke**, the leader of the task force of 50 **HighWire**-affiliated publishers. "We think this will present librarians with

a better alternative to the **Big Deal**, one that provides them with similar savings, but does not lock them into titles they do not need." "We are excited to participate in this collaborative effort, as we recognize that shrinking library budgets are increasingly going to the 'easiest' buy, rather than the highest quality titles," says **Meg McGough**, Marketing Manager of the **Histochemical Society**. "With the wealth of research by the scholarly community that is hosted at **HighWire** today, we strongly support this group effort to simplify the purchasing process for each of our independent society journals." <http://intl.highwire.org>

Wow! That's it for now. Happy New Year!
Yr. Ed.