

The implications of usage statistics as an economic factor in scholarly communications

By Heather Morrison

Abstract

Usage statistics for electronic resources are needed, and highly desirable, for many reasons. It is encouraging to see the beginnings of quality, reliable usage data. This data can form the basis of economic decisions (selection and cancellation) that make a great deal of sense in the context of the individual library. However, the cumulative effects of such decisions could have serious implications for scholarly communications. For example, the journals of small research communities could easily be vulnerable to mass cancellations, and might fold. Fortunately, open access provides an alternative. The question of whether the impact of local decisions on scholarly communications as a whole should be taken into account in collection development policies is raised. The possibility that usage statistics could form the basis for a usage-based pricing system is discussed, and found to be highly inadvisable, as usage-based pricing tends to discourage usage.

The Need For, and Development of, Usage Statistics

There are, and always will be, real needs for usage statistics for information in the electronic format. Authors, publishers, and funding authorities, as well as libraries themselves, need to know whether or not costly resources are being used. A lack of usage can alert a library to a resource that has not been set up properly. Low usage rates may indicate a resource that needs better promotion. High usage is an indication of the value of a resource to the library's users. The timing of usage statistics tells us the hours that library users are active, which can inform other decisions such as scheduling of virtual reference hours.

Usage statistics can also provide information to help us to better understand users and their information-seeking behavior and needs. Through usage statistics, Ohiolink discovered that about 15% to 35% of the titles received through the "big deals", which no Ohiolink library had previously subscribed to, were actively used. (Gatten and Sanville 2004). Statistics have helped the open access Stanford Encyclopedia of Philosophy (SEP) to make a case for funding for ongoing open access, by providing evidence of heavy usage, and also by providing details about who is using the resource. In other words, statistics indicate that the SEP is used not only by the professional philosophy researchers for whom it was designed, but also by many departments across many campuses; thus, the case was made that the SEP was a good resource, and that libraries, as well as

philosophy departments should financially support the project. (Zalta et al 2005).

Important strides have been made toward the development and standardization of usage statistics for electronic resources. This is largely due to the efforts of the International Coalition of Library Consortia (ICOLC) and the COUNTER (Counting Online Usage of Networked Electronic Resources) project (<http://www.projectcounter.org/>). ICOLC developed a set of guidelines, "Guidelines for Statistical Measures of Usage of Web-Based Information Resources" (ICOLC 2001). COUNTER further developed the ICOLC guidelines, and provided a means of auditing compliance with the standards.

Thanks to these efforts, librarians are now beginning to see usage statistics, based on these standards, which are comparable across resources and platforms. A great deal more work needs to be done, however. There are still many vendors and publishers who still need either to adopt usage statistics reporting mechanisms, or that need to bring the quality of their statistics up to the COUNTER standard. Nevertheless, there are currently enough quality usage statistics available, that this is now a factor in making financial decisions such as the cancellation, and retention of journals.

The remainder of this article will explore the potential impact of usage statistics as an economic factor in scholarly communications. Selection and cancellation decisions based upon usage statistics may have one set of implications when viewed from the perspective of the individual library or library consortium. These same decisions may have a totally different set of implications when viewed from the perspective of scholarly communications as a whole. The availability of quality usage statistics raises the possibility of developing pricing models based on usage. Factoring in usage has some advantages in developing pricing models in the short term, as a transitional measure. In the longer term, usage-based pricing is not optimal as an economic basis for scholarly communications, most notably because usage-based pricing provides an economic disincentive to use. For example, usage-based pricing creates a situation where a cash-strapped university could save money by canceling research-based assignments and/or information literacy programs at early undergraduate levels.

Usage Statistics and the Individual Library or Library Consortium

Usage statistics makes it possible to calculate the cost per use on a database or title by title basis. This kind of cost analysis has been very useful for Drexel University in their shift from print to electronic-only journals, for example. Carol Montgomery, Dean of Libraries Emeritus and Research Professor, College of Information Science and Technology at Drexel, has done a comparison of the cost-per-use for e-journals (for Drexel, these averaged from one dollar to six dollars per use) with the cost-per-use of print (e-journals averaged two dollars per use; unbound print issues were six dollars per

use, and bound print volumes were thirty dollars per use. (Montgomery 2004). This cost-per-use analysis facilitated the move from print to electronic-only for Drexel University.

Hahn and Faulkner (2002) have shown how usage-based metrics can be used by the individual library not only to make informed decisions about cancellations, and to help faculty understand cancellation decisions, but also to develop benchmarks to help determine a rationale for future purchases. Gatten and Sanville (2004) analyzed usage statistics at the individual library and consortial level, and found that aggregated usage statistics are a reasonable method for a consortium to retreat from the “big deal”, if financial, or other factors, made that necessary.

This approach to selection and cancellation makes a great deal of sense at the individual library or consortium level. If libraries cannot afford to purchase everything, it makes a great amount of sense to prioritize the titles that library patrons are actually using.

Usage Statistics and Scholarly Communications as a Whole

What happens to scholarly communications as a whole, if library decisions based on usage become standard, and journals continue to rely on subscription income? Consider the implications of such decisions in relation to open access, conservatism in science (that is, the tendency to favor a predominant viewpoint and filter out new evidence that does not fit), important but less popular or less-adequately funded academic areas, small research communities, and titles in different languages.

Over ninety percent of publishers allow authors to self-archive a copy of their own work (SHERPA 2005). Authors’ tendencies to self-archive vary by discipline. Physics, for example, has a strong tradition of self-archiving, starting with preprints, in the arXiv e-print archive (<http://arxiv.org/>). In some subdisciplines, such as high-energy physics, the rate of author self-archiving approaches 100%. Researchers will likely have read many of the articles as preprints before their formal publication. It then makes sense that this would have an impact on usage statistics for purchased resources. In physics, this has not made any difference to subscriptions. However, is it possible that in other disciplines, a failure to take into account the usage rates of articles that are openly accessible could result in many libraries canceling journals, even though the articles are very much used?

Thomas Kuhn, in “The Structure of Scientific Revolutions,” (Kuhn 1962/1970) described one form of scientific advance as a process of revolution from one paradigm, or set of beliefs, to the next. Concepts that fit within a prevailing paradigm are readily accepted, while concepts that do not fit, are considered to be anomalies, and thus, discounted. Picture then, the usage statistics of a journal focusing on topics that fit within a prevailing paradigm, as compared with the usage statistics of a new journal startup reflecting the concepts of what, all else being equal, could become the next paradigm at some point in

the future. It seems plausible that the usage statistics would be higher for journals that fit within the accepted paradigm, and lower for journals outside of that paradigm. If usage statistics are used as the basis for selection and retention decisions, could one type of result be a reinforcement of an inherent bias toward conservative concepts in science?

At any given time, some areas of scholarly endeavor are likely to be more popular and/or better-funded than others, regardless of their underlying merit. A current example is the present emphasis on science, technology, and medicine, with relatively lesser emphasis and funding for the humanities. This does not necessarily reflect any inherent lesser significance of the humanities, but rather reflects current societal values favoring more financial support of technologically-based fields.

Within any given discipline, some areas are likely to be more popular, or better funded than others. Consider, for example, the implications of making decisions about canceling medical journals at the time when a particular health crisis is occurring. For example, did outbreaks of Severe Acute Respiratory Syndrome (SARS) have an impact on the usage of biomedical journals dealing with the basic science of virology? If the world economy took a downturn during a similar crisis, and many libraries as a result were canceling journals, would these journals be protected, perhaps at the expense of other basic biomedical journals of equal importance in the longer term? What if an economic downturn occurred at a time when the world's attention was on non-viral medical factors – could key journals in virology be cancelled, perhaps hampering the progress of research just before the next viral crisis? These are cautionary questions that library managers must keep in mind when using usage statistics, without considering other factors, in their collection development decisions.

For a variety of reasons, the global research community in any given discipline may be small, or may be large, depending on various factors. Heart disease, for example, is a major killer throughout the world. The research community investigating its causes, prevention and treatment, is huge. Core heart journals are likely to be well used, wherever they are available. Many illnesses are less common, or occur primarily in isolated geographic areas. The research communities in these areas are likely to be much smaller, and journals devoted to such less common problems are likely to be more vulnerable to cancellations based on usage statistics. If journals in such an area are relying on subscription income, and many libraries cancel their subscriptions due to similar low usage patterns, the journals may fold. Opportunities to publish in these areas could decrease, which could lead to fewer researchers pursuing research in these areas. The end result could be a decrease in diversity of research, which has a concrete impact on real people, in many fields, such as medicine.

Usage-based selection decisions have a particular significance in relation to titles in languages other than English. A conversation with the author's professors at the

University of Alberta in the 1970's may be instructive. It was the time of the Cold War, and a serials cancellations process was underway. The debate at the time was about the cancellation of journal titles produced behind the Iron Curtain. From the usage point of view, this made a great deal of sense. These journals were written in languages (e.g. Russian and Ukrainian), that few people at the University of Alberta could read. Most of those who could read these languages were foreign language specialists, and likely not interested in, or sufficiently familiar with the scholarly disciplines that these journals covered, to find any particular journal title useful. Using articles in these journals required an expensive translation process that was rarely undertaken. This scenario was no doubt repeated at research libraries throughout North America. When the library focuses on the needs of the University of Alberta and its library clients, canceling journals that receive little or no usage makes eminent sense. What happens, though, when a great many libraries, all facing the same financial pressures, and coping with the same serials crisis, make basically the same cancellation decisions? What happens to journals, publishers, and authors when many libraries choose to target their particular journals for cancellation? What happens to the desire for cross-cultural communication, keenly felt during the Cold War, when one of the most hopeful avenues, scholar to scholar communication, decreases or disappears?

This is not merely an historical problem. Given the difficulties libraries are facing purchasing even the most necessary scholarly information for our clients, how are librarians, as a profession, doing today with collecting journals in other languages, from other countries and cultures? In the future, with China expected to become an economic superpower, will important research journals be published in Chinese only? If so, will libraries with few Chinese readers simply not purchase these journals, anticipating little use?

There is more than one approach to this question of language and journals. Libraries could decline to collect titles in languages that the majority of its patrons do not understand. Given sufficient financial resources, libraries could provide translation services. Another option would be for our educational institutions at various levels to choose to prioritize and strengthen the teaching of different languages, perhaps as a requirement, along with other academic disciplines. This latter option not only offers the potential for enriching our understanding of the world, but it also provides us with a better foundation for competing in a future world where important research results may not necessarily be uniformly reported in one language only.

To summarize this section, if journals are relying on subscription income for financial survival, and if libraries, faced with an ongoing serials crisis, are making selection and cancellation decisions on the basis of usage statistics, there are some potentially very serious implications for scholarly research. These kinds of decisions could lead to the cancellation of journals whose articles are well used, but in their open access form. This approach could also lead to a more conservative, popularity-based, scholasticism that is

less diverse in topic, language, and culture. While research is needed to confirm these possibilities, there are enough obvious reasons to give pause for thought before too many libraries begin to rely solely upon usage statistics in their selection and cancellation decisions.

If these assumptions are correct, the collective effect of these kinds of decisions, which make so much sense at the individual library level, have a potentially very unfortunate effect on scholarly communications as a whole. What then, is the remedy? Does it make sense for libraries to include consideration of the overall impact on scholarly communications in their collection development policies?

The good news is that open access not only can, but almost certainly would, counter most of these trends. The new paradigm might make it difficult to publish in traditional journals, or to start up a new traditional-style journal, with an existing publisher. With open source publishing software available, this research community can easily begin their own open access journal, and leave the question of reading and accepting their ideas with the reader. The issues are more difficult for research communities whose journals may be subject to cancellation, however converting to an open access model, or starting up new journals is an option for these communities as well. Journals that are in languages that are less likely to attract a significant subscription base can opt for open access as the best means to enhance the impact of their authors.

The danger of usage-based pricing

The ready availability of quality, reliable usage-based data raises the possibility of pricing based on usage. At face value, usage-based pricing does seem fair. Those who use a resource heavily pay the most, smaller users pay less. Indeed, there is much to say for considering usage when developing pricing models. Usage data can come in handy, for example, to determine the relative value of a resource for different types or sizes of libraries, and price accordingly. One example, using an FTE-based pricing model, would involve comparing the relative usage of resources at two-year colleges as compared to that at four-year universities. A resource that is used somewhat less at two-year colleges in general could be weighted to 75% FTE for colleges, while a resource that is used a great deal less at two-year colleges could be weighted at 50% FTE for these colleges.

There is much to be said for offering usage-based pricing, or the “pay by the drink” model, on an optional basis, when some libraries are unable to afford needed subscriptions. For obvious reasons, this is much better than no access at all.

However, if a pricing model based on usage were to become prevalent, there are some real dangers, as there are disincentives to use with usage-based pricing.

As Andrew Odlyzkow, Director, Digital Technology Centre, University of Minnesota, referring to internet usage pricing models, characterized it: "Usage-sensitive pricing is effective. The problem is that many of its effects are undesirable. In particular, such pricing lowers demand, often by substantial factors" (Odlyzkow 2001). For example, when AOL switched from usage-based to flat pricing for its users in 1996, usage tripled. This effect has been replicated in other countries and cultures. Research has shown that, with internet usage, even small charges discourage use, even if the charges are small enough that even heavy usage would be less than flat pricing.

While this research is based on Internet, rather than on print information resources and on individuals, rather than libraries, it makes sense that the same principles would apply to libraries and institutions as well. Picture, for example, a cash-strapped university looking for ways to cut the budget. With usage-based pricing, eliminating research papers at the first- or second-year level, eliminating the hands-on or exercise-based portion of an information literacy program, or scrapping an information literacy program altogether, would all be ways to achieve cost savings.

If the cost of use is known, there is a danger that a cash-strapped library will pass the cost along to the user, resulting in the direct disincentives to the user that Odlyzkow describes. This has been the tendency for many libraries with interlibrary loans, an area where libraries themselves have implemented usage-based fees deliberately, in order to limit demand (Budd 1989). Clinton (1999) discusses about how libraries in the United Kingdom have implemented user fees for interlibrary loans, to discourage what they see as indiscriminate use of the service.

With a print-based collection, users are free to browse to their heart's content. As a researcher, the author has often browsed extensively, often in journals not obviously related to the research topic, looking for new approaches or research methods, or possibly knowledge from one discipline that might have implications in another. If libraries move to electronic-only collections and pay on the basis of usage, this kind of cross-disciplinary research might well be perceived as costly. Readers and researchers might be discouraged from browsing for the sake of curiosity, and be asked to limit their reading to what might be clearly justifiable economically. Learning and certain types of research, such as interdisciplinary research, would suffer.

To conclude, pricing based upon usage appears to not be optimal for scholarly research, due to the likelihood of it discouraging use.

Summary and Conclusion

There are real needs for quality usage statistics, and it is encouraging to see some developments in this area, thanks largely to ICOLC and Project COUNTER. There are some benefits to considering usage in the economics of scholarly communication, particularly for the individual library, and as an informational measure to determine levels for other pricing models.

However, there are some real potential pitfalls if usage becomes prevalent as the basis for selection and cancellation decisions. There is reason to suspect that the cumulative effect of such decisions, made separately by many libraries, could create a tendency towards an overall increase in scholarly conservatism, the loss of important but less popular or less well-funded areas of research, detrimental effects on smaller research communities, and less linguistic and cultural diversity. Journals allowing open access options such as self-archiving also could be adversely affected. Happily, open access not only can, but also almost certainly will, counter many of the unfortunate effects of such decisions. The question of whether broader implications for scholarly communications as a whole should be incorporated into collection development policies is also being raised.

The possibility that usage statistics will form the basis of a usage-based pricing system has also been examined, and found to be inadvisable, as usage-based pricing tends to discourage usage.

Economics is concerned with the allocation of scarce resources which have potentially competing alternative uses. This is one of the most basic principles of economics. Consumption has an impact in determining what is produced, and for whom. Raising prices of products can control consumption by consumers who either have less desire for a particular product, or who have less ability to pay. (Allen 1967, p. 8-9).

The scholarly journal article in the electronic form does not fit within the realm of economics, as there is no reason to see a scholarly journal article as a scarce resource. An openly accessible article can be downloaded by millions, and its value will not at all be depleted. There are other kinds of goods which can gain value by creating a false scarcity; for example, commercial movies in electronic form. This does not fit with the model for scholarly knowledge, however, as scholarly knowledge, unlike goods and services produced primarily for profit, gains in value the more that it is used.

Science works in a series of steps, or blocks, which build upon one another. If one researcher finds a next step, the more researchers who read the results and build on them,

the faster that the research community as a whole, can all advance to the next step. Consider, for example, the cancer researcher. When research is concluded and the results are published, we could be one step closer to a cure, treatment, diagnosis, or basic understanding of how cancer works. The more people who find out about this step and move forward to the next step, the sooner we can all reach the ultimate goal (a cure, treatment, etc.). There is no value to be gained for the researcher in withholding this information. There is nothing to be gained from a pricing model that will tend to result in reasons for discouraging potential users from reading an article.

References

- Allen, C.L. *The Framework of Price Theory*. Belmont, California: Wadsworth Publishing Company, 1967.
- Budd, John M. "It's not the principle, it's the money of the thing". *The Journal of Academic Librarianship* 15, September (1989): 218-22.
- Clinton, Pat. "Charging users for interlibrary loans in UK university libraries – a new survey". *Interlending & Document Supply*. 27:1 (1999): 17.
- Gatten, Jeffrey N. & Tom Sanville. "An Orderly Retreat from the Big Deal: Is it Possible for Consortia?" *D-Lib Magazine* 10:10, (October 2004).
<http://www.dlib.org/dlib/october04/gatten/10gatten.html>
- Hahn, Karla L. and Lila A. Faulkner. "Evaluative Usage-based Metrics for the Selection of E-journals". *College & Research Libraries* 63:3 (May 2002): 215-217.
- ICOLC. *Guidelines for Statistical Measures of Usage of Web-Based Information Resources (Update: December 2001)*.
<http://www.library.yale.edu/consortia/2001webstats.htm>.
- Kuhn, Thomas. *The Structure of Scientific Revolutions*. 2nd edition. Chicago: University of Chicago Press, 1962/1970.
- Montgomery, Carol. *Presentation*. XXIV Annual Charleston Conference: *All the World's A Serial*. (2004).
- Odlyzkow, Andrew. *Internet pricing and the history of communications*. Revised version (February 8, 2001).
<http://www.dtc.umn.edu/~odlyzko/doc/history.communications1b.pdf>
- SHERPA *Publisher copyright policies & self-archiving*.
<http://www.sherpa.ac.uk/romeo.php>. (April 14, 2005).
- Zalta, Edward N.; Colin Allen, Uri Nodelman, Daniel McKenzie. *Stanford Encyclopedia of Philosophy*. *Open Letter to Librarians*.
<http://plato.stanford.edu/fundraising/librarians.html> (April 17, 2005).

