

University of Illinois at Urbana-Champaign
University Library
E-Journal Use Report

Prepared by the Task Force on Journal and E-Journal Use:

Tina Chrzastowski, Chair

Lisa German

Betsy Kruger

Chris Quinn

Lynne Rudasill

David Ward

Paula Watson

Lynn Wiley

*Comments by John Weible of Library Systems Department are inserted in the document text. His name appears in brackets before each comment.

Submitted to Karen Schmidt, Associate University Librarian for Collections, 2002

Introduction

As research libraries continue to expand access to electronic resources and to expend larger portions of their budget on e-resources, it becomes more critical to measure use of these products. Are these e-resources serving users well? Is the money invested in these resources well spent? Measuring the use of electronic resources, whether databases or full-text journals, is one performance measure that can help answer these questions. High frequency of use generally indicates users are finding and using a valuable electronic product. Users return again and again to sources that provide needed information or access, while they tend not to use resources that do not serve their research needs. However, low use of a resource could indicate many things. For example, low use could mean that the product is not being found by users, that users do not know how to use the product, or that they don't know the value or application of the product. User access and interface is critical to the introduction and uses of electronic resources. Without use data, the library cannot begin to identify high- and low-use electronic resources, or to address the use patterns of these products.

Use data, when combined with cost data, can help to determine the cost effectiveness of electronic products. Expenditure data for electronic products are another critical piece of an ongoing evaluative program to review e-resources. These data can also inform decisions on product enhancements and upgrades.

I. Charge to the Task Force

The Task Force on Journal and E-Journal use was created in early 2001 by AUL for Collections Karen Schmidt. The committee includes Tina Chrzastowski (Chair), Lisa German, Betsy Kruger, Chris Quinn, Lynne Rudasill, David Ward, Paula Watson and Lynn Wiley. The Task Force was charged with studying the acquisition of journal use statistics in the UIUC libraries in both print and electronic formats. A copy of the original charge is found in Appendix I.

II. Annotated Recommendations

1. Collect, cumulate, and post electronic resource use statistics on a staff-accessible web site.
 - The Task Force recommends full disclosure of the use of electronic resources when available. Use data should be made available at a staff-accessible web site, while links to use data found at publisher web sites are also encouraged. We support the regular evaluation of electronic products, and use made of e-resources should be a part of any review program.
2. E-product use statistics should be collected from sources that include both locally captured and publisher captured statistics.
 - There are two types of e-resource use data: locally collected and publisher collected. The Task Force, with the aid of John Weible, successfully regularized the collection of electronic journal use statistics through the Library's proxy server in early 2001. Although not a perfect solution to this type of data collection (since users can easily bookmark links and avoid the proxy), these data establish a benchmark of use and provide at least some use data when publishers are unable to do so. Where publisher use statistics are available, the local use counts provide a comparison that will help to determine how many users bookmark e-journal URLs or to confirm use levels provided by publishers.

[John Weible] Clarification: While our proxy server can be the source of some usage logs relevant to electronic resource use, it is not the source I emphasized in our earlier discussions. We have three internal sources of usage data for web-accessible databases, each of which can only show part of the situation. Even if combined, they don't show all of the usage data that the vendor can collect.

[John Weible] Specifically, these are the three sources:

- [John Weible] E-Journal access log: We began collecting usage data last year, after discussing the situation with this Task Force. That data currently counts the clicks from the Gateway E-Journal database listings, with a granularity of counts per journal per day. When/if the proposed new Gateway E-Resource system is employed, we can collect the same statistics for the non-journal items as well. Long-term, this is the best one to query since from it we could readily produce title- or vendor-specific totals for any date range desired. Since extraneous data is not stored, we can keep much more historical data online than the two following logs allow. This (rather than the proxy server) is also the "regularized" collection you should mention above.
- [John Weible] Web server standard log: Because most Gateway links to e-resources go through redirecting scripts, it is possible to query the huge access logs for web server, looking for clicks toward specific electronic resources. This data is currently more complete than the above, but it is much more time-consuming to extract from the access log.
- [John Weible] Proxy server access log: Logs every web page request sent through it. These could theoretically be used to analyze more detailed things than our normal web server's logs, but only for the remote users. Most on-campus users do not get their e-resource traffic routed through this machine.

3. All electronic product licenses should require publisher-captured and publisher-produced use statistics.
 - The Task Force recommends that every electronic journal or database license signed by the University of Illinois Library contain language requiring the producer or publisher to provide use statistics in a timely manner. Preference is for monthly statistics, but in some cases quarterly statistics will be acceptable.
 - [John Weible] Additionally, we should specify in what form those statistics should be supplied. They should be electronic, not paper, for example. Additionally, what electronic formats are allowed? And what method of transmission to us?
4. The Library Systems Department should become responsible for posting, collecting, and cumulating e-resource use statistics.
 - One centralized unit must be responsible for the collection, collation, posting and maintenance of e-resource use data. The Task Force recommends that the Library Systems Office be given this responsibility.
 - The Library Systems Office is recommended due to the expertise of the staff, who are experienced in electronic data and web access, and they also are presently collecting local use data. If possible, the Head of Library Systems should be designated the recipient of all use data generated by publishers/producers.
 - [John Weible] While the above recommendation is possible and a worthy goal, without changes to the processing of electronic resource purchasing, maintenance, and other issues, it will become a nightmare for us to maintain. Unless there is a single unit responsible for coordinating such agreements and purchases with ALL of the vendors, the details of getting this sort of information will likely remain too cumbersome to make it feasible.
 - [John Weible] Also note that as mentioned in #3, the details of what kinds of files and formats are acceptable need to be agreed upon.
 - In addition to the use data, the Task Force recommends that the Library Systems Office also provide regularly-updated data showing the number of e-products available to patrons. These would include e-journals, databases, reference sources and books. In order to estimate the use being made within the library universe, the size of the universe must be known.
 - [John Weible] For locally-collected statistics, this will be easier and more complete if the new, more comprehensive e-resource repository is implemented. Some of these resources are completely outside the influence of Systems, and they aren't tracked in any central repository.
5. The suggested collation time period is monthly, with running totals and annual cumulations. A year is January-December.
 - In as much as possible, data should be regularized to monthly collations, with annual cumulations. The "year" is defined as January to December.
6. The web site of electronic product use statistics should include:
 - Electronic journals (by title and by # of uses, cumulated monthly and annually)

- Electronic texts (by title and by # of uses, cumulated monthly and annually)
 - Article and Index Databases (by title and by # of uses, cumulated monthly and annually)
 - Other Electronic Resources (encyclopedias, etc.) (by title and by # of uses, cumulated monthly and annually)
 - Document Delivery (by title and by # of uses, cumulated monthly and annually) (Link from IRRC site)
 - Article Interlibrary Loan (by title and by # of uses, cumulated monthly and annually) (Link from IRRC site)
- In addition to the data by format and title, use data by vendor (publisher) would also be useful in analyzing cost/use benefit. Vendor cumulations (for example all Wiley or all Elsevier electronic journals) would make it easier to determine cost effectiveness of package purchases where journals titles outside owned print subscriptions are purchased.
7. If available, use by IP (Internet Protocol) addresses should be included.
- Use data that can delineate by IP address is very useful. Even if the system can only determine two types of access, library IP and non-library IP, those data are useful. Another type of IP access reporting that can be helpful is the number of unique IP addresses accessing the resource. The Task Force recommends license language include use data by some form of IP measure.
 - [John Weible] You should be aware that client IP addresses are increasingly unhelpful and often misleading for this sort of purpose. Worldwide, the use of several technologies are increasing which cause client IP's to change more often or to be masked from view by the servers which would be collecting such statistics. Specific technologies include DHCP (dynamic host configuration protocol), web proxy services, web caches, some firewalls, NAT (network address translation), and web anonymizer services. Most of these are already commonplace on this campus, both inside and outside the networks hosting library machines. Some ISPs even employ systems which can cause a user's apparent IP address to change during a session.
 - [John Weible] In addition, there is legitimate concern that analyzing usage of resources by client IP address can lead to invasion of privacy. Some library IT units and other IT units with similar data are choosing to make only aggregate host IP address information available for analysis. This can still usually provide enough detail to determine departments or service providers where the users are, but not enough to encroach on anyone's privacy.
 - [John Weible] Most of these same issues and arguments apply to DNS name logging and analysis, since there is often a one-to-one correspondence between an IP and a DNS name.
8. Performance measures other than e-resource use should be considered as contiguous and comparative. There are a number of performance measures that relate to the use of electronic (and print) resources. These include (but aren't limited to) photocopier statistics, patron counts, publicity or user education for the resource, and use counts from local web sites. In addition, other informative data include cost of the resource, number of copies held, format

availability, local holdings, and location. The Task Force recommends an examination of these types of background data that might explain why use levels are either lower or higher than expected. In addition, a use study of print journals is being tested in a few libraries for possible adoption library-wide. These data can also be collated and combined with other performance measures to create a composite picture of library use.

- In addition, as suggested by the ARL E-Metrics study, the above- mentioned data could create performance measures such as the number of virtual library visits compared to all library visits or the percentage of online journals uses compared to photocopies (and print uses when those data are available).
- [John Weible] Recent discussions with Paula and the AULs have indicated their desire for us to build a web-based repository for this kind of data as well as the usage data with which you're primarily concerned. All of this fits into a larger project which should be conceived and specified in further detail.
- [John Weible] I believe that expectations and operations have evolved to the point that we need to begin building a more well-defined and cohesive intranet for Library employees. The statistics repository is only one big area that belongs in such an intranet. Partly because of some technical changes, we have also begun rethinking how to collect and analyze usage log data from ALL of our web servers. At this time, there are various web pages which, because of content and sometimes restricted access, qualify as parts of a Library "intranet." We need to build a single authentication/authorization entry point to that diverse collection of resources to make it easier to manage and use. FINAL Notes: In conclusion, I need to point out one significant hesitation for commitment to these statistics projects. In all discussions about this, the hope is for the Systems Office to maintain the repository. The details of operations have not been discussed, so it's not clear what the expectations are. Here are various possible expectations for the statistics projects, only some of which have been explicitly stated, but all of which need to be made clear.
 - Expectation #1: LSO sets up data storage space and builds a web site with interface that allows downloading vendor-supplied statistics files by library employees.
 - Expectation #2: We additionally provide access to the LSO-collected statistics, such as web log or e-resource use data.
 - Expectation #3: A requested feature is to provide a method for entering manually-collected statistics (like turnstile counts) into a database via the web site.
 - Expectation #4: LSO receives files from dozens or hundreds of vendors and puts them into the repository when received.
 - Expectation #5: Do we then also have to nag all of the vendors when they don't supply them?
 - Expectation #6: Wouldn't some of the vendors make us use their non-standard process to go query and/or fetch the data from their systems? This would complicate the process, and make it more likely that we couldn't automate it.

- Expectation #7: Is someone expecting that we'll take the various data files from vendors and convert the data into a common format or database, so that the data can be more readily compared?
- Expectation #8: Then if we did Exp. #7, is providing Microsoft Access/Excel access to the resulting statistics database via ODBC sufficient, or do people expect a web-based query and analysis tool as well?
- Concerns:
 - Exp. #1, #2, and #3 seem reasonable for us to undertake. Those features are primarily up-front work to build the tools, with just a small impact on our labor afterward. Exp. #4 could be acceptable, but only if a standard, streamlined method of receiving documents is established, and if the concerns about needing a single office in the Library that maintains relations with the vendors are addressed. Exp. #5 through #8 would require significant and ongoing labor for us to support. We don't have the staff to undertake that kind of commitment and maintain current services.

III. Appendices

Charge to the Task Force

Background: The UIUC Library has excellent examples of journal use studies that have proven to be effective tools for making informed decisions on the selection and cancellation on journal titles. The studies that have been performed to date provide the groundwork for further studies on journal use throughout our collections. The studies look at only some of our journal collections, however, and we do not have established and regular ways to manage the data on electronic use that are available to us.

Because of the expense and importance of journals to all of our collections, a regular use assessment program that incorporates a much larger number of departmental libraries and collection sites is desirable. In addition, we receive a wide variety of statistics on electronic journal use that are not systematically collected or reviewed. These latter statistics are especially valuable to us now as we make important decisions about electronic access to journals. In a library with so many distinctive collections and differing policies on journal circulation, and with the wealth of electronic use data that are available to us, how can we provide for a regular and relatively easy way of assessing use of journals in all formats? How can we incorporate this into the critical decisions on acquisition, cancellation and document delivery that our collection specialists are called upon to make everyday? This is the impetus for creating the Task Force on Journal and E-Journal Use. The Task Force is charged with addressing the following areas:

1. Using a small number of libraries, demonstrate how we can expand journal use studies into areas of the collection where studies have not previously been conducted;
2. Examine the challenges that arise for collections for different disciplines and with different storage and access (E.g., some collections rely on the stacks for journal storage. How can a journal study be designed to address this situation? How do different circulation policies affect journal use studies and how can circulation data be used in different situations?)

3. Pull together the variety of e-journal statistics that are available to us and suggest ways in which we can report and use these data in a meaningful and timely manner.
4. Develop a standardized list of use statistics that are meaningful to the UIUC library which include all types of journal use, including but not limited to, print and electronic journals, document delivery, and ILL. Outcome: Make recommendations on one or methodologies for regular assessment of paper journal use studies that – if possible - do not require unusual staffing or computing needs; catalog and make available in a centralized location the many electronic use reports that are available to us; incorporate inter-library loan and document delivery into use reporting Time frame: A preliminary report should be done by September 1, 2001.

Statistical sample from proxy server

- <<http://www.library.uiuc.edu/transacts/transacts.asp> > (available to UIUC Library Staff only)

Statistical sample from American Chemical Society Journal Package use report

- <<http://www.library.uiuc.edu/chx/usestats.htm>>

ScienceDirect sample usage report

- (under construction)

Sample graphs incorporating cost and use data (under construction) Elsevier Electronic and Journal Usage and Cost Statistics

- <https://www-s1.library.uiuc.edu/administration/collections/estats/elsevier_cost_per_use_tina.htm> (available to UIUC Library Staff only)

IV. Bibliography / Suggested Readings

Luther, Judy, "White paper on electronic journal usage statistics, " Council on Library and Information Resources, October 2000, 25 pages.

<<http://www.clir.org/pubs/reports/pub94/contents.html>>

Shim, Wosnik; McClure, Charles R.; Fraser, Bruce T, et al, "ARL E-metrics phase II report. Association of Research Libraries, October 2001.

<<http://www.arl.org/stats/newmeas/emetrics/phasetwo.pdf>>