

Statistics

1. Definition and scope

Numerical counts measuring key activities in the use of electronic resources -- notably searches, items viewed, sessions, and turnaways. Depending on the analyses to which these counts will be subjected, additional levels of granularity may be tracked: e.g., full-text PDFs vs. full-text HTML, or full content units vs. descriptive records. These counts are typically used to evaluate the effectiveness of a service or resource, to compare usage across several services, or to compare usage of resources across target communities.

Counts *can* be maintained by any party to the transaction, including the content provider, the gateway (intermediary search engine), or the library itself. Deciding which party to the transaction *should* report these counts is a key issue for discussion as well as determining if there are new transaction types to be defined (e.g. remote search).

2. Issues list (for discussion at meeting)

- what to count (agreeing on what the important measures are, and reaching unambiguous, system-independent definitions of those measures)
- who counts it (determining which party in the search/results display process has primary responsibility for reporting usage)
- when to count it (establishing guidelines to better define a "search" or "retrieval" in the context of a metasearch activity when multiple sources are often searched at once.)

3. Related work to be drawn upon

Project COUNTER (Counting Online Usage of NeTworked Electronic Resources) "COUNTER Code of Practice" (December 2002) --

<http://www.projectcounter.org>

An international effort building on many national and individual type of library eMetrics initiatives, COUNTER's recently-issued Code of Practice (http://www.projectcounter.org/code_practice.html) has been endorsed by many of the leading electronic journal publishers, aggregators and library organizations. The Code of Practice includes definitions of activities to be measured (Section 3 / Table 1), identifies the organization responsible for reporting usage (see below), specifies the format in which data should be reported (section 4), and in future will include a process for auditing compliance. Activities counted include searches, sessions, turnaways, full-text article requests (=displays), and "successful item requests" at other levels of detail (TOCs, abstracts, references). *COUNTER is the only eMetrics effort to grapple with the problem of multiple organizations counting and reporting the same activity.*

7.3 Aggregators, gateways and hosts

Many, perhaps the majority, of online searches, are conducted using gateways or aggregators, rather than on the site of the original vendor of the item being sought. This presents special challenges for the collection of meaningful usage statistics. Section 3 of the Code of Practice specifies where responsibility lies for the recording and supplying of usage statistics when an intermediary aggregator or gateway is involved (see Section 3, Table 1, Terms 3.1.3.1, 3.1.3.2, 3.1.3.3, 3.1.3.4 and 3.1.3.5). The five scenarios defined therein describe protocols for delivery of the requested page to the customer:

- Direct from the vendor's server

- Direct from an aggregator
- Referred from an aggregator or gateway
- Via a gateway
- Referred to an aggregator or gateway

From Section 3, Table 1, (emphases added):

3.1.3	Source of page	
3.1.3.1	Direct from vendor's server	Delivery of content to the user is from the vendor's own service/site, to which the user has direct access.
3.1.3.2	Direct from an aggregator	Delivery of content to the user is from an intermediary (a gateway that is also a host), using its own store of publishers' content. Gateway is responsible for recording and supplying usage statistics for full-text requests direct to the customer and also, where contractually permitted to do so, to the vendor. (In this case the vendor may not add the 'gateway' usage figures to those recording usage of content delivered by the vendor direct to the customer)
3.1.3.3	Referral from an aggregator or gateway	Delivery involves the gateway sending the end user from the gateway's site to the vendor's site for the requested content. Vendor is responsible for recording and supplying full-text usage statistics to the customer. Gateway may also supply usage statistics to the customer, but must report them separately from those covering its delivery of full-text direct to the customer
3.1.3.4	Via a gateway	Delivery of content is via a gateway, which requests the content from the publisher and delivers it to the user in the context of the gateway service. Responsibility for collecting and supplying usage statistics to the customer is the same as in 3.1.3.2 above.
3.1.3.5	Referral to an aggregator or gateway	In this case an index or abstract service refers the customer to the gateway for full-text. In this case the full-text is delivered according to one of scenarios 3.1.3.2, 3.1.3.3 or 3.1.3.4 above, and the recording and supplying of usage statistics to the customer is as specified in each of these cases.

Association of Research Libraries "Measures for Electronic Resources" ARL E-Metrics Project (2000-2001) -- <http://www.arl.org/stats/newmeas/emetrics/>

Among its other objectives, this project identified several measures of electronic resources (collection counts and usage) that libraries should collect and analyze. The complete set of recommended measures are presented at http://www.arl.org/stats/newmeas/emetrics/module_3_files/frame.htm . Usage measures relevant to this committee's work include sessions (U2), searches (U3) and "items requested" (U4), described on slides 18-23 of the above-cited presentation.

NISO Z39.7-2002 "Information Services and Use: Metrics & statistics for libraries and information providers--Data Dictionary" (Draft Standard for Trial Use, June 2002) -- <http://www.niso.org/emetrics/>

Of particular interest are section 7.9.1, Electronic Collections (usage); Appendix A, Methods of Measurement, and Appendix B, Measuring the Use of Electronic Library Services. Usage measures in 7.9.1 that are relevant to this committee's work include units/records examined, queries (searches and menu selections), sessions and turnaways.

International Coalition of Library Consortia (ICOLC) "Guidelines for Statistical Measures of Usage of Web-Based Information Resources" (revised December 2001) -- <http://www.library.yale.edu/consortia/2001webstats.htm> .

Originally issued in 1998 and now endorsed by consortia representing thousands of academic, statewide and national libraries, the ICOLC Guidelines were the first significant attempt to standardize the statistics reported by electronic resource vendors. Recommends counts of sessions, queries, menu selections, turnaways, and various types of "full-content units" displayed.

4. Possible approaches and/or solutions

- what to count: as noted in the summary of related work above, consensus appears to have been reached as to what statistics are needed, and the COUNTER Code of Practice includes fairly detailed definitions and guidance for counting and reporting. Is further work required here?
- who counts it: see COUNTER Code of Practice, section 3.1.3.1-5 above. Oliver Pesch notes:

One of the issues that came up with the COUNTER group was defining who should count what. For example, if an e-journal gateway redirects the user to full text should the gateway count that as a full text retrieval as well as the publisher? The decision was that the party doing the actual delivery of full text to the user (in this case the publisher) would count the full text retrieval and the gateway would count it as a link to full text.

In the case of metasearch products we have a similar issue. A library will get statistics from their metasearch engines as well as from Gale, ProQuest, EBSCO, etc. We should provide some guidance as to how the stats can be combined so that activity on metasearch product can be measured as can activity at the content provider sites without double-counting. I believe this is an important part of the discussion. Perhaps the result will be defining new elements within COUNTER (e.g. "remote search") and guidelines for their use.

- when to count: in discussions for COUNTER and other search initiatives it is generally assumed that when multiple databases are searched in a single search action of the user, a "search" transaction is attributed to each database. Does this hold in a metasearch environment when a broadcast search technique is used? If a metasearch engine performs a preliminary search to determine the relevance of a given resource, should this be counted? Perhaps we can put forth some guidelines for best practice in counting these and other activities, like retrieval and submit to COUNTER for inclusion in their code of practice.