

ARL E-Metrics Project Activities Related to Vendor Statistics

1. Usage Reports Comparison: July-September, 2000.
2. ARL Meeting with Vendors: March 14, 2001 in Denver
3. Field Testing: May-June, 2001
 - 12 vendors agreed to participate. 8 responded.
 - Requested data as well as documentation.
 - Data sent to 2-3 libraries and the study team.

Table 1. Vendor Statistics Field-testing: Data Formats and Documentation

Vendors	Data Format	Availability of Documentation (Definition & Data Collection)
Academic Press	txt, Excel	n/a
Bell & Howell	Zip (Excel, txt, PDF)*	Yes
Ebsco	Txt	Yes
Gale Group	Csv	Yes
Lexis-Nexis	Zip (csv), Word, txt	Yes
NetLibrary	zip(txt), csv	Yes
Science Direct	Txt	Yes
SilverPlatter	Csv	n/a

n/a: not available from the vendor during the field testing.

Table 2. Key ICOLC Statistics Included in the Vendor Reports

Vendors	Items requested	Searches	Sessions	Turnaways
Academic Press!	Fulltext, reference, abstract, table of contents	Yes	Yes	n/a
IDEAL	Fulltext, abstract, citation	Yes	No	n/a
Bell & Howell	Fulltext, abstract	Yes	No	n/a
Ebsco	Fultext, abstract	Yes	No	n/a
Gale Group	Fulltext, citation & abstract, hits, views, print station	Yes	Yes	Yes
Lexis-Nexis	Fulltext, document retrievals	Yes	No	n/a
NetLibrary	Pageview, browse, checkout, dictionary use	Yes	Yes	Yes
Science Direct	Fulltext, abstract	Yes	Yes	n/a
SilverPlatter	Fulltext, abstract	Yes	Yes	Yes

n/a: not applicable.

Summary of Vendor Statistics Field Testing

1. Are the data provided to libraries reliable?

Since the field-testing dealt with only one month's worth of data, it is difficult to answer the question. However, we have not heard from the field-testing libraries of any unusual discrepancy between the field-testing data and data they received before the field-testing. The study team realizes that just comparing data from the same vendors will not give us a satisfactory answer. During the course of writing this report, we came

across an email message from a major database vendor acknowledging errors in their usage reports. This suggests that libraries are really not in a good position to know what exactly goes into the vendor reports. Some unusual numbers or patterns are relatively easy to identify. But consistent under (or over) counts are harder to detect.

One way to deal with the reliability issue is to collect data that libraries can generate in-house. For example, some libraries have set up a redirect webpage for external databases to count the number of attempted logins to licensed databases. This kind of data give gives clues that can be used to cross-check vendor-supplied numbers. Also, the library community needs to consider concrete ways (e.g., third party validation) to ensure consistent and reliable reporting from vendors, or at least should demand from vendors better documentation of the data collection and filtering process.

2. Are the data comparable across libraries, products, and vendors?

Use of different system parameters (e.g., time-out), application of different assumptions about user behavior (e.g. how to treat or count multiple clicks on the same document within a session), and the lack of adequate explanation in vendor documentation regarding specific definitions and data collection and filtering processes all contribute to the problem. Therefore, we conclude that it is largely impossible to compare data across vendors, and we recommend that comparison be limited to data from the same vendors. We believe that the comprehensive standardization of usage statistics and data delivery methods (e.g. file format and data arrangement) cannot be easily achieved in the short-term. Those are long-term goals toward which vendors and libraries need to work together. The ARL community should continue to make progress in this area by working amongst themselves and with the database vendor community.

3. Is the data easy to obtain and manipulate?

We believe that the data provided from the vendors studied are easy to obtain and manipulate. Most vendors offer several data formats including text format (e.g., comma separated file) and spreadsheet format (e.g., MS Excel) in addition to standard HTML format for easy viewing in web browsers. Also, many vendors offer an ad-hoc report generation facility whereby libraries can customize the fields they want to examine and set desired time periods.

However, we estimate that processing vendor reports from multiple vendors can become a considerable burden on libraries, in terms of time and staff efforts, as the formats and data arrangements vary considerably from vendor to vendor. We strongly recommend that vendors report standardized usage statistics, such as the ones recommended by the ICOLC and those found in part 4 of Phase II report (<http://www.arl.org/stats/newmeas/emetrics>). These should appear in the standardized column and row arrangements and include a separate report that contains vendor specific additional data.

4. Does the data provide meaningful information about the usage of networked information resources?

Usage statistics currently being provided by vendors give useful information regarding the utilization of external subscription-based information services. Libraries use data for a variety of purposes: usage trends over time, justification for expenditures, cost analysis, modification of service provision. Related to the issue of the value of data is trustworthiness (reliability) of data. Also, there is some concern over the lack of user-related information in usage statistics.