

Information Use Management and Policy Institute

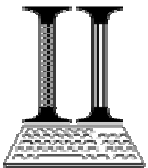
E-METRICS: Preparing Your Library to Collect Network Statistics (E-metrics)

By

John Carlo Bertot <bertot@lis.fsu.edu>
Associate Professor and Associate Director
Information Use Management and Policy Institute
School of Information Studies
Florida State University, Tallahassee, Florida 32306
850.644.8118 phone
850.644.9763 fax

For

NISO Committee AY, Library Statistics Standard Z39.7
January 2004



School of Information Studies
Florida State University
Information Institute
<http://www.ii.fsu.edu>

Note: This appendix intends to provide librarians with some guidance on managing the e-metrics collection process within their libraries. The appendix is not comprehensive, but rather selective in its presentation of e-metrics, issues, and assistance. Each library is different in implementation of electronic services and resources; thus, librarians should view this appendix's contents as general guidance that will likely require changes for application in a specific library setting.

PREPARING THE LIBRARY FOR COLLECTING NETWORK STATISTICS (E-METRICS)

General Considerations

1) Assemble a team: Choose an overall statistics coordinator for your library, if there is not one at the present time. It may be useful to enroll staff to help coordinate local data collection activities, collect the data, and prepare associated reporting and commentary. It is best that a single individual have ultimate management responsibility for the data collection process, though collection actually will occur throughout the library.

2) Determine which statistics your library will collect: Not all of the network statistics apply to your library, or your library may simply not be able to collect some statistics for a variety of reasons. It is important to determine which statistics your library will collect and report as part of a management process. This document describes ALL the network statistics; focus on the ones that your library will collect. Table 2 summarizes the collection points and approaches for the network statistics. We understand that not all library systems have branches. If your library is such a library, then consider all “branch level” data collection to mean system level collection.

3) Determine how your library will collect the data: The statistics require different intervals for collection – some use sampling approaches, others are “always on.” How you determine the collection period is up to the library.

4) Develop local data collection procedures: Identify the local, step-by-step, procedures necessary to collect the chosen network statistics. Who is going to do what, when, why? The below table may be useful:

E-metric:		
Task	Staff Member (identify by job title)	Hours to collect data (in quarter hour increments)
Collection procedure:		

5) View the process with a constructively critical eye: Keep the big picture in mind – what would you as the local librarian in charge of collecting these data want to know in advance so that the data are collected efficiently and used effectively?

Selected E-metrics and Collection Issues.		
E-Metric	Location of Collection	Collection Process
Workstations/Users		
Public Access Workstations	Branch level	Determine at one point in time annually during fiscal year period
Public Access Workstation users	Branch level	One-week sample, , estimate annual count
Sessions		
Commercial Services Sessions	System level/vendor	Likely monthly, aggregated to get an annual count
OPAC Sessions	System level/vendor	Likely monthly, aggregated to get an annual count
Rejected Sessions (turnaways)	System level/vendor	Likely monthly, aggregated to get an annual count
Searches/Menu Selections		
Commercial Services Searches (queries)	System level/vendor	Likely monthly, aggregated to get an annual count
Library Collection Searches (queries)	System level	Likely monthly, aggregated to get an annual count
OPAC Searches	System level/vendor	Likely monthly, aggregated to get an annual count
Units/Records Examined		
Library Collection Full-Content Units Examined	System level	Likely monthly, aggregated to get an annual count
Commercial Services Full-Content Units Examined	System level/vendor	Likely monthly, aggregated to get an annual count
Library Collection Descriptive Records Examined	System level	Likely monthly, aggregated to get an annual count
Commercial services descriptive records examined	System level/vendor	Likely monthly, aggregated to get an annual count
OPAC Descriptive Records Examined	System level/vendor	Likely monthly, aggregated to get an annual count
Virtual Visits		
Virtual Visits	System level	Likely monthly, aggregated to get an annual count
Virtual Reference		
Virtual Reference Transactions	Branch or system level	One-week sample
Instruction		
Formal User Information Technology Instruction	Branch level	Count instruction sessions throughout month, aggregate to get annual count
Point-of-Use Information Technology Instruction	Branch level	One-week sample, estimate annual count

Workstations/Users

Public Access Workstations (Branch Level Statistic)

Definition: Annual count of the total number of library owned public access graphical workstations that connect to the Internet for a dedicated purpose (to access an OPAC or specific database) or multiple-purposes. This statistic is counted and collected for each participating branch, if applicable.

Procedure:

1) Count and report the number of graphical workstations with Internet access (no matter the speed or type of connection) that are made available to the public as of the end of the field test period for each participating branch, if applicable. Include bookmobile, should your library have one that is Internet-ready.

Computers in computer labs used for public instruction if graphical and connected to the Internet should be counted. Public access graphical workstations that connect to the Internet that are used by both staff and the public should be counted if the workstation is used by the public for at least half of the hours during an average week that the library is open to the public. Reference desk computers used by staff to assist the public should not be counted.

Public Access Workstation Users (Branch Level Statistic)

Definition: Annual count of the number of users of all of the library's graphical public access workstations connected to the Internet computed from a one-week sample.

Procedure:

1) Select a one-week period during the test period. One week equals the number of hours the library is open over a consecutive seven-day period. Note: Data may be collected for more than one week and averaged but report this change in procedure when submitting the data to the project web site.

2) Prepare a written, step-by-step plan for collecting this data element. Identify the dates of the week(s) chosen to collect these data. The number of users may be counted by observation (continuous or every X minutes), manually using registration sign up sheets, via computer software, or through the "circulation" of workstations through your library's circulation system. You decide which method to use. Indicate whether you observed (and for how long), used sign up sheets, or used software to collect the data for this element. If you used software please indicate the name, approximate cost, address, phone and web page of the software used.

3) Count the number of users of all of the library's graphical public access workstations connected to the Internet during the chosen sample week.

Count each user that uses the graphical public access workstations connected to the Internet, regardless of the amount of time spent on the computer. A user who uses the library's workstations three times a week would count as three users in the count. Internet use includes all types of usage including WWW, e-mail, telnet, chat, etc. The study team recognizes the potential difficulty of determining whether a user on a multi-purpose (cd-rom access, word processing, etc.) workstation is using the Internet. Do not include staff use of these workstations.

4) Obtain a total figure of users for the week (or an average weekly use figure if you counted users over a two-week period) and report that number for each participating branch, if applicable. If you collect the user data over a two-week period, for example during the first week 70 users were counted, 80 users were counted the second week, the average number of users would be 75 (obtained and reported by adding week one's users to week two's $[70 + 80 = 150]$ and dividing by the number of weeks surveyed $[150 \div 2 = 75]$).

Sessions

Definition: A session is defined as a successful request of an online service or library's online catalog. It is one cycle of user activities that typically starts when a user connects to the service or database and ends by terminating activity that is either explicit (by leaving the service through exit or log-out) or implicit (timeout due to user inactivity).

Commercial Services Sessions (System Level/Vendor Statistic)

Definition: A session is defined as a successful request of a commercial service (e.g., online database). It is one cycle of user activities that typically starts when a user connects to a database and ends by terminating activity in the database that is either explicit (by leaving the database through log-out or exit) or implicit (timeout due to user inactivity). (ISO 2789, Annex A, modified to exclude OPAC sessions). Note 1: For multiple databases compiling several individual databases further information should be provided as to the separate databases hosted. Note 2: In some cases, e.g. database use inside the library, several users one after the other might make use of the same workstation, and sessions could not be separated. In most systems, a session is cut off after a specified time of non-use, thus avoiding part of the problem. The average timeout setting would be 30 minutes. If another timeout period is used this should be reported. Browser or proxy caching will be likely to reduce the number of requests registered in log files.

OPAC Sessions (System Level/Vendor Statistic)

Definition: A session is defined as a successful request of the library's online catalog. It is one cycle of user activities that typically starts when a user connects to the OPAC and ends by terminating activity in the OPAC that is either explicit (by leaving the database through log-out or exit) or implicit (timeout due to user inactivity). In some cases, e.g. OPAC use inside the library, several users one after the other might make use of the same workstation, and sessions

could not be separated. In most systems, a session is cut off after a specified time of non-use, thus avoiding part of the problem. The average timeout setting would be 30 minutes. If another timeout period is used this should be reported. Browser or proxy caching will be likely to reduce the number of requests registered in log files. (ISO 2789, Annex A; modified to exclude commercial services).

Rejected Sessions (Turnaways) (System Level/Vendor Statistic)

Definition: A rejected session (turnaway) is defined as an unsuccessful log-in to an electronic service by exceeding the simultaneous user limit. (ISO 2789, Annex A) Note: Failure of log-in because of wrong passwords is excluded.

Procedure:

- 1) Request from commercial database vendor(s)
 - a) Count the number of started sessions to each database for the field test month; and
 - b) Calculate the total sessions to host by adding the number of sessions from each database.

- 2) Obtain from OPAC network administrator or vendor
 - a) Count the number of started sessions to the OPAC for the field test month; and
 - b) Calculate the total sessions to host by adding the number of sessions from each day.

- 3) Obtain from OPAC administrator, network administrator and/or vendor
 - a) Count the number of rejected session to each database or OPAC

Searches/Menu Selections (Queries)

Definition: A search is defined as intending to represent a unique intellectual inquiry whether conducted through a search form submitted to the server or through the use of menu selections (e.g. browsing a list of subjects.)

Commercial Services Searches (Queries) (System Level/Vendor Statistic)

Definition: A search is defined as intending to represent a unique intellectual inquiry. Typically a search is recorded every time a search request is submitted to the server. Limited to commercial services. Mistyped search strings do not represent unique intellectual inquiries. Include menu selection searches. Exclude spider/crawler searches. (ISO 2789, Annex A).

Library Collection Searches (Queries) (System Level/Vendor Statistic)

Definition: A search is defined as intending to represent a unique intellectual inquiry. Typically a search is recorded every time a search request is submitted to the server. Limited to library

electronic collection services. Include menu selection searches. Exclude spider/crawler searches. Excludes OPAC searches. (ISO 2789, Annex A)

OPAC Searches (Queries) (System Level/Vendor Statistic)

Definition: A search is defined as intending to represent a unique intellectual inquiry. Typically a search is recorded every time a search request is submitted to the server. Limited to the library online catalog service. Include menu selection searches. Exclude spider/crawler searches. (ISO 2789, Annex A)

Procedure:

- 1) Request from licensed database vendor(s), or obtain from OPAC network administrator or vendor
 - a) Count the number of searches performed in each online library collection, commercial service, or OPCA for the field test month; and
 - b) Calculate total searches in all hosts by adding the total number of searches in each database.

Units/Records Examined

Definition: Content in the electronic collection that is delivered to a user. The sub-categories that follow provide for a detailed breakdown by type of content delivered (full-content unit, or descriptive record) and system delivering the content (Library Collection, Commercial Service or OPAC). Note: Please see 7.6: Loans and Document Delivery.

Library Collection Full-Content Units Examined (System Level/Vendor Statistic)

Definition: Number of full-content units from library electronic collection examined, downloaded, or otherwise supplied to a user. Exclude OPAC or commercial services (i.e., online databases).

Commercial Services Full-Content Units Examined (System Level/Vendor Statistic)

Definition: Number of subscription service full-content units examined, downloaded, or otherwise supplied to user, to the extent that these are recordable and controlled by the server rather than the browser. (ICOLC Guidelines, December 2001). Note 1: Journal articles – by journal title with ISSN and title listed. Note 2: Ebooks – by book title with ISBN and title listed. Note 3: Reference materials – by content unit appropriate to the resource (e.g., dictionary definitions, encyclopedia articles, biographies, etc.). Note 4: Non-textual resources – by file type as appropriate to resources (e.g., image, audio, video, etc.). (ICOLC Guidelines, December 2001)

Procedure:

1) Request from network administrator and licensed database vendor(s).

a) Count the number of views in each database for:

- Full text articles,
- Full text pages,
- PDF articles,
- PDF pages,
- Abstracts,
- Citations,
- Text only,
- Text and graphics; and

b) Calculate total views to host by adding the total views in each database. Consider using the table below.

Worksheet for the # Items Examined Using Subscription Services.			
Type of View:	Database 1	Database 2	Database 3
# Full text articles			
# Full text pages			
# Abstracts			
# Citations			
# Text only			
# Text/Graphics			
# PDF articles			
# PDF pages			
Total Views to Host			

Library Collection Descriptive Records Examined (System Level/Vendor Statistic)

Definition: Number of descriptive records concerning the library's electronic collection delivered to a user. Determined by the record type appropriate to the resource, e.g., abstract, archive, and index. Exclude OPAC or commercial services (i.e., online databases).

Commercial Services Descriptive Records Examined (System Level/Vendor Statistic)

Definition: Number of descriptive records concerning the library's commercial services delivered to a user. Determined by the record type appropriate to the resource, e.g., abstract, archive, index. Exclude OPAC or library collection descriptive records (i.e., library website services and collections).

OPAC Descriptive Records Examined

Definition: Number of descriptive records from the library's online catalog delivered to a user. Exclude commercial services or library collection descriptive records (i.e., library website services and collections).

Procedure:

- 1) Request from network administrator and licensed database vendor(s).
- 2) This item may also appear as part of the **Full-Content Units Examined** count. Consult the network administrator or licensed database vendor(s) to confirm to avoid duplicate counting.

Virtual Visits

Virtual Visits (System Level Statistic)

Definition: A user's request of the library web site from outside the library premises regardless of the number of pages or items viewed. (ISO 2789, 3.3.25) Excludes web site visits from within the library. Note: This statistic is the equivalent of a session for a library's website. As such, there is a need to exclude various actions (e.g., hits, downloads) by users during any given visit.

In the case of a user visit to a library web site a user who looks at 16 pages and 54 graphic images registers one visit on the Web server. Due to various web server issues and differing software this measure is an *estimate* of the visits to the web site. One definition (from the *MS Site Server* manual) of a virtual visit is: "A series of consecutive requests from a user to an Internet site. If your log file data includes referrer data, then new visits begin with referring links external to your Internet site. Regardless of whether or not you have referrer data, if a user does not make a request after a specified time period, the previous series of requests is considered to be a completed visit." Another log analysis software provider, *WebTrends*, defines a visit using the phrase "user session:" "A session of activity (all hits) for one user of a web site. A unique user is determined by IP address or domain name. By default, a user session is terminated when a user falls inactive for more than 30 minutes." An alternative approach is to assign each visitor unique authentication tags that are attached to each transaction.

Count of visits to the library via the Internet with a breakdown by:

- *# Internal virtual visits:* Visits while library users are in the library using public access Internet workstations (excludes library staff and staff workstations);
- *# External virtual visits:* Visits while library users access the library remotely (excluding the visits made by library users within the library using the public access Internet workstations); and,
- *Total # virtual visits:* A total count of both internal and external virtual visits.

Only report, however, the # External virtual visits.

Procedure:

- 1) Identify all electronic sources of visits to the library. This may involve activity that takes place on more than one computer server. The library may own some of the computer servers, another local government agency may own some others, or an Internet Service Provider (ISP).
- 2) Separate the various sources of virtual visits into staff internal, public internal, and public external. Two common approaches are using IP Address or some form of authentication tagged to each transaction. Exclude staff internal from the counts for this measure where possible.
- 3) Develop strategies for collecting the necessary data from each of these sources of virtual visits. Different software may be needed to measure each electronic source of virtual visits. In some cases, the library may calculate the virtual visits using one or more log analysis software packages. In other cases, the external owner of the computer server or service (the Internet Service Provider) must provide the data. Discussions may need to be held with these service providers to obtain the needed data. In still other cases, computer monitoring software may be appropriate.
- 4) In the case of a library web pages housed on library server: Identify, configure and install appropriate log analysis software. Determine log analysis software definition that corresponds to the virtual visit definition. Note: All log analysis software may not track virtual visits the same way so the count obtained will necessarily be an estimate. Arrange with the server technical staff for regular (e.g., monthly) reporting of internal staff visits (for your own internal use if interested) internal library user visits at the various public access Internet workstations, external library user virtual visits and total virtual visits (internal public visits plus external visits, excluding staff use). Run the log analysis software.
- 5) In the case of library web pages housed on an Internet Service Provider's (ISP) server: Identify log analysis software the ISP uses. Determine the definition of "visit" used by the log analysis software that corresponds to the virtual visit definition with the assistance of the ISP. Arrange with the ISP for regular (monthly) reporting of internal staff visits (for your own internal use if interested) internal library user visits at the various public access Internet workstations, external library user virtual visits and total virtual visits (internal public visits plus external visits, excluding staff use).

Virtual Reference

Virtual Reference Transactions (Branch or System Level Statistic)

Definition: Annual count of the number of virtual reference transactions conducted via e-mail, website, or other network-based medium designed to support virtual reference. (ARL E-metrics; Bertot, McClure, Ryan). Note: Includes questions either received or responded to.

Procedure:

- 1) Select a one-week period during the test period. Note: Data may be collected for more than one week and averaged but report this change in the procedure when submitting the data to the project web site.
- 2) Prepare a written, step-by-step plan for collecting this statistic. If your library offers virtual reference services and it is centralized through a single location, simply report the number of virtual reference transactions at the end of the week. If virtual reference is a distributed feature throughout the library system and its branches (if applicable), then you will need to aggregate the number of virtual reference transactions per branch.
- 3) Count the number of electronic reference requests received during the week. Be sure to report only transactions that occur virtually – both in receipt or answer. Report an electronic reference transaction as you would a face-to-face reference transaction. Thus, for example, one e-mail request may contain several reference questions taking varying time to complete. For example, one e-mail request contained 1 ready-reference question and one reference question that took 10-15 minutes to answer. Count the number of questions not the number of requests. So in the example you would report 2 as the number of electronic reference transactions.
- 4) Report the number of virtual reference transactions for the sample week – either for the system if this is a centralized service or for each branch that offers the service. Should you conduct the count over more than the one week period, you will need to obtain an average weekly number of virtual reference transactions by adding the total number of virtual reference transactions each week and dividing by the number of weeks this statistic was surveyed. For example, during the first week 7 electronic reference requests were counted, 9 were counted the second week. An average of 8 electronic reference requests is obtained and reported by adding week one's transactions to week two's [$7 + 9 = 16$] and dividing by the number of weeks surveyed [$16 \div 2 = 8$].

Instruction

Formal User Information Technology Instruction (Branch Level Statistic)

Definition: A count of the number of users instructed and the hours of instruction offered in the use of information technology or resources obtainable using information technology in structured sessions – either delivered in the library using a computer lab or other instructional setting or delivered electronically through online-based instruction. (Bertot, McClure, Davis).

Procedure:

- 1) Include only instruction in the use of information technology or resources obtainable using information technology. Examples of user instruction include use of the web, Internet searching, use of public access Internet workstations or personal computers, subject-based resources

available on the Internet, social implications of information technology (e.g., filtering and the public library).

2) Count all users attending and the session length of formal, structured lectures, public meetings, or courses in the use of information technology or resources obtainable using information technology the library offers, the library contracts for, or that use library facilities. A sign up sheet may be the most appropriate technique. Consider using the tally sheet shown below. In the count, include the number and duration of online training sessions (e.g., online tutorials or contacted training services such as those offered by ElementK – <http://www.elementk.com/>). To obtain online training session counts will likely require the assistance of the account/system administrator.

Weekly Formal User Instruction Tally Sheet.				
Library:		Date Week Begins:		
Date	Sponsor	Instruction Subject	# of Users	Session Length (in minutes)
Totals:				

3) A user need not be a registered library user. A single individual may attend multiple training sessions of the same or different types, each of which is counted. So if a single individual attended multiple training sessions (even if he/she repeated particular course), you would count each of those attendances and time intervals.

Point-of-Use information technology instruction (Branch Level Statistic)

Definition: A count of the number of users instructed and the hours of instruction offered in the use of information technology or resources obtainable using information technology in unstructured sessions at the impromptu request of users.

Procedure:

1) Select a one-week period during the test period. Note: Data may be collected for more than one week and averaged but report this change in the procedure when submitting the data to the project web site.

2) Prepare a written, step-by-step plan for collecting this statistic. The below table may facilitate the capturing of informal/point-of-use training data.

Daily Point-of-Use Instruction Tally Sheet.		
Library/Branch:		Date/Day:
Session	# of Users (May be more than 1)	Session Length (in minutes)
Session 1		
Session 2		
Session 3		
Session 4		
Session 5		
Totals:		(Total hours):

3) Count the number of users and hours (measured in 5-15 minute increments as decided by the library) of point-of-use instruction offered during the sample week period chosen. Include only the informal/impromptu instruction services provided in the use of information technology or resources accessible using information technology. Examples include use of the WWW, Internet searching, use of public access workstations or personal computers, subject-based resources available on the Internet.