Developing national data collection models for public library network statistics and performance measures: Project update

John Carlo Bertot and **Charles R. McClure**
Information Use Management and Policy Institute, School of Information Studies,
Florida State University, USA

Abstract

Public library networked statistics and performance measures are important indicators of the use, uses, and users of networked services that public libraries offer their patrons. Individually, these factors provide libraries with the ability to incorporate network usage data into key decision making processes and planning activities. Together, through a systematic approach to the collection and reporting of public library network statistics on a national scale, these factors provide significant data regarding use and usage trends throughout U.S. public libraries that inform policy makers, researchers, and library professionals as to what types of network activities are occurring in public libraries and how those uses can lead to public library role changes in the networked environment.

The research presented in this paper is made possible through the support of a U.S. Institute of Museum and Library Services National Leadership Grant (NR-00009). Additional study information can be found at: http://www.ii.fsu.edu/Projects/IMLS/index.html

Introduction

Public library networked statistics and performance measures are important indicators of the use, uses, and users of networked services that public libraries offer their patrons. Access to data that identify what networked services are being used by whom and when can:

- Enable local library directors and state library agencies to compete for resources with other local and state organizations as by documenting the range, extent, and impact of library-provided networked services.
- Assist public libraries make a strong case for federal or local community support for technology and information infrastructure by documenting their Internet-based services and resources.
- Facilitate the transition from traditional library use measures such as circulation, reference transactions, interlibrary loans, etc., to network measures that describe the nature and use of library-based network activities and resources.
- Assist libraries in administrative, management, and planning activities.
- Allow individual libraries, states, and regions to effectively compare themselves to others in terms of Internet development, costs, provision of services, connectivity, and use.
- Provide libraries a means through which to assess, compare, and make decisions regarding collections resources in general and future collections development decisions in particular.

Individually, these factors provide libraries with the ability to incorporate network usage data into key decision making processes and planning activities. Together, through a systematic approach to the collection and reporting of public library network statistics on a national scale, these factors provide significant data regarding use and usage trends throughout U.S. public libraries that inform policy makers, researchers, and library professionals as to what types of network activities are occurring in public libraries and how those uses can lead to public library role changes in the networked environment.

Background

With the assistance of a 1998 National Leadership Grant from the U.S. Institute of Museum and Library Services (IMLS), the authors conducted a study to develop national public library network statistics and performance measures. Through this study, the authors developed, defined, and field-tested a series of public library network statistics and performance measures for ultimate collection at the national level (see Figure 1). Overall, the statistics and measures developed in this study provide the means to track various public library networking capabilities and activities such as the type and level of Internet connection; types, public access workstations; available network-based services, such as databases, online reference, and web-based services (e.g., digital libraries); and staff and user instruction.

Upon completion of the network statistics and performance measure project, the authors received a 2000 National Leadership Grant from IMLS to assess the feasibility of and develop a national model for collecting public library network statistics and performance measures. Partners in the current study include the U.S. National Commission on Libraries and Information Science (NCLIS) and the National Information

Standards Organization (NISO). Based on the project findings and model field test (discussed in more depth below), the researchers will recommend to IMLS, the public library community, state library administrators, and others, possible approaches to the collection of public library network statistics on a national scale. These recommendations will be included in the final project report to IMLS in May 2002.

The authors note that there are a number of national and international efforts underway that continue to research the library network statistics and performance measure environment. This paper summarizes, but does not review, selected efforts in Figure 2.

Developing A National Public Library Network Statistics Data Collection Model

There are a number of network statistics and performance measurement data collection issues that require resolution (Bertot, McClure, and Ryan, 2000). Among these issues are the:

- *Range of Sources of Network Data within Individual Library*. No two libraries have the same information technology infrastructure, configuration, or systems implementation. This creates a substantial challenge for the collection of the same data from libraries using similar (but different) technology in various configurations.
- *Data Quality*. Librarians want accurate, credible trustworthy, valid, and reliable data that describe the use and uses of their networked resources and services. However, there is a false expectation that machine-generated or captured data (e.g., online database sessions, web visits) are exceptionally accurate. This is not the case, as the quality of such data depends on a number of factors. Indeed, Figure 3 demonstrates the difficulty of capturing a simple database session count. All nationally collected and reported data related to libraries and services are best seen as estimates – even those that are currently collected (Library Research Service, 1995).
- *Limits to Longitudinal Data*. Longitudinal data are useful to track trends within a library and as a check for unusual spikes or bad data. But the rapidly changing nature of information technology will have a substantial impact on the life cycle of the network statistics and performance measures. It is unclear as to just how longitudinal network statistics will be given the need to change what they capture and how as technology changes.
- *New Data Collection Techniques*. Network measures require researchers and professionals to consider the benefits and/or necessity of using new data collection techniques including traditional quantitative methodologies (surveys, or Likert scale surveys of user satisfaction with network services) in new ways; less familiar qualitative (e.g., focus groups, interviews) methods; adapting traditional methodologies (e.g., pop-up Web-based surveys); and creating new methodologies (e.g., Web-based transaction log analysis) to capture network usage data. In some cases, to promote timely and responsive measures it may make sense to rely on carefully developed samples at the local, state, and national levels rather than 100% population responses. In other cases, sequencing data collection, in which a question is not asked annually but every two or three years may be appropriate to reduce local data collection burden.
- *Ability of Local Libraries to Collect Network Measures*. In order to attain national network statistics and performance measure data, it is necessary to collect the raw data at the local library outlet level. It remains unclear as to the ability of the library outlets to collect such data.
- *Preparation and Training Necessary*. Collecting data on network measures will require preparation and library staff training to be successful. There are a number of training topics that need attention including the identification of the range and diversity of technology generating network measures; the notion that at least for the near term estimates, samples, and the lack of long term longitudinal network data may be the norm; and introduction to new data collection techniques and how they may be applied to collecting network measures relevant to local libraries.
- *Training in New Data Analysis Techniques Necessary*. Training in how to analyze and interpret these new network measures (some more than others) will be necessary at all levels. For example, training librarians to download pre-formatted data into a standard spreadsheet and then do some basic analysis. In addition, few of those interviewed outside of some systems librarians knew how to effectively use network analysis data. In the case where the library has systems staff, network data may only be used for internal technical purposes. But often, these technical experts have not seen the utility of this data for wider administrative purposes such as demonstrating use, showing need, garnering funding.
- *Training in the Analysis and Use of the Network Data Reported Necessary*. Librarians have spent decades convincing local governing boards that circulation counts, attendance records, reference transactions, etc. that go up annually are a "good thing." Now that these and other traditional counts are stagnant or declining in many cases, librarians have to re-educate governing boards that web hits,

electronic reference questions, full text downloads, and other indicators are as or more important than the traditional measures.

- *Partnerships – Maintaining Control, Obtaining Data*. Partnerships, both formal and informal, are a way of life for public libraries. But they can create problems for the collection of needed network measures when the technology, network, or databases are not owned by the library (i.e., online database vendors, consortia). It is important for libraries to foster reporting agreements with external entities to ensure that libraries receive use data for services to which they subscribe or can access through subscriptions by other entities (e.g., state library agency).

These issues, at a minimum, require attention and an acceptable level of resolution for it to be possible to develop and collect national public library network statistics and performance measures.

Figure 1. Public Library Network Statistics

Public Access Workstations

 # Public access workstations
Annual count of the 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.

 # Public access workstation users 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.

Maximum speed of public access Internet workstations Indication of the maximum bandwidth of public Internet access, e.g., less than 56kbps, 56kbps, 128kbps, 1.5mbps, etc.

Databases

 # Full text titles available by subscription
Report: Serial titles, Other titles, Total titles
Count of the number of full text titles that the library subscribes and offers to the public computed one time annually.

 # Database sessions
Total count of the number of sessions (logins) initiated to the online databases. Definition adapted from proposed ICOLC standard http://www.library.yale.edu/consortia/webstats.html.

 # Database queries/searches
Total count of the number of searches conducted in the library's online databases. Subsequent activities by users (e.g., browsing, printing) are not considered part of the search process. Definition adapted from proposed ICOLC standard http://www.library.yale.edu/ consortia/webstats.html.

 # Items examined using subscription services
Count the number views to each entire host to which the library subscribes. A view is defined as the number of full text articles/pages, abstracts, citations, and text only, text/graphics viewed. Definition adapted from proposed ICOLC standard http://www.library.yale.edu/consortia/webstats.html.

Electronic Services

 # Virtual reference transactions
Annual count of the number of reference transaction using the Internet. A transaction must include a question received electronically (e.g., via e-mail, WWW form, etc.) *and* responded to electronically (e.g., e-mail).

Public service time spent servicing information technology
Report: Information technology staff, Paid public service staff (Professional Librarian, Paraprofessional), Volunteer, & Total
Annual count of the staff hours spent in servicing information technology resource and service activity in public service areas computed based on a one week sample.

Virtual Visits

 # Virtual visits to networked library resources
Report: # Internal virtual visits, # External virtual visits, # Total virtual visits
Count of visits to the library via the Internet. A visit occurs when an external user connects to a networked library resource for any length of time or purpose (regardless of the number of pages or elements viewed). Examples of a networked library resource include a library OPAC or a library web page. 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.

Figure 1. Public Library Network Statistics continued

Instruction

User information technology instruction
Report: # Users instructed, # Hours of instruction
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, informal, and electronically delivered instruction sessions conducted or sponsored by the library.

Staff information technology instruction
Report: # Staff instructed, # Hours of staff instruction
Annual count of the total number of staff instructed and the number of hours of formal instruction in the management or use of information technology or resources obtainable using information technology.

Figure 2. Selected Network Statistics and Performance Measure Initiatives

International Coalition of Library Consortia (ICOLC)
ICOLC is an international coalition of predominantly research libraries (some of which are sponsors of the ARL e-metrics project) interested in pursuing standard network statistics and reporting systems regarding database vendor data. ICOLC first published its proposed standards and definitions in November 1998 and is currently considering revisions to those standards. Additional information on the ICOLC initiative is available at http://www.library.yale.edu/consortia/webstats.html.

International Standards Organization (ISO)
Through the ISO Technical Committee 46 (Information and Documentation), subcommittee 08 (Statistics and Performance Evaluation) members of ISO have been revising both general library statistic standards and incorporating network statistics and performance measures into the statistical data collection efforts of participating libraries (multi-type). As of July 2001, the U.S., through NISO, rejoined the ISO effort after a one-year absence. Recent balloting efforts resulted in the passage of the proposed ISO library statistics (document ISO/DIS 2789) although a number of voting members provided substantial comments on the statistics. Additional information on this and other ISO efforts are available at http://www.iso.ch/iso/en/ISOOnline.frontpage.

European Community-sponsored Equinox project
The Equinox project focused on developing library performance and quality measures. In particular, the project aimed to further develop existing international agreement on performance measures for libraries for the electronic library environment as well as develop and test an integrated quality management and performance measurement tool for library managers. The project identified a number of performance indicators that, in some cases, have been integrated into the ISO library statistics initiative. Additional information on the Equinox project is available at http://equinox.dcu.ie/.

LibEcon project
A European initiative, LibEcon focuses on the collection of economic and other library-related data from predominantly European libraries. For its Millennium Study, the survey incorporated selected network statistics developed by the IMLS study as well as ISO activities. Additional information on LibEcon is available at http://www.libecon2000.org/.

Council on Library and Information Resources (CLIR) initiative
CLIR investigated the issues surrounding network statistics primarily from an online database data perspective. The initial study, conducted during 1999 and 2000, resulted in the publication of a white paper entitled White Paper on Electronic Journal Usage Statistics. It is the understanding of the study team that the work begun through this effort continues. Additional information on this initiative is available at http://www.clir.org.

Publisher and Libraries Solution Committee (PALS)
This recent initiative, operating through the auspices of the Joint Information Systems Committee (JISC), is exploring the data needs of libraries from publisher provided online usage statistics. A Vendor-based usage statistics working group has been developed to explore the issues involved regarding online vendor statistics in a more in-depth fashion. This group met in June 2001 in the United Kingdom to pursue further network statistics needs of libraries from vendors. Study team members are in contact with this group. Additional information at: http://www.jisc.ac.uk/curriss/collab/c6_pub/#uswg.

Figure 2. Selected Network Statistics and Performance Measure Initiatives continued

Association of Research Libraries (ARL) E-Metrics
Beginning in July 2000, this project seeks to identify, define, and standardize a set of network statistics and performance measures for ARL libraries. The intent is to develop a core set of network statistics and performance measures through which ARL libraries can manage their networked resources and services, plan future network resources and services, and benchmark themselves against other ARL libraries. In addition, the project seeks to provide network statistics and performance measures in an outcomes-based context. Additional information on the project is available at: http://www.arl.org/stats/newmeas/emetrics/index.html.

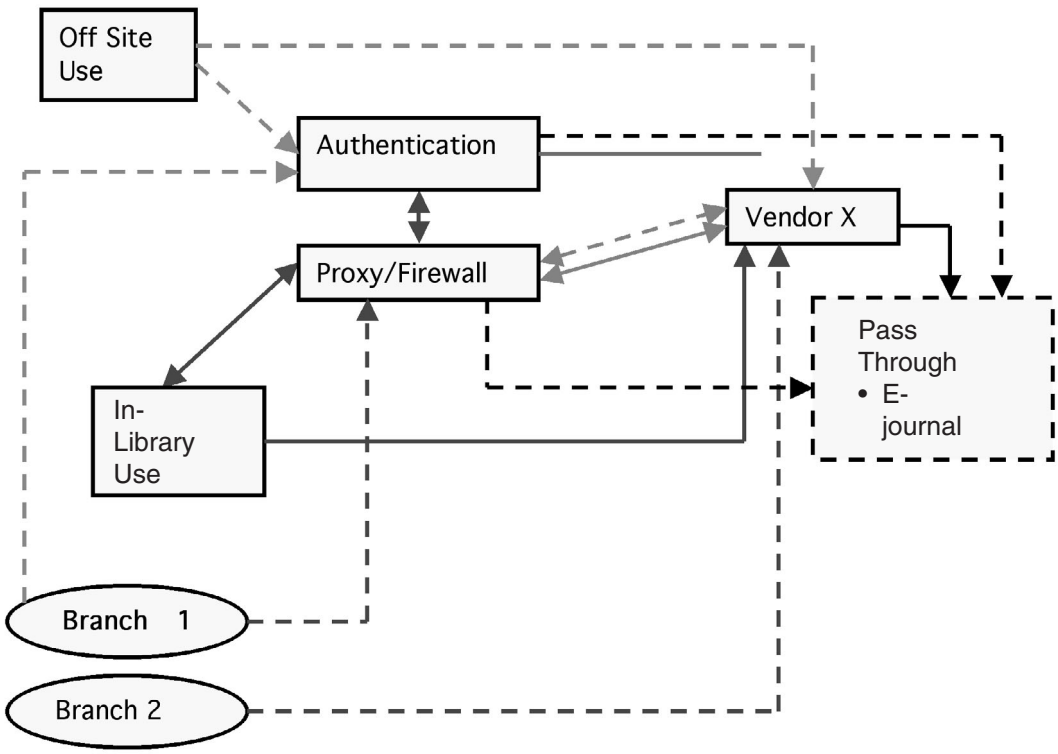

Figure 3. Capturing an Online Database Session

Possible Models for National Data Collection

The ability to resolve several of the issues identified above lies, in part, with the type and nature of the national public library network statistics and performance measure system adopted by library professionals, researchers, and policy makers. Research by the authors – as well as numerous formal and informal interviews, focus groups, and discussions with library leaders, researchers, and policy makers – suggests that there may be numerous approaches to the development of a national network statistics and performance measure collection, reporting, and analysis system. These include:

Extending the current National Center for Education Statistics (NCES), National Commission on Libraries and Information Science (NCLIS), state library agency, and public library Federal State Cooperative System (FSCS) collaborative approach for annual public library data collection.

In this model, public library data on selected statistics are passed from public libraries to state library agencies up to NCES for compilation, analysis, and reporting. All 50 states plus the District of Columbia and U.S. Territories participate in the process.

At present, there are approximately 50 data elements collected and reported through the FSCS process (e.g., operating budgets, FTEs, circulation). It is necessary to propose new elements through an administrative procedure, and element adoption requires the vote of at least 26 state data coordinators (SDC, personnel located in state library agencies) with a three-year phase-in on inclusion by all 50 states, Washington, D.C., and U.S. Territories. The adopted elements then go into the following year's survey form for collection. The time from element vote to adoption to collection can be as long as three years.

Over the last several years, the FSCS group undertook at least two efforts to adopt a variety of network statistics. The most recent effort occurred in December 1999. While some network statistics were voted in, a vast majority – including some network statistics contained in Appendix A – failed to get the necessary 26 votes for adoption. [Note: state library agencies voted in early 2001, however, to include the online database statistics on their annual surveys beginning with the 2002 reporting year].

Developing a lead states and libraries approach to data collection and reporting.

For a variety of reasons, it may not be feasible for all public libraries and state library agencies in the nation to simultaneously adopt and report data on a set of network statistics and performance measures. However, research by the authors demonstrates that there are a number of states (20+) that indicate their willingness and/or desire to collect at least a core set of network statistics and performance measures from the public libraries within their states. The same research shows that, while a state library agency or a number of public libraries within a state may not be willing or able to collect network statistics, lead public libraries within states find it imperative to collect such data for a variety of decision making, management, and reporting purposes. In this model, lead public libraries and state library agencies adopt, collect data, analyze data, and report data on a core set of network statistics and performance measures. The lead state library agencies and public libraries also serve as an incubator for developing, defining, and reporting new network statistics and performance measures.

Creating an ongoing sampling design to generate national estimates.

This model employs a sampling approach for a variety of data collection activities to use with public libraries, state library agencies, and library network consortia. The intent of this approach is to develop a sample that would enable the generation of national estimates of core set of network statistics and performance measures from public library, state library agency, and library network consortia. Such a model would permit the targeting of network statistics appropriate to the level of data collection – library, state library agency, library consortia – as well a framework for modifying or creating new statistics and performance measures on an as needed basis. It would be possible to engage in the data collection process on a regular (e.g., annual, biannual) and/or ad hoc (e.g., as necessary) basis.

Adopting a combination approach to network statistic and performance measure data collection.

The above data collection models are not mutually exclusive. Rather, it is possible to combine aspects of the FSCS, lead state/library, and sampling approaches to collect, analyze, and report public library network statistics and performance measures so as to provide nationally aggregated network statistical data.

A key aspect of the research project is to determine which model or aspects of the above models – including models and/or approaches not yet identified – are appropriate under what circumstances for the development, definition, collection, analysis, and reporting of national public library network statistics and performance measures.

National Model Field Test

The study team is preparing to conduct a national model data collection field test that would both continue the test of the network statistics and perform-

ance measures and test a method for data collection and reporting. The field test would occur in October 2001, so as to provide state library agencies, consortia, and public libraries with adequate time to prepare for the field test. The goals of the field test are to:

(1) Create a fast response approach to the development, collection, analysis, and reporting of network statistics and performance measures;

(2) Foster an environment of constant change;

(3) Implement a reasonably burden free data collection and reporting process for public libraries, state library agencies, and library consortia;

(4) Work with non-library partners to gain access to library network data (e.g., vendors, consortia, state library agencies); and

(5) Produce national estimates of public library network service uses and usage.

To do so, the field test will:

- *Use a selected states, public library, and consortia approach*. The field-testing approach presented to the public library and state library agency community was on of an initially purposeful approach in which states, and public libraries and consortia within those states, would be selected based on the extent to which the states were already collecting/planning to collect network statistics and performance measures, annual state library database expenditures, known consortia within the states, a varied composition to the state's public library landscape along rural/urban, poverty, and population of legal service area demographics (as found in the annual public library surveys as well as poverty data held by the study team and NCLIS), and a willingness to participate in the study.

- *Select the states*. Based on the above criteria, the study team identified 14 states to participate in the field test (Alabama, Colorado, Georgia, Illinois, Kentucky, Louisiana, North Carolina, Ohio, Pennsylvania, Texas, Virginia, Wisconsin, and Wyoming). To date, ten states have agreed to participate in the field test.

- *Use multiple data collection points*. In the networked environment, network-based services are provided directly or indirectly to or by public libraries. Thus, it is necessary to collect network use and usage data from multiple entities to provide an estimate that reflects actual public library network services. For example, state library agencies are increasingly providing statewide access to online databases that users can access either through public library facilities or remotely, and consortia provide their library members with access to online databases that users can access either through public library facilities or remotely. For there to be an accurate reporting system (though still an estimate, as discussed in Bertot, McClure, and Ryan, 2001), the consortia and state library agencies need to report the online database usage statistics for the public libraries. These usage statistics need to be reported in combination with the public library network usage data so as to get a better sense of the total public library use of, in this case, online database resources. The other network statistics (e.g., public access, virtual visits, Instruction) would be collected by individual public libraries and then reported to the state library agency.

- *Create a reporting system*. The field test will parallel the FSCS process, but use a selected sample-based approach in which

  o Public libraries collect and report their network statistics usage data to the state library agencies (likely the SDC). Unlike the initial field test of the network statistics and performance measures during 1999-2000 during which public libraries or consortia were assigned specific statistics to test and report, this field test will ask public libraries to collect all the statistics that apply to their organizations – e.g., public access, virtual visits, instruction, databases, and electronic services. Not all libraries provide instruction, have a website, or subscribe to online databases, thus they would not report instruction, virtual visit, or database data.

  o Consortia report their public library online database usage data to the state library agencies (likely the SDC). [*A key issue to resolve is whether these consortia or state library agency-based online database usage data should be disaggregated and reported by individual public libraries or kept in an aggregate "public library usage" state. Disaggregation would require substantially more effort, and it is unclear if such effort provides additional benefit over an aggregated reporting of the usage data*.]

  o State library agencies separate out public library online database usage data. It may also be the case that state library agencies serve as the state library facility. It is unclear as to whether public access workstation data, or other services related data such as virtual visits or electronic reference, would be reported as part of the public library national data. It is more likely that state library agencies will need to consider the reporting of such data on the state library agency annual survey.

- State library agencies report network usage data to a central data collection agency. For the purpose of the field test, the Information Institute at Florida State University will act as the central data collection agency. The Information Institute will also generate reports based on the data collection effort so as to identify sample reports and reporting formats. As part of the field test, the study team will identify issues, burden, and other management requirements for the data collection effort. This study team will then present this information to NCLIS (who is actually one of the study partners), the National Center for Education Statistics (NCES), SDCs, state librarians, consortia managers, and public library directors. Based on feedback from these entities, the study team will make final adjustments and recommendations as to the national data collection and reporting processes.

This approach will provide substantial feedback on the ability of libraries, consortia, and state library agencies to report the network statistics data; the process by which data are reported, and the ability of the reported data to describe public library network services and resources use. In addition, the results of the field test will identify what the characteristics of national data collection system should be given the network statistics environment.

Clearly, such a national data collection system will need to:

- *Create a fast response approach to the development, collection, analysis, and reporting of network statistics and performance measures*. A key criticism of the FSCS process is the time lag between the development of data elements and the eventual reporting of those elements. For a variety of reasons, it can take four years under the current FSCS process from development to reporting of statistics (to be fair, the FSCS group undertook changes in its bylaws recently to expedite the element adoption and reporting process). By the time the NCES releases the public library data reports, the data are often outdated. This is particularly problematic in the networked environment in which any network statistics and performance measures will likely remain relevant for two-three years.
- *Foster an environment of constant change*. Gone are the days of statistics and performance measures that last for decades. The networked environment is such that change in technologies and the implementation of those technologies is rapid. Thus, the statistics and performance measures that capture network data will necessarily undergo constant modification. It is imperative, therefore, that the model for national library network statistics and measures foster an environment of flexibility, change, and creativity in the creation, collection, and reporting of statistical data.
- *Implement a reasonably burden free data collection and reporting process for public libraries, state library agencies, and library consortia*. It is clear that data reporting requirements imposed on public libraries are arduous. It is also clear, however, that network usage statistics are increasingly important to professionals, researchers, and policy makers. Thus, it is necessary to develop a data collection and reporting system that provides maximum benefit for minimal effort.
- *Work with non-library partners to gain access to library network data*. Increasingly, key network usage data is out of the public library, state library, and library consortia domain. Examples include online database usage, Internet service provider (ISP), and telecommunications carrier (e.g., bandwidth consumption) data. It is critical to the measurement of library network services that the national data collection activities develop reporting partnerships with, minimally, the online database vendor, ISP, and telecommunications carrier communities.

Undoubtedly, there are other characteristics necessary for a national public library network statistics data collection system that the field test will identify.

Next Steps

The development of a set of national public library network statistics (see Figure 1) was an important first step that provides public libraries with the ability to describe the use, uses, and users of their networked resources. It is critical, however, to aggregate individual public library network statistics data collection efforts in order to inform policy makers, practitioners, researchers, and others as to the shift in public library services and resources, the use of their services and resources, and the impact such uses have the public library as a community-based organization. Doing so will enable public libraries to report – at the local, state, and national levels – the contributions they make to their communities through their network-based services. Ultimately, this study will recommend possible approaches to accomplish these objectives.

References

Bertot, J.C., McClure, C.R., and Ryan, J. (2001). *Statistics and Performance Measures for Public Library Networked Services*. Chicago, IL: American Library Association.

Bertot, J.C., McClure, C.R., and Ryan, J. (2000). Developing National Library Network Statistics and Performance Measures for U.S. Public Libraries: Issues, Findings, and Recommendations. *Performance Measurement and Metrics: The International Journal for Library and Information Services*, 1(1): 15-42.