

Managing Data Collection for Real-Time Reference: Lessons from the AskERIC Live! Experience

Yvonne Belanger
R. David Lankes
Pauline Lynch Shostack

SUMMARY. Successful management of a real-time digital reference service requires effective and efficient data collection to provide accurate and useful information about service processes and outcomes. This article describes the data collection procedures developed by AskERIC for its real-time reference service; in particular, this article will describe how these procedures address limitations, gaps, and inaccuracies in the data as passively captured by software. Finally, particular issues and considerations for service managers about various statistical measures within the real-time environment are discussed. [Article copies available for a fee from *The Haworth Document Delivery Service*: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2003 by *The Haworth Press, Inc.* All rights reserved.]

Yvonne Belanger (yvonne@iis.syr.edu) is Project Manager for the Virtual Reference Desk Project and former AskERIC QA Project Manager; R. David Lankes (rdlankes@ericir.syr.edu) is Director of the Information Institute of Syracuse; and at the time of submission, Pauline Lynch Shostack (pauline@askeric.org) was the AskERIC Project Coordinator; all at the Information Institute of Syracuse, 621 Skytop Road, Suite 160, Syracuse, NY 13244.

The AskERIC Project is a sponsored program of Syracuse University funded by the U.S. Department of Education.

[Haworth co-indexing entry note]: "Managing Data Collection for Real-Time Reference: Lessons from the AskERIC Live! Experience." Belanger, Yvonne, R. David Lankes, and Pauline Lynch Shostack. Co-published simultaneously in *Internet Reference Services Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 8, No. 1/2, 2003, pp. 137-148; and: *Virtual Reference Services: Issues and Trends* (ed: Stacey Kimmel, and Jennifer Heise) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2003, pp. 137-148. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

<http://www.haworthpress.com/store/product.asp?sku=J136>

© 2003 by The Haworth Press, Inc. All rights reserved.

10.1300/J136v08n01_12

KEYWORDS. Virtual reference, digital reference services, statistics, evaluation, real-time reference, AskERIC Live!

INTRODUCTION

Successful management of a real-time digital reference service requires effective and efficient data collection to provide accurate and useful information about service processes and outcomes. Any library or AskA service contemplating the launch of or seeking to evaluate a virtual reference service is faced with a number of questions of what statistics are useful to capture, and how to go about capturing these statistics. The following description of the procedures developed by the AskERIC service¹ illustrates some of the challenges and considerations facing managers of virtual reference services and outlines methods for and suggestions for how and why to capture certain information about the service. Finally, future improvements that would support the development and management of quality digital reference services are discussed.

THE AskERIC Live! SERVICE

Since November 1992, AskERIC has provided digital reference services via e-mail to over 250,000 parents, teachers, policymakers, and educators. For several years AskERIC contemplated and planned the addition of a real-time reference service to complement the services and resources currently offered. After a series of pilots, AskERIC Live! was launched in October 2001.² From October 2001 to June 2002, AskERIC Live! conducted over 1,100 real-time reference transactions.

General Service Characteristics

AskERIC Live! assists patrons with questions about educational theory and practice; the majority of patrons of the service (62%) are provided assistance by means of the ERIC database or the resources available on either the AskERIC Web site or another ERIC Clearinghouse's Web site. AskERIC Live! uses 24/7 Reference software. 24/7 Reference is a project of the Metropolitan Cooperative Library System (MCLS), supported by Federal LSTA funding and administered by the California State Library.³ Currently offered three hours each weekday afternoon (1:00 p.m. to 4:00 p.m. Eastern time), the volume for the service averages three sessions per hour during available hours, with a mean weekly volume of approximately 40 sessions per week for the pe-

riod of January to May 2002, a volume that represents approximately 5% of the total volume of digital reference handled by the AskERIC service.

In keeping with the distributed service model of AskERIC's successful e-mail reference service, patrons are assisted by AskERIC Central staff members at the ERIC Clearinghouse on Information and Technology, as well as by librarians and subject specialists at other participating ERIC Clearinghouses around the country. The same AskERIC Central personnel that manage AskERIC's e-mail reference service act as administrators for the real-time service. As a participant in the 24/7 Reference network, AskERIC also has the ability to transfer patrons with queries outside the scope of the AskERIC service to public librarians participating in the 24/7 Reference consortium.

Data Collection: Informal Needs Assessment

In addition to policy development and staff training, an important goal of the initial pilot tests of the AskERIC Live! service was to determine how to gather data about the new service, and then to evaluate the extent to which the automated statistical reports available from the software could gather and report this data accurately and effectively. During these pilots, it quickly became apparent that regardless of which software package was selected for the service, additional procedures would need to be implemented to collect data beyond that automatically recorded and reported by the software. The following section discusses some of the inherent limitations of the automated reporting and describes AskERIC's procedures for the collection of meaningful, useful, and accurate data about the service.

DATA COLLECTION PROCEDURES: A TWO-FOLD APPROACH

The data collection strategy used by AskERIC is based upon three primary service management considerations: ease of collection, reliability, and utility. In order to fully and accurately gather all of the information needed about the service, AskERIC employs two separate instruments: a detailed Web-based transaction log form, and an Excel workbook containing general statistics as recorded by the shift administrator. These two mechanisms serve different purposes, are simple to use, and provide a significant amount of useful and reliable information beyond what the 24/7 software provides. Some of the information recorded is redundantly collected by the software as well (e.g., was the patron's Web browser limited in its ability to fully utilize the software's interactive capabilities?); however, it has proven to be much simpler and less time consuming to actively collect this data using a radio button on the Web form

log rather than to conduct analysis of the transaction logs at a later date. Continuous data collection has some definite advantages in terms of gathering a large sample size of data for analysis as well as avoiding the difficulty of determining a sampling procedure or defining a "typical" time period for a nascent service, and has been advocated for effectively monitoring, managing and improving the quality of reference services.⁴

Instrument #1: Web-Based Log Form

Upon completion of the transaction, the AskERIC staff member uses a Web-based log form to report session-level statistics (see Figure 1). The log form writes this session-level information to an Access database for later synthesis and analysis. To support aggregation of information about the service at various units of analysis, the form includes space for the librarian's name, ERIC clearinghouse affiliation, and date of session (preset by default with today's date). Information is recorded about session length, technical information, whether or not a follow-up was required or problems encountered, and type of assistance provided. The form also provides a structure for content analysis of the session, including the AskERIC controlled vocabulary as well as additional descriptive keywords. Finally, this log form provides free text space so that staff can record any issues or problems encountered during the session. Despite the amount of detail collected about each session, this form takes only a minute or two to fill out and is not considered burdensome by the AskERIC staff. Using a browser that supports auto-fill of fields speeds data entry, as do typical Web form features such as radio buttons, check-boxes and drop-down menus.

Instrument #2: Administrator Spreadsheet

In addition to the session-level information recorded by librarians using the Web-based log form, the administrator for each shift uses a simple Excel spreadsheet for daily recording. The shift administrator keeps a count of sessions per librarian, with sessions transferred to the 24/7 Reference network counted separately from sessions completed by AskERIC staff. Although the software obviously can also provide a count of the number of sessions per librarian, the difficulties with this statistic as recorded by the software will be discussed in the next section. In addition, the administrator notes the number of hours of service in order to provide a context for interpretation of other measures of service volume. It is also helpful to include commentary about any unusual incidents during that shift that may have affected the hours of service, e.g., severe network problems or a shortened shift due to a holiday closing. In

FIGURE 1. AskERIC Live! Log Form

AskERIC Live! Log Form

First Name CH

Last Name Duration (minutes)

Date - -

Mode
 Basic Advanced

Closing Code
 Completed Follow-up Technical Problems Out of Scope

AskERIC Category Details

Category 1

Category 2

Keywords

Type of Assistance Provided

ERIC Search Web Search
 AskERIC Site Explanation
 CH Site Other

Comments

order to balance workload volume from the e-mail reference service with time spent on the virtual reference desk, the administrator records the number of hours spent by staff at other clearinghouses "staffing" the real-time desk. Finally, patrons may access the AskERIC Live! service from any one of five locations on the AskERIC Web site; the referring URL as patrons enter the AskERIC Live! queue is noted to provide a final piece of context for service volume numbers. Because more than one access point is provided, this information offers not only insight into which links are most effective in attracting patrons to the service, but information about the point at which patrons seek assistance in navigating resources on the AskERIC Web site.⁵

CHALLENGES TO ACCURATE DATA COLLECTION

Within the context of managing a virtual reference service, there are issues inherent to the real-time service delivery mode that present significant challenges for accurate data collection. Several statistics are either not captured at all or not measured accurately by the software used by AskERIC; limitations exists in other software products currently in use for real-time reference.⁶ Table 1 offers a side-by-side comparison of the statistics captured by AskERIC using their Web-based session log form and the administrator's Excel spreadsheet, along with an indication of whether or not the software captures that information, and some commentary regarding each measure. Service managers attempting to decide what kind of information to collect may find this information helpful in determining the usefulness of some of these service measures.

Below is an elaboration of a few of the issues and challenges in accurately capturing statistics in an accurate and meaningful way, along with the rationales behind some of the solutions implemented by the AskERIC service to address these difficulties.

Number of Sessions

Many software products used for real-time reference, especially those built from commercial Web contact center software such as 24/7, collect and report statistics regarding the number of real-time sessions per librarian over any specified time period. However, AskERIC has found it to be more useful to collect this statistic separately, because the numbers of sessions reported by the software logs frequently do not reflect the true number of reference transactions. Users may become temporarily disconnected yet quickly reconnect and continue the session; although the software would judge this to represent two separate sessions, it seems clear this situation represents only one refer-

TABLE 1. Statistics Collected

Descriptive Statistic	Collected within software?	Collected by AskERIC?	Explanation
Session length	Service Time or Handle Time	Yes—Web session log form	Problems with accuracy—see discussion below.
Number of reference transactions	Number of Sessions or Calls Handled	Yes—Administrator shift spreadsheet	Problems with accuracy—see discussion below.
Number of patrons transferred to 24/7 network	Calls Handled With Transfer To or Transfer From	Yes—Administrator shift spreadsheet	Ease and simplicity of accurate manual collection
Number of hours of service	No	Yes—Administrator shift spreadsheet	Provides important context for service volume data
Narrative comments about session by librarian	No	Yes—Web session log form	Encourages staff to report problems and actively contribute ideas about service development and quality improvement of service delivery
Patron browser capability ("Basic" or "Advanced" mode)	Yes	Yes—Web session log form	Determines what mode of interaction should be emphasized during librarian training
Topic of patron question according to AskERIC controlled vocabulary	No	Yes—Web session log form	Software does not currently support the use of rich vocabularies for content analysis
Additional topic keywords describing session	No	Yes—Web session log form	Supports better content analysis and information for vocabulary development efforts
Resources used during session	Transcript, including URLs visited	Yes—Web session log form	Checkbox on session log form is simpler than manually mining stored transcripts; Integrated data mining tools could support automated transcript analysis
ERIC Clearinghouse affiliation of AskERIC staff member	No	Yes—Web session log form	Allows for aggregation of statistics by service unit (ERIC Clearinghouse)
Patron wait time before connecting with librarian	Average Queue Time	No	AskERIC can bring on additional staff to handle excess load; wait time is not currently a concern

ence transaction. Recent improvements to the eGain software underlying the 24/7 Reference product have improved the software's ability to accurately count transferred sessions, but the ability to connect the pieces of an interrupted transaction is still lacking.

From a management perspective, the minimal burden of maintaining a

simple tally of the true number of sessions per librarian and overall number of patrons served is less of a concern than the more problematic lack of accurate information about such a fundamental measure of service volume. Measuring the number of transactions in this way ensures that only complete exchanges are counted, and that multiple counts are not recorded for the same transaction.⁷

Session Length

One of the most useful pieces of information about service processes from a management perspective may be session length (defined here as the amount of elapsed time between making the initial connection with the patron and closing the session). As measured by AskERIC staff using the Web-based session log form, the average session (service-wide) lasts 16 minutes. Accurate information about average session lengths, as well as outlier session lengths, supports decision-making about staffing and scheduling, provides information about staff workload, and informs policy development. It would be interesting to learn whether patterns exist in session lengths, and whether they are consistent within and across various digital reference organizations, but these kinds of comparisons can only be made if the data are collected in a consistent and reliable manner.

For a number of reasons, "session length" or "handle time" as reported by the software is not reliable. A variety of ordinary situations create discrepancies between the session length as recorded by the software and the "reality" of session length as experienced by the librarian and patron. As mentioned above, the capability to transfer patrons between librarians and between consortium partners is a valuable feature, but one which has historically caused problems in the accurate measurement of session lengths. For example, the version of the 24/7 Reference software initially used by AskERIC Live! would record a twenty minute transaction for both Librarian A and Librarian B in a scenario where Librarian A spent five minutes negotiating a query with a patron before transferring them to Subject Expert B for a fifteen minute exchange. Recent software improvements include more accurate logging of transferred sessions and have therefore improved the ability of the software to correctly measure the actual time spent with the patron in this kind of scenario. However, additional improvements are still needed in this area. Inaccurate session lengths are frequently recorded when patrons disconnect prior to the mutually agreed upon conclusion of the reference interaction, either due to technical problems in establishing a connection or the familiar virtual reference phenomenon of the "disappearing patron."⁸ Sessions that are not properly closed due to librarian error or technical problems may also result in passively recorded session

lengths that are clearly not accurate. AskERIC has found it more efficient to simply have librarians record the information upon conclusion of a session rather than struggle with cleaning up the data as recorded by the software to filter out or remove the invalid or inaccurate session lengths recorded.

OTHER MANAGEMENT CONSIDERATIONS FOR DATA COLLECTION

Staff Support

To maintain their support for data collection efforts, managers should keep reference staff informed of the planned use of any statistics they are asked to collect.⁹ AskERIC staff members are actively involved in the ongoing development of the log form and the uses of the information collected, and are encouraged to provide feedback about the questions and options included on the log form.

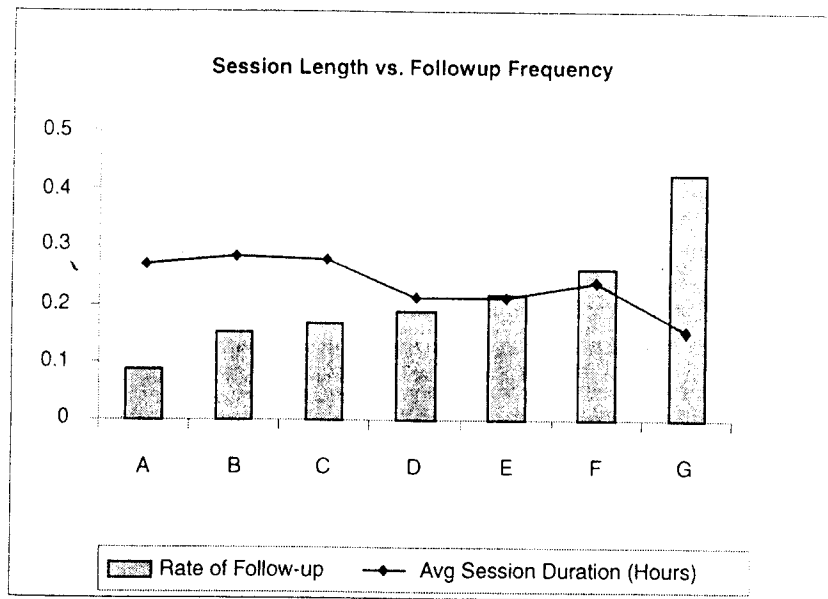
Potential Uses of Collected Data

Methods for assessing and improving digital reference services are beginning to emerge. McClure et al. have created a scheme for measuring and evaluating digital reference services that advocates data collection in up to four areas for use in benchmarking and standards development: outcome measures (quality of answers); process measures (effectiveness and efficiency); economic measures; and user satisfaction.¹⁰ AskERIC's statistical data collection procedures address a significant number of outcome and process measures. In addition, AskERIC's attention to the accuracy of data collected on the amount of staff time spent on the service can also provide the groundwork for future economic analysis of the service. In order to address outcome and user satisfaction measures in service evaluation, AskERIC also utilizes periodic pop-up point of use surveys and employs transcript reviews.

Even without the richer information available in user surveys or reference transcripts, the simple statistics collected by the Web session log form can be potentially useful for policy development or quality improvement efforts. For instance, Figure 2 represents a thought-provoking comparison between average session length and the frequency of the need for follow-up via email after the real-time transaction for seven staff members.

In-depth analysis could verify the existence or strength of a correlation between session length and frequency of follow-up. However, from a service management point of view, even the simple comparison shown here may be useful for quality improvement in a number of ways. Service managers can use

FIGURE 2. Duration of Sessions



this type of information to launch a discussion among personnel of service goals and policies; some services might aim to reduce the need for follow-up, while others might wish to reduce the average session length. Managers might also use this type of information to gain insight into the true amount of staff time typically required to completely assist digital reference patrons. Finally, managers could even consider the comparison between session length and staff time on an individual level to investigate the impact of particular staff development programs (e.g., training on Web searching techniques). Ideally, software and data collection tools would support combining this information with patron satisfaction data obtained from formal surveys or aggregated unsolicited, to combine process-oriented measures with outcome-oriented measures. As services grow and mature, the range and sophistication of their data collection efforts should also develop.¹¹

CONCLUSION

Lankes, Gross, and McClure note that digital reference practitioners and the vendors who serve them must pursue ways to integrate and embed quality

standards and assessment data into the processes, software, and infrastructure of digital reference.¹² The development of processes and procedures for embedding ongoing assessment and evaluation into digital reference services is a challenging yet essential task.

For the virtual reference service manager, reliable and detailed statistics about service performance can be a valuable tool in gaining insight into the data for policy development or improvement of service delivery. The data collection procedures AskERIC has put in place for their real-time service analysis process measures and provide support for economic analysis of the service. In the case of AskERIC as well as other services, software tools must support comparing and cross-referencing service performance data, such as that captured by AskERIC's Web-based session log form, with other richer data available, such as transcripts and user surveys. Digital reference service providers as well as software developers must do more work to create and develop tools to support innovative strategies for collecting reliable and data about digital reference services.

NOTES

1. AskERIC Web site. Available: <<http://askeric.org/>>.
2. AskERIC Live! Available: <<http://askeric.org/QA/userform.shtml>>.
3. 24/7 Reference. Available: <<http://www.247ref.org/>>.
4. R. Aluri, "Improving Reference Service: The Case for Using a Continuous Quality Improvement Method," *RQ* 33, no.2 (1994): 220-236.
5. The AskERIC Web site logs an average of 6 million hits (90,000 unique visitors); to keep service traffic in line with current service capacity, links are distributed among pages which receive relatively moderate rather than heavy traffic.
6. For a further discussion of the functionality and limitations of software proposed for real-time reference, see S. Coffman, "We'll Take It From Here: Further Developments We'd Like to See in Virtual Reference Software," *Information Technology and Libraries* 20, no. 3 (September 2001): 149-53.
7. J. Bertot, C.R. McClure, and J. Ryan. *Statistics and Performance Measurement of Public Library Networked Services*. (Chicago: American Library Association, 2001).
8. Matt Marsteller and Paul Neuhaus, "The Chat Reference Experience at Carnegie Mellon University." (2001). Available: <http://www.contrib.andrew.cmu.edu/~matthewm/ALA_2001_chat.html>. Accessed: July 11, 2002.
9. H. White, "Measurement at the Reference Desk," *Drexel Library Quarterly* 17(1), 3-35.
10. C.R. McClure and R.D. Lankes, "Assessing Quality in Digital Reference Services: A Research Prospectus." (2001). Available: <<http://quartz.syr.edu/q/Overview.htm>>. Accessed: July 15, 2002.

C.R. McClure, R.D.Lankes, M. Gross, and B. Cholco-Devlin, *Statistics, Measures, and Quality Standards for Assessing Digital Reference Library Services: Guide and Procedures*, 2002.

R.D. Lankes, C.R. McClure, and M. Gross, "Cost, Statistics, Measures, and Standards for Digital Reference Services: A Preliminary View," *Library Trends* 51, (2003): Forthcoming.

Exploring the Synchronous Digital Reference Interaction for Query Types, Question Negotiation, and Patron Response

Matthew R. Marsteller
Dianne Mizzy

SUMMARY. The authors explored synchronous digital reference transcripts, using the first full year of *Chat* reference data from Carnegie Mellon University. Synchronous digital reference transcripts are the text of the interchange between the librarian and the patron when they converse using chat software via the Internet. The article is focused on three stages of the synchronous digital reference interaction. The first stage is the query posed by the patron, the second stage is the questioning employed by the librarian in response to the patron's query, and the third stage is the patron's response to the librarian's questioning. Results indicate that reference interviews occurred 64% of the time with an over-

Matthew R. Marsteller (matthewm@andrew.cmu.edu) is Physics and Math Librarian, Engineering and Science Library, Room 4400, Wean Hall, Carnegie Mellon University, Pittsburgh, PA 15213-3890. Dianne Mizzy (dianne@mizzy.com) is Public Service Librarian, Hillman Library, University of Pittsburgh, G-22 Hillman Library, University of Pittsburgh, Pittsburgh, PA 15260.

The authors would like to recognize the invaluable assistance of Elaine Rubenstein of the University of Pittsburgh's Office of Measurement and Evaluation of Teaching. Her work and advice on the statistical analysis was deeply appreciated.

[Haworth co-indexing entry note]: "Exploring the Synchronous Digital Reference Interaction for Query Types, Question Negotiation, and Patron Response." Marsteller, Matthew R., and Dianne Mizzy. Co-published simultaneously in *Internet Reference Services Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 8, No. 1/2, 2003, pp. 149-165; and: *Virtual Reference Services: Issues and Trends* (ed. Stacey Kimmel, and Jennifer Heise) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2003, pp. 149-165. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

<http://www.haworthpress.com/store/product.asp?sku=J136>
© 2003 by The Haworth Press, Inc. All rights reserved.
10.1300/J136v08n01_13