Pomerantz, J., Nicholson, S., Belanger, Y., & Lankes, R. D. (in press). The current state of digital reference: Validation of a general digital reference model through a survey of digital reference services. *Information Processing & Management.*

The Current State of Digital Reference: Validation of a General Digital Reference Model through a Survey of Digital Reference Services

Jeffrey Pomerantz
Ph.D. Candidate
School of Information Studies
Syracuse University
jppomera@syr.edu
(315) 443-4924
fax: (315) 443-5448

Scott Nicholson
Assistant Professor
School of Information Studies
Syracuse University
srnichol@syr.edu
(315) 443-1640
fax: (315) 443-5673

Yvonne Belanger
Project Manager
Information Institute of Syracuse

Syracuse University
yvonne@iis.syr.edu
(315) 443-9703
fax: (315) 443-5448

R. David Lankes
Director
Information Institute of Syracuse
Syracuse University
rdlankes@ericir.syr.edu
(315) 443-3640
fax: (315) 443-5448

Abstract

This paper describes a study conducted to determine the paths digital reference services take through a general process model of asynchronous digital reference. A survey based on the general process model was conducted; each decision point in this model provided the basis for at least one question. Common, uncommon, and wished-for practices are identified, as well as correlations between characteristics of services and the practices employed by those services. Identification of such trends has implications for the development of software tools for digital reference. This study presents a snapshot of the state of the art in digital reference as of late 2001 – early 2002, and validates the general process model of asynchronous digital reference.

Keywords: Digital reference, Virtual reference, Models

Introduction

The traditional face-to-face reference transaction has been an object of study for nearly as long as reference service has been offered in libraries. The model of the traditional reference interview that has developed over time is a structured conversation, and a number of models exist of specific steps within the reference transaction. Since the invention of the World Wide Web, there has been a great increase in the number of reference services utilizing asynchronous electronic communication media to conduct the reference transaction. Several models exist that describe specific steps in this asynchronous reference transaction. Some of these models describe processes similar to processes in the traditional reference interview, while others describe entirely new processes. Few models exist that describe the entire transaction. Despite variations in these conceptualizations of the digital reference transaction, all digital reference services perform many of the same functions when managing electronically-received reference questions.

This study begins from the general process model of asynchronous digital reference presented in Figure 1. This model is derived from Lankes (1998) and the Virtual Reference Desk Project (VRD)'s AskA Software specifications document (Virtual Reference Desk Project, 1998). This model consists of 5 steps:

1. Question Acquisition includes all issues related to the process of obtaining information from a user. This includes not only the user's question, but also question categorization and user identification information, via email, web forms, chat, or embedded applications.
2. Triage is the assignment of a question to a reference or subject expert. This step may be automated or conducted via human decision support. Triage

also includes the filtering out of repeat or out-of-scope questions.

3. Answer Formulation includes all actions taken by the expert to generate a response to a question, including sending the response to a reviewer or directly to the user. Factors for creating "good" answers such as age and cultural appropriateness are included in this step.

4. Tracking is the quantitative and qualitative monitoring of repeat questions for trends. Tracking allows the identification of "hot topics," and may indicate where gaps exist in the collection(s).
5. Resource Creation involves the use of tracking data to build or expand collections to better meet users' information needs.

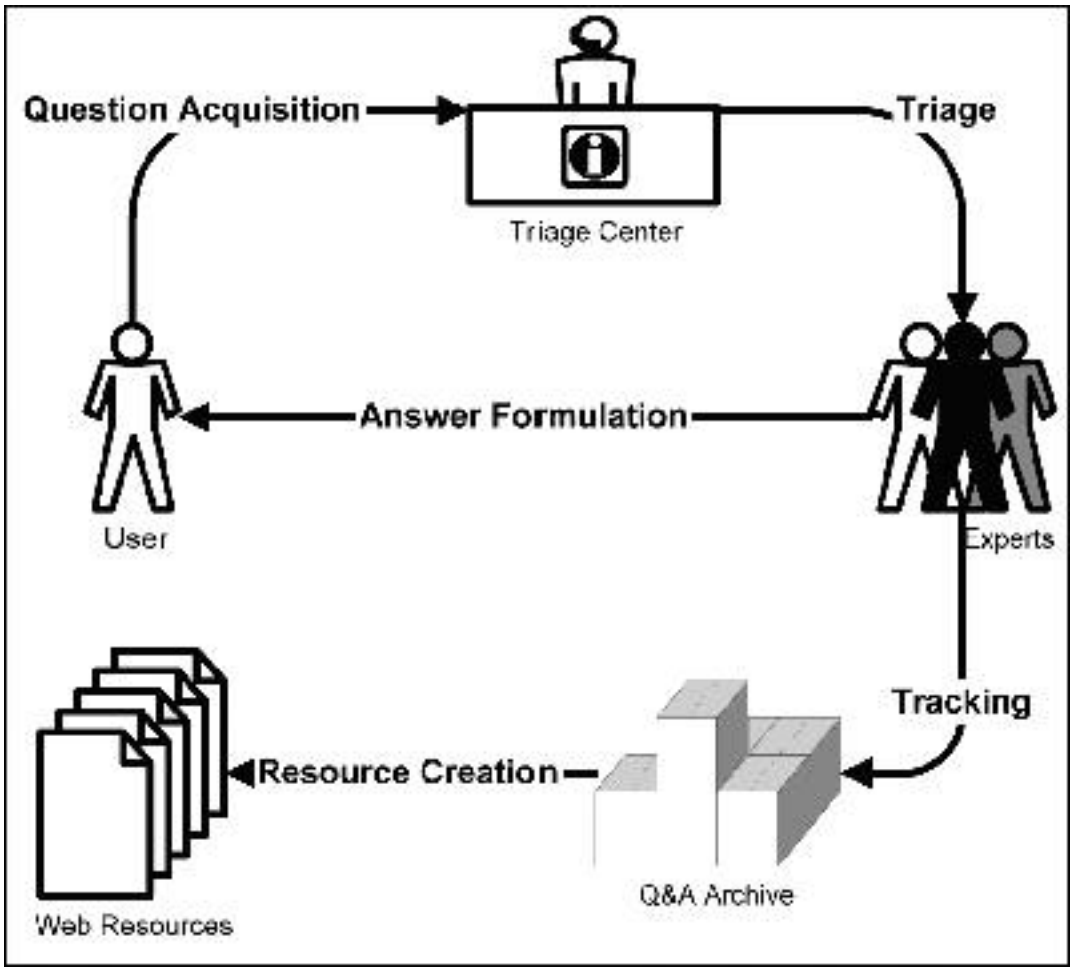


Figure 1 : General Digital Reference Model

This process model is presumed to be generally applicable to all asynchronous digital reference services, though different services employ variations of the processes at each step. Using the triage step as an example, this process may be automated, a human “filterer” may assign questions to experts, or experts may select their own questions from a pool of currently unanswered questions (McClennen and Memmott, 2001). Some services may even skip steps; for example, not all services may archive questions or answers to create resources.

The Virtual Reference Desk Project’s AskA Software specifications document was created as a model to guide the design process for a software application to manage questions received electronically by a digital reference service. The application that was developed under the aegis of the Virtual Reference Desk Project is called the Incubator. During the fall of 2001 the Virtual Reference Desk project received funding under NSF 01-55, the National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) program, to develop a second version of this software that would reflect the current best practices in asynchronous digital reference. The goal of this study was to validate and, if necessary, expand the existing general process model of digital reference as a precursor to creating new specifications for this software application.

Digital reference service will not be defined here, as that has been addressed well and in depth elsewhere (Lankes, 1998; Janes, Carter, and Memmott, 1999). The population of interest for this study was digital reference services of all types: any service utilizing asynchronous electronic communication media to conduct the reference transaction. These services may be affiliated with any sort of a library – public, academic, or special – or may be unaffiliated with any library. Lankes (1998) refers to services of this latter type as “AskA” services, “such as Ask-A-

Scientist" (p.9), since most services of this type specialize in a particular subject: for example, art (Ask Joan of Art), education (AskERIC), mathematics (Ask Dr. Math), oceanography (Ask Shamu), etc.

This paper describes a study conducted to determine the paths digital reference services take through the process model. This study sought first to describe common and uncommon practices in digital reference; second, this study sought correlations between characteristics of services and the practices employed by those services. Identification of such trends has implications for the development of software tools for digital reference: the most common practices must be supported, while the least common practices may be dropped from software specifications when compromises need to be made. Additionally, patterns between types of services and practices employed by those types of services enable customization packages relevant to services' specific needs and uses. Finally, any steps or processes suggested during the course of the study not reflected in the general process model have implications not only in software development but also for revisions to the basic process model.

The research questions for this study are:

1. What processes do digital reference services employ when managing electronically received reference questions?
2. What are the most and least commonly employed processes?
3. Which, if any, processes typically occur in combination?
4. Which, if any, processes or sets of processes are typically employed by which types of digital reference services?

Literature on Question Management in Reference Services

Few models exist that encompass the entire reference transaction as it takes place at a reference desk in a physical library. However, many models exist of specific steps in the reference transaction, or present the reference transaction from the perspective of a specific stakeholder. This section will not present a comprehensive review of these models. Superb comprehensive reviews have been undertaken in greater length than is possible here, by Richardson (1995), Katz (1997), and Bopp and Smith (2001). Instead, this review will present only those models that most directly influenced the conceptualization of this study.

One of the best known models is Taylor's (1968) model of "prenegotiation decisions by the inquirer" (p.181). This model begins at the point at which an individual realizes that he or she has "a certain incompleteness in his picture of the world" (p.180), perhaps even before he or she has recognized this incompleteness as an information need *per se*. This model then proceeds through the individual's decision-making process of *whether* to ask for information (as opposed to searching for information him- or herself), *who* to ask for information (ask a colleague or go to the library), and ultimately *how* to search for information in the library (search him- or herself or ask a librarian for assistance). Taylor's model is relevant to this study in that it presents the various steps in the reference transaction as points at which decisions must be made. The decisions in Taylor's model are made by the patron; by contrast, the decisions in this study are those made by the reference service.

Another model that presents the reference transaction as a series of decision points was developed by Robinson (1989). One of the few that encompasses the entire reference transaction, Robinson's model approaches the transaction from the perspective of cost analysis of a reference transaction, specifically matching

"the level of resources to the level of service" (p.46). Beginning at initiation of the reference interview, this model includes estimation of the difficulty of the question and types of resources required to answer it, the actual use made of those resources in finding an answer, the delivery of that answer to the patron, and the evaluation of the service provided. Robinson's model comes closest to a general process model, as it begins with the acquisition of a question by a reference service, and ends with the provision of an answer to the patron. While Robinson's model spans the entire reference transaction, it differs from this study in that it is specifically concerned with the reference transaction as it takes place at a reference desk.

An early model of the reference transaction as it takes place online was developed by the Internet Public Library, and was presented in a simplified form by Michael McClennen at the 2001 Virtual Reference Desk conference (McClennen, 2001). This model depicts the process of an online reference transaction using a flow chart that shows the states in which it is possible for an electronically-submitted reference question to be (e.g., accepted, claimed, overdue, answered, etc.), and the processes that move a question from one state to another (e.g., accept, claim, due date passed, answer, etc.). Another similarly algorithmic depiction of online reference transactions can be found in MathNerds' flowcharts of their algorithms for assigning problems (MathNerds, 2001). These flowcharts present the process from both the patron and the expert's points of view. A third model along these lines is Kresh's (2000) conceptual model of the Collaborative Digital Reference Service (CDRS). Rather than being a flowchart representing states and state changes, this model presents the entire workflow involved in the answering of digital reference questions from the point of view of the service itself, rather than from the more limited perspective of patron or librarian.

The similarity of all of these models (including the general process model of digital reference presented in Figure 1) is an indication that models of the process from a number of different perspectives (service, expert, user, etc.) are remarkably consistent across services. The primary purpose of the general digital reference process model is to carry over all crucial elements of the desk reference transaction to the digital environment. The secondary purpose of this model is to suggest steps in the transaction that may be modified or customized to maximize the potential and capabilities of the digital medium.

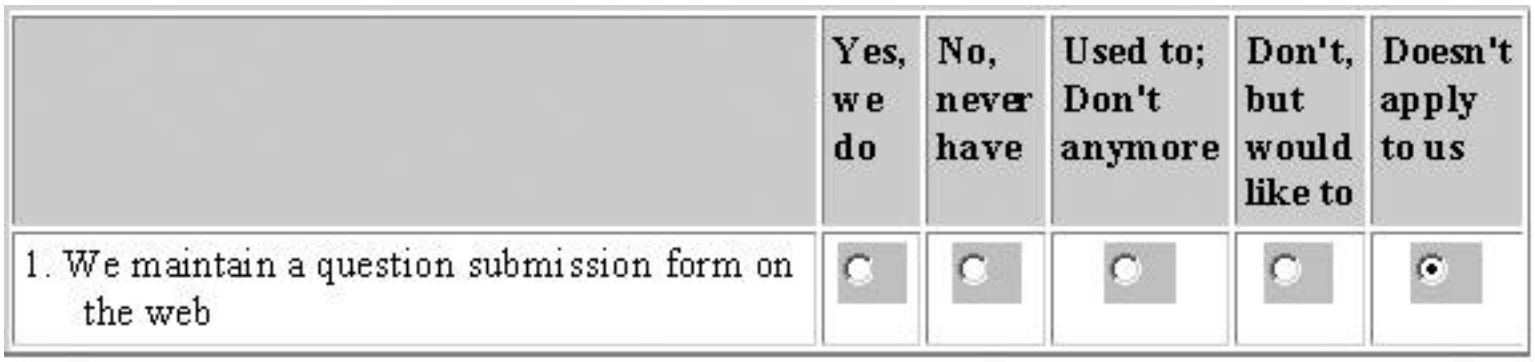
Methodology

In order to determine the processes that digital reference services employ in managing electronically received reference questions, a survey was created based on the general digital reference model (a copy of the survey instrument is available on request from the authors). Each decision point in this general model provided the basis for at least one item on the survey; this breakdown allowed each point in the model to be validated. For each item, respondents were asked if their service performs the process described. The mutually exclusive options provided for each process were:

1. The service performs the process,
2. The service does not perform the process and never has,
3. The service used to perform the process but no longer does,
4. The service does not perform the process but would like to or plans to, or
5. Not applicable.

For example, in the Question Acquisition step of the model, the first process is

the input of questions via a question-submission web form. The item “We maintain a question submission form on the web” was created based on this process. Figure 2 shows this item, as a typical example of the items in the survey.



	Yes, we do	No, never have	Used to; Don't anymore	Don't, but would like to	Doesn't apply to us
1. We maintain a question submission form on the web	○	○	○	○	◉

Figure 2 : A sample item from the survey

Some decision points involve processes that may be performed in a variety of ways. For items based on these decision points, respondents were asked to specify how their service performed the process specified. Figure 3 presents two examples of such items.

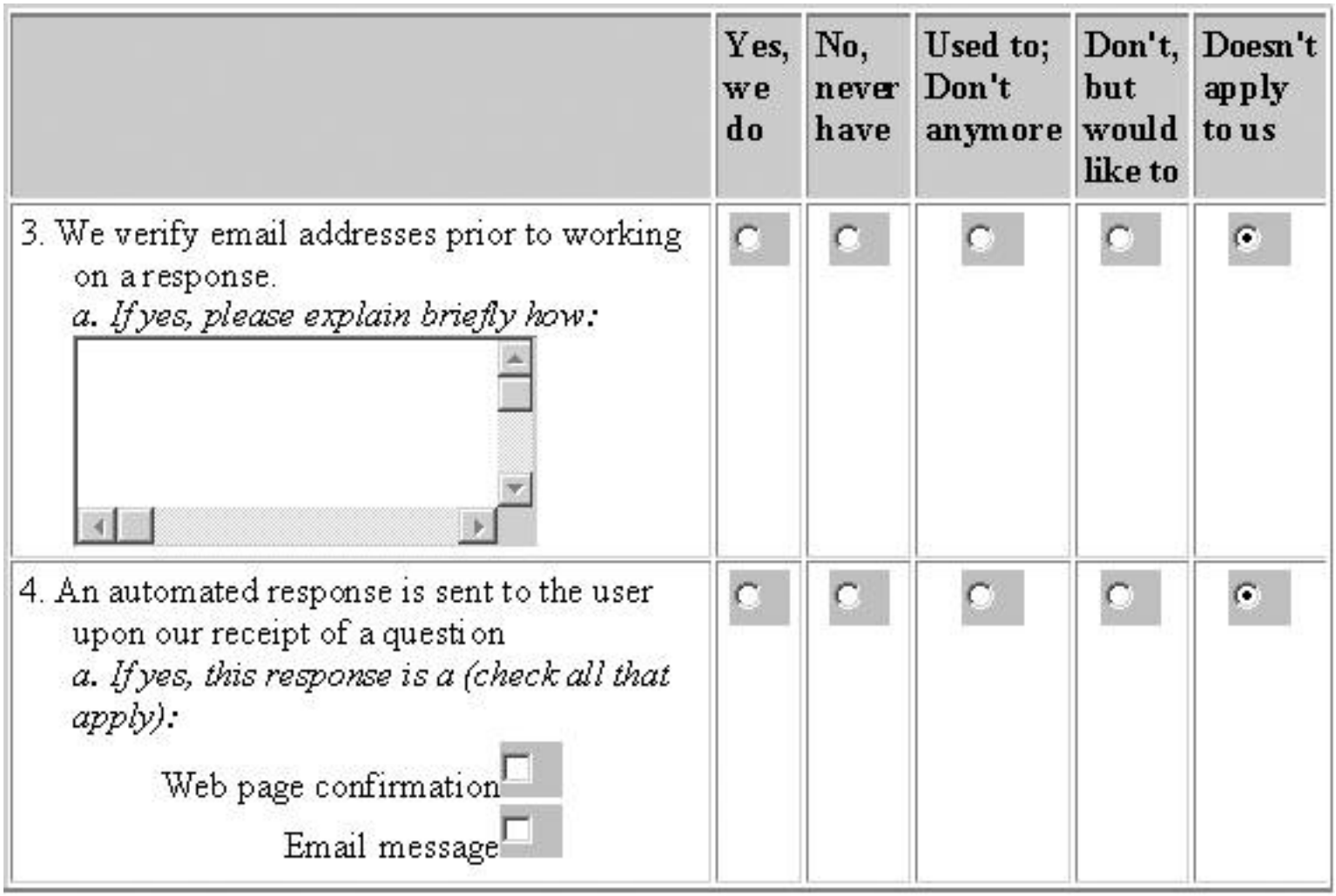


Figure 3 : Two survey items concerning processes that may be performed in a variety of ways

Respondents were solicited through a flyer given to all registrants at the 2001 Virtual Reference Desk conference, as well as through a posting to the DIG_REF listserv. The Virtual Reference Desk conference is currently the only conference in dedicated to the theory and practice of digital reference, and is a forum for the presentation of the state of the art in the field, in services of all types. The DIG_REF listserv is “a forum to help set an agenda for redefining reference services in the Internet context” (http://www.vrd.org/Dig_Ref/dig_ref.shtml). Thus these two venues are the voice of the field of digital reference as it exists today. Registration at the 2001 VRD conference was 430, and at the time of the posting, there were 2,114 individuals subscribed to the DIG_REF listserv. There

is some overlap between attendees at the VRD conference and subscribers to DIG_REF. There were also some digital reference services represented at the VRD conference by more than one individual, and it is likely that the same is true for DIG_REF. Between these two venues, however, the majority of the digital reference services in North America, and several from overseas, were solicited for this survey.

The researchers received 49 responses to the survey. Of these responses, two were eliminated, as one was a duplicate and one contained no data. Thus the pool of data analyzed for this study consisted of 47 responses.

Data Analysis

These 47 responses came from 47 different digital reference services; respondents were asked to provide the URL of their service, allowing the researchers to be certain that there were not multiple responses from a single service. Thus each of the 47 responses represents a service. Like any survey for which participation is solicited “in public,” as it were, these 47 respondents are self-selected.

Additionally, the researchers had no control over the types digital reference services that responded to the survey. It is unclear how many digital reference services exist. The Virtual Reference Desk Project maintains a list of AskA services called the “AskA+ Locator” (http://www.vrd.org/locator/subject.shtml), which, as of this writing, contains over one hundred services. Bernie Sloan maintains on his personal website a list of over 90 email-based reference services offered by public and academic libraries (http://www.lis.uiuc.edu/~b-sloan/e-mail.html). It is important to note that neither of these lists claims to be

comprehensive, and it is therefore impossible to know how many services are *not* listed. By extension, it is impossible to know what percentage or segment of the total population of interest to this study is made up by the 47 responding services. As a result, these responding services may not be representative of all existing digital reference services. These facts limit the generalizability of this study. This limitation is discussed in the Discussion section of this paper.

Respondents were asked to reply to the survey items with the entire service in mind. The level of analysis for this study is therefore the service, and not any individual or role within the service. The respondent services fall into the following categories:

- 47% academic libraries
- 21% public libraries
- 21% AskA services
- 11% unidentified

The respondent services span the spectrum of possible types of digital reference services (with the exception of special libraries). This study therefore presents a snapshot of the processes employed by the responding digital reference services during the period of time that this survey was administered from mid-November 2001 – January 2002. While it is not possible to know if any of the responding services instituted any changes in their practices or policies during that brief span of time, the researchers were careful to schedule the administration of the survey so that it did not correspond with a product release for any of the several commercially available software applications designed to enable web-based digital reference, or any of the applications created by various digital reference services (Lagace and McClennen, 1998).

One-Item Analyses

The first step in analyzing the data from the survey was to determine common and uncommon practices – that is, what practices in managing the digital reference process were employed or not employed by a majority of the services surveyed.

Some of the most widely employed practices are:

- Sending an automated response to the user by email, upon receipt of a question (96%)
- Asking the user for an email address on a question submission web form (90%)
- Maintaining a question submission web form (83%)
- Responding to questions by email (80%)

These findings are consistent with the findings a number of other studies. Janes, Carter, and Memmott (1999) found that 65%, and White (2001) found that 71.4% of digital reference services affiliated with academic libraries elicited questions via a web form. Goetsch, Sowers, and Todd (1999) found that 78% of ARL libraries elicited questions via a web form. Janes, Hill, and Rolfe (2001) found that 60% of AskA services, both commercial and non-commercial, elicited questions via a web form. These figures vary, which may be an artifact of the different respondent pools surveyed in these five studies. However, a trend is clear: web forms are being used overwhelmingly often as the user interface for submission of questions to digital reference services. This trend is not confined to the world of digital reference either. According to the Congress Online Project, members of Congress are moving away from email addresses and towards web-

based interfaces for receiving communications from constituents: 66 out of 100 Senators and 226 out of 440 House Members "are not using public e-mail addresses, and are directing constituents to their Web sites to send messages" (2002, MEMBERS ARE TURNING OFF E-MAIL ADDRESSES AND TURNING ON WEB FORMS section, ¶ 1).

This study's finding that 90% of services that maintain a question submission web form ask the user for an email address on that web form is also consistent with the findings of other studies. Janes, Hill, and Rolfe (2001) found that 85% of AskA services require an email address for submission of a question, and White (2001) found that 100% of digital reference services affiliated with academic libraries that maintain a web form for question submission, ask for the patron's email address on the web form. The requirement of an email address on a web form is consistent with the finding that 80% of services respond to questions by email.

Some of the practices employed by a minority of services are:

- Responses are not reviewed for quality and/or accuracy before being sent to or posted for users (19%)
- Patrons cannot pick up their responses on the web (as opposed to being sent an email containing the text of the response) (23%)
- If questions are assigned to particular experts, these individuals do not have the ability to reject a question (so that it must be reassigned to a different expert) (37%)

The low percentage of services reviewing responses for quality and/or accuracy is consistent with the preliminary findings of the Assessing Quality in Digital

Reference Services project, reported on by Gross, McClure, and Lankes (forthcoming). Preliminary findings of this project indicate that many digital services are developed "without plans for evaluation." As a result, there is a lack of data not only on the quality of responses provided by digital reference services, but also on the value of performing such evaluation at all.

The low percentage of services that allow patrons to pick up their responses on the web stands in stark contrast to the number of services that elicit questions via a web form. While digital reference services may elicit questions on the web, this study found, as mentioned above, that the overwhelming majority respond to questions by email. This finding may indicate that the web forms used for question submission are generating emails that are sent to the service's email inbox. This supposition is supported by Goetsch, Sowers, and Todd (1999), who found an even more dramatic contrast between media of submission and pick-up of questions: they found that 97% of ARL libraries that elicit questions via a webform deliver the question by email.

Interestingly, this trend is not confined to the world of digital reference either. According to the Congress Online Project, "an estimated 25% of House offices now answer e-mail with e-mail" (2002, MORE OFFICES ANSWER E-MAIL WITH E-MAIL -- WHAT ARE YOU WAITING FOR? section, ¶ 1) – indicating that the web forms used by constituents are generating emails that are sent to the House Member's offices. Worse, "most offices continue to treat e-mail like postal mail, replying with stamped letters rather than e-mail" (Goldschmidt et al, 2002, INTRODUCTION section, ¶ 4). It seems that, in the adoption of these new online technologies, the trend is to adopt the new technology for the public face of the service (submission of questions) first, and only later to adopt the technology for

use by the service itself (answering of questions).

This discrepancy between the number of services that utilize the web for question submission and answer pick-up speaks to the small number of services using web-based interfaces for managing the entire digital reference process. As mentioned above, there are several commercially available software applications designed to enable web-based digital reference, as well as others that have been developed by various digital reference services. Two of these applications developed by digital reference services are the Incubator, developed under the aegis of the Virtual Reference Desk Project, and QRC, developed by the Internet Public Library (IPL) (Lagace and McClennen, 1998). As of this writing, the Incubator is used by 6 services, including the VRD itself, and QRC is used by 5 services, including the IPL (Michael McClennen, personal communication). One of the commercially available software applications, LiveAssistance, is used by "about 25" services (Sarkar, 2002).

Many of the practices employed by services were at one extreme or the other – an overwhelming majority of services either did or did not employ certain practices. On the other hand, the distribution of "wish list" practices – those practices that a number of services do not employ but plan to or wish they could – did not provide a clear result, indicating that services were divided over what practices were desirable future goals. Some of these "wish list" practices are:

- When a question is received, a knowledge base of previously answered questions is automatically searched (35% wish that their service has this capability)
- The progress or state of a question is automatically tracked (e.g., new, assigned but not answered, in progress, overdue, etc.) (25%)

- Previously answered questions are stored in a knowledge base (15%)

The second and third of these "wish list" practices – tracking the state of a question and storing previously-answered questions in a knowledgebase – are functions that may be implemented in many web-based digital reference applications. Indeed, "track[ing] the progress of individual questions" was one of the problems that led to the development of QRC in the first place (Lagace and McClennen, 1998).

The first "wish list" practice – automatically searching a knowledgebase of previously-answered questions – is a function for which there seems to be a great deal of desire in the digital reference community, but which is as yet in the early stages of its adoption in this community. Information Retrieval (IR) systems match queries with documents – in digital reference, however, some or all of those documents may be previously-answered questions. Some digital reference services – such as the MadSci Network (http://www.madsci.org) and Ask Dr. Math (http://mathforum.org/dr.math) – maintain public, searchable archives, in which previously-answered questions are returned as search results. Bry (2000) explains that when the user submits a question to MadSci, a CGI script searches the archive for potential answers. Bry states that "approximately 63 percent of questions are matched with archived files" – however, "only 25 percent of users deem their questions answered by this process (15 percent of all submitted questions)" (p.118). Perhaps it is because only a quarter of the questions submitted to a digital reference may be adequately answered automatically, that only 6% of services currently employ automation to answer questions. Nevertheless, there is clearly a demand in the digital reference community for development of a reliable IR or automated question-answering system. Indeed, at

the time of this writing, the authors are working on a study, funded under the NSF program 02-054 (National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL)), that seeks to discover: 1) what types of questions may be answered automatically and what types require human intermediation, and as a corollary, 2) when a question is sufficiently different than any previously-asked question that it cannot be answered with an archived response.

These trends come from one-variable analyses of the data, as each is the result from a single item on the survey. The next analysis that the researchers conducted was to combine two and three variables, to discover combinations of practices employed by services. Given that there were nineteen items on the survey, some with multiple parts, there are a great number of possible permutations of practices that could be presented here. All of these will not be presented. Only the combinations that are particularly interesting will be presented, either because their results are unexpected, or because they are illustrative of the state of the art of digital reference service, or because they support the findings of other studies.

Multiple-Item Analyses

This discussion begins with the receipt of a question by a service. Of the 45 services that ask for an email address on their webform, 13 (29%) verify that the email address is valid prior to working on a response. 25 (56%) do not verify the email address and never have. 5 (11%) wish that they had this functionality. (Services that receive questions only via email can reasonably assume that the email address from which the question came is valid.)

Of the 17 services that automatically generate a response when a question is received, 8 (47%) generate that response in the form of an email message, 4 (23.5%) generate that response as a webpage only, and another 4 (23.5%) generate that response as both an email message and as a webpage. This automated response is not necessarily an answer to the user's question (indeed, only 3 (6%) of all services responded that when a question is received, a knowledge base of previously answered questions is automatically searched). The survey did not ask what the automated response is, if it is not an answer to the user's question. However, the authors are familiar with several digital reference services that automatically generate some form of an acknowledgement of receipt of the question.

Once an expert formulates an answer to a user's question, there are two ways that the user can receive that answer: the service can send it as an email to the user, or the user can come to the service and "pick up" the answer on the web. Of all responding services, 43 (91%) responded that they send the full text of the answer in an email, while the remaining 4 (9%) send a "pickup" notice for a response posted on the web. Interestingly, 6 (13%) of services responded that they have the ability for their patrons to pick up their responses on the web – it seems that some services put the burden on the user to return to the service to check if their question has been answered. Another 6 (13%) of services responded that they wish they had the ability for their patrons to pick up their responses on the web. As mentioned above, there are several commercially available software applications designed to enable web-based digital reference, as well as others that have been developed by various digital reference services. It would seem that, given the apparent desire for such applications, this is a technology which is as yet in the early stages of its adoption in the digital reference community.

As stated above, only 3 (6%) of services automatically search a knowledge base of previously-answered questions when a new question is received. Interestingly, however, a far greater percentage of services 20 (42%) responded that they store previously-answered questions in a knowledge base. As discussed above, these findings reflect the practice of such digital reference services as the MadSci Network and Ask Dr. Math, which have their archives of previously answered questions publicly available on their website, thus making this archive a resource for their users. Other services may maintain an archive of previously answered questions only for the use of their experts. There may also be other uses made by services of their question-and-answer archive.

At a reference desk it is possible for the librarian to clarify the patron's information need by engaging the patron in a reference interview. By contrast, in the practice of digital reference, the initial question along with any information gathered at the time of that initial question submission is typically all that the librarian has. Digital reference services have found that asynchronous media do not lend themselves well to question negotiation: Carter and Janes (2000) report that if an expert replies to a user's question with a request for clarification, 30% of users do not ever reply with that clarification. Judging by the authors' conversations with digital reference experts at a number of other services, Carter and Janes' finding is a remarkably small percentage. The ability to determine if an incoming question is a "follow-up" to a previous question is therefore an important function in digital reference triage. However, only about one-third (actually 17 (36%) of services responded that they have the ability to determine if an incoming message is a follow-up. Of these services that can detect follow-ups, 13 (77%) assign follow-up questions to the individual who responded to the

original question. This survey question had a second part that asked how the service determined if a question is a follow-up: only 12 (25%) of services indicated that this determination was performed by an automated process; the remaining 35 (75%) indicated some form of human intervention.

Another important function in digital reference triage is ensuring that a question reaches the expert who is best suited to answer it. This assignment may be performed by the service, or the service may allow experts to select questions themselves (as in services that store questions in a "triage area" (Lankes, 1998, p.137)). 28 (60%) of services responded that they assign questions to particular experts, while 13 (28%) responded that they do not and never have assigned questions. Of the 28 services that assign questions:

- 20 (71%) employ a human to perform this assignment,
- 4 (14%) employ an automated process, and
- 1 (4%) employ both a human and an automated process.

Additionally, of the services that assign questions:

- 11 (39%) assign questions on the basis of the subject expertise,
- 8 (29%) assign questions on the basis of both subject expertise and the load of questions received that day, and
- 1 (4%) assign questions on the basis of both subject expertise and the geographical location of the expert or the patron.

These findings support the results of a Delphi study conducted by Pomerantz, Nicholson, and Lankes (forthcoming) to determine factors that affect the process of sorting and assigning reference questions received electronically by digital reference services, both to experts within the service and between services.

Pomerantz, Nicholson, and Lankes discovered that there are 15 factors that are important in this decision-making, the top three of which are: (1) Subject area of the question, (2) The service's area(s) of subject expertise, and (3) The expert's area of subject expertise.

Cluster Analysis

In order to better understand the responses to this survey, data mining tools were employed to analyze the findings. While the results of many of the survey questions are clearly interpretable, some results are ambiguous. However, the researchers believed that there were groups of service types that would provide guidance in interpreting these ambiguous results. Thus, data mining was used to cluster the respondents by looking for common patterns of responses. Data mining is unlike traditional statistics in that one begins with the quantitative creation of several possible solutions, and then either testing or qualitative deduction is used to select the most appropriate model for the situation.

Clementine, a data mining tool published by SPSS, was used for the clustering. The K-Means method was selected, which allows the user to specify a number of clusters and the program will create the best combination of respondents to meet that requirement. This method works in a multi-dimensional space, with one dimension assigned to each answer – thus creating a polythetic clustering scheme, in which clusters are formed based on multiple characteristics. K-means is an agglomerative method of cluster analysis: it begins by arbitrarily choosing one centroid per cluster; a centroid is a point in the multidimensional space that represents a combination of answers to all the questions. It then examines a record, assigns the record to a cluster, and then adjusts the values of the centroid

to create as much distance as possible between the clusters. The next record is assigned to a cluster and the centroids are again adjusted. This process is repeated until further adjustment of the centroids does not improve the distinction between clusters.

The centroid is the mathematical center of a cluster, representing the collection of all respondents in that cluster; the entities in that cluster may have slightly different attributes than the centroid. Just as the average height of a group of people may not be the height of any one person in the group, the centroid may not have the same set of attributes as any of the respondents in the cluster. This is similar to the notion of “fuzzy sets,” proposed by Zadeh (1965). Viewed from the perspective of classification, the “centroid” of a fuzzy set is the “prototype,” a hypothetical entity that possesses all of the “perceived attributes” (Rosch, 1978, p.35) of that category.

Clementine was used to create several arrangements of the records using different numbers of clusters. The research team then discussed what information was produced by using each arrangement, and determined that three clusters produced an arrangement that was the most logical and provided a distinction without being too finely-grained. Interestingly, these three clusters are approximately the same size, containing 12 (26%), 16 (34%), and 19 (40%) services. In performing this data analysis, survey responses were grouped together, as follows:

- The service performs the process, and The service does not perform the process but would like to or plans to,
- The service does not perform the process and never has, and The service used to perform the process but no longer does, and
- Not applicable or the question left blank.

In order to more easily discuss and understand the clusters, the researchers named them: the “High Tech/Low Touch” group employs the most automation and the least human intermediation, the “Low Tech/High Touch” group employs the most human intermediation and the least automation, and the “High Tech/High Touch” group employs a balance of both. Table 1 shows these three groups of services, and the percentage of each that responded positively (the service does or plans to perform the process) to questions having to do with the service’s use of automation.

Table 1: The three groups of services and key survey responses

Process	High Tech/ Low Touch (n = 16)	Low Tech/ High Touch (n = 19)	High Tech/ High Touch (n = 12)
Maintains a webform for question submission.	94%	89%	83%
Verifies email addresses prior to working on a response.	63%	11%	67%
Automatically generates a response to the question.	69%	16%	92%
Has the ability to detect follow-up questions.	81%	16%	67%
Automatically sorts questions to experts.	19%	5%	8%
Stores question-answer sets in a knowledge base.	75%	16%	100%
Automatically searches a knowledge base when a question is received.	88%	0%	50%

Patrons can pick up their responses on the web.	19%	21%	42%
Automatically tracks the progress or state of a question.	50%	26%	67%

The “High Tech/Low Touch” group (34% of the total number of services) relies heavily on automation throughout the process of managing questions. This group is composed primarily of AskA services, and some academic and public libraries. What unites the services in this group is that they all utilize a high-tech approach to providing digital reference service. The AskA services in this group are among the highest-volume digital reference services in existence (e.g., the VRD network and AskERIC), and several of the academic and public libraries are among those ARL members with the greatest number of reference queries reported during the past decade (http://fisher.lib.virginia.edu/arl/). It makes sense that such high-volume services would have a high-tech approach to providing digital reference: a service such as AskERIC (http://www.askeric.org) that receives an average of 700 questions per week would require an unfeasible amount of manual labor to process all of those questions. Such services have, by necessity, had to develop methods for automating as many processes as can be automated.

The “Low Tech/High Touch” group (40% of services) relies heavily on human intermediation to handle questions throughout the entire digital reference process. This group is composed of small- to medium-sized academic and public libraries. Another study currently underway by one of the authors is finding that digital reference services at many academic and public libraries receive a low volume of questions, and that these questions are often answered by only one librarian (Pomerantz, in preparation). As a result of having such low volume, these

services may not have had the need to automate many of their processes; manual handling of questions has likely been sufficient.

The “High Tech/High Touch” group (26% of services) employs automation for some steps and human intermediation for some steps of the digital reference process. Which steps are automated and which human-intermediated is different for each service. This group, like the “High Tech/Low Touch” group, is composed primarily of AskA services, and some academic and public libraries. These services span the range from high to low volume, but are bound together by their selective use of and balance between automation and human intermediation to meet the unique requirements of the service.

Discussion

Due to the researchers’ lack of control over the respondent group, as well as the small size of the sample, the researchers claim only a limited degree of generalizability of these findings. As stated above, it is unclear how many digital reference services exist; it is therefore impossible to estimate the percentage or segment of those services represented by this study’s sample. It is, by extension, also impossible to estimate the number of digital reference services that are affiliated with public or academic libraries, or that are AskA services, unaffiliated with any library at all. Without this data, it is impossible to make any claims about the representativeness of this study’s sample. All that this paper claims is that it presents a snapshot of the current state of the art in digital reference, as performed by a subset of services from the venues that represent the field of digital reference as it exists today.

Previous analyses of digital reference service practices have concentrated on the type of the service (academic, public, AskA) (for example, Janes, Carter, and Memmott, 1999; Garnsey and Powell, 2000; Janes, Hill, and Rolfe, 2001). Perhaps the most significant finding of this study has been the utility of a more complex grouping scheme based upon functional, rather than organizational, characteristics. The groups described here cluster services according to their use of automation in the process of providing asynchronous digital reference. This finding speaks to the current state of the art in digital reference: while the practices employed by digital reference services may be more or less universal, the means by which these practices are achieved appear to be "evolving" along three distinct paths.

These three paths do not, however, seem to be leading to three entirely unrelated species of digital reference services. Rather, what these results show is that digital reference services tend to cluster around three points on a spectrum of technology use, ranging from highly automated to entirely human-intermediated services.

These three types of digital reference services appear to be the result of services adopting "packages" of technologies, practices, and policies: once a specific technology or practice is adopted, other related technologies and practices are dictated or precluded by that choice. For example, services that maintain a web form for question submission are more likely to automatically generate a response to submitted questions. Services that send answers to users via email do not have a pressing need for web-based answer-pickup functionality. Services that assign questions to experts require criteria by which this assignment is performed (whether that assignment is performed manually or automatically). Services that

store question-answer sets in a knowledge base have the capability to search that knowledge base for previously-answered questions.

Additionally, different services have different institutional policies, and these policies may dictate or preclude the use of certain technologies and practices. For example, as a matter of institutional policy, a digital reference service may be required to respond to every question received within a certain number of business days; the shorter this timeframe, the greater the need for automation. On the other hand, it may be the policy of a digital reference service to allow experts to self-select questions; in this case, the service may automate the filtering out of repeat or out-of-scope questions, but there is no need to automate the assignment of a question to an expert.

In all three types of services, however, there are certain practices that have been adopted by a majority of digital reference services. Similarly, there are certain practices that have not been adopted by a majority of services – though time will tell whether those practices will never be adopted or will in time be adopted, technology permitting. The practices that have been widely adopted are those that are common on the Web as a whole: using email and maintaining web forms. Interestingly, the practices that have not been widely adopted are those that would reduce the amount of control experts have over the selection and answering of questions: assignment of questions to experts and quality review of answers.

Conclusion

This study allowed for the possibility that the general process model presented above is not complete. By having a number of open-ended questions, as well as a

question that asked for “additional comments,” the survey allowed respondents to suggest steps and processes not reflected in the model. No additional steps or processes were suggested by the respondents, however. This fact strongly indicates that the general process model is complete for the current state of the art in asynchronous digital reference, i.e., email- and web-based digital reference only.

An obvious omission of the general process model is the fact that it does not take into account the processes involved in “live” or “real time” reference, e.g., chat environments, instant messaging, and graphical co-browsing. These technologies had not yet widely impacted the practice of digital reference when the general process model was developed. The past few years have seen rapid development in technology supporting collaborative, synchronous reference service, and rapid acceptance of this technology in reference services of all types. The revision of the general process model to include “real time” reference will be a useful future direction for research.

Another useful future direction for research will be to redo this survey in one or two years’ time. This study captures a snapshot of the state of the art in digital reference at the time that this survey was administered. While digital reference is well established as a service, the technology employed by digital reference is still developing rapidly. As such, it is reasonable to assume that the picture presented in this study is likely to change in the future. While the general model of digital reference will continue to hold true, the types of services that exist, and the distribution of services employing different processes at different steps will change. It would therefore be useful to perform a longitudinal survey of the changing state of the art in digital reference. Alternatively, a study similar to this

one could be performed to focus in depth on one particular type of service (public, academic, or AskA, or High Tech/Low Touch, Low Tech/High Touch, or High Tech/High Touch, or some other subdivision), and as such survey the state of the art for one specific segment of the community of digital reference services.

This study has done two things: First, it has presented a snapshot of the state of the art in digital reference as of late 2001 – early 2002, which is of interest for historical reasons. Second, this study has tested one of the few general models of digital reference, shown it to be valid, and provided details about the practices employed in the different steps. This model is generally applicable to all digital reference services, though different services employ different practices at different steps. This model can serve as a guide for future research on digital reference, as well as future development of digital reference services.

Bibliography

Bopp, R. E., & Smith, L. C. (2001). *Reference and Information Services: An Introduction*. Englewood, CO: Libraries Unlimited, Inc.

Bry, L. (2000). Simple and Sophisticated Methods for Processing Large Volumes of Question and Answer Information through the World Wide Web. In R. D. Lankes & J. W. I. Collins & A. S. Kasowitz (Eds.), *Digital Reference Service in the New Millennium: Planning, Management, and Evaluation* (Vol. 6, pp. 111-123). New York: Neal-Schuman Publishers Inc.

Carter, D. S., & Janes, J. (2000). Unobtrusive Data Analysis of Digital Reference Questions and Service at the Internet Public Library: An Exploratory Study. *Library Trends, 49*(2), 251-265.

Congress Online Project. (2002). *Congress Online: Special Report: E-mail Overload In Congress - Update*. Washington, DC: The Pew Charitable Trusts, the George Washington University, and the Congressional Management Foundation. Retrieved 3 September 2002, from Congress Online Project Web site: http://www.congressonlineproject.org/080702.html.

Garnsey, B. A., & Powell, R. R. (2000). Electronic Mail Reference Services in the Public Library. *Reference & User Services Quarterly*, *39*(3), 245-254.

Goetsch, L., Sowers, L., & Todd, C. (1999). *Electronic Reference Service* (SPEC Kit 251). Washington, DC: Association of Research Libraries.

Goldschmidt, K., Folk, N., Callahan, M., & Shapiro, R. (2002). *E-mail Overload in Congress: Managing a Communications Crisis*. Washington, DC: The Pew Charitable Trusts, the George Washington University, and the Congressional Management Foundation. Retrieved 3 September 2002, from Congress Online Project Web site: http://www.congressonlineproject.org/email.

Janes, J., Carter, D., & Memmott, P. (1999). Digital Reference Services in Academic Libraries. *Reference & User Services Quarterly*, *39*(2), 145-150.

Janes, J., Hill, C., & Rolfe, A. (2001). Ask-An-Expert Services Analysis. *Journal of the American Society for Information Science and Technology*, *52*(13), 1106-1121.

Katz, W. A. (1997). *Introduction to Reference Work: Reference Services and Reference Processes* (Vol. 2). New York: McGraw-Hill.

Kresh, D. N. (2000). Offering High Quality Reference Service on the Web: The Collaborative Digital Reference Service (CDRS). *D-Lib Magazine*, *6*(6). Retrieved 3 September 2002, from D-Lib Magazine Web site: http://www.dlib.org/dlib/june00/kresh/06kresh.html.

Lagace, N., & McClennen, M. (1998). Questions and Quirks: Managing an Internet-Based Distributed Reference Service. *Computers in Libraries, 18*(2), 24-27. Retrieved 3 September 2002, from Information Today Web site: http://www.infotoday.com/cilmag/feb98/story1.htm.

Lankes, R. D. (1998). *Building & Maintaining Internet Information Services: K-12 Digital Reference Services*. Syracuse, NY: ERIC Clearinghouse on Information & Technology.

Lankes, R. D., & Kasowitz, A. (1998). *AskA Starter Kit: How to Build and Maintain Digital Reference Services*. Lanham, MD: Scarecrow Press, Inc.

MathNerds. (2001). *About MathNerds*. MathNerds. Retrieved 3 September 2002, from MathNerds Web site: http://www.mathnerds.com/about/about.asp.

McClennen, M. (2001, 12-13 November). *A Process Model for Digital Reference.* Paper presented at the 3rd Annual VRD Conference, Orlando, FL.

McClennen, M., & Memmott, P. (2001). Roles in Digital Reference. *Information Technology and Libraries, 20*(3), 143-148.

Pomerantz, J. (in preparation). *Question Types in Digital Reference: An Evaluation of Question Taxonomies.* Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.

Pomerantz, J., Nicholson, S., & Lankes, R. D. (forthcoming). Digital Reference Triage: An Investigation using the Delphi Method into the Factors Influencing Question Routing and Assignment.

Richardson Jr., J. V. (1995). *Knowledge-Based Systems for General*

Reference Work: Applications, Problems, and Progress. San Diego: Academic Press.

Robinson, B. M. (1989). Reference Services: A Model of Question Handling. *RQ, 29*(1), 46-61.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Sarkar, D. (2002, 21 August). Library Encourages Patrons to Chat. *Federal Computer Week.*

Taylor, R. S. (1968). Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries, 29*, 178-194.

Virtual Reference Desk Project. (1998, 29 September). *Virtual Reference Desk AskA Software: Decision Points and Scenarios*. Retrieved 3 September 2002, from Virtual Reference Desk Web site: Virtual Reference Desk Project. Available: http://www.vrd.org/Tech/AskA-sw.PDF.

White, M. D. (2001). Diffusion of an Innovation: Digital Reference Service in Carnegie Foundation Master's (Comprehensive) Academic Institution Libraries. *The Journal of Academic Librarianship, 27*(3), 173-187.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control, 8*, 338-353.