

A Virtual Expression of Need: An Analysis of E-mail Reference Questions

Wendy M. Duff and Catherine A. Johnson

Abstract

This paper analyzes e-mail reference questions posed to archivists in order to enhance our understanding of how users of archives seek information. This study analyzed 375 e-mail reference questions submitted to provincial, federal, university, city and special archives in order to determine, from the users' own words, how users formulate reference requests to archives. Understanding what elements the archives' client uses to describe his or her information need enables the creation of more relevant archival descriptive tools. According to this analysis, people used proper names, dates, places, subject, form, and, occasionally, events when composing their information request. As archives move toward a greater presence on the World Wide Web, archivists should design electronic information systems that account for the information seeking patterns expressed in e-mail reference requests.

Introduction

Archival principles and standards guide archivists in organizing and providing access to records. In the 1980s and early 1990s, the Society of American Archivists' *Archives Personal Papers and Manuscripts* and the Canadian *Rules for Archival Description (RAD)* codified archival practice and established the elements needed to represent a *fonds* or collection. The primary purpose of creating descriptive tools is to help users locate relevant material. However, the content of and means of access to these descriptions are based on archivists' perceptions and models instead of the point of view of the users.

Only recently have archivists begun to consider the users' point of view in evaluating access to their collections.¹ As archives move toward a greater

The authors would like to thank Marlene Van Ballegoie, Elaine Toms and Joan Cherry for their help and advice in conducting this study, David Bearman for his insightful comments on this paper, and the archives that contributed their e-mail messages.

¹Robert Spindler and Richard Pearce-Moses, "Does AMC Mean Archives Made Confusing? Patron understanding of USMARC AMC Catalog Records," *American Archivist* 56 (Spring 1993): 330-41; Wendy Duff and Penka Stoyanova, "Transforming the Crazy Quilt: Archival Displays from a Users' Point of View," *Archivaria* 45 (Spring 1998): 44-79.

presence on the World Wide Web, more users with less archival experience will want to gain access to archival collections. In order to accommodate this increase in demand for access to collections, it is necessary to ensure that users are able to search for information as independently as possible. Therefore, good archival systems will try to fulfill most users' requests without requiring that archivists intervene between the user and the system. To do this, archivists will have to know what types of information the user considers important when conducting a search to ensure that these elements are included in descriptions of materials and on search interfaces.

A useful way to determine what elements a patron uses to describe an information need is to examine a user's request in his or her own words. This has been done in libraries by studying verbatim transcripts of reference interviews between librarians and patrons. Since these interviews are usually conducted face-to-face or over the telephone with an archivist or librarian, one is able to make assumptions about the information request based on verbal clues and in the case of face-to-face interviews the patron's body language and other non-verbal cues.² Using e-mail reference questions to analyze users' requests, however, allows researchers to get a picture of the information need without it being filtered through a second party. The e-mail reference question is a representation of an information need, expressed in the words of the user without prompting by either an archivist or a reference request form. The e-mail request, therefore, contains the elements that the *user* considers relevant in his or her search for information. A knowledge of what elements people use to describe their information need and how they structure their requests will make it possible to design better research aids, including automated information retrieval systems, finding aids, and web interfaces, that will guide the user to the information she or he wants. This, in turn, will increase the independence of the researcher in both the "real" and the "virtual" archives.

Organization of the Paper

This paper analyzes e-mail reference questions posed to archivists in order to enhance our understanding of how users of archives seek information. First, a brief literature review will discuss previous research into users' queries and, specifically, reference queries posed to archives. This is followed by a discussion of the methodology and coding criteria used in this study. The data analysis describes how the questions were analyzed and the software used. The findings section delineates the research questions and discusses the results of the analy-

²Eileen G. Abels, "The E-mail Reference Interview," *RQ* 35 (1996): 345–58; Diane L. Fishman, "Managing the Virtual Reference Desk: How to Plan an Effective Reference E-Mail System," *Medical Reference Services Quarterly* 17 (Spring 1998): 1–10.

ses. Finally, the discussion section summarizes the findings and suggests their implications for the design of archives' web pages and information retrieval systems.

Literature Review

To date, e-mail reference service has been examined mainly in the context of libraries. While libraries are predominantly concerned with providing access to information, archives are primarily interested in providing access to their particular collections. For instance, in discussing e-mail reference service, libraries are mainly concerned with finding answers to their users' questions, whether the answers are found within the library, in other institutions, or on the Internet.³ Other concerns involve how to incorporate electronic reference into traditional reference service and the virtual reference interview.⁴ Some studies have examined e-mail reference questions to determine the types of questions asked, but few have examined the content of e-mail messages for clues as to how patrons seek information, or to determine the elements they use when forming their requests⁵.

Some researchers, however, have used other methods to better understand the information seeking behavior of humanities scholars. In the early 1990s Marcia Bates and others, involved in the Getty Online Project found that these scholars included names of individuals as subjects 74 percent of the time, geographical name 37 percent of the time, and date 26 percent of the time when formulating queries.⁶ Helen Tibbo also underlines the importance of cataloguers providing access to manuscript collections through the "facets of description that best describes their materials such as time, place, and topic."⁷ In an interesting study of the information behavior of history Ph.D. students,

³ Nettie Lagace & Michael McClennen, "Questions and Quirts: Managing an Internet-based Distributed Reference Service," *Computers in Libraries* 18 (February 1998): 24-27; Bernie Sloan, "Service Perspectives for the Digital Library: Remote Reference Services," *Library Trends* 47 (Summer 1998): 117-143.

⁴ Amy M. Kautzman, "Digital Impact: Reality, the Web, and the Changed Business of Reference," *Searcher* 7 (March 1999): 18-24; Fishman, "Managing the Virtual Reference Desk: How to Plan an Effective Reference E-Mail System;" Elizabeth Davenport, Rob Proctor & Ana Goldenberg, "Distributed Expertise: Remote Reference Service on a Metropolitan Area Network," *The Electronic Library* 15, (August 1997): 271-278; R. David Lankes, "AskA's: Lesson Learned from K-12 Digital Reference Services," *Reference & User Services Quarterly* 38 (Fall 1998): 63-71.

⁵ Abels, "The E-mail Reference Interview" Sloan, "Service Perspectives for the Digital Library: Remote Reference Services;" Fishman, "Managing the Virtual Reference Desk: How to Plan an Effective Reference E-Mail System;" Lara Bushallow-Wilbur, Gemma DeVinney, & Fritz Whitcomb, "Electronic Mail Reference Service: A Study," *RQ* 35 (Spring 1996): 359-71.

⁶ Marcia J. Bates, Deborah N. Wilde, and Susan Siegfried, "An Analysis of Search Terminology Used by Humanities Scholars: the Getty Online Searching Project Report Number 1," *Library Quarterly* 63 (January 1993): 15.

⁷ Helen Tibbo, *Abstracting, Information Retrieval and the Humanities: Providing Access to Historical Literature*, (Chicago & London: American Library Association, 1993), 83.

Charles Cole described the practice of name collecting.⁸ Cole attributed this practice, which was evident in nearly half of his sample group (19 out of 45 students), to a strategy to “induce pattern formation in the data associated with the names that may lead to an original thesis.” The technique involved the student noting the name of an individual or a company each time she came across it in her research material. By “scan reading” she is then able to slice through the material and focus in on that name whenever it appears in the text. Through this method, relationships between individual companies or persons and events become clear and patterns emerge. If proper names are a common access point into primary source material for historians, then it would be interesting to learn if this is also a common practice among other user groups of the archives. There have been three recent studies that have deconstructed reference questions to find out more about this practice.

David Bearman examined the “presentation language” of archival users in order to learn what kinds of questions archives and museums information systems answered, how the answer was used, and their criteria for success.⁹ The queries in his study were collected in one day from both staff and the general public on “inquiry forms” at eighteen participating archival repositories. The questions were categorized into three types: questions that could not be answered by an information retrieval system, procedural questions, and research-based questions. The facets, or elements, of each category of question were also identified.

A study of users’ questions asked at archives in Quebec, either in person, by e-mail, or by telephone, examined only those complex research-based questions requiring the intervention of an archivist after other sources of information had been consulted.¹⁰ The study identified three types of research questions: general subject, specific subject, and specific form questions. Gagnon-Arguin found that the association of standard elements, like date and place, with a particular question category, such as general subject, could help model research strategies that would assist researchers in becoming more independent of the archivist.

Another study, conducted by Karen Collins, examined reference questions asked at two institutions maintaining historical photographic collections.¹¹ In this study, the queries were collected on forms that prompted the user “for information pertaining to the subject, place, time, visual characteristics,

⁸ Charles Cole, “Name Collection by Ph.D. History Students: Inducing Expertise,” *Journal of the American Society for Information Science* 51 (March 15, 2000): 444–55.

⁹ David Bearman, “User Presentation Language in Archives,” *Archives & Museum Informatics* 3 (Winter 1989/90): 3.

¹⁰ Louise Gagnon-Arguin, “Les questions de recherche comme matériau d’études des usagers en vue du traitement des archives,” *Archivaria* 46 (1998): 86–102.

¹¹ Karen Collins, “Providing Subject Access to Images: A Study of User Queries,” *American Archivist* 61 (Spring 1998): 36–55.

emotional or subjective qualities, or other requirements of the image(s) sought.”¹² Eight question categories and the elements associated with each category were identified: generic subject, specific subject, time, place, genre, visual features, physical format, and creator/provenance. The elements associated with generic subject, for example, included persons, geographical features, objects/things, activities, and concepts.

Although Bearman and Gagnon-Arguin’s studies included mail and e-mail requests, their findings also included the data gained from the mediated in-person and telephone requests. Because our study examined only e-mail requests, it is likely to be more representative of the user’s unmediated expression of need. The importance of examining the user’s query without the intervention of any mediating influence was recognized as early as 1968 by F. W. Lancaster: “It appears that the best request statements (i.e., those that most clearly reflect the actual area of information need) are those written down by the researcher in his own natural-language narrative terms.”¹³ By investigating all types of e-mail questions sent to archives and conducting statistical analyses to reveal the relationships between the elements of the queries, the current study hopes to reveal which types of information the user chooses to construct her own information need.

Method

In spring 1999, a message was sent on two archives listservs, ARCAN-L (Canadian) and ARCHIVES AND ARCHIVISTS (American), asking archives to forward e-mail reference questions sent to their institutions. All queries came from users outside the archives and therefore might not represent all types of questions posed to an archival information system. Nor was there any control over the types of questions that the institutions submitted. Some archives sent very few e-mail queries, so it is quite likely that they filtered out some of the questions they felt were not appropriate. For instance, one institution predominantly sent requests for copies of land grants, while other institutions sent only a few similar kinds of service requests. It would, therefore, be incorrect to assume that the results reported in this study represent the true proportions of each type of e-mail request sent to the archives. Preselection by some of the archives, may have skewed the types of questions sent to include the more obvious reference or resource discovery-type questions.

The choice was made to study e-mail questions because they afforded the opportunity to examine users’ requests in their own words. When making

¹² Collins, “Providing Subject Access to Images,”

¹³ F. W. Lancaster, “Evaluation of the MEDLARS Demand Search Service,” (Washington, D.C.: Government Publishing Office, 1968), 102 cited in Abels, “The E-mail Reference Interview,” 356 n. 10.

reference requests, either in person or over the telephone, to a reference librarian or archivist, or to an information retrieval system, a user will refine his or her search request several times depending on the feedback from the person or the system. E-mail requests and letters are different in that the writer knows she needs to include as much information as possible because of the lack of immediate feedback.¹⁴ This may or may not be so true of e-mail requests, which might be sent with the expectation of a quick response, i.e., immediate feedback. In this study, it is not known how quickly the e-mail requests were responded to or how often the requests were refined through the prompting of the archivist. Only 18 out of 361 questions were identified as follow-up questions responding to information or requests made by the archivist.

In all, eleven institutions representing provincial, federal, university, city, and special archives responded, forwarding 375 questions within a six-month period. The response varied across the institutions. One institution sent everything it received, two institutions sent a sample of questions they received, and others sent questions sporadically over the allotted time period. In some cases, the e-mail questions were preselected by the institutions to include only those that were considered resource discovery-type questions (see below for definition). Of the 375 questions sent, only fourteen were considered unusable because they were either not questions or they dealt with job-related issues rather than archives-related matters (asking an archivist to speak at a conference, for instance). Some of the archives edited the messages to remove any personal information. The remaining messages were put through a filter to remove all identifying information by automatically “X”-ing out proper names or places.

Coding of Questions

Each question was read and coded separately by the two authors of this paper. The coding was then discussed jointly. When there was no disparity in the coding terms used, the coding was accepted. If there was a difference, it was discussed until agreement on coding was reached. When patterns emerged that identified new question categories and elements, the earlier questions were reexamined to see if they needed to be recoded.

Each of the questions was coded for three things: type of request and the elements that fell into either the “wanteds” (i.e., the information desired) component or “givens” (i.e., the information already known) component of the

¹⁴ Mail and e-mail requests are also more likely to represent a “compromised” need, as defined by Taylor [R. S. Taylor, “Question—Negotiation and Information Seeking in Libraries.” *College and Research Libraries* 29 (May 1968): 178–94], since they are composed with some expectation of what the system can deliver. In an in-person or telephone reference interview the archivist can often determine through probing what the user really wants, whereas in mail or e-mail requests the archivist often has to take the request at face value.

query. Coding was carried out using NVivo, a qualitative data analysis software. The coding of each request was based on Denis Grogan's schema for categorizing reference questions asked in libraries.¹⁵ He identified eight types of questions asked at the reference desk, falling into two broad categories: limited help questions and open-ended questions. The limited help questions included administrative/directional, author/title, and fact-finding. The open-ended questions included: material-finding, mutable questions, research enquiries, residual enquiries, and unanswerable questions. Although we began thinking about our questions in these terms, we soon discovered that some of the categories were not relevant to our sample and other types of questions emerged from the data. In our schema, the limited help questions included: Administrative/Directional, Material Finding, Specific Form, Known Item, and Service Request. The more open-ended questions included Consultation, which corresponds closely to Grogan's research enquiries, and User education, which is a combination of Grogan's mutable and residual enquiries. Our final categories of questions and definitions are:

1. Administrative/Directional:

These questions involved requests for information about costs of photocopying, hours, location and phone numbers of the archival institution as well as policies and procedures of the institution; for instance permission to access materials, permission to publish and how to cite sources.

Example: *I would like more information on the photographs that the archives hold in Store[.] [C] an they be copied by someone visiting the archives? can they be placed on someone's history and genealogy pages? with or without the Archives' permission as long as mention of the XXX is on the page?*

2. Fact-finding:

In this type of question, the user wants a specific answer to a specific question and expects the archivist to supply it. The user is not looking for a source where the answer might be found, but the actual answer to the question.

Example: *Can you find for me the degree that XXX XXXXXXX earned from the University of XXXXXXXXXXXXXXX in 1923? Also can you confirm that he was a Rhodes XXXXXXX for that same year. Many thanks.*

3. Material-finding:

The user wants to know whether the archives has any sources about a particular person, place or event. The user usually has a good idea of what she or he wants and is often able to give specific dates, places, and activities to further identify the requested information.

¹⁵Denis Grogan, *Practical Reference Work*, 2nd ed., (London; Library Association Publishing, 1992).

Example: *I am doing some historical research on XXXXX-XXXX Community College. I was wondering if you have any historical information on any of their schools? They opened up in Vancouver in 1913 and in Victoria in 1913. They also owned and operated an aviation school in 1928–1930. If you have any info can you please e-mail me or call me. Thank-you very much.*

4. Specific Form:

The user wants to know if a specific record type or form is available at that archives; i.e., birth registries or military records.

Example: *Please tell me who I write to, for copies of all my XXXXX XXXXXX military records, I have the XXX. He was a pilot during WWI, from XXX XXXXXXXXXXXX Thank you.*

5. Known Item:

The user gives the name of a specific or known item and wants to know if it is located in that archives.

Example: *I am searching for a copy of a XXXXXXXX'X magazine from XXX 19, 1962. Are you able to help me locate one?*

6. Service Request:

This involves a request for a specific service like a photocopy or inter-library loan. The request is almost always accompanied by a title, call number or citation for the document to be copied or loaned.

Example: *I was searching your web site and am interested in acquiring copies of two of the land grants. They are:
XXXXXXXX XXXXXXXX, vol. 17, page 242, grant no. 788, NB reg. date 1836/06/22
XXXXXXXX X XXXXXXXX, vol. 62, page 0, grant no. 10362, NB reg. date 1862/06/17*

7. Consultation:

In this type of question the user requires advice from the archivist which calls on her or his specific knowledge of the resources. For example, a questioner might ask whether a particular fonds was likely to contain the type of material the user is looking for as well as a recommendation of where the user might find the required information.

Example: *I am an Italian student of the university of XXXXX. I chose the title on my thesis: the XXXXX'X XXXXXXXXXXXX temperance XXXXX in the 1920s–30s. Now I am in XXXXXXXX and I have found much secondary sources and bibliography. My idea was to look for papers of an activist of the WCTU not yet studied or not enough studied, but here I have just found reports, minute books, journals, and not diaries, letters and so on.*

Can you please tell me where I can find this sort of materials and where? I know that probably there is something about XXX XXXXXX in XXXX XXXXXX, or that one important speaker was Mrs. XXXXXX XXXXXX. Do you think I have some possibilities to find what I am looking for here in XXXXXX, in XXXXXX or in some other parts of XXXXXX?

Thanks and, please, I am in XXXXXX just until the first of XXX, can you answer me as soon as possible?

8. User Education:

The user has only a vague sense of what records she or he wants access to and needs advice on where or how to get started. There is usually a “how” statement involved—how do I find the information or how do I get started.

Example: *I am looking for information on my great-great grandfather, who was an explorer, timber cruiser and involved in the gold rush in XX in the second half of the nineteenth century. I believe he was the main person responsible for opening up the area around XXXXXXXXXXXX XXXX, on XXXXXXXXXXXX XXXXXX. I know that XXXX'X XXXX is named after him. His name is XXXXXX XXXX, and he was born some time around the 1840s in the XXXXX XXXXXX, XXX. He was famous for his saying “by the lovely dove.”*

I would be extremely grateful if you could find out anything more about him, or even how I could come and do some research.

Grogan also discusses Gerald Jahoda’s taxonomy of reference queries based on what the user wants and what she or he “gives” to further define the “wanted.”¹⁶ We used this taxonomy in coding the reference questions. Each question was examined in order to identify the terms that described the information need (the “wanted”) and the terms that described what the questioner knew about the information need (the “givens”). These terms were then categorized into the types of information wanted and given; i.e., personal name, place, form, etc. According to Jahoda, “by categorizing the givens and wanteds from query statements into general descriptors, we can group similar queries together, and can often approach their answer searches in similar ways.”¹⁷ For instance, in a question requesting the birth record for Mary Brown born in Kent County, New Brunswick, in the 1850s, the birth record would be the “wanted”, and the personal name, place, and date would be the “givens.” Supplementary information, describing the object of the request but not vital to its identification, was also coded but not used in this analysis. Each question, therefore, was coded by question category and the type of “wanteds” and “givens”.

¹⁶ Grogan, *Practical Reference Work*, p. 46.

¹⁷ Gerald Jahoda, and Judith S. Braunagel, (1980). *The Librarian and Reference Queries: a Systematic Approach*. (New York: Academic Press, 1980), 11.

Example of Coding

I would like some information on XXXX XXXXXXXXX. A photographer, a painter and a writer born in 1906. She lived in the XXXXXXXX. I would like to know if you have any information concerning her life history, her photographs, paintings or any books that were published.

Query 1:

<u>Type of Question:</u> Material Finding	
<u>Wanted</u> Biographical Information	<u>Givens</u> Personal name Occupation Date Place

Query 2:

<u>Type of Question:</u> Material Finding	
<u>Wanted</u> Bibliographic citation	<u>Givens</u> Personal name Form-image Form-book

This question was, in fact, a two-part question which we coded as two separate queries. Both were categorized as material-finding requests. In the first query, the “wanted” element is biographical information, and the “givens” elements are personal name, occupation, and date. The second query in this request is for citations (bibliographic information) to any photographs, paintings, or books (forms) that were published by the subject individual (personal name).

Data Analysis

The elements of the questions were not defined *a priori* but grew out of the analysis. Using NVivo software, we coded 183 different elements that users employed to describe their “wanted” and “givens.” Elements describing similar concepts were later collapsed into more general categories, and some elements

that were used rarely (less than five times in any question category) were eliminated in the final analysis. For instance, over forty different record types were identified among the 'givens', including birth certificates, land grants, and sound recordings, which were later collapsed into one category designated as "form". Similarly personal name, corporate name and name of an entity were collapsed into one "name" category.

Each of the eight question categories was analyzed separately. In order to show the relationship between the "wanteds" and the "givens," a cross-tabulation between the "wanted" elements and the "given" elements was performed. The results of this analysis were then incorporated into a table with the 'wanteds' in rows as cases and the "given" variables in columns. In this way we were able to make an association between what the user wanted and the element most often used to describe the information need.

Institutions

Eleven archival institutions forwarded e-mail reference questions. In all, 375 questions were sent and, of these, 361 were considered usable for the project. These questions generated 392 queries.

As Figure 1 shows, almost half (47 percent) of the questions came from the two provincial archives in Canada. One of these institutions sent us every

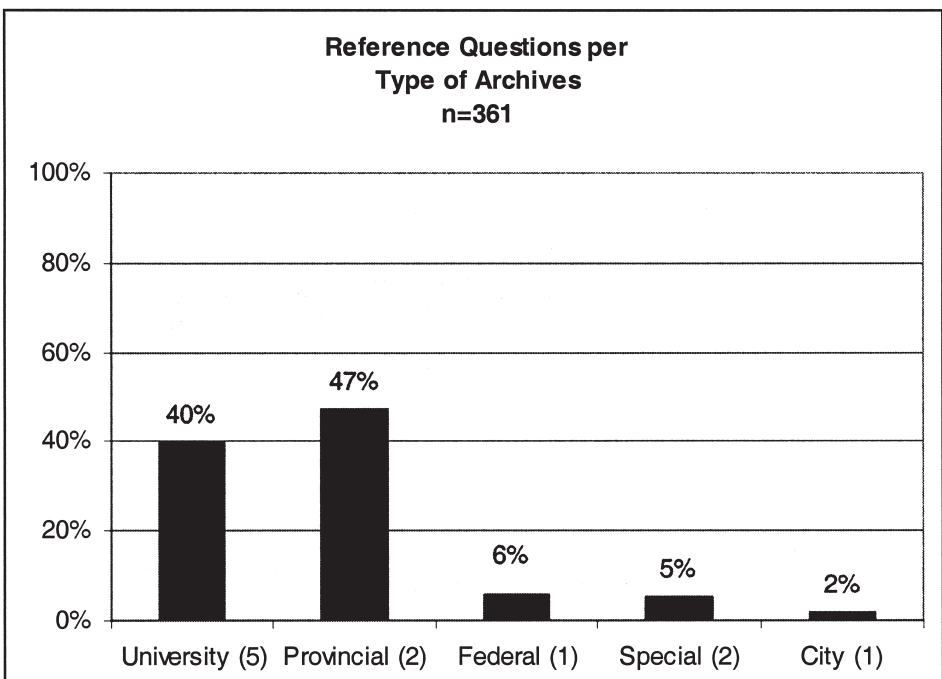


FIGURE 1. Reference Questions Per Type of Institution

question it received in the six-month period, amounting to nearly 40 percent of all the questions received. Five university archives supplied 40 percent of the messages, with the other messages coming from one federal (6%), two special (5%) and one city (2%) archives.

Findings

The purpose of this study was to determine, from the users' own words, how users formulate reference requests to archives. Understanding what elements the archives' client uses to describe his information need enables the creation of more relevant archival descriptive tools. Our research questions were: 1) What types of questions do archives users ask? 2) What terms do people use when expressing an information need? and 3) Do the types of terms people use differ depending on the category of question asked and the type of information wanted?

What Types of Questions Do Archives Users Ask?

We identified eight categories of questions from the e-mail reference requests sent to the archival institutions. Since some of the questions contained more than one query, the total number of queries amounted to 392 for the 361 questions analyzed. Figure 2 indicates the breakdown of question

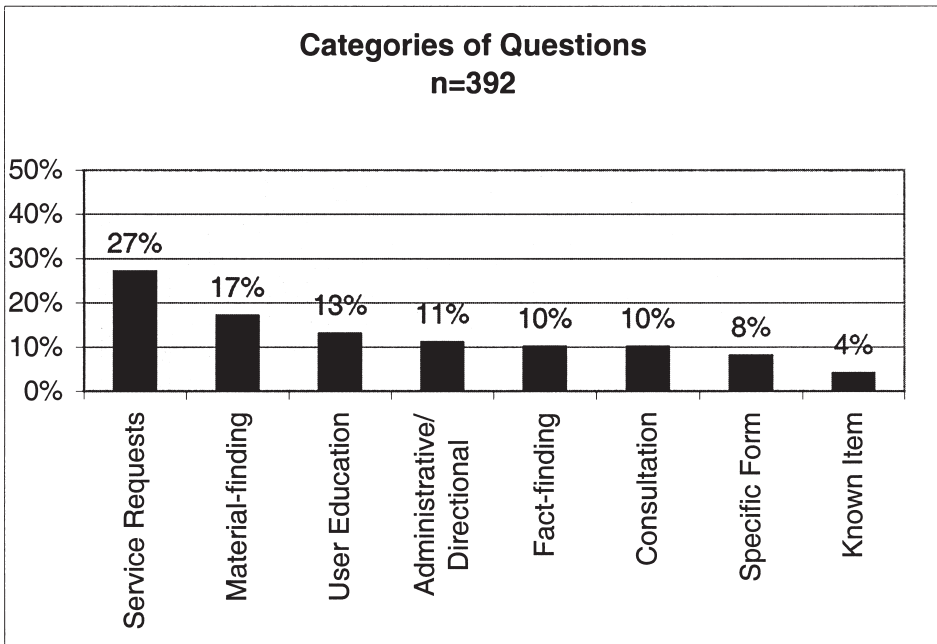


FIGURE 2. Breakdown of Question Categories Asked of the Archival Institutions

categories identified from the data. The most frequently asked type of question in the survey was a service request (27%), followed by material finding (17%), and user education (13%). Most of the service request questions originate from one institution, the provincial archives, that supplied almost 40 percent of all requests. Whether the other archives also received this type of question, which usually amounted to a request for photocopying or interloaning material, but chose not to submit them to the survey, is unknown.

Material finding, specific form, and known item questions share similar characteristics; they are closed-ended questions asking for the location of specific items that could likely be answered by an information retrieval system. Because of this, we collapsed these three categories into one category which we labeled resource discovery questions. The resource discovery category then became the most frequent type of question category (29 percent), followed by service requests (27 percent), user education (13 percent), administrative/directional (11 percent), fact finding (10 percent) and consultation (10 percent).

What Terms Do People Use When Expressing an Information Need?

To identify the terms people used when expressing their need, the questions were first grouped by type of question, and then the type of information needed was analyzed. People supplied different types of information when asking different types of questions. Furthermore, as shown in Figure 3, people

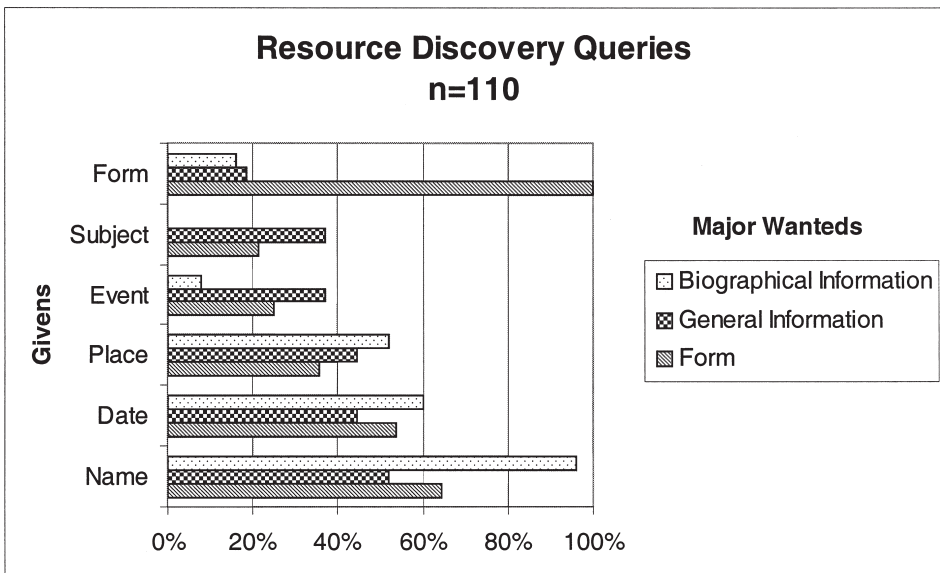


FIGURE 3. Resource Discovery Questions' Major Wanted and Givens

wanting different types of information used different terms to express their information need.

Resource Discovery Queries

The resource discovery queries were analyzed in order to relate the type of information wanted with the type of information given. The three major “wants” in this category were form, general information, and biographical information. The most common elements, or givens, that described each of these wants were name, date, place, event, general subject, and form.

As Figure 3 illustrates, a personal name was given when the user wanted biographical information 99 percent of the time, while it was given much less often when the user wanted a specific form (69%) or general information (52%). Not surprisingly, users who requested a specific form of material provided the type of form 100 percent of the time, compared to only 16 percent of the time when requesting biographical information or general subject (9%). While general subject was present in both the general information (37%) and form (21%) requests, it was not given in any of the requests for biographical information. Similarly, an event was rarely given in a request for biographical information (8%), but was included in 37 percent of the general information queries and 25 percent of the form requests. Both date and place were given consistently across all the three types of requests.

Service Requests

Service request queries formed an unusual category that appears to be directly related to the presence of finding aids and indexes on the World Wide Web. Most of these queries, usually for photocopies of land grants, came from one institution that had an index to these grants available on its web site.

Figure 4 indicates the main types of information included in a service request question. In all cases the request was accompanied by at least the call number or title of the item, and in 50 percent of the cases, the exact citation. Date and name were less frequently included, while place, *fonds*, subject, and source where the citation was found were rarely (less than 10% of the time) mentioned.

Consultation Queries

When making consultation queries, users usually wanted either general information or a recommendation.

As Figure 5 shows, when seeking a recommendation from the archivist, users gave a proper name (74% of the time), a date (58%), a general subject or topic (42%), a place name (32%), a specific form (21%), and an event (11%).

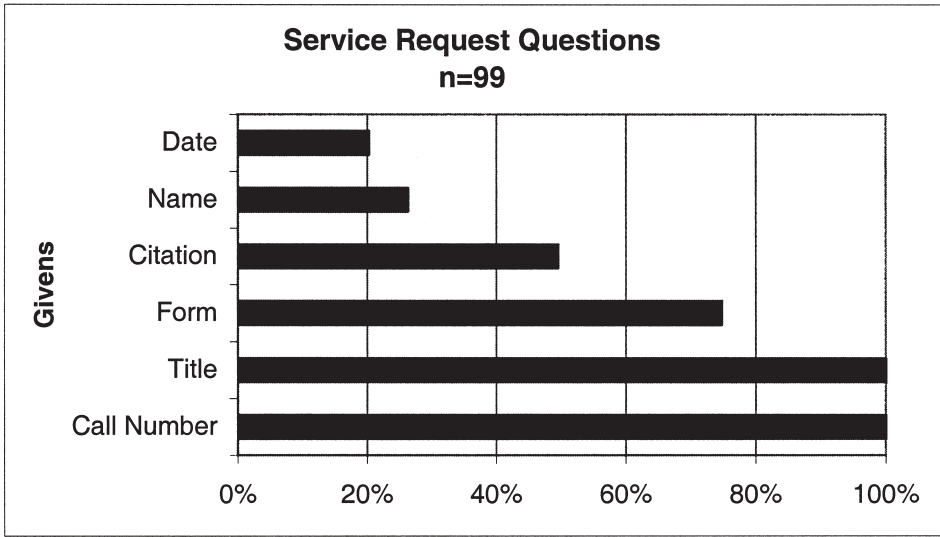


FIGURE 4. Main Given Elements in Service Request Questions

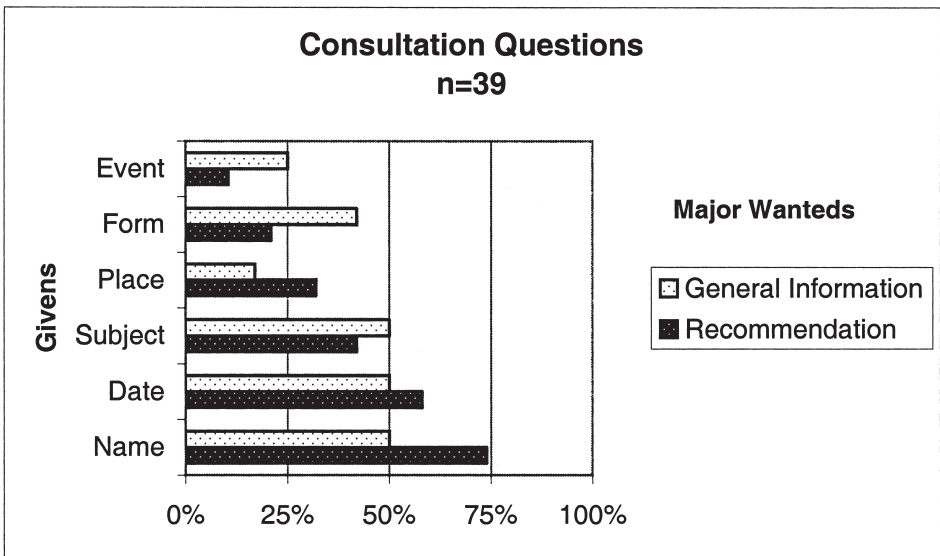


FIGURE 5. Major Givens and Wanted Elements in Consultation Questions

When requesting general information, they provided a proper name only 50 percent of the time, but were more likely to mention a specific form (42%) and an event (25%) than was the case when asking for a recommendation. Although the givens elements were mentioned in differing proportions when describing these information needs, the types of information given to describe the information need were the same in consultation questions which asked for either a recommendation or general information.

User Education Queries

User education queries were determined subjectively. From the tone of the question and the lack of a clearly defined request, it was decided that the user needed some general instruction about the types of records kept in the archives and how to find them. More than half of the user education queries wanted biographical information. It is speculated that these requests generally came from people just beginning to do genealogical research. In this category, although the users tended to have specific information relating to the name, place, and date of the requested information, they did not usually ask for the specific form or source that might have given them the information they were looking for. For instance, although an activity (usually a birth, death or marriage) was mentioned in relation to the biographical request more than half of the time, the actual form (birth, death, or marriage records) was not often used to describe the information wanted. In fact, this type of request was characterized by the amount of information accompanying it, much of it supplemental to the actual information “wanted.” Often the request itself was not explicitly stated but assumed within a general request, e.g., “I would be extremely grateful if you could find out anything more about him [great-great-grandfather], or even how I could come and do some research.”

Relation of Terms Used to Information Needs

In the resource discovery queries, the terms used to express an information need vary somewhat by frequency depending on the type of question asked and the information wanted. However, users gave proper names and dates for the majority of questions and they mentioned places, events, general subject, and specific forms to a lesser extent. In the consultation queries as well, name, date, general subject, place, form, and event were used to define the information need. Service requests were unusual because the patron was not seeking information about something but rather requesting a specific service for which she/he almost always provides some element of a citation (call number or title). Name, date, place, and general subject were also used to describe the object, but less than a quarter of the time. User education queries also used the elements of name, place, and date to describe their information need, but also included activity (marriage, death, birth, etc.) to further define what they were looking for. Activity was not a major component of the other query categories.

Discussion

Consistent with previous studies, this research found that people used proper names, dates, places, subject, form, and, occasionally, events when

composing their information request. Although the types of questions were categorized somewhat differently, Bearman found the most frequently mentioned elements were form, name, citation, place, subject, and date.¹⁸ Similarly, the “facets” identified by Gagnon-Arguin were subject, place, date, occupation, event, and form.¹⁹ Karen Collins’ study, although quite different in design, also identified name, place, date, and event as major elements in users’ queries.²⁰ An information retrieval system should enable the user to search and retrieve information using these types of terms. Furthermore, authority control systems that standardize proper name, place, and form genre terms would probably improve retrieval. General subject terms are not as important, since they were given only 37% of the time for general requests and 21% of the time for form requests. The reasons for this may be that many users were looking for information about people rather than subjects, or they were searching for specific forms because they knew they would provide certain types of information. On the other hand, users may turn their requests for a general subject into a consultation request because the system fails to provide the information needed. Consultation queries provided a general subject for half of the requests for general information and for 42% of the requests for recommendations. The questions which asked if a particular *fonds* was relevant to a particular information need may indicate that the description of the *fonds* is insufficient, resulting in the need for the advice of an archivist.

Requests for services, particularly photocopying of documents, seem to be related to archives adding indexes to their web sites. If an order form is not available for the user to make the request, which was the case for the institution receiving the bulk of the service request queries, an e-mail message has to be composed for each request. In addition, almost 15% of all service requests asked how much the particular service would cost. Although this information may have been available somewhere on the archive’s web site, it was not linked to the indexes in an obvious way that would have made the information clear to the user. Hence, when archives add indexes to their web sites, it would be easier for the user, as well as more efficient for the archives, to include a form for ordering documents that also includes all the relevant information needed to make the request.

The user education queries also raise many questions. Whether we categorized a query as a material finding or a user education question was based to a large degree on our assumptions about what kind of questions an information retrieval system could answer. In all cases the user had some kind of information that modified the information need, such as where a person lived and

¹⁸ Bearman, “User Presentation Language in Archives,” 6.

¹⁹ Gagnon-Arguin, “Les questions de recherche comme materiau d’etudes des usagers en vue du traitement des archives,” 94–95.

²⁰ Collins, “Providing Subject Access to Images: A Study of User Queries,” 48–49.

an approximate date. They were unsure, however, about what kinds of records to consult to find what they wanted. An expert system that asked a number of questions, directing the user to relevant types of records, would help the user become more independent in fulfilling his or her information need.

Conclusion

This study examined e-mail reference questions to determine the types of questions patrons ask and the elements they use to express their needs. While the questions themselves may not be representative of all the questions posed to archival institutions, their contents do shed light on the types of terms a patron might use to search an archival database and the type of information he or she might want. A follow-up study might be conducted where the archival institutions are asked to submit all their e-mail reference questions in order to get a more accurate picture of the kinds of questions sent to archives. These virtual expressions of need are a rich source of data that can help archivists improve their retrieval systems and better meet the needs of their users.