Understanding Academic Patrons' Data Needs through Virtual Reference Transcripts:

Preliminary Findings from New York University Libraries

by Margaret Smith[1], Jill Conte[2], Samantha Guss[3]

Abstract

New York University (NYU) Libraries has an extremely high-volume chat reference service. This popularity presents a unique opportunity for gaining insight into library patrons' conceptualizations of their data reference needs and how these needs are changing. Through analysis of three years' worth of chat transcripts, we began to explore user needs and familiarity related to locating secondary data and statistics, performing data analysis, and using existing data services. Ultimately, we focused our analysis on requests for census data. This article discusses, in detail, the methods, preliminary results, limitations, and proposed next steps of our investigation. Our final goal is to contribute to the growing body of knowledge about how information needs are conceptualized and articulated, and how this knowledge can be used to improve data reference in an academic library setting.

Keywords: academic libraries, data reference, grounded theory, virtual reference services, chat transcripts

Introduction

NYU Libraries serves the NYU 'Global Network University', the main campus of which is situated in Greenwich Village, next to Washington Square Park, in Lower Manhattan. The NYU Polytechnic School of Engineering is housed nearby, in downtown Brooklyn, and NYU has portal campuses in Abu Dhabi and Shanghai, as well as 11 smaller global academic centers where students study away for a semester or year. NYU enrolls approximately 45,000 students (half of whom are undergraduate students), and employs approximately 3,000 teaching faculty. Bobst Library is the flagship of the NYU Libraries' system, with 12 publicly accessible floors, 6 million volumes, and seating for 3,000.

The library's urban location and proportionately small seating capacity, combined with the area's above-average commute time and a user community spanning the globe, lead to high demand for NYU Libraries' virtual library services. Our chat reference service is extremely busy; we receive approximately 15,000 chat transactions annually, 30-40 a day on average, mostly occurring between the hours of 9am and midnight, New York City local time. The average duration of a chat conversation is 16 minutes.

This popularity offers a unique opportunity for gaining insight into library users' conceptualizations of their data needs and how these needs are changing. Through analyzing three years' worth of chat reference transcripts, we began to explore user needs and familiarity related to locating secondary data and

---

… patrons and librarians often use the word 'data' casually when discussing databases or information in general.

---

statistics, performing data analysis, and using existing data services, focusing on the way patrons initially ask data questions. While existing scholarship has addressed the theory and practice of data reference (Gerhan, 1999; Kellam and Peter, 2011), very little empirical research to date has qualitatively explored users' articulations of their data needs (Wang, 2013). This project is unique in that it employs transcripts of actual reference transactions, as opposed to user

surveys (Read, 2007), as the basis for analysis. Furthermore, such a high-volume chat reference service, which is staffed by data specialists and non-specialists alike, offers an opportunity to assess how the service as a whole handles–and can better handle–data reference.

Research Method: Grounded Theory

Because little research to date has been done on how users conceptualize and articulate their data needs, we chose a grounded theory approach, which is an exploratory, iterative methodology. This inductive approach seemed well suited for our purposes, as we did not start out with any particular hypothesis or hypotheses, but we knew that we had a rich data set. In grounded theory, researchers constantly 'move back and forth' between data collection and analysis (Bryant and Charmaz, 2007), resulting coincidentally in data refinement and conceptual categorization that leads to increasingly theoretical insight (Payne and Payne, 2004; Bryant and Charmaz, 2007).

On our first pass at analyzing the chat transcripts, we used the process of 'open coding' and 'memoing' (Grounded Theory Institute, 2014) to look for common patterns and to recognize and establish emerging themes. From there, we developed nascent codes and descriptors to start categorizing the data; codes were applied to relevant portions or passages of transcripts, while descriptors were applied to entire transcripts. We used the process of 'constant comparison' (Grounded Theory Institute, 2014) to scrutinize and further develop codes and descriptors as we applied them.

During this initial phase, we communicated on a regular basis through memos and real-time meetings to discuss observations, to deliberate over the shape of the emerging coding/descriptor schema, and to consider strategies that would better focus the data set. This iterative process of collaborative inquiry–i.e., observation, analysis, deliberation, and refinement–likewise marked each subsequent phase of our investigation, as the data collection and analysis processes described below demonstrate.

While we remain in the exploratory stage of our investigation, using a grounded theory approach will allow us over time to move from coding, categorizing, and comparing concepts to building an overarching theory that we can then marry with existing literature on the topic (Grounded Theory Institute, 2014).

Data Collection and Analysis

Due to the iterative nature of grounded theory, most of our data collection and analysis processes were inextricably entwined. Initially, we collected three years' worth of chat reference transcripts, as text files, from LibraryH3lp, our chat service provider. We then used two main tools to compile our data: FileLocator Pro, to retrieve transcripts containing data-related keywords, and TextCrawler, to remove system-generated librarian identifiers.

To analyze these transcripts, we used Dedoose, a web-based application developed to perform mixed-methods analyses in the social sciences. Dedoose allowed us to categorize each transcript using controlled descriptors–for example, to indicate whether a transcript should be included or excluded from a sample, and also to apply qualitative codes to excerpts of text within the transcripts. These descriptors and codes could then be cross-tabulated, analyzed, and visualized in various ways. In combination, these tools–FileLocator Pro, TextCrawler, and Dedoose–were extremely

effective for selecting a sample of transcripts; for protecting the privacy of individuals involved; and for classifying and analyzing the transcripts within a sample.

The process of gathering data-related reference transactions, however, was a non-trivial task. Even generating a starting search strategy required careful consideration of disambiguation. For example, we quickly realized that a search for the phrase, 'number of', would also retrieve results where a librarian or patron mentions the call number of a book. After a few minor tweaks to minimize these mismatches, our search strategy settled on this:

> data OR statistics OR stats OR GDP OR demographics OR census OR mortality OR GIS OR quantitative OR numeric OR SPSS OR Atlas.ti OR Atlas OR NVivo OR qualitative OR vivo OR "Data Services" OR data.services@nyu.edu OR "Data Service Studio" OR "data.service@nyu.edu" OR STATA

This limited the number of transcripts substantially, but still retrieved an immense number of transcripts that were not data-related. For example, patrons and librarians often use the word 'data' casually when discussing databases or information in general. Additionally, there were quite a few hits where the patron was asking for help locating or accessing a book or article that had one or more of our search terms in its title, yet the resource itself was not data-related (e.g., a quantitative study related to nursing). There were also cases where the physical space of our Data Services department was referenced, but not in regard to data needs (e.g., complaints of an unruly patron or broken computer in that area).

In order to ensure that the sample contained as many data-related results as possible, we read through the transcripts, looking for actual relevance to data, and assigned an inclusion or exclusion descriptor to each one. Even so, we ended up with 950 data-related transcripts from just one year's worth of transcripts. So we further refined our inclusion/exclusion criteria to omit those data-related transcripts involving 'known item' questions, such as a patron asking for help locating a specific financial report that contained data they had found via Google. While sometimes these patrons seemed clearly interested in the data that the report contained, it was often difficult to say whether this was definitively the case, or whether they were more interested in the report as a whole. We applied these new descriptors to the sample. At this point, 633 transcripts remained, a large proportion of which still involved questions about specific databases for business and financial information.

At a loss for ideas of other wide-sweeping exclusions we could make, we made a first pass at creating descriptors and codes for the transcripts in this sample. We read through them separately, coming up with lists of descriptors/codes that seemed potentially relevant, such as which specific resources were mentioned, the general subject area of the query, how accurate the librarian's response was (on a numeric scale), and how satisfied the patron seemed (on a numeric scale). We then discussed our experiences as a group and quickly realized the overwhelming effort that would be necessary to apply multiple, quantitative descriptors to a sample of this size. We decided to drop nearly all of the descriptors, and instead, apply codes within the text of each transcript, indicating the presence of different characteristics, like 'inaccurate answer' or 'patron satisfaction'. This was a speedier process, and we were able to make better progress in creating, discussing, and assigning codes.

Although we were now making more progress, we discovered that the sample did not include as many juicy, in-depth data reference questions that we had hoped to explore. After a few more code-refining group discussions, we introduced a new code that indicated simply which transcripts were compelling. We focused on these transcripts, looking for patterns that might help us come up with a new iteration of our search strategy. In doing this, we were surprised by how many reference questions we received that were explicitly related to United States and international census data, and, conveniently, it seemed like these questions tended to be the more in-depth exchanges that we were after.

We completely revised our search strategy, so that it included the terms that were frequently used in these interactions:

> census OR Factfinder OR "Social Explorer" OR "American Community Survey" OR "Fact Finder"[4]

This strategy retrieved 147 results across all three years of transcripts, although, of course, there are some caveats to the 'meaningfulness' of this search. For example, it only captures use of the word 'census', so sometimes questions are included which merely involve the concept of a census or patrons may ask for known items, other than censuses, that happen to have the word 'census' in the title. It also relies on user and librarian understanding of when to consult a census: sometimes the user is wrong, sometimes the librarian is wrong, and our sample includes both of these cases. Furthermore, this strategy omits census-related questions where the patron's information need was not sufficiently explored or understood, such that a census would have been an appropriate suggestion on the part of the librarian, but the transaction never got that far.
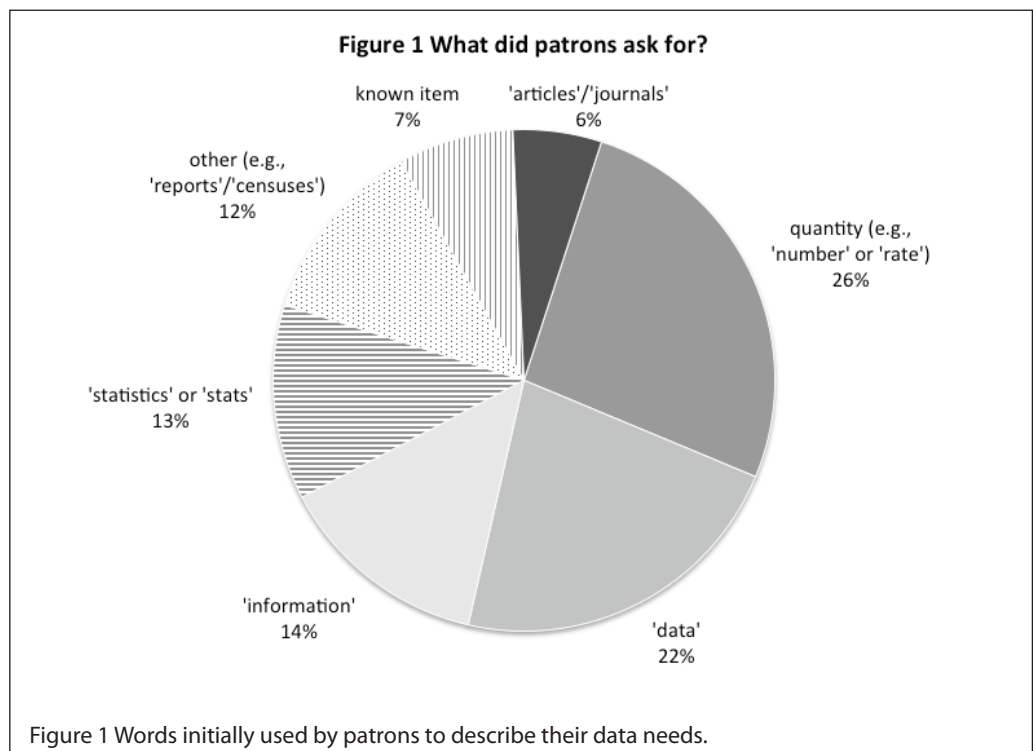
about how the users conceptualized data. In consulting the library and information science literature for other studies on how users formulate information requests, we came across an article that examined reference questions submitted to archives staff via email (Duff and Johnson, 2001). We expanded the scope of our coding beyond the patron's initial statement of need, categorizing the overall kinds of information given and wanted by the patron, as Duff and Johnson had done.

Preliminary Findings

Below is a quantitative and qualitative snapshot of some of the observations and themes we have been able to extract from the data thus far using the iterative processes of coding and categorization.

General Observations

Not all patrons asked for 'data' in the data reference questions we identified. In fact, users invoked various terms to describe their data needs. Figure 1 breaks down the frequency of language that patrons used to communicate their need for data.

Figure 1 Words initially used by patrons to describe their data needs.

For each transcript in this new sample, we started by examining only the patron's opening question, unnegotiated in any way by the librarian. We made observations about more easily categorizable and quantifiable aspects, like what time period was requested, as well as more qualitative, nuanced observations on the phrasing used by the patron. As before, we separately compiled lists of our observations; these ended up being extremely similar. Where there was no difference in what was observed, we created a corresponding code. Where disparity occurred, we discussed potential options and implications until consensus was achieved. We then applied this coding scheme to the transcripts.

We were interested in exploring further the qualitative aspect of the users' questions, potentially using this to develop theories
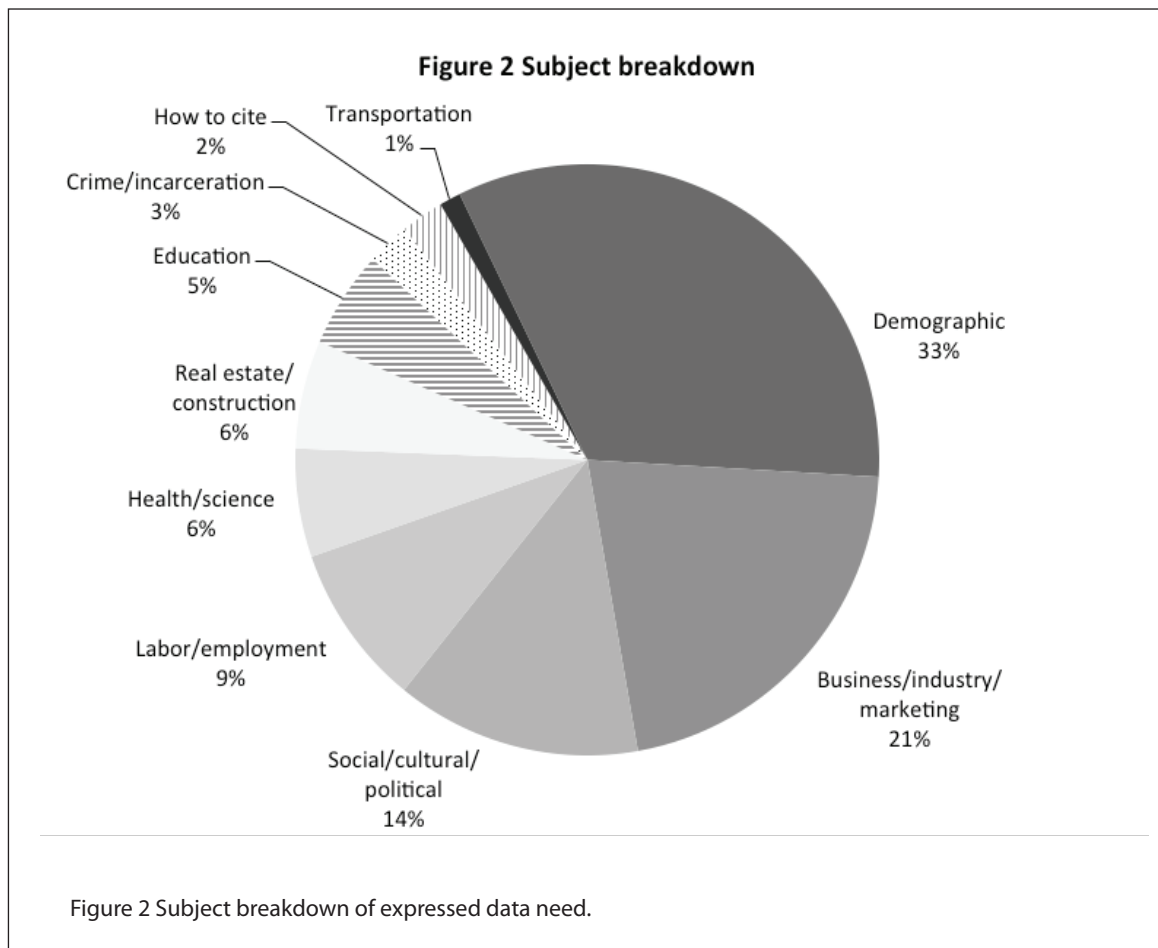
Roughly one quarter of users did ask explicitly for 'data'. Another quarter of users used alternative language that implied that they were looking for quantitative or numeric information, while a third quarter asked for either 'information', 'statistics', or 'stats'. The remaining quarter of users asked for specific publications types that possibly contained data, e.g., journal articles, research reports, or books.

Some patrons were very specific about temporal and geographic aspects of their data needs, while others were not. In some cases, this information was freely given in their opening statements; in others, such details emerged through a reference interview. Overall, 49% of users voiced data needs that included a specific time period; of those, 38% sought historical data or data from a range of years, while 9% sought the 'most recent' data available.

In contrast, only 4% of users indicated a specific time scale (e.g., annual, decadal). 82% of users asked for data from a specific geographic location; of those, 68% sought United States data and 27% sought New York City data. 79% of users described data needs that included a particular geographic scale; of those, 32% sought city-level data, 17% sought country-level data, and 12% sought neighborhood-level data. In many cases, it was difficult to know exactly which geographic scale a patron actually needed unless it was expressed at the most granular level. For example, a user asking for New York City data may have actually needed data on Harlem (a neighborhood within New York City), which they may have thought–correctly or incorrectly–would be findable in the city-level data set.

The nature of patrons' data needs also varied across subject area, as Figure 2 demonstrates.

of data reference as well as demand for specialized data and/or subject expertise in our sample.

Emerging Themes

Data analysis is still ongoing, but a number of themes have emerged that are worth further exploration. Although there are many interesting themes related to patrons' question topics, librarian responses, and general characteristics of the interactions, the ones described below focus on patron behavior, and specifically on how patrons pose their initial questions to the librarian.

The Easiest/Fastest Way

The first theme describes when a patron specifies that they are not only looking for data or statistics, but specifically for a faster or more efficient way than they can devise on their own. Several examples appear below:

Figure 2 Subject breakdown of expressed data need.

Patron: I'm wondering what is the most efficient way to find NY Census data from 1840-1940...I just need general numbers/demographics

—

Patron: hi - i'm trying to figure out how many Italians immigrated to the US at the end of the 19th-early 20th century Patron: is there an easy way to find this?

—

Patron: Hello, I am trying to locate health statistics for the borough of Brooklyn from the census. Can

Nearly one third of all the data queries we identified were in reference to demographic data, while roughly one fifth were in relation to business, industry, and marketing data. Together, demographic and business data reference questions constituted the bulk of our data set.

Lastly, 36% of the transcripts we identified showed 'referral activity'. This means that they had been transferred between different librarians within NYU's LibraryH3lp system, that the librarian had consulted with another librarian during the course of the chat, or that the librarian had given the user another librarian's contact information for follow-up. This suggests the collaborative nature

you suggest a link? The census is a bit convoluted and I am a bit rushed.

By asking the question in this manner, the patron could be implying that they believe they have the ability to find what they are looking for if only they had enough time to do it. Along the same lines, they could also be phrasing their question this way to 'save face'–that is, to make it seem to the librarian like they are more confident about their searching abilities than they really are. The patron could also be admitting that they know that what they need is likely to exist, but know that they lack the skills to find it.

Ask (For) A Librarian

Instead of asking for help finding data, several patrons instead asked directly for a person who might know the answer to their question. For example,

> Patron: Hello- is there someone who is great with using the Census website?

—

> Patron: Hi - where would I find someone who knows about Gov Docs?

—

The patrons who asked their questions this way showed a fairly sophisticated understanding of the library's reference service; that is, they understood the concept of specialist librarians, that many data and statistics questions go beyond the realm of general reference, and that there are librarians on staff who specialize in data and statistics areas. Of course, it is difficult to know the patron's true mindset in phrasing a question like this, but it could be read as either benevolent (indicating to the general reference librarian that it is ok if they do not know how to help with a very specialized question) or impatient (immediately asking for a specialist knowing that communicating with the generalist may not be a good use of time).

'Am I In The Right Place?'

On the other hand, many patrons began their conversation with the librarian by admitting their inexperience with the reference service model in asking a first question about whether or not the librarian might be able to help them, or verifying what they might expect to receive from the librarian. Here are a few examples of this:

> Patron: I'm looking for information regarding United States annual steel production as far back as possible, to present
> Librarian: ok
> Patron: would you be able to help me find that info? perhaps recommend some material

—

> Patron: Hey I have to find some figures on topics based on cities, if I were to tell you some of these topics do you think you can give me a hunch on where to start or which databases would be helpful?

—

> Patron: I have a question about citing US census data?
> Patron: (I'm not sure if that's something you could help me with)

Interestingly, this patron could potentially have the same spectrum of intentions as the savvier patron who asked for a librarian above. By expressing doubt about whether the librarian can help, they again make it ok for the librarian to say they do not know how to help (or to give basic help or make a referral) and it potentially saves time by making sure they are asking the question to the right person in the right place.

Authority

Another common theme arose, relating to the authority and reliability of sources the patron had already found–a theme that will be unsurprising to anyone who does any reference or information literacy instruction. For example:

> Patron: Hi, I need an academic source that establishes the years for all living generations. Could you help me find a reputable source?

—

> Patron: hello! im looking for demographics on southern brooklyn (birth rates, sex, age population). we are not allowed to use wikipedia as a resource

—

> Patron: i can't seem to find what i want
> Patron: is indexmundi.com a reliable source?

—

> Patron: I am researching the recents stats of homelessness in NYC
> Patron: how can I find accurate numbers?

This theme suggests a more substantial knowledge gap for the patron–the lack of ability to evaluate the reliability and authority of a source–but also the wherewithal to acknowledge this gap and ask for help. It is difficult to tell from most chat transcripts whether these patrons were interested in authority for the sake of an assignment (i.e., their instructor told them they can only use authoritative sources) or for the sake of having reliable data for their own projects or needs, but it is likely that both types are represented.

'Where' vs. 'How'

Another interesting distinction that emerged was that some patrons ask for 'where' to find the data they need while others ask 'how' to find it. For example:

> Patron: I was trying to find demographic information from 1980 to 1990 for Far rockaway, ny
> Patron: where should I look

—

> Patron: Hi, do you know where can I find the total number of college students in specific cities

Versus:

> Patron: Hello, I need to find cities in US where people need to use public transportation a lot
> Patron: Do you know how can I find the data?

—

> Patron: i want to find the revenue number of taobao.com, an ecommerce website in China
> Patron: could you show me how to find the numbers? thank you

While this could simply be a result of different manners of speaking (rather than something deliberate and worth analyzing), it could also reveal clues into the way different patrons are thinking about their data needs and questions. Both patrons seem to assume that the data they need exists, but the one who asks 'where' also seems to believe that once they know where to look, then the process of extracting or accessing it and understanding what it means will be easy or at least doable. This patron could be a more experienced data user, or could be overestimating their abilities. The patron who asks 'how' is acknowledging that they do not know how to approach searching and possibly also does not know what to do

once the desired data are found. Looking at whether a patron asks 'where' or 'how' may also tell us something about where the patron is in the research process, for example, if they are looking for data or statistics to support an argument that they have already made, or if they are in a more exploratory stage.

Unanswerable

Finally, we will explore a broader, more complex category of patron questions that we have chosen to classify as unanswerable for some reason or another. This does not mean that the question is not legitimate or should not have been asked, only that the way that it was asked makes it impossible to answer at face value. Essentially, these questions are ones that require a good reference interview on the part of the librarian, and looking closely at the original phrasing of the question gives us interesting insight into how the patron was thinking about the information need and approaching it for the sake of the librarian. There are several flavors of the unanswerable theme, which are discussed after each example, below.

> Patron: Hello - I have been searching a statistic for two days and I have been unsuccessful and running out of time :(can you help me?
> Patron: i am trying to find the uninsured rate (for healthcare) in Canada - and cannot for the life of me find it
> Patron: I know Canada has universal health care but I need a solid statistic within the past 5 years of those citizens that are uninsured

In this case, the patron is asking for something that they admit should not logically exist: if Canada has universal health care, then there should not be any uninsured Canadian citizens, and therefore no statistics on the number of such citizens. Even so, the patron clearly has an information need; it is reasonable to assume that they are aware of this logical fallacy, so the librarian's job is to help clarify that need and then help fulfill it. This is, in fact, what happened over the course of this chat conversation. It could be that the patron had spent sufficient time on this project such that when they asked the question, they forgot that the librarian would not have the same context to understand what was meant by this query.

The example above also hints at the patron's challenge in operationalizing concepts into variables that are likely to exist and be available. This was observed many other times too, for example:

> Patron: Hello I'm currently working on a project about the changing face of Jersey Street in New Brighton, Staten Island. How would you advise that I find out the culture of crime in that area from 1950s through now?
> Librarian: hi there
> Librarian: are you looking for books? articles? statistics?
> Patron: stats please

In order to find statistics on the 'culture of crime' in a certain area, this patron will need to decide how the concept should be defined and measured first.

> Patron: how would I find the specific ethnic breakdown and class breakdown of East Los Angeles? I need information on that specific region

Likewise, while there exist some standardized definitions for collecting data on ethnicity (though these can and should be scrutinized), there are no similar standards for data on 'class' in the United States. This patron will need to clarify what they mean by 'class breakdown' before they can find statistics about it in East Los Angeles.

Other users asked for things that were simply unlikely to exist or be available publicly or through library databases, for example:

> Patron: I need statistics for US tomato consumption in 1840s, 1850s...thru 1900. USDA stats start in 1886

Many reference librarians will read this patron's statement as a successful search: the patron identified the correct authority most likely to have the data if they exist; however, since that authority does not have them, the answer is that those data almost certainly do not exist. There could be some additional discussion about proxy variables or other creative places to look and maybe this patron's need could still be satisfied, but the interesting part is the difference in how the patron thought about this question versus the way a professional librarian would.

Most of the queries that fall into this category also raise questions about what background work the patron has already done and whether they might be better served by looking for books or articles instead of data sources.

Limitations and Next Steps

We acknowledge, of course, that our approach has limitations. While chat transcripts allow us to look back at reference transactions in a way we never could with in-person reference, we also do not have any feedback about the experience from the patron or the librarian. As a result, it is very difficult to truly know what the patron really wanted, or whether the patron or librarian considered the interaction successful. Furthermore, the concept of a successful interaction is complicated by the fact that user satisfaction or dissatisfaction does not necessarily equate to a correct or complete answer from the librarian. For example, is an interaction successful if the librarian determines that the desired data exist, but only in PDF format, and then the user leaves discouraged? Or if the librarian gives an answer that is wrong or incomplete but the patron is happy with the answer? Additionally, the set of transcripts we extracted may be incomplete, because it is difficult to identify transactions where neither the patron nor the librarian recognized a data need, which may be among the most interesting interactions.

There are many additional themes in the chat transcripts in our data set; this investigation is a preliminary exploration of how patrons ask data-related questions. More themes– and their relationships to one another–will be discussed in future publications.

A grounded theory approach suggests that the next phase of this project will be to begin exploring the relationships between themes and determining what this data set is a study of (Grounded Theory Institute, 2014). From the themes already uncovered, we have several pressing questions:
- Are these themes specific to census-related questions? Are they even specific to data-related questions? Or are they more generalizable to all chat reference?
- Is there a relationship between any of these themes and the overall success or failure of the reference interaction?

• How do these examples fit into established models of 'question-asking'?

Once we have built a theory or theories from the data, the final step will be to integrate them into the established literature and articulate how our work moves the conversation forward, possibly adding to a growing body of knowledge about the librarian's role in supporting the data lifecycle.

In addition to the theoretical advantages of understanding our users, there are practical aspects of this inquiry as well. This project gives us a rare opportunity to look closely at some of the problems our users and librarians are having with data in reference transactions and to think about how we can improve our services for the benefit of all. In better understanding the kinds of queries we receive, and the ways data needs are conceptualized and articulated, we hope to build better data research guides for our patrons and improve the training, scripts, and guides available to the librarians staffing the service.

One clear way to improve service is to offer training to library staff on how to use open-ended questions during the data reference interview. As evidenced by questions classified within the 'unanswerable' theme, users often have an incomplete understanding of how to operationalize concepts into variables that could be found in existing data sources. Training that allows staff to practice asking the kinds of open-ended questions that will help users and librarians move toward a shared understanding of what the user needs, and what exists, will translate into more effective data reference interactions.

Our analysis also shows that users struggle with questions related to the reliability and authority of data sources. This could be communicated efficiently through an online guide showing the who, how, and why of data creation, collection, and distribution, as well as strategies for evaluating sources. Making this kind of a convenient takeaway available allows librarians to more easily seize a teaching moment, and enrich and expand the learning experience beyond the immediate data reference interaction. These guides are especially valuable because they make it easier for generalists staffing the service to convey specialized information. These are just two possible ways to improve service based on our preliminary findings. As demand for secondary data grows across academic disciplines, strengthening the data reference piece of a larger reference program that is staffed by specialists and generalists alike ensures the future health and relevance of academic reference services.

References

Bryant, A. and Charmaz, K. (2007) The SAGE Handbook of Grounded Theory. [Online] London: SAGE Publications Ltd. Available from: http://srmo.sagepub.com/view/the-sage-handbook-of-grounded-theory/SAGE.xml. [Accessed: 7th July 2015].

Duff, W. and Johnson, C. (2001) A virtual expression of need: an analysis of e-mail reference questions. The American Archivist. [Online] 64 (1). p.43-60. Available from: http://americanarchivist.org/doi/abs/10.17723/aarc.64.1.q711461786663p33. [Accessed: 13th July 2015].

Gerhan, D. R. (1999) When quantitative analysis lies behind a reference question. Reference & User Services Quarterly. [Online] 39 (2).

p.166-176. Available from: http://www.jstor.org/stable/20863727. [Accessed: 7th July 2015].

Grounded Theory Institute. (2014) What is grounded theory? [Online] Available from: http://www.groundedtheory.com/what-is-gt.aspx. [Accessed: 7th July 2015].

Kellam, L. and Peter, K. (2011) Numeric Data Services and Sources for the General Reference Librarian. [Print] Oxford, UK: Chandos.

Payne, G. and Payne, J. (2004) Key Concepts in Social Research. [Online] London: SAGE Publications Ltd. Available from: http://srmo.sagepub.com/view/key-concepts-in-social-research/SAGE.xml. [Accessed: 7th July 2015].

Read, E. J. (2007) Data services in academic libraries: assessing needs and promoting services. Reference & User Services Quarterly. [Online] 46 (3). p.61-75. Available from: http://www.jstor.org/stable/20864696. [Accessed: 7th July 2015].

Wang, M. (2013) Supporting the research process through expanded library data services. Program. [Online] 47 (3). p.282-303. Available from: http://www.emeraldinsight.com/doi/abs/10.1108/PROG-04-2012-0010. [Accessed: 7th July 2015].

Notes

1. Margaret Smith is Physical Sciences Librarian at New York University and can be reached by email at margaret.smith@nyu.edu.

2. Jill Conte is Librarian for Sociology, Psychology, and Gender & Sexuality Studies at New York University and can be reached by email at jill.conte@nyu.edu.

3. Samantha Guss is Social Sciences Librarian (for Political Science, International Studies, Geography & the Environment, Government Information, and Data & Statistics) at the University of Richmond and can be reached by email at sguss@richmond.edu. She was previously Data Services Librarian at New York University from 2009-2014.

4. American Factfinder is the United States Census Bureau's online tool for accessing data. The American Community Survey is a demographic survey that complements the United States Decennial Census. Social Explorer is a commonly used commercial database that repackages U.S. Census and other data.