

NCknows Report
Usage Statistics
From February – December 2004

14 February 2005

Prepared by:

Jeffrey Pomerantz

School of Information and Library Science
University of North Carolina at Chapel Hill
CB 3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
pomerantz@unc.edu

Lili Luo

School of Information and Library Science
University of North Carolina at Chapel Hill
CB 3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
luolili@email.unc.edu

Charles R. McClure

School of Information Studies
Florida State University
101 Louis Shores Building
Tallahassee, FL 32306-2100
cmclure@lis.fsu.edu

Executive Summary

This report contains some simple statistical analyses of the NCKnows project, from its launch on 16 February 2004 through 30 December 2004. There are three evaluation questions that this analysis was designed to answer:

1. What patterns are emerging in the volume of questions received by NCKnows?
2. What patterns are emerging in the volume of questions received and handled by individual libraries?
3. What patterns are emerging in the demographics of the patrons using NCKnows?

Results of this analysis show that the NCKnows service is receiving a great deal of use, and primarily among the user community to which it is geared: North Carolinians. The service shows a healthy growth, and a reasonable division of labor given the disparities in the amount of experience with virtual reference possessed by the member libraries.

Introduction

This report contains some simple statistical analyses of the NCKnows project, from its launch on 16 February 2004 through 30 December 2004. This report contains analyses of the volume of questions received by the NCKnows service, and handled by NCKnows librarians, according to various sets of criteria.

NCKnows (www.ncknows.org) is a collaborative state-wide chat reference service, coordinated by the State Library of North Carolina's Library Development Section (statelibrary.dcr.state.nc.us). Libraries of all types are participating: academic, public, and government; in research universities and community colleges, urban and rural, large and small. NCKnows was launched in February 2004, and as of this writing NCKnows is in an 18-month pilot phase.

From the outset of the NCKnows project, the evaluation was designed to take three perspectives: evaluation from the perspective of (1) NCKnows patrons, (2) the individual libraries participating in NCKnows, and (3) the entire NCKnows collaborative effort. This analysis allows evaluation from the points of view of the entire NCKnows effort.

There are three evaluation questions that this analysis was designed to answer:

1. What patterns are emerging in the volume of questions received by NCKnows?
2. What patterns are emerging in the volume of questions received and handled by individual libraries?
3. What patterns are emerging in the demographics of the patrons using NCKnows?

Statistical analyses such as those presented in this report, though simple compared to other data analysis efforts, are critical to the evaluation of a virtual reference service. Tracking data on the use of the service by month, day, library, and other criteria, is important so that the service will be able to make informed decisions concerning the most effective deployment of resources. Demographic data on the patrons (collected here and also in the patron exit survey, which will be presented in a separate report) is important to understand who makes up the service's user community. Statistical analyses are also essential to determine the degree to which user information needs are being met and the responses by reference staff in meeting those needs. Finally, usage data may be tracked longitudinally, providing useful measures of the growth of the service over time.

Methodology

The 24/7 application stores the full text of all chat sessions, and metadata about sessions. This metadata includes such data points as the patron's login time and date and the queue through which the patron logged into the service, the duration of the patron's wait time in the queue, the duration of the chat session, the patron's username, the librarian's username who interacted with the patron, and the resolution code with which the chat session was closed.

This data may be downloaded from the 24/7 application in Microsoft Access format, one file per month. This was done, and the monthly files were merged to form one “master file” containing data spanning February through December 2004. The analyses presented in this report were performed on the data in this master file, using Microsoft Access’ query functionality.

Some chat transcripts were filtered out as inappropriate for inclusion in the data analysis. The NCKnows service was launched on 16 February 2004, so all chat transcripts from 1-15 February were removed. All transcripts during this period of time were test sessions, conducted to test the 24/7 application, or were training sessions between two librarians gaining experience with the use of the application. All transcripts (after 16 February) that were closed with the resolution code TEST were also filtered out of the master file. Unfortunately resolution codes are not used consistently and some training sessions were not closed with the TEST resolution code. Because there is no way to automatically filter these transcripts out of the master file, some test transcripts are undoubtedly included in the set of transcripts that were analyzed for this report. For the same reason, of course, it is impossible to know how many of these test transcripts there are.

Results

The master file of NCKnows transcripts, with all transcripts prior to 16 February 2004, and all transcripts closed with the resolution code TEST filtered out, included a total of 4,417 transcripts.

Table 1: Sessions handled by NCKnows

	Number of sessions	Percentage of sessions(%)
Total # NCKnows sessions via all queues	8,567	100
Total # sessions handled by NCKnows, overall	4,090	47.74
Total # sessions handled by 24/7 network, overall	4,477	52.26
Total # sessions handled by NCKnows, during NCKnows hours, on weekdays	3,653	71.16
Total # sessions handled by 24/7 network, during NCKnows hours, on weekdays	1,461	28.84
Total # sessions handled by NCKnows, during NCKnows hours, on weekends	373	61.54
Total # sessions handled by 24/7 network, during NCKnows hours, on weekends	233	38.46

When counting questions that come in through all of the NCKnows queues during NCKnows' hours of service (10 AM – 8 PM on weekdays, and 1 – 6 PM on weekends), NCKnows librarians handle approximately 60-70% of the sessions, and the 24/7 network handles approximately 30-40%.

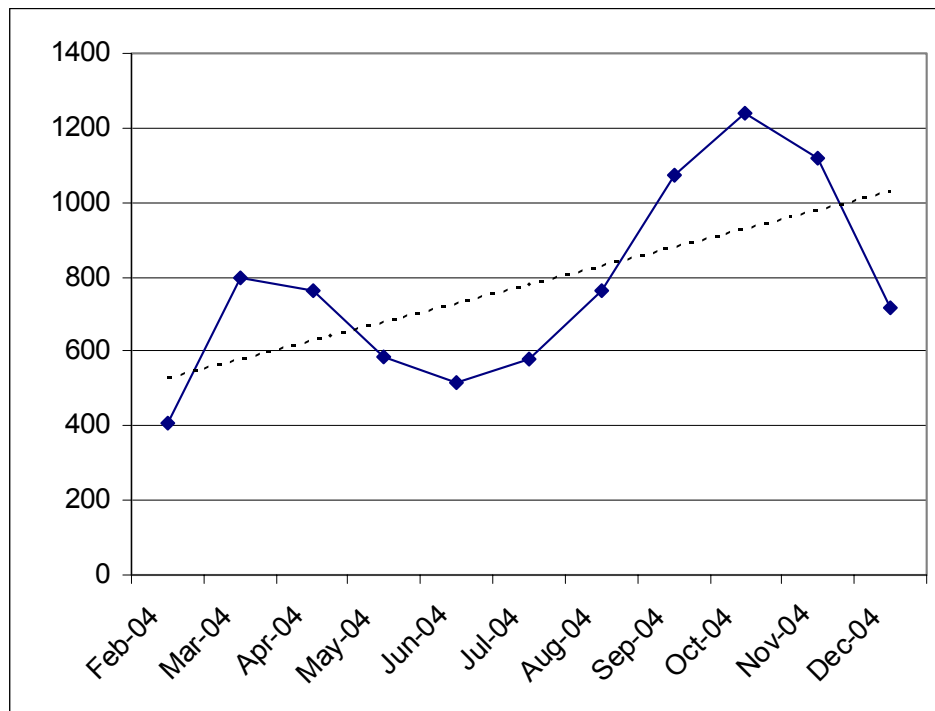
When counting questions that come in through all of the NCKnows queues during all hours of the day, NCKnows librarians handle approximately 48% of sessions, and approximately 52% are handled by the 24/7 network.

In the September 2004 statistics report, this split was 45/55%. After three months, the number of chat sessions handled by NCKnows increased by 3%. Although this is only a slight increase, it indicates that NCKnows is progressing on picking up more questions from their queues. One possible explanation for this change is: since PLCMC canceled its own contract with 24/7, the questions that used to be handled by the 24/7 network are now split between NCKnows and 24/7.

Table 2: Number of Sessions per Month (n = 8,567)

Month	Number of sessions
February	410
March	798
April	764
May	586
June	518
July	579
August	762
September	1,073
October	1,237
November	1,120
December	720

Figure 1: Number of Sessions per Month



The volume of chat sessions handled by NCKnows in February is low because the service was launched mid-month, so only 2 weeks of data is represented, instead of the 4 weeks

for other months. Thus the volume of chat sessions shows an overly optimistic spike between months 1 and 2. The pattern of rise and fall in NCKnows' question volume, however, is typical: other studies of the volume of questions received by virtual reference services, both academic and public, show a fluctuation corresponding to the academic calendar (see, for example: Bailey-Hainer, 2005; Bushallow-Wilber, DeVinney, & Whitcomb, 1996; Sloan, 2001). The overall trend is towards increased volume over time, as shown by the dotted trendline. The trendline indicates that NCKnows' volume has been increasing at an average of approximately 50 questions per month.

Table 3: Sessions per Day of the Week (n = 8,567)

Day of week	Percentage of sessions
Sunday	9.98
Monday	17.95
Tuesday	18.92
Wednesday	17.79
Thursday	15.94
Friday	11.67
Saturday	7.74

Figure 2: Sessions per Day of the Week

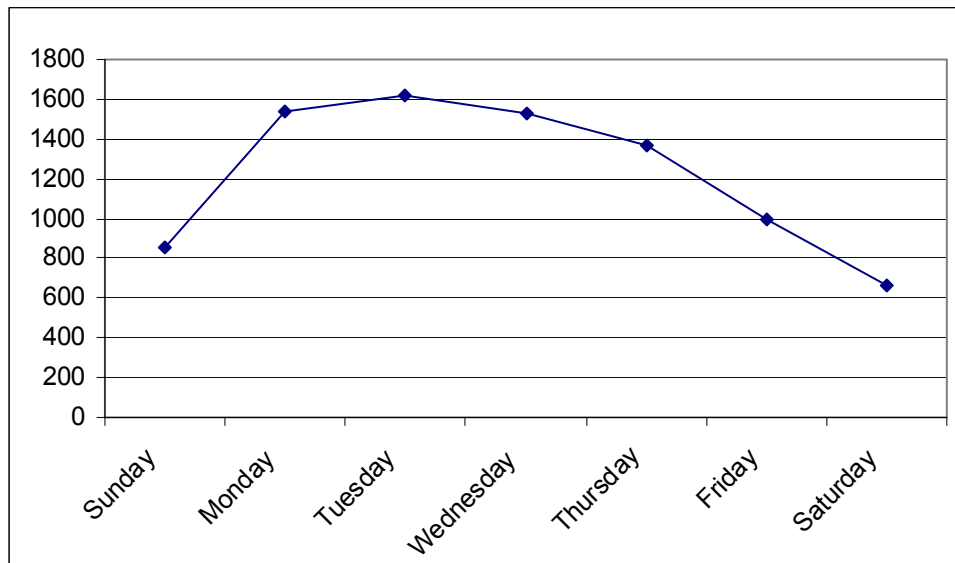


Table 3 presents the volume of questions received by day of the week, in total from the launch of the service through the end of December. Monday, Tuesday, and Wednesday see the heaviest and an almost equal volume of chat sessions, followed closely by Thursday. Volume then drops off on Friday, and is lowest over the weekend. This is consistent with other studies of the volume of questions received by virtual reference services (see, for example: Hill, Madarash-Hill, & Bich, 2003).

Table 4: Sessions per Time of Day, Weekdays (n = 7,049)

A. Hour	B. % of Total	C. NCknows % (n = 3,704)	D. 24/7 % (n = 3,345)
Midnight - 1 AM	1.25	0.00	100.00
1 - 2 AM	0.64	0.00	100.00
2 - 3 AM	0.38	0.00	100.00
3 - 4 AM	0.35	0.00	100.00
4 - 5 AM	0.62	0.00	100.00
5 - 6 AM	0.16	0.00	100.00
6 - 7 AM	0.45	0.00	100.00
7 - 8 AM	1.15	0.00	100.00
8 - 9 AM	3.11	0.46	99.54
9 - 10 AM	4.81	9.73	90.27
10 - 11 AM	6.55	66.88	33.12
11 AM - Noon	8.00	77.30	22.70
Noon - 1 PM	6.61	74.25	25.75
1 - 2 PM	7.56	70.73	29.27
2 - 3 PM	8.58	71.90	28.10
3 - 4 PM	8.31	76.45	23.55
4 - 5 PM	7.73	65.32	34.68
5 - 6 PM	7.16	75.45	24.55
6 - 7 PM	6.10	65.58	34.42
7 - 8 PM	5.93	67.70	32.30
8 - 9 PM	4.70	4.53	95.47
9 - 10 PM	4.13	0.69	99.31
10 - 11 PM	3.49	0.00	100.00
11 PM - Midnight	2.21	0.00	100.00

Table 5: Sessions per Time of Day, Weekends (n = 1,518)

A. Hour	B. % of Total	C. NCknows % (n = 386)	D. 24/7 % (n = 1132)
Midnight - 1 AM	1.05	0.00	100.00
1 - 2 AM	0.79	0.00	100.00
2 - 3 AM	0.53	0.00	100.00
3 - 4 AM	0.20	0.00	100.00
4 - 5 AM	0.13	0.00	100.00
5 - 6 AM	0.20	0.00	100.00
6 - 7 AM	0.20	0.00	100.00
7 - 8 AM	0.72	0.00	100.00
8 - 9 AM	2.31	0.00	100.00
9 - 10 AM	3.36	3.92	96.08
10 - 11 AM	4.74	2.78	97.22
11 AM - Noon	5.27	1.25	98.75
Noon - 1 PM	7.51	7.02	92.98
1 - 2 PM	7.38	59.82	40.18
2 - 3 PM	8.89	56.30	43.70
3 - 4 PM	8.70	67.42	32.58
4 - 5 PM	8.10	62.60	37.40
5 - 6 PM	6.85	61.54	38.46
6 - 7 PM	5.73	0.00	100.00
7 - 8 PM	6.59	0.00	100.00
8 - 9 PM	7.11	0.00	100.00
9 - 10 PM	5.73	0.00	100.00
10 - 11 PM	4.35	0.00	100.00
11 PM - Midnight	3.56	0.00	100.00

Column B: The percentage of all chat sessions received during the hour in column A. The numbers in this column add up to 100%.

Column C: The percentage of chat sessions handled per hour by NCKnows librarians.

Column D: The percentage of chat sessions handled per hour by 24/7 librarians.

NCKnows receives 91.56% of questions on weekdays, and 9.44% on weekends.

The important comparison to make in tables 4 and 5 is between columns C and D. Recall from Table 1 that NCKnows is handling 48% and 24/7 is handling 52% of sessions that come in through all of the NCKnows queues. This is true overall, but not necessarily for any given hour of the day. During the hours when NCKnows is not providing service, 24/7 handles 100% of the incoming sessions, which is to be expected. During the hours that NCKnows is providing service (10 AM – 8 PM on weekdays, and 1 – 6 PM on weekends), NCKnows handles as much as 71% of the incoming sessions on weekdays and 62% on weekends. Conversely, however, 24/7 handles as much as 29% on weekdays and 38% on weekends.

Figure 3: Sessions per Time of Day, Weekdays

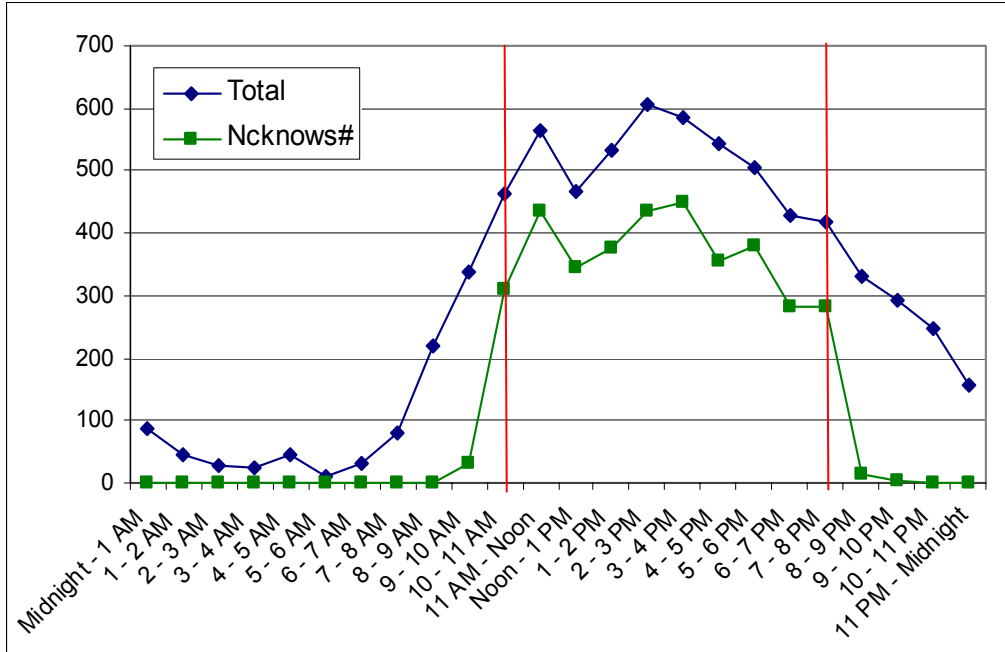


Figure 4: Sessions per Time of Day, Weekends

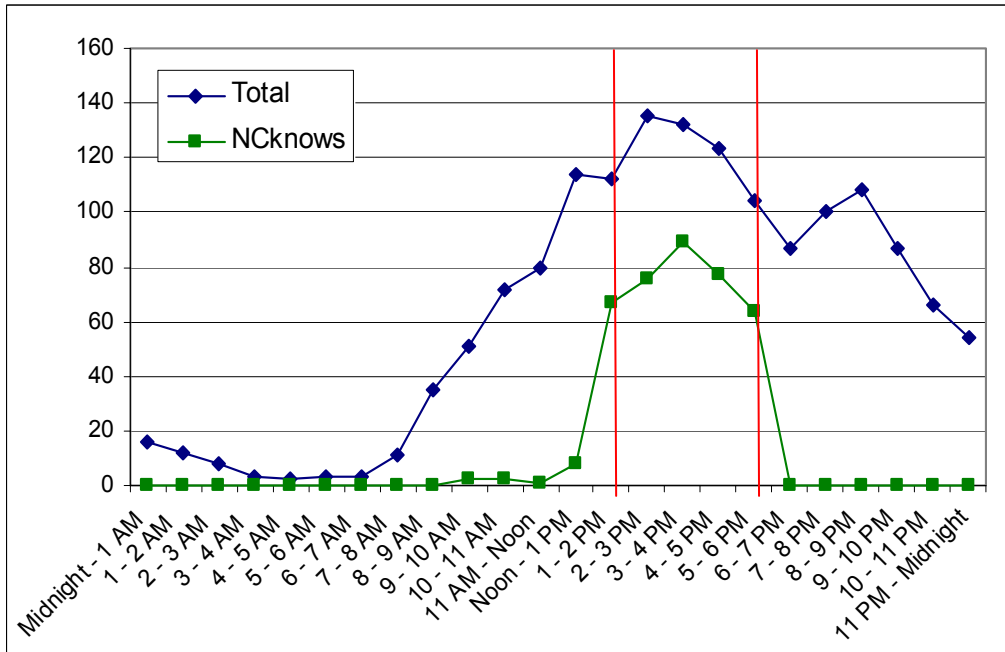


Table 6: Net asker / Net answerer

A. Library	B. % sessions incoming via library's queue (n = 8,567)	C. % sessions answered by librarians in library (n = 4,090)
Appalachian State University – NCknows	0.85	1.47
Appalachian State University	0.63	
AB Tech	2.54	4.40
Brevard College	0.18	0.17
Caldwell County Public	0.05	0.02
Central Carolina Community College	0.82	1.10
Central Piedmont Community College	4.94	6.01
Cumberland County Public Library	6.85	7.70
Davidson County Community College	0.33	0.32
Davidson County Public Library	0.58	0.66
Edgecombe Community College	0.20	0.17
Forsyth County Public Library	2.46	3.08
Greensboro Public Library	5.30	6.82
Guilford Technical Community College	0.04	0.05
Haywood County Public Library	1.62	2.20
Hickory County Public	0.61	0.81
High Point Public Library	0.74	1.00
Johnston Community College	0.16	0.20
Madison County Library	0.14	0.17
McDowell Technical Community College	0.02	0.02
NC_A	0.08	0.10
NC_P	0.18	0.27
North Carolina A&T	1.75	1.83
NCLIVE	1.19	1.39
New Hanover County Public Library	0.05	0.05
Pfeiffer University	0.49	0.61
Piedmont Community College	0.04	0.07

Pitt Community College	0.23	0.29
Public Library of Charlotte & Mecklenburg County	40.30	26.45
Polk County Public Library	0.35	0.49
Sampson-Clinton Public Library	0.14	0.24
State Library of North Carolina	6.30	6.45
University of North Carolina – Chapel Hill	2.00	1.81
University of North Carolina – Greensboro – NCKnows	2.08	9.49
University of North Carolina – Greensboro	3.96	
University of North Carolina – Wilmington	0.74	0.73
Union County Public Library	0.06	0.05
Wake County Public Libraries	6.22	7.41
Wake Technical Community College Library	4.14	4.91
Wake Forest University School of Medicine Library	0.39	0.66
Wayne County Public Library	0.27	0.29

Column B: The percentage of all chat sessions received by NCKnows that were received via each library’s queue. These numbers are a measure of each library’s contribution to NCKnows’ total volume of questions.

Column C: The percentage of all chat sessions received by NCKnows and handled by NCKnows librarians that were handled by librarians in each library. These numbers are a measure of each library’s contribution to providing answers for NCKnows.

Each row represents the traffic brought into NCKnows and handled by each library. If $B > C$ in a row, then the library is a “net asker”: the library is bringing a greater percentage of sessions to NCKnows than they are answering (e.g., PLCMC). If $C > B$, then the library is a “net answerer”: the library is answering a greater percentage of sessions than they are bringing (e.g., Greensboro Public Library).

Some libraries have two rows (e.g., ASU and UNC-G). This is because these libraries have two separate 24/7 queues, one for NCKnows and one for their own chat service. These “individual” chat services are staffed during different hours than the library staffs 24/7. Because each 24/7 queue has a unique name, it is possible to separate the chat sessions that were received by these libraries via these two separate queues. Because librarians in these libraries do not have separate usernames for the times that they are staffing these two separate queues, it is not possible to separate the chat sessions that were handled by librarians during the hours in which one service or the other is staffed.

To get an accurate picture of the net asker / net answerer status of these libraries, add the values in the two cells in column B and compare the sum with the value in column C.

Clearly PLCMC is the greatest net asker by far. This is a reflection of the fact that PLCMC has been offering chat reference service since February 2002 (a full two years longer than NCKnows), and has marketed the service during that time, and has thereby built up a sizeable user base. On the one hand, this might be interpreted as unfair, that PLCMC is not pulling its weight, or that other libraries are picking up PLCMC's slack. This is, however, not at all the case. The values in column B are percentages of all 8,567 chat sessions received by NCKnows, while the values in column C are percentages of only the 4,090 chat sessions handled by NCKnows librarians. In a sense, it is an unfair comparison, since NCKnows librarians are handling only 48% of the chat sessions received by NCKnows (see Table 1). To be fair, 24/7 is the largest net answerer of all.

Throughout the duration of this project, PLCMC has consistently received the greatest volume of questions of any of the participating libraries. PLCMC's net asker status is on the decline, however: the percentage of all chat sessions received by NCKnows that were received via PLCMC's queue dropped from 42% to 40% between September and December 2004. Although this is not a big change, it confirms our speculation that as the NCKnows service matures and engages in marketing efforts over time, the difference between PLCMC's and other libraries' net asker statuses will diminish.

As PLCMC's net asker status is on the decline, PLCMC's net answerer status is on the rise: from 8.06% to 26.45% between September and December 2004. One possible explanation is that PLCMC librarians can answer more questions than other librarians in the same period of time, as they have had two years more experience with answering chat reference questions than other librarians involved in NCKnows. Another possible explanation is that when PLCMC cancelled their separate account with 24/7, many of the questions that were formerly handled by 24/7 network now have to be answered by PLCMC or by NCKnows. It's therefore reasonable to see a big increase in the volume of questions handled by PLCMC at this early stage after PLCMC has fully merged with NCKnows.

Table 7: Percentage of Users with North Carolina or Other Zip Codes (n = 7,679)

Month	Percentage of sessions
North Carolina	75.78
Other States	5.44
Indeterminate	18.78

When a patron logs into NCKnows, one of the initial fields that the patron is asked to fill out is Zip Code. 76% of patrons type zip codes that are within North Carolina, and only less than 5% from states other than NC. Zip code could not be determined for 19% of patrons: this includes patrons who did not submit anything in the Zip Code field (what is captured in this situation is the value “anonymous”), as well as patrons who submitted meaningless values (for example, 4- or 6-digit numbers, or non-numeric values).

The zip code field was captured for 7,679 chat sessions. It is not clear why the zip code field was not captured for all 8,567 chat sessions, since even if the patron did not submit anything in this field, what would have been captured would be the value “anonymous.” For whatever reason, the 24/7 application did not capture all zip code submissions.

There have been a few patrons who have submitted locations outside the United States in the Zip Code field. These patrons were removed from this analysis of zip codes, since they obviously could not submit a zip code, but it was still possible to determine where they were located. These patrons were from the following nations: Cameroon, China, Norway, 2 from France, and 4 from Canada. The patrons from Cameroon and China have been interviewed following up on their chat sessions. The results from these interviews will be presented in a separate report.

It is impossible to know how reliable this zip code data is. Although there is little incentive to do so, it is as easy for a user of a chat-based reference service to lie about his or her location as to tell the truth. NCKnows has no mechanism to determine the veracity of a user’s response to this question, but patrons may not understand this. It would not be difficult for a patron outside of North Carolina to look up North Carolina zip codes, if they believed that it was required to log into NCKnows. The evaluation team believes that this is unlikely, however. Web users are used to interacting with web forms these days, and when something is required on a web form it is usually highlighted in some way. The zip code field on the NCKnows login page is not indicated as required, so there is no reason for patrons to think that they must have a North Carolina zip code in order to log in. Further, although the patrons from Cameroon and China came across NCKnows via Google searches, it is unlikely at this point that NCKnows is known by many patrons outside of North Carolina. After marketing efforts have been launched for NCKnows, however, this may change.

Table 8: Percentage of Users who are Library Card Holders (n = 3,089)

Month	Percentage of sessions
Yes	71.41
No	9.04
Indeterminate	19.55

71% of patrons submitted either “yes,” some variation of yes, or their actual library card number and 9% submitted “no” or some variation of no. 20% could not be determined: this includes patrons who did not submit anything in the Library Card Number field (what is captured in this situation is the value “anonymous”), as well as patrons who submitted meaningless values (for example, a string of all 0s, or non-numeric values).

Again, it is impossible to know how reliable this library card holder data is. The 24/7 application does not check a library card number against any sort of database of library card holders; indeed, the 24/7 application does not even require that the patron submit an actual card number. It would be difficult for a patron to make up a library card number, but a patron could simply say that they do have a card when they do not. Again, the evaluation team believes that this is unlikely. Because the library card holder field is not indicated as required, there is no reason for patrons to think that they must have a library card.

Recommendations

Recommendation: Extend the hours that NCKnows librarians staff the service. NCKnows receives the greatest volume of questions during its regular hours of service. NCKnows also receives a lesser, but still significant volume of questions for a few hours before the service’s regular hours, and for a few hours after. Librarians employed by 24/7 handle all of the chat sessions during NCKnows off-hours, and this equals 58% of the chat sessions that come into NCKnows. This percentage could be lowered if NCKnows’ hours of service are extended (see Figures 3 & 4). The evaluation team recommends that NCKnows should extend its hours of service to:

- 7 AM – midnight on weekdays, and
- 8 AM – 1 AM on weekends

Recommendation: Join the 24/7 network. During the hours that NCKnows is providing service, 24/7 handles as much as 29% on weekdays and 38% on weekends (see the discussion of Tables 4 & 5). Note that “24/7” refers to two different entities: “24/7 the company of contracted librarians” and “24/7 the network of affiliated libraries.” PLCMC used to be a member of the 24/7 network, but NCKnows is not. At present, only 24/7-the-company may handle NCKnows sessions, while 24/7-the-company and 24/7-the-network handled PLCMC sessions.

In an analysis of PLCMC's chat traffic conducted in January 2004 (that is, prior to the launch of NCKnows), the evaluation team discovered that PLCMC librarians handled 16%, and the 24/7 network handled 84% of the sessions received by PLCMC. Of these sessions, 86% were from users in North Carolina (67% from users in the Charlotte area alone), and 8% were from users outside of North Carolina. PLCMC made out well in this: for a comparatively minimal investment in supporting users outside of the PLCMC's target user community, the PLCMC chat service more than quadrupled the volume of transactions that they were able to handle during their hours of service. And this figure takes into consideration only the 2 hours per day, 5 days a week during which the PLCMC's service is staffed. Taking the full 24 hours per day into consideration, the PLCMC chat service has increased the volume of transactions that they were able to handle per day many times over.

As was discussed in the Transcript Peer Review Report, NCKnows librarians were rated slightly but not significantly higher than 24/7 staffers. Thus there is little evidence that the responses from NCKnows librarians are better or worse than those responses from 24/7 employees. And consider further: many 24/7 employees are librarians, but some are not. But members of the 24/7 network are librarians, in other libraries, so to join the 24/7 network would mean to increase the percentage of sessions that are handled by librarians, rather than by 24/7 employees. It thus makes sense for NCKnows to join the 24/7 network. Benefits will include the ability to handle greater volume of chat sessions with minimal additional impact on NCKnows librarians, and an increase in the percentage of sessions that are handled by librarians.

Recommendation: Market the NCKnows service aggressively. This recommendation is not borne out of the findings from this statistical analysis, but is instead borne out of the previous two recommendations. If the hours of NCKnows were to be extended, and NCKnows were to join the 24/7 network, it would be necessary for NCKnows to bring in sufficient volume of sessions to make those investments worthwhile. In order to bring in an increased volume of sessions, it will be necessary to perform marketing and PR campaigns for the NCKnows service.

Recommendation: There should be clearer guidelines regarding the proper use of the Resolution Codes at the conclusion of chat sessions. This issue was discussed in the Recommendations section of the Transcript Peer Review Report, and so will not be discussed further here.

Conclusion

The NCKnows service is receiving a great deal of use, and is showing a healthy growth, consistent with the documented growth of other virtual reference services. While there are some inequities in the division of labor in terms of bringing in and handling chat sessions among the member libraries, these inequities are reasonable given the disparities in the amount of experience with virtual reference possessed by the member libraries, and are therefore likely to even out over time, as member libraries gain user bases.

The use that the NCKnows service is seeing is primarily among the user community to which it is geared: North Carolinians. The primary user community for NCKnows is therefore the same as the primary user communities of all of the member libraries. It is even possible that NCKnows users are a superset of the user communities of each individual library – in other words, NCKnows may be reaching patrons that all member libraries collectively had not been reaching previously. This is not something that can be answered from the data analyzed in this report, but will be investigated in the analysis of the patron exit survey and follow-up interviews, which will be presented in a future report.

References

Bailey-Hainer, B. (2005). Virtual Reference: Alive & Well. *Library Journal*, 130(1), 46-47. <http://www.libraryjournal.com/article/CA491140>.

Bushallow-Wilber, L., DeVinney, G., & Whitcomb, F. (1996). Electronic Mail Reference Service: A Study. *RQ*, 35(3), 359-369.

Hill, J. B., Madarash-Hill, C., & Bich, N. P. T. (2003). Digital Reference Evaluation: Assessing the Past to Plan for the Future. *Electronic Journal of Academic and Special Librarianship*, 4(2-3). http://southernlibrarianship.icaap.org/content/v04n03/Hill_j01.htm.

Rogers, E. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.

Sloan, B. (2001). *Ready for Reference: Academic Libraries Offer Live Web-Based Reference. Evaluating System Use* (webpage). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. <http://www.lis.uiuc.edu/~b-sloan/r4r.final.htm>.