Chat Transcript Analysis Project (CTAP)

A Review of Work Completed and Findings

Kindra Morelock, Team Lead
Jessica Alverson
Jennifer Schwartz
Ana Lucic

Phase 1: Background

In 2017, the Reference, Instruction, and Academic Engagement (RIAE) Department at the DePaul University Library embarked on an analysis of our IM Reference Chat Transcripts. RIAE has been providing a virtual chat reference service since 2009 but no action had been taken to analyze the content of the transcripts in order to determine what kinds of questions patrons were asking through the service. LibraryH3lp was DePaul University Library's chat provider from 2009 - late 2017.

Total chats declined from 9,569 in FY15 to 7,596 in FY17, a 21% decrease over four fiscal years, see Chart 1. There are at least two contributing factors to this decline. First, the overall DePaul University enrollment has declined over this same period, from 23,539 to 22,769 students, a 3% decrease. Enrollment in the School for New Learning, whose students tend to be some of the most frequent library users, declined from 1,207 to 905, a 15% decrease. Second, the way statistics were collected and reported by LibraryH3lp changed during 2015. Prior to the change, each chat - no matter how short - was counted as its own transaction. After the change, short chats that were ended prematurely or accidentally, and that were part of continued conversations between an individual user and the librarian, were combined and were no longer counted separately.

While there may be additional reasons why the number of chats has declined, RIAE decided to look into the content of the chat and to see if there were some patterns that might emerge.
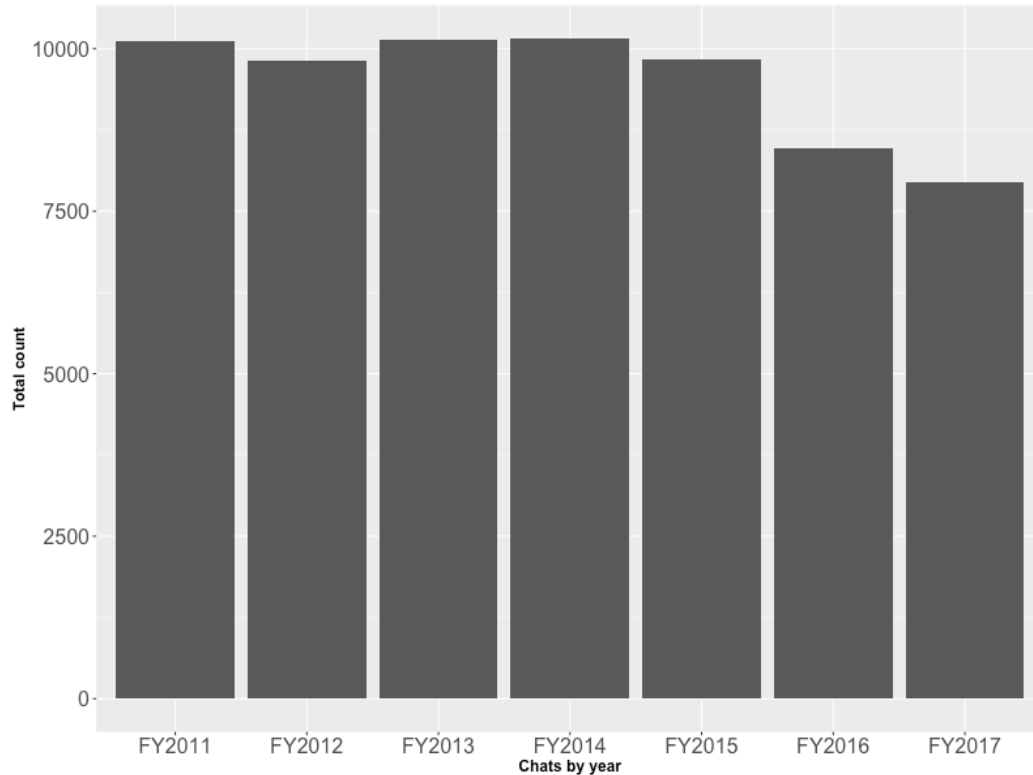
Chart 1: Count of chats by Fiscal Year

Phase 1: Methodology

Chat Transcript Coding

The first phase of the project began in summer 2017. Kindra Morelock first performed some basic metadata analysis: number of total chats, number of chats by month, number of chats by hour, etc. She then extracted a representative sample from fiscal years 2014 - 2017 using a 95% confidence level and a confidence interval of eight. 147 chats were extracted for FY2014 and FY2015; 148 for FY2016 and FY2017.

Once the chats were extracted, they were divided into approximately seven different documents and given to those who had agreed to code the chats according to the coding schema that had been agreed to beforehand.

Named Entity Recognition Analysis

Chats frequently mention vendors and databases (EBSCO, Proquest, Lexis Nexis) but also services such as ILLIAD or WRD (WRD instruction session). We were interested in establishing which databases, vendors and services patrons most frequently refer to in the chats. We focused on the patron text only and ignored the lib rarian's responses as the response would

frequently repeat or mirror the question asked by the patron. Named entity extraction, among other things, can help us identify people's names, dates, locations, and organizations in the text. Although services (WRD , ILLIAD) are not organizations, they are frequently recognized as organizations in the text. We used [Stanford Named Entity Recognizer](#) to recognize and extract the category "ORGANIZATION" i n the chats.

Phase 1: Results

Chat Transcript Coding Results

After the chats were coded, Kindra inputted the information into the quantitative analysis software, NVivo.

Each chat was labelled with the appropriate RHEAD scale level. The RHEAD scale was developed at DePaul to track the type and complexity of questions asked by our patrons at the research help service points.. RHEAD scale 1 indicates a basic directional question; 2 indicates a more complex directional question, 3 indicates a basic reference question, 4 indicates a more complex reference question, and 5 indicates an extended professional exchange that takes place over several interactions.

The results show that the greatest number of chats we receive are actual reference questions (RHEAD 3 and RHEAD 4). RHEAD 4 questions declined between FY2015 and 2016. RHEAD 1 questions rose 50% from FY2014 - 2017. See Table 1 for more details.

	2014	2015	2016	2017	Totals	Percent of the Total	Percent Change from 2014 - 2017
82 : RHEAD 1	14	18	25	21	78	14.8%	+ 50%
83 : RHEAD 2	24	24	15	25	88	16.8%	+ 4%
84 : RHEAD 3	43	43	56	41	183	34.8%	- 5%
85 : RHEAD 4	50	50	37	37	174	33.1%	- 26%
86 : RHEAD 5	2	0	0	0	2	0%	N/A

Table 1: FY 2014 - FY 2017 representative chat coding sample: RHEAD scale

We have also determined that October, February, and May are traditionally the busiest months on IM reference, while December, July, and August see very little traffic (see Chart 2).
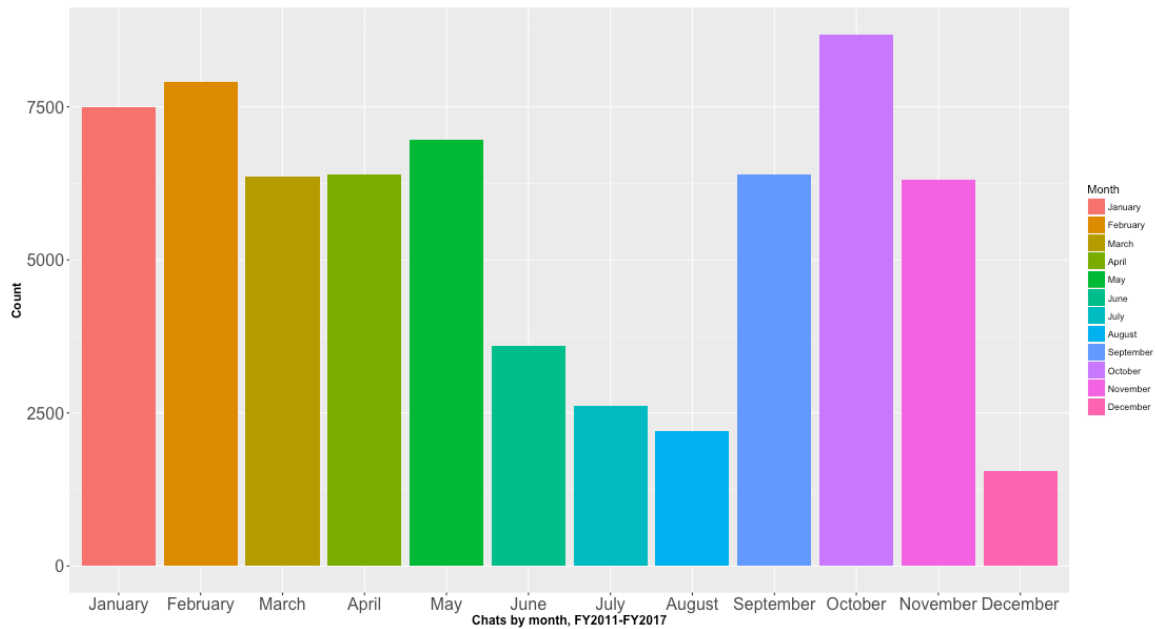

Chart 2: Count of chats by month, FY2011 - FY2017

Chart 3 stacked bar graph shows the percentages per month and per fiscal year from January until December. September, October, January, and February comprise approximately 50% of our total chats for the year. Summer months get fewer chats with the chat activity going up again from September until November. Although FY2016 and 2017 saw a drop in the overall number of chats (Chart 1), the proportion of chats by month does not vary greatly over the seven fiscal years (when divided by the total number of chats).

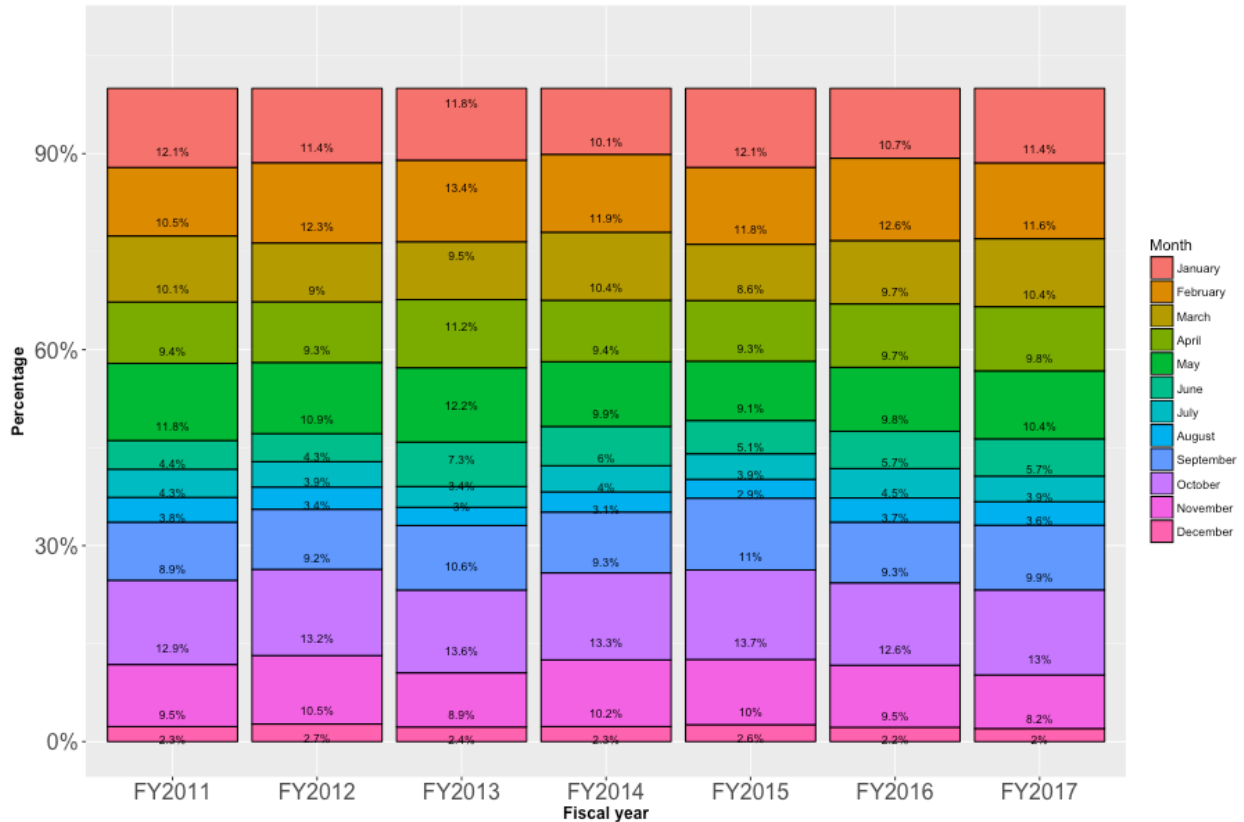Chart 3: Stacked bar chart indicating percentage of chats per fiscal year

Named Entity Recognizer Results

A total of 931 organizations have been mentioned by patrons in the chats from FY2013 - FY2017. Not surprisingly, DePaul is the most frequently referred to entity in the chats and, for this reason, was not included in Chart 4 that represents the 25 most frequently mentioned organizations in chats during this period.
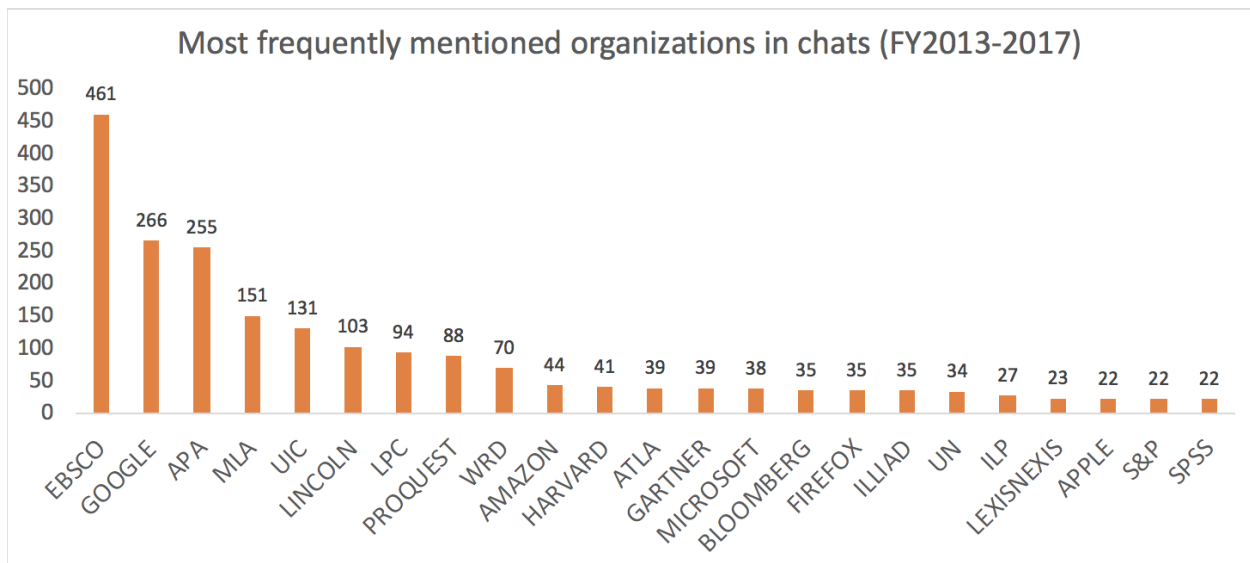
Chart 4: 25 most frequently referred to organizations in chats (FY2013 -2017)

As Chart 4 indicates, one of the most frequently referred entities in our chat sample is EBSCO. Google and APA feature most prominently after EBSCO with MLA following closely by.

Ana plans to focus on the analysis of chats that mention EBSCO as well as APA and MLA and establish the nature of these chats: are the patrons asking for help identifying an EBSCO database? Are they asking for links to APA and MLA tutorials, or are particular APA journals being referenced or the MLA International Bibliography? A few examp les that use APA and MLA in the chat are included below:

"*not sure how to cite in APA 6th edition*"

"*Oh some articles will let me view the article cited in various formats. this one didn't. any recommendation? oh actually i do need it APA sorry! so I guess I am good then with that*"

"*Oh, just one last question. It seems my instructor would like for us to use APA citations. I thought with research papers, we were to use MLA. any thoughts there?*"
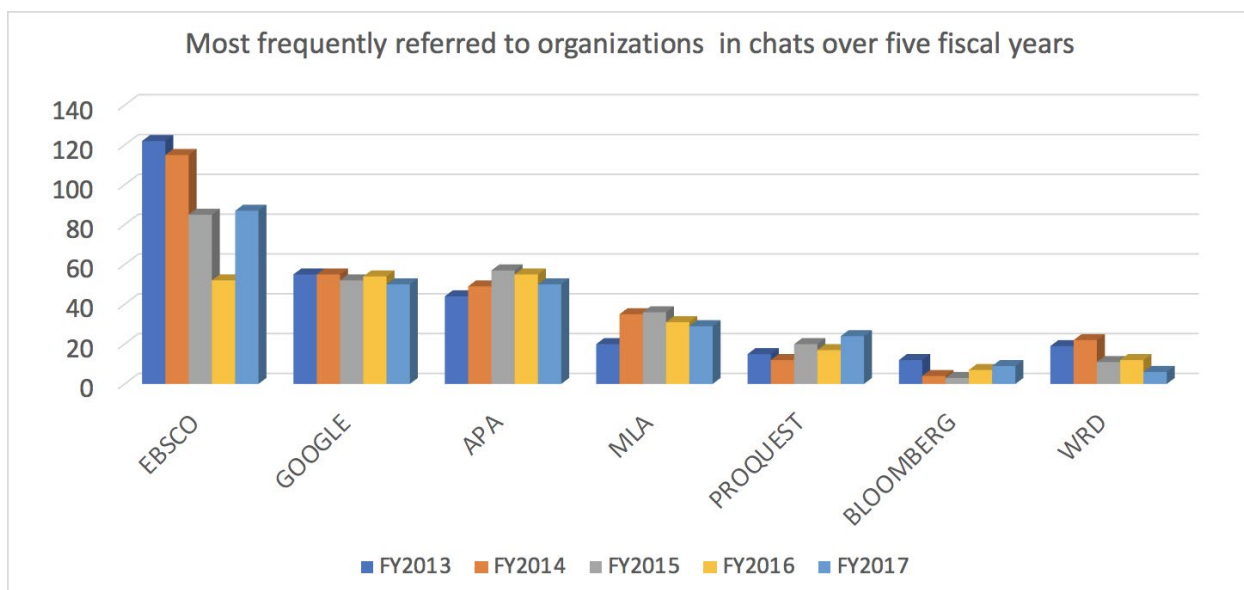
Chart 5: Most frequently referred to organizations over the f ive years (FY2013-2017)

A similar type of analysis can be conducted with any of the referenced entities. Chart 5 indicates that EBSCO references over the years vary with FY2016 seeing a significant drop. APA references show a rise over the period FY2013 -2015 and then a slight drop. We also see a slow but steady progression of references to Proquest in chats during the analyzed period.

A more detailed analysis of named entities in the chat transcripts is planned for f uture work. Patrons refer to different organizations in the chat environment differently; they may use acronyms or misspell or write differently the organization's name (e.g. Lexis/Nexis, LexisNexis, Lexis-Nexis). We still need to establish whether *EBCO* or *EBASCO* refer to EBSCO or some other organization. Future analysis will require a more thorough consolidation of named entities. Present analysis has standardized the references by converting all organizations to uppercase and by expanding a number of acronyms.

Phase 2: Background

Once Phase 1 was completed, we decided to look at only the chats that originated from the DePaul University Library homepage: https://library.depaul.edu. See Figure 1 for image of the website and "Chat Now" button.
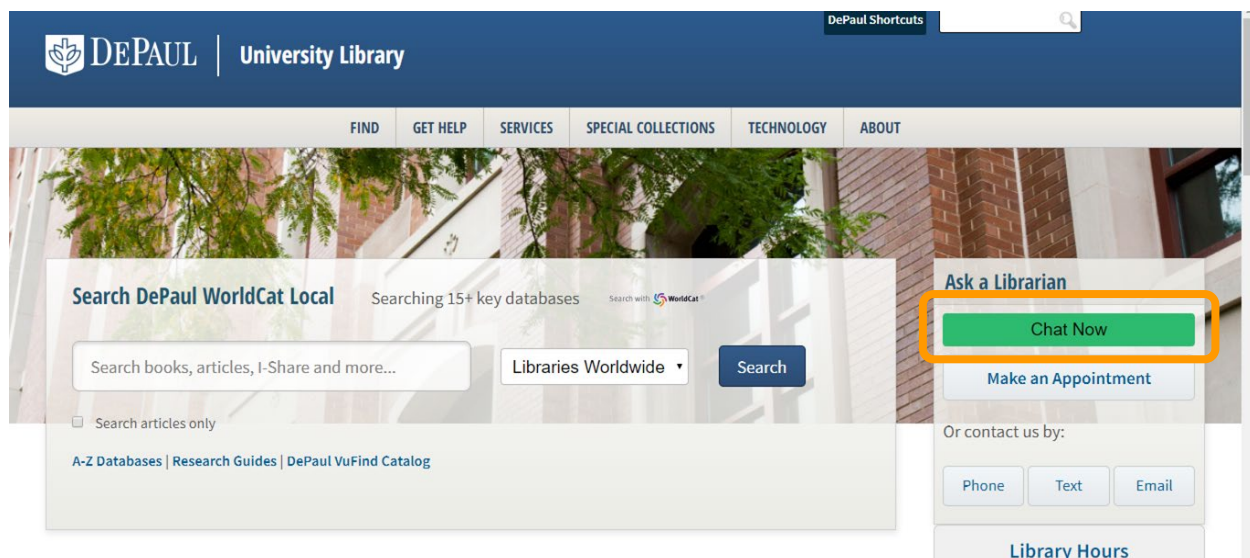
Figure 1: DePaul University Library homepage (accessed 2018 August 2)

Unsurprisingly, this is by far the page from which the highest number of chats originate due to the Chat Now button's prominence on the landing page, as well as the fact that in the first -year writing library instruction session that RIAE librarians teach, the chat function is highlighted and is generally demonstrated from the homepage.

The CTAP team came up with this research question: "What is the initial question that patrons have that causes them to click on the Chat Now button on the homepage?"

Phase 2: Methodology

Kindra Morelock extracted the full transcripts from any chat in FY2017 that originated from the library homepage URL. There were 2,620 chats that she then evenly split into four different

files. The assigned codes were extracted from annotate.co and the files were restructured using Python and then visualized/analyzed using Tableau.

Phase 2: Results

The results of the homepage chat analysis mirrored the results from Phase 1. Approximately 28.7% of the chats originating from the homepage are categorized as "Research Help," which generally corresponds to a RHEAD scale 4. Known Item Searches (abbreviated in our codes as KIS) are generally comparable to a RHEAD scale 3. One hypothesis was that there would be more chats about general library information, such as library hours from the homepage and fewer research help questions, but that is not the case. See Charts 6 and 7 for details.

Homepage Chats - Tag Frequency - FY17

Tags	Number of Records
Research Help	689
KIS Book	179
KIS Article	150
GLI Library Policies	136
MISC	115
GLI Hours	107
Citation	99
IShare	89
GLI Other	85
Non-Library Question	75
GLI Library space	72
GLI Library Privileges	63
KIS Other	62
ILLIAD	58
KIS Database	54
ARES	53
GLI Equipment	53
KIS Journal	43
KIS Ebook	40
Printing	39
GLI Billing and Fines	31
KIS Textbook	22
GLI Borrower ID	21
KIS Media	20
KIS Software	17
Noise problem - LPC	17
Noise problem - Loop	1

Tags (group):
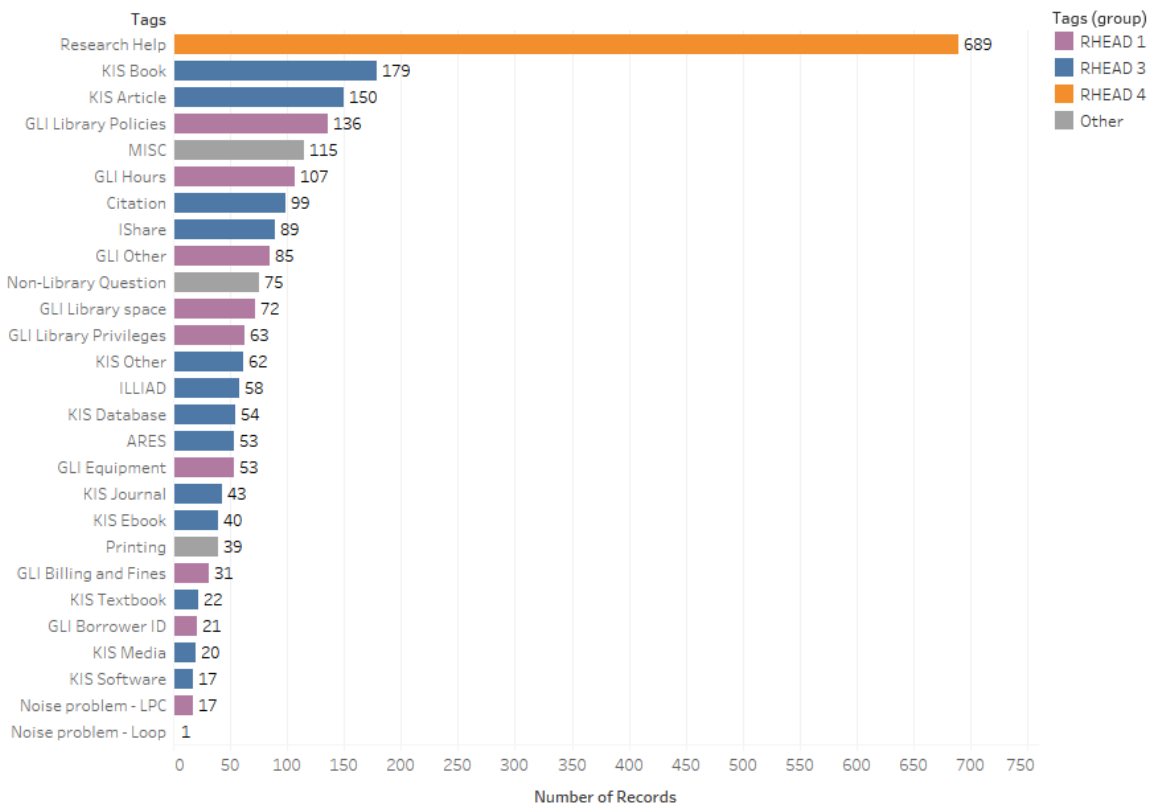- RHEAD 1
- RHEAD 3
- RHEAD 4
- Other

Chart 6: Number of chats originating from the DePaul University Library homepage by tag

Homepage Chats by Category - FY17

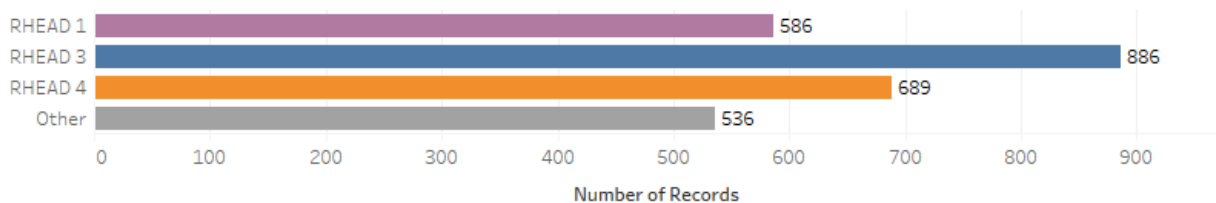| Category	Number of Records
RHEAD 1	586
RHEAD 3	886
RHEAD 4	689
Other	536

Chart 7: Number of chats originating from the DePaul University Library homepage aggregated by category

Homepage Chats by Category - FY17

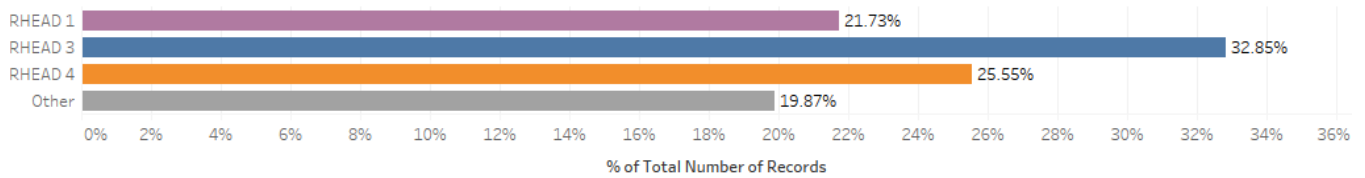| Category	% of Total Number of Records
RHEAD 1	21.73%
RHEAD 3	32.85%
RHEAD 4	25.55%
Other	19.87%

Chart 8: Number of chats originating from the DePaul University Library homepage aggregated by category, percentage of total

Contrasting this is the 9% of RHEAD scale 4 and 5 face-to-face questions we receive at either the Loop or Lincoln Park Research Help Desks out of 18,902 transactions. See Chart 9 and 10 for details.
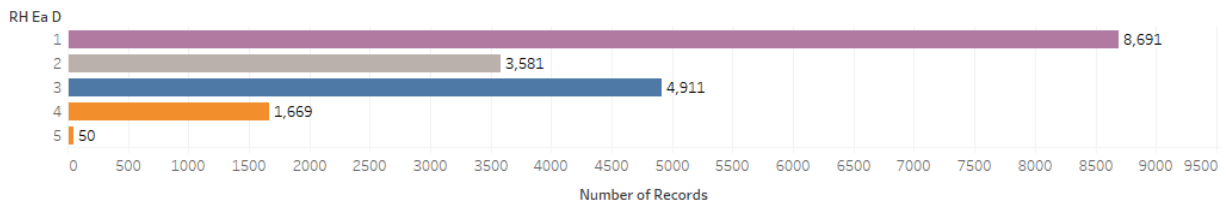
RHEAD Scale - Desk Transactions - FY17

RH Ea D

	Number of Records
1	8,691
2	3,581
3	4,911
4	1,669
5	50

Chart 9: Total Research Help Desk questions by category

RHEAD Scale - Desk Transactions - FY17

RH Ea D

	% of Total Number of Records
1	45.98%
2	18.95%
3	25.98%
4	8.83%
5	0.26%
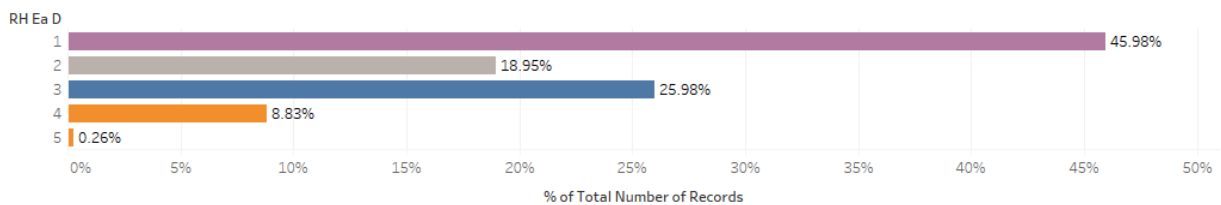
Chart 10: Total Research Help Desk questions by category, percentage of total

Discussion

The Chat Transcript Analysis Project began as a way for RIAE to learn about what patrons were asking over their IM chat service. It is a unique data source that gives a glimpse into the content of both the questions that the patron asks as well as the responses that library staff give to the varied queries. It is a point-in-time glimpse into the various pain points that users have with library services. RIAE staff have anecdotally had ideas of what patrons asked over chat, but this is the first large-scale project aimed at analyzing the transcripts and finding out where our patrons need the most help with regards to library services. This information can be used to pre-emptively give them that help, through increased visibility on the website, additional FAQ's, or other means.

This analysis has demonstrated the success of the IM chat reference service in providing research help to patrons at point of need and irrespective of physical location. Based on our coding of FY14-17 transcripts, 68% of the chats from FY14-17 are either RHEAD scale 3, 4, or 5, indicating true reference questions. The ability to use a chat service instead of face-to-face communication allows for greater access for those with disabilities, distance learning students, commuter students, or new students who feel uncomfortable with face-to-face interactions with a professional.

There are significant differences between the types of questions we receive at either campus at the research help desks and through IM chat reference. This can be in large part attributed to the mode of the service: IM chat patrons are not likely to ask where the bathroom is located or to use a stapler, for example. New students may feel shy about approaching a professional; the current generation of college students are likely as or more comfortable with communicating through technological mediation than they are face-to-face.

However, we are still experiencing a drop in IM chat transactions. Some theories of why the chat numbers have dropped include:

1. Enrollment declines, particularly in the School for New Learning
2. Patrons are finding what they need more easily on the website
3. Changes to the way that LibraryH3lp counted individual chats
4. Possible changes to vendor interfaces where the chat widget disappeared or is no longer prominent
5. Low patron satisfaction - fewer repeat patrons

Theories 2 and 5 could be evaluated through doing UX testing and/or post-chat surveys. For Theory 4, the EIM working group, in conjunction with Technical Services, is looking at options for DePaul branding on vendor databases and ways to get back to the DePaul Library website (or embedding the chat widget directly into the database).

Conclusions
- The DePaul University Library's IM chat reference service is very successful in providing research help to patrons at point of need and irrespective of physical location
- The types and proportion of questions we get over chat vs. at the desk is different, most likely because of the mode of service

- Because so many chats are research oriented, PRTs who are manning the chat service, particularly after hours, must be well trained and refer to a librarian when necessary
- Additional staffing with professional librarians during October and February should be considered

For Future Investigation

- Compare with findings after WorldCat Discovery goes live [no referring widget]
- Compare with a full year of LibChat
- Where are patrons when they chat -- on or off of campus? If they are in the Library, why do they choose chat vs. walk up assistance?

Appendix A - Codes by Year

	2014	2015	2016	2017	Totals
1 : 2nd Floor Renovation	0	0	0	0	0
2 : Academic Department	0	0	0	1	1
3 : Alumni	2	0	2	0	4
4 : ARES	4	3	0	1	8
5 : Can't find information on website	2	5	3	3	13
6 : Computer Literacy Issue	8	3	3	3	17
7 : Conflicting information or access	0	0	3	0	3
8 : Database downtime	0	0	0	0	0
9 : Serials Solutions	0	0	0	1	1
10 : WorldCat Local	0	0	1	3	4
11 : Department (Academic)	1	0	0	0	1
12 : Discard chat	9	12	12	12	45
13 : E-Books	5	9	4	5	23
14 : E-Resources Ticket	3	1	0	4	8
15 : General Library Information	0	4	1	1	6

16 : Borrower ID	7	0	0	0	7
17 : Hours	8	5	3	4	20
18 : Library Location	0	0	2	0	2
19 : Other	20	15	17	16	68
20 : Policies	12	18	25	21	76
21 : ILLiad	9	16	5	12	42
22 : I-Share	8	8	8	9	33
23 : Access to Non-DePaul e-books	1	0	1	0	2
24 : Account creation	2	1	7	1	11
25 : Request materials	4	5	13	6	28
26 : Known Item Search	0	1	0	0	1
27 : Article	17	17	15	13	62
28 : Book	17	21	23	14	75
29 : Database	6	4	5	3	18
30 : Journal	0	4	3	5	12
31 : Media	2	1	3	1	7
32 : Other	4	4	2	3	13

33 : Librarian doesn't respond	5	2	1	4	12
34 : Library Equipment	0	3	0	0	3
35 : Non-DePaul patron	3	4	5	2	14
36 : Non-Library Question	2	0	1	7	10
37 : Patron Exit	1	0	0	0	1
38 : Doesn't respond	7	8	4	6	25
39 : Midstream	1	4	1	2	8
40 : Referral	13	22	15	22	72
41 : referring_URL	1	0	0	0	1
42 : About	1	0	0	1	2
43 : about_the_library	0	1	1	0	2
44 : hours	0	0	0	0	0
45 : locations	1	5	2	1	9
46 : Find	9	1	5	2	17
47 : A-Z_list	1	0	8	7	16
48 : collections	1	0	0	0	1
49 : microfilm_microfiche	0	0	0	0	0

50 : database	1	0	0	0	1
51 : EBSCO	14	14	8	11	47
52 : ProQuest	1	0	1	0	2
53 : serials_solutions	2	9	3	4	18
54 : worldcat	26	32	25	27	110
55 : Get_Help	3	0	0	2	5
56 : askalibrarian	0	9	14	16	39
57 : FAQ	0	0	0	1	1
58 : research_guide	17	8	10	9	44
59 : homepage	46	55	56	52	209
60 : no_referring_URL	16	11	8	4	39
61 : Services	0	0	0	0	0
62 : borrowing	1	0	0	0	1
63 : course_reserves	2	2	0	0	4
64 : library_instruction	0	0	1	2	3
65 : policies	0	0	1	0	1
66 : resources_current_students	0	0	3	1	4

67 : rooms	0	1	0	4	5
68 : technology	0	0	1	4	5
69 : SPCA	1	0	0	0	1
70 : Research Consultation	1	0	3	2	6
71 : Research Help	0	0	0	1	1
72 : Choosing a topic	0	3	2	2	7
73 : Citation help	2	5	1	1	9
74 : Finding a source(s) on a topic	38	36	33	34	141
75 : Identifying a database	3	14	9	2	28
76 : Other	4	5	2	2	13
77 : Primary source	1	1	1	1	4
78 : Search strategies	17	17	10	11	55
79 : Whether or not a source is scholarly	3	0	2	3	8
80 : Reserves	10	1	1	5	17
81 : RHEAD	0	0	0	0	0
82 : 1	14	18	25	21	78
83 : 2	24	24	15	25	88

84 : 3	43	43	56	41	183
85 : 4	50	50	37	37	174
86 : 5	2	0	0	0	2
87 : Textbooks	2	2	0	0	4
88 : Truncated Chat	2	7	3	5	17
89 : Type of answer provided	0	0	0	0	0
90 : Instructional	76	76	49	50	251
91 : Providing the answer (little to no instruction)	57	58	85	70	270
92 : WiFi Usage	2	1	2	0	5
93 : WorldCat Local interface	0	6	5	2	13
94 : Can't find known item	0	3	1	1	5