HAL open science

Swimming the Channels: An Analysis of Online Archival Reference Enquiries

Joseph Pugh, Christopher Power

To cite this version:

Joseph Pugh, Christopher Power. Swimming the Channels: An Analysis of Online Archival Reference Enquiries. 15th Human-Computer Interaction (INTERACT), Sep 2015, Bamberg, Germany. pp.99-115, 10.1007/978-3-319-22698-9_8 . hal-01609418

HAL Id: hal-01609418

https://inria.hal.science/hal-01609418

Submitted on 3 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Swimming the channels: an analysis of online archival reference enquiries

Joseph Pugh, Christopher Power

HCI Research Group, Department of Computer Science
University of York, UK, YO10 5GH
{jjp513, christopher.power}@york.ac.uk

Abstract. Archives of historical and cultural data, such as the UK's National Archives, receive huge volumes of enquiries from users. These have been seldom systematically studied, despite the obvious benefits to the organisations concerned and interaction designers. The literature looking at the spectrum of remote communications carried out by a modern archive is sparse. Similarly, there is a lack of information about the problems users are having with archival information systems, and no information on the distribution of problems or where in users' information seeking journeys they occur. This paper reports on a mixed-method study using content analysis and grounded theory to address these gaps in the literature. The results of the study indicate that users primarily are encountering problems knowing where to start looking or where to look next in their information seeking journeys. Further, these problems seem to create a deep anxiety or uncertainty in archive users which drives them to seek reassurance and guidance from human archivists, who will provide the type of disambiguation and support that current information systems do not. The paper closes with implications of this work on the future prioritisation of design practice and research in online archives.

Keywords. Archives, reference enquiries, content analysis, grounded theory social media, email, information seeking

1 Introduction

Archives of historical and cultural data, such as the National Archives in the UK (TNA), have millions of active users worldwide who approach them for a variety of different reasons. Whether it is students and professional researchers investigating the history of a nation, or members of the curious public mapping their family trees, each of them comes to archives with a need to find information. In the age of digitisation of documents and other resources, more often than not they are trying to find this information through online search engines and other interactive applications. These archival systems are difficult to navigate partly because archives are hard to navigate: they are a "crazy quilt" of inconsistent organisation and wildly varying metadata [6]. Some records may have extensive descriptions, approaching their full text, others may have lengthy lists of names to support genealogical researchers.

However, many records have extremely terse, general descriptions which do not assist keyword searchers ("Correspondence", "Despatches" or the dreaded "Miscellaneous"). The challenge of users finding what they want is further complicated by the range of different systems available in any one archive, let alone across archives, that users have to learn, remember and apply in the right context.

As a result of this complexity, patrons of archives often have to contact archives directly for assistance. Just as interaction with the resources has changed in the digital era, the interactions with archives as organisations have changed. As opposed to coming into the archive, remote users try to access documents online, and will often put enquires to staff through a variety of different digital channels from telephone to email and onward into social networking. As an example, TNA fields over 100,000 research enquiries annually, with the majority of those enquiries now coming from remote users.

However, there is currently little research on what types of questions are being asked of archives through these different channels by their remote user base. Whereas libraries have extensively studied their interactions with users, there is no such tradition in archives and thus relatively scant information, despite a surge in interest by archives in user centred services in recent years.

Due to this lack of characterisation, it is difficult to judge how the different channels are serving either users or archives. Are all of the channels serving the same function or do particular types of enquiries come to specific channels? If the latter, then the characterisation of these queries could provide insight to archives as to how best to manage services across those channels.

More interestingly, these enquiries may offer an indication of the needs of users in regard to the key activity of *information seeking* in the archives and could provide indications of where users are getting "stuck" in their information seeking journey. Are they looking for help in interpretation? Or are they just trying to find where to locate a document? Characterisation of these types of enquiries will lead to a better understanding of what is working and not working in terms of online information seeking in archives.

In this paper we present a qualitative study involving a content analysis of the enquiries that come to a large national archive that offers multiple channels for user engagement. In addition to this, we include an analysis of a popular online genealogy forum, Rootschat, for comparison. This latter dataset is intended to improve the external validity of the analysis of the archive based enquiries, and help identify if the dataset is representative of more general problems in information seeking in archives. In the next sections we will motivate this problem from the relatively sparse literature available on the subject of archival interactions with their users, and then describe the overall design, conduct and results of the study.

2 Literature Review

There is no question that archivists have come late to the field of user studies. Since the early 1980s when Elsie Freeman pointed out that archivists had "never examined

systematically who our users are" [7] there have been many calls for further research [11]. But a residual disinterest in user studies amongst archives professionals has ensured that the discipline knows far less about the behaviour and needs of its users than do librarians [8]. What progress has been made has appeared either in unpublished work commissioned directly by archives or published by a relatively small group of investigators. Foremost amongst them are Wendy Duff and her collaborator Catherine Johnson who together carried out seminal work in 2002 and 2003 when they examined the behaviour of two of the main user groups in archives, namely genealogists [5] and historians [4]. No collection of the views and strategies of these users in an archival setting had been available before, though researchers such as Samuel Wineburg had looked at how historians interpret documents in more artificial settings [17]. The same year, Elizabeth Yakel and Deborah Torres published their work on 'archival intelligence', for the first time seriously addressing the question of what skills an archival researcher needed to have [18]. Prior to this work archivists had little except their own experiences to guide them in trying to understand what their users needed.

The move from paper correspondence to enquiries received by email encouraged, almost simultaneously, Kristin Martin [13] and Duff and Johnson [3] to examine and categorise these enquiries. In Martin's case, her aim was assess the impact of email on the content of these enquiries within a single repository. In Duff and Johnson's case the aim was to inform the development of future digital archival information systems. The pair hoped that by categorising enquiries across a group of archives and examining the types of information supplied by users, they would uncover "clues as to how patrons seek information". Martin predicted that as more information about holdings became available online, users would form more specific enquiries relating to those holdings.

More recently, Kirschoff et al have described the increasing profusion of archival information systems, digital libraries and other cultural databases as "digitisation islands" [9]. Archival users must undertake the difficult task of moving between these unconnected islands but little literature exists on how skilled they may be at achieving this. In fact, in spite of extensive library literature on offering digital services, the archival literature on topics such as answering enquiries via IM or "livechat" is virtually non-existent with Gary Brannan's paper discussing such sessions at West Yorkshire Archive Service a rare exception [1].

One might argue that libraries and archives are similar, and thus lessons from the library literature would carry forward to archives. However, there is insistence from authorities in the archival sphere, such as Mary Jo Pugh [14], that archival reference 'encounters' differ strongly from those in libraries, and thus are worthy of study in their own right:

> "*The most significant difference is that reference encounters in libraries are usually short and voluntary, each devoted to a single question. In contrast, reference transactions in archives are more likely to be substantive, obligatory and continuing.*"

Is this view of archive interactions correct in the online space? When interacting with users online do archives maintain more substantive and continuing interactions? Today, archives are strongly focused on providing user centred services yet answers to these questions are not clear from current research.

A further theme that runs through the, admittedly small, user-based literature for archives is the study of information seeking behaviours and the problems encountered. For example, for family and other historians, it is impossible for their work to progress if they cannot find what they are looking for – indeed, to a great extent, the products of such searches *are* their work. Andrea Johnson examined the information seeking behaviour of over 500 archival users in the course of her doctoral work. Johnson summed up the three main problems encountered by these users as "where shall I look?", "what shall I say?" and "what is that?" [8].

Currently, it is very difficult to see if and how existing search and other information systems are meeting the needs of archive users, and subsequently, little information to drive future design. While Duff and Johnson's work provides a useful model for studying enquiries, their research was carried out on email exchanges between users and archivists. This valuable contribution is now over ten years old, and with the introduction of new means of remote online communication with archives being introduced in that time, there is no information about how this changes and shapes the interactions of users with the archive through these channels. Andrea Johnson ends her paper with a call for more empirical research in order to resolve the question of how to best support "the interaction between the user and digital archival material".

For these reasons, the study reported in this paper examines enquiries to the National Archives to address these gaps in the literature. The contributions of this study will allow us to know where and when users are encountering problems, through what channels they pose what kinds of questions to archivists, and what are the drivers of those requests. This will be of value to interaction designers as it will allow for prioritisation of putative design innovations and research that could help users address the most common problems at the appropriate point in their information seeking journeys.

3 Method

This mixed method qualitative study consisted of a combination of content analysis [10] and a grounded theory [15] of a variety of different sources of existing user-generated data stored at the National Archives in the UK and a selection of forum posts on the popular genealogy community forum Rootschat.

The content analysis used two existing coding schemes to answer the following two questions:

- What are the differences in the types of enquiries users make through different online channels?
- When users have information seeking enquiries, at what stage of information seeking behaviour are they in?

Further, the grounded theory analysis of the content was conducted using an open coding scheme, with the aim to understand what the drivers were behind information seeking enquiries.

3.1 Data collection

The research team, in cooperation with TNA, identified four main channels through which remote enquiries are received and handled: email, telephone, livechat and Twitter. In addition, a sample of enquiries to the Rootschat family history forums was also taken for comparison to the TNA channels. Each channel had data sampled from different time periods. This was to avoid the skewing of the data towards particular events or announcements from the National Archives - for example, the announcement of new records relating to the centenary of World War I would likely skew all channels towards requests for data from that era. Data was sampled in an incremental way with researchers coding data as it was retrieved. For each channel, when coding of the enquiries stabilised such that proportion of posts allocated to each code in the coding schemes did not change, sampling was ended for the channel.

At TNA, there is a Contact Centre which provides an email and telephone service. The purpose of this Contact Centre is to provide a positive advice service for callers, to enable them to pursue their research and to point users in the right direction with some suggestion but not to carry out their research for them, which is a service carrying a charge. The Contact Centre is staffed by a mix of administrative workers and records specialists.

TNA receives 1,000 – 2,000 emails per week to the enquiries inbox. These are responded to by staff working in shifts in the Contact Centre with a total of 37,613 responses logged in 2012/13. In order to manage this volume, the Contact Centre staff make heavy use of a collection of email templates maintained on the Archives' intranet. For example, in 2012 roughly 4,800 enquirers (nearly one in six) received a form response personalised by a link to a page on the National Archives website deemed relevant by the member of staff responding.

For this study, 150 emails received by the National Archives Contact Centre were examined. This comprised a sample from 31st May 2013 and another sample from January 15th and 16th 2014. In each case the Contact Centre were requested to send whatever had been received on a given date up to the number requested after filtering for spam messages. Requests for naturalisation records are handled via a separate web form containing structured fields and were therefore omitted from this study.

For the telephone channel, TNA collected a large amount of telephone enquiries in 2008 as part of their own internal auditing procedures. 51 telephone enquiries were analysed from four samples taken from between 27th February and 10th April 2008.

TNA staff run four hours of live chat sessions four days a week. Transcripts from fifty live chat sessions were analysed with twenty-two livechat sessions from 20th February 2014, three additional sessions from 21st February 2014, twenty three session from 3rd April 2014 and two session from 4th April 2014. The transcripts were anonymised to remove the names of the both the archives interlocutor and the user.

TNA has been using Twitter since July 2009 and has over 60,000 followers @uknatarchives. Fifty-five Twitter conversations were collected, covering the period from 3rd January to 24th May 2014. These conversation were collected by selecting every tweet where TNA participated in a discussion with the expectation that these would be have the highest likelihood of being in response to questions

Rootschat is one of the world's largest freely accessible family history forums and the largest in the UK. It contains about 4.5m posts, has around 200,000 registered members and records about a quarter of a million unique visitors a month. The site comprises a large number of different forums. The sample examined here was generated by collecting the top two posts from the most recently edited threads on 12th January 2014 and 2015 from a random selection of twenty-five (of forty) different English county boards to represent a snapshot of enquiry activity on this site.

For each data sample, each unit of communication (e.g. 1 email, 1 livechat session, 1 tweet conversation, 1 telephone session, 1 forum post) may contain more than 1 enquiry. Hereafter, for purposes of clarity and comparison, each channel is discussed in terms of the number of enquiries present within them as opposed to these units of communication. Further, not all enquiries are considered to be information seeking enquiries. For example, a copy request is not indicative of an information seeking activity. Accordingly, when talking about information seeking enquiries we see an expected dip in numbers. The total number of enquiries encountered for each channel is presented

Table 1. Final numbers of *enquiries* and *information seeking enquiries* identified in samples.

	Email	**Telephone**	**Livechat**	**Twitter**	**Rootschat**
Enquiries	153	51	63	57	71
Information Seeking Enquiries	115	36	56	21	64

3.2 Content Analysis

A content analysis was carried out on all of the above data with two researchers coding samples of the data and checking for reliable application of the codes. There were three different ways that the data was coded in order to answer the research questions.

The data was first coded using a version of the codes of the Duff and Johnson study. [3] This coding scheme allows us to characterize the types of enquiries that came in through each channel. The following are the codes used with a brief summary of what each means:

- **Administrative/Directional:** Enquiry asks about administrative information. Examples include costs of photocopying, opening hours or directions.
- **Fact-finding:** An enquiry requiring a specific factual answer.

- **Material-finding:** An enquiry about where to find sources about a particular person, place or event.
- **Specific form:** An enquiry about if a particular source type is available (e.g. census, military service records).
- **Known item:** An enquiry to know if a specific item known by the individual is held at the archive.
- **Service request:** An enquiry for a specific service provided by the organisation (e.g. copying).
- **Consultation:** An enquiry asking for advice that calls on the archivists' specific knowledge of resources. For example, does a specific series contain material relevant to the researcher.
- **User education:** An enquiry where the user has "vague sense" of the record they want or wants to know "how to get started". These may be statements that start with "how do I?"

In some instances, the enquiries did not fit into this classification scheme. These enquiries were subsequently labelled as *New*, with their contents analysed for possible common themes. In most cases these sets were quite small, but where there were patterns we discuss more information about these *New* types of enquiries in the results.

While the Duff and Johnson coding scheme provides a view of the type of enquiry in relation to the services offered by the archive, it does not provide insight into what users submitting the enquiries are really trying to achieve and in particular gives no view of progress in situations where users were in the process of information seeking.

A literature scan, a full reporting of which is beyond the scope of this paper, revealed many, many models of information seeking have been developed by a range of experts [16]. One model, proposed by Marchionini, seems a good fit for the archival domain [12] as it appears to include the key states for people working with information systems or collections. It also lacks the oversimplified 'one critical path' of a number of information seeking models which excessively privilege forward progress. It also seems to map well to the limited work with groups of archival users which has been carried out. Marchionini's "choose a search system" (sometimes "select source") is clearly analogous to Andrea Johnson's "where shall I look". [8] "Formulate query" is "what shall I say" and "what is that" is "extract information", which is the phase relating to the use of the information products (the documents) located in information seeking.

Using this model, a second coding was devised that characterises the stage of information seeking the user was at when the enquiry was made. The following were the codes used from the Marchionini model:

- **Define Problem**: An enquiry where the user cannot (or does not) define clearly what they are looking for.
- **Choose a search system / select source**: An enquiry where the user seeks direction as to where to begin a search. What types of information/source are available?

- **Formulate query**: An enquiry where the user has a clear sense of what they are looking for but is unable to generate the query terms required to meet their information need.
- **Execute query:** An enquiry where the user requests search be undertaken for them as they are unable to unwilling to do so.
- **Examine results:** An enquiry where the user is trying to understand the result set. Users may seek clarification or reassurance of their own understanding of the results, or the relevance of the results.
- **Extract information:** An enquiry about specific records where the user is trying to make sense of the record. This could include technical problems with document access (e.g. failed downloads).

3.3 Grounded Theory

A grounded theory method [15] was used to identify key categories, themes and patterns that were within the data. This method was undertaken without a pre-conceived hypothesis or theory regarding what the drivers were behind the enquiries. An open coding scheme, grounded in the data, identified key features of the enquiries and the attendant problems being experienced by users and solutions proposed by archivists.

4 Results

The following are the results from the content analysis and grounded theory analysis.

4.1 Content Analysis

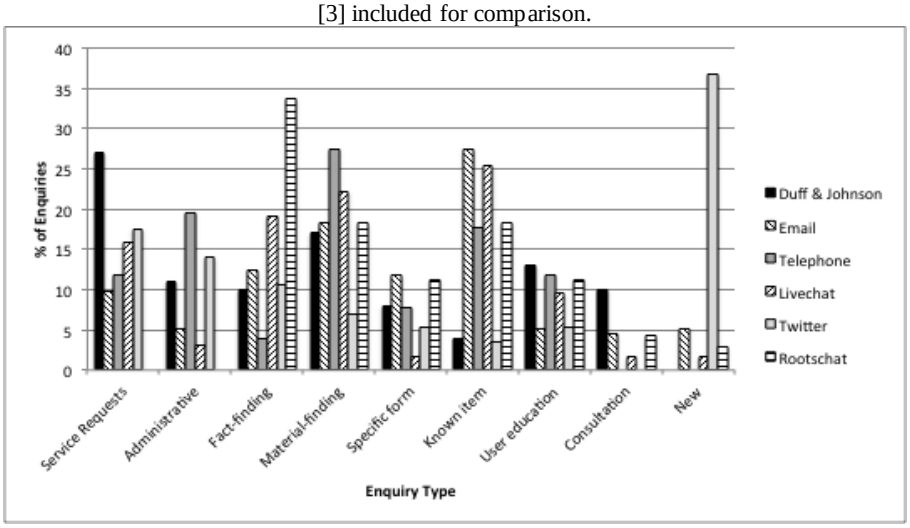
The results of the application of the Duff and Johnson coding scheme are presented in **Table 2**. Further, in **Fig. 1** the percentages of each enquiry type are presented by channel alongside the results from Duff and Johnson for purposes of comparison.

Table 2. Coding of enquiry types for each enquiry channel using the Duff and Johnson [3] coding scheme.

Category	Email	Telephone	Livechat	Twitter	Rootschat
Service Requests	15	6	10	10	0
Administrative	8	10	2	8	0
Fact-finding	19	2	12	6	24
Material-finding	28	14	14	4	13
Specific form	18	4	1	3	8
Known item	42	9	16	2	13

User education	8	6	6	3	8
Consultation	7	0	1	0	3
New	8	0	1	21	2
Total	153	51	63	57	71

Fig. 1. Percentage of enquiry types for each enquiry channel. Results from Duff and Johnson [3] included for comparison.



Applying Duff and Johnson's coding to TNA emails produced some similarities between our data set and theirs. For example 10% of Duff and Johnson's enquiries were considered to be fact finding and the proportion is 12% in the TNA sample. 17% were deemed to be material finding and the proportion is 18% in the TNA sample. However some categories are very different. TNA received many fewer administrative and service enquiries and its largest block of enquiries were known item requests (27.5%). In Duff & Johnson's sample only 4% of requests were known item. It must be emphasised that it is not required that a user should be able to reel off a catalogue reference (COPY 1/400/254, say) in order to "know" an item. If they are accurately describing a discrete record that really exists (the will of Tobias Box, the log of HMS Brilliant) or should exist but perhaps has not survived, then it has been classed as a known item.

For the telephone-based enquiries, we see spikes of activity in the areas of administrative requests and material finding. In contrast to more modern channels, the telephone was the only channel in which material finding exceeded known item requests.

While the main intention of the livechat sessions is to support research enquiries, we see that they in fact cover a much broader range of topics. Their profile actually looks very similar to the email sample. We can see, for example, similar proportions of known item requests (27% to 25.4%). We do see an increased proportion of service requests and fact-finding requests. This is not unexpected in a real-time medium: people are looking for quick answers to immediate queries that they have.

Twitter appears qualitatively different from any of the other media, and in particular highlights the *New* category introduced into the Duff and Johnson coding. Whereas in the other channels we largely were able to code within the Duff and Johnson coding scheme, Twitter deviated strongly from this trend. In other channels, there were very few outlying enquiries, but with Twitter nearly 37% of enquiries to which TNA responded could not be categorised in this more traditional model. Many of these messages proved to be observations not questions, corrections to other users misconceptions (referring to the archive as an authority) or consisted of positive comments about TNA. These exchanges seem very distant from Mary Jo Pugh's characterisation of archival interactions [14] as "substantive, obligatory and continuing."

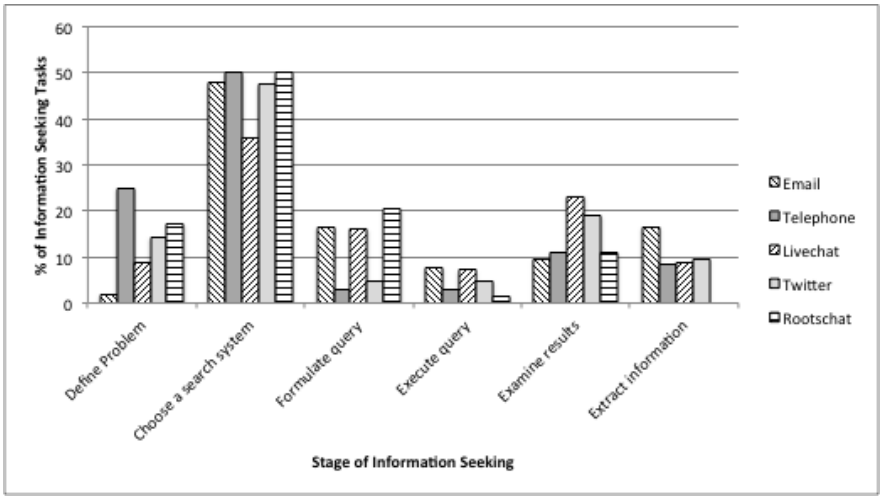
Turning to the coding using the Marchionini model of information seeking, the results are presented in **Table 3**.

Fig. 2 shows the relative percentages of each code for each sample.

Table 3. Coding of the information seeking stage for each enquiry channel using the Marchionini [12] model based coding scheme.

Information Seeking Stage	Email	Telephone	Livechat	Twitter	Rootschat
Define Problem	2	9	5	3	11
Choose a search system	55	18	20	10	32
Formulate query	19	1	9	1	13
Execute query	9	1	4	1	1
Examine results	11	4	13	4	7
Extract information	19	3	5	2	0
Total	115	36	56	21	64

Fig. 2. Percentage of enquiries in each information seeking stage for each channel.



Using the Marchionini coding, in the majority of channels almost half of enquiries seemed to relate to the *choose a search system* phase; that is to say they were primarily concerned with asking "where do I look" type questions. Even in the livechat channel, where it is at its lowest percentage, we see that over a third of the sample are in this stage. Due to the dominance of that stage of information seeking, it is perhaps unsurprising that later stages have much lower percentages. There are relatively small numbers of issues in the "examine results" and "execute query" phase. If most enquiries indicate people do not know where to look, then the chances of them every reaching the point where they can execute and examine their queries is much lower.

4.2 Grounded Theory

A variety of different features in the enquiries were coded in the open coding scheme. On the enquirers' side, the codes encapsulated facts about why they conducted their queries, such as the users overall research question or reasons for their search. The actions users had taken in the system, such as previous searches tried, result sets checked, or sometimes reporting null results (e.g. "nothing comes up") or expectations of content not present (e.g. "there must be thousands of other files"). We also observed instances where users did not appear to understand how the archive was organised, such as advising the archivist, incorrectly, that documents belonged somewhere else. The codes also captured misunderstandings about the archives' services, such as requesting downloads of very large (sometimes non-digitised) volumes of material. We also captured any indicators of the overall user experience or emotional

content as described by the enquirer. This often included feelings of frustration and expressions of anxiety.

On the side of the archivist interlocutors, the solutions proffered were coded by the mechanism by which they were resolved and associated information. These mechanisms included redirection of enquirers to prepared research guides on a web page, or a specific part of the archive, or a different archival catalogue or even an external organisation. In some cases, in particular when things could not be resolved, referrals to other colleagues were also observed and coded.

As the codes were developed, they were compared to each other for their content in an attempt to identify overall patterns of behaviour and subsequent impacts of that behaviour. These patterns of statements and enquiries were grouped together to form higher-level themes that were subsequently compared to the literature and each other to identify what the central core theme(s) were within the data.

After conducting the grounded theory analysis, a number of key themes emerged with one formal theory about what was driving the enquiries.

Archivist as Google. The first theme that emerges from the data is that the abundance of search engines and their related systems (e.g. websites), guides, manuals and other help mechanisms appear to not be sufficiently integrated to help users in their information seeking. While each of these different mechanisms for finding information are in themselves valuable, there is no mechanism to prompt users to go to particular pieces of help within the system. The absolute divide between searching the collections and searching guidance about those collections is immensely problematic.

This was evidenced by the large number of the enquiries across the channels which were handled by directing the enquirer to existing online record guidance. For example, two different livechat enquirers are both looking for unit war diaries relating to battalions of the Durham Light Infantry:

> *"I cannot find anything on your website....do you keep these please?"* (Livechat Enquirer 14)

> *"...have been trying to locate it but without success"* (Livechat Enquirer 3)

The war diaries are available at the archive, and through the primary Discovery search portal but more importantly, information on how to find unit war diaries is prominently displayed in a number of different places on the National Archives website. As a result, the archivist is easily able to direct the enquirers to the correct guidance. But the purpose of that guidance is to support users in finding the documents themselves and to free up archivists to focus on more difficult information seeking problems.

Situations where the archivist was 'acting as Google' in directing people to known resources was present in all channels. This type of interaction has an impact on archive resources in two ways. First, it impacts their ability to undertake more complex data seeking tasks, where individuals have hard to find or hard to access materials.

Secondly, the resources that people are being directed to have been invested in to address specifically these types of requests.

Ability to survey the terrain. When we look at the Marchionini model coding in combination with the grounded theory analysis, an interesting theme begins to emerge. The vast majority of enquiries that are coming from users fall squarely in the category of "where do I look?" This is a somewhat surprising result: given the availability of search technology we might assume that knowing where to look is a problem already solved.

One possible interpretation is this is all idiosyncratic to the National Archives. However, when we look at Rootschat, we see similar types of enquiries being made. Fact-finding, material finding or known item searches, users are predominantly getting stuck knowing where to look. This indicates to us that this is not unique to TNA.

A further possible explanation is a general laziness of the user base in trying to find out where to search. We all know that it is easier to ask someone else to undertake a task than to put in the effort ourselves. Within the dataset, such seeming lack of effort by users does occasionally appear:

> *"I am new to family research, please advise if it is possible to search the records online from Australia and what is required to help me do this."* (Email Enquirer 60)

> *"How do I get info on my grandad in Ww1 please? I have a name, DoB? Is there an email or tel number."* (Twitter Enquirer 33)

However, only a very small number of the enquiries across the channels appear to display this lack of effort. So, while this idea has face validity, we would expect a lot more of this type of interaction if it were the key reason for the enquiries. Further, when looking at the user base this explanation seems unlikely. People who are investigating their family histories, one of the largest groups now using archives, are doing so because they want to, not because they have to do so, and it seems unlikely they would hand over the task without making reasonably strenuous efforts. Trained professional historians also seem unlikely to 'outsource' their research in this way.

Other users seem to have all the information they need to conduct a search, yet still seem unclear even as to how to begin:

> *"Trying to trace my Grandfather ...who served in the Sherwood Foresters. Notts & Derby.Regt. in the Great War. How do I use the web site to get any information on him?? Got his Reg no [xxxxx]."* (Email Enquirer 42)

With the aid of a service number, many relevant records are easily available to this user. Similarly, an email enquiry about passage to Australia:

> "*My dad was in the Merchant Navy from 1948 - 1955. The only info. I have was that he was in Townsville, Australia in 1953 and he settled in South Africa in 1955.*

Is there any way of finding his record of service...His name was [xxxxx], DOB 30.11.1927." (Email Enquirer 45)

We must ask ourselves, why, with an abundance of search systems to choose from, are these users contacting someone for instructions in place of trying to search for themselves?

The answer to this question perhaps lies in situations where users have tried to find information, but have ended up back near the start of the Marchionini model. There were several enquiries throughout the channels that indicate that users are seeking reassurance that what they have found is correct:

"Hi, was wondering if you could help find my Grandad's naval records and medals on here, I found him on your website ADM [yyyyy], the name [xxxxx]. I find it very confusing. Can you help me?" (Livechat Enquirer 30)

Furthermore, other users indicate that they do not trust their own findings:

"I am searching for records of my father's s service in WW1, he received his RAC pilots licence, aged 16, in 1916 and his wings from the Royal Flying Corps later that year. Despite flying for over 600 hours there is no record of his war service with the RFC. I presume that this is because there are no RFC records for the period late 1916 - to early 1918, but would be grateful if you could confirm this?" (Email Enquirer 97)

"When I was young I was told that my Grandfather ([xxxxx]) fought in WW1 with the Irish Guards. I found this person and army number [yyyyy] in the Irish Guards the information I paid for has not helped me in my quest to find if this is in fact my Grandfather, it tells me the same name but not where he lived or other useful information which indeed would confirm if this is the same person am looking for." (Email Enquirer 89)

Consider the case of Email Enquirer 97 above. This user is in fact correct to be suspicious of their lack of search progress (the documents are available); but they have no mechanism to know that a path has already been laid out for them in the research guides provided by TNA. To return to Kirchhoff's metaphor of digitised islands, the enquirer suspects he has passed his island by but there are no signposts to tell him if he is right or wrong - just as there were no signposts on any part of his journey.

Compare this to Email Enquirer 89 who is unable find information to confirm which individual is his grandfather. In this case, the information may be contained in a different part of the archive, but there is no way for the enquirer to know this is the case. If the enquirer understood the archive's structure, that might help their situation, but there are few mechanisms within existing search systems at TNA or at other archives that help teach users about the structure of archives as part of their ongoing search. This leads to frustration and anxiety on the part of the searcher. They are on

one digitised island, and cannot even see the other islands in the distance they need to reach.

In the case of Email Enquirer 45, we have an even worse situation in that the individual finds the overabundance of choice so daunting that they are unable to even know where to begin their journey even though they appear to have all the information needed to successfully navigate to the end.

In all of these situations, the archival information systems have removed the archivist from the traditional assistive role they play in the physical reading room environment. This *disintermediation* [2] means that where systems fail to assist users in understanding the next step, the user must fall back on contacting the archivists through the channels available to them, even if it means trying to determine where to start.

Seeking Reassurance. There are a large number of enquiries where users indicate they have doubts and anxieties, often caused by a lack confidence either in what they have found or of what to do next. They then appear to seek reassurance from the archivist.

The disintermediation between user and archivist means that there is no opportunity for the user to seek reassurance that they have done the right thing and the archive's systems, primarily focused on finding materials, are not designed to fulfill this role. This, in turn, prompts users to try to engage with archivists through the various remote channels. At times, this is in regards to records, as seen in previous examples. In other cases, people seek to confirm facts they suspect but do not know:

> *"I am aware that some libraries and archives require letters of recommendation to access certain resources, and was wondering if the National Archives follows any similar practices….My time in London is limited, and (as I'm sure you'll understand) I am anxious to have everything necessary (including letters) before I arrive."* (Email Enquirer 6)

Interestingly, this theme fits with prior research. Duff and Johnson's work with mid-career historians [4] exposed the stressful nature of negotiating an unfamiliar archival collection. Their participants described the experience as an "anxiety attack", "overwhelming" and referred to the "fear" they felt at having missed something in the course of their work. These same fears ring true in our own data.

Considering all of these themes, we drew out a formal theory of what drives enquiries to the archives: the disintermediation of the archivist creates anxieties in users regarding their search progress.

5 Implications for design

From the results of content analysis using the Duff and Johnson coding, we see that there is definite variability in how different channels are used for enquiries. This

indicates, at least in the case of TNA, there are distinct benefits to running these different channels. Users are opportunistically using the channels in different ways, and sometimes in ways that would not be predicted.

Looking at the content of the messages, the channels shape the enquiries in some way: either by drawing in users with certain forms of problem or by encouraging the framing of enquiries in certain ways. For example, livechat users can explore problems in a more naturalistic, conversational way than the Twitter users but these users lack the more casual, unplanned and spontaneous opportunities for interaction afforded by social networking.

Martin [13] predicted that there would be shifts in the types of questions that would be asked over time due to the availability of online archive tools, and this does seem to have happened, at least when comparing our data to the 11 archives studied by Duff and Johnson [3]. We do indeed see that the proportion of known item enquiries appears to have increased substantially, with users providing much more specific enquiries. This increase appears in email, livechat and telephone. Twitter users appear concerned with other topics than individual collections items.

This means that archives are in a better position to accommodate user needs than they have ever been. They are receiving, across a spectrum of channels, large numbers of detailed and specific enquiries, sufficient to identify the most significant issues faced by the main body of their users and attempt to resolve them. The use of models materially assists in this process. Overwhelmingly, plotting these enquiries against the Marchionini model, tells us that users need more support within the archival interactive systems to improve their ability to choose a search system and navigate between digitised islands of archive data. Knowing 'where to look' remains a major problem for users, and they fall back on human support to solve this problem instead being able to rely on the system to aid them in being more effective or efficient in their information seeking tasks.

Kirchhoff et al. [9] advocate creating portals linking the islands together to solve this pervasive problem of users not knowing where to go. Interestingly, TNA now provides a form of portal, or federated search, through its Discovery search system that indexes archives across the UK and a number of other formally separate databases. But Discovery remains only one of a cluster of databases with which the historian or genealogist is likely to need to engage and in the data collected from TNA users it is clear that they still have difficulty knowing where to look. The lesson for online archives is that portals in and of themselves will not necessarily improve the users' ability to know where to start or where to go next in their search activities. In fact, creating such portals may exacerbate the problem further by offering so much choice in one place as well as adding another proverbial island.

As digitisation efforts continue (even accelerate) and further archival collections appear online, interaction designers must proactively seek solutions to solve this problem of knowing where to look. The results indicate that it is inadequate for support to exist somewhere separate from search systems; it must be proximal to and integrated within the system supporting the users' current task(s). This support might take the form of contextual help based on recognising broad types of queries ("looking for a person?"). It might make use of dynamic term suggestion, more thoroughly scoped

search, result clustering (beyond the faceted browsing already available), wizards or relevance feedback.

However, what our results do not provide is the means by which to choose which of these design interventions, if any, are appropriate in specific contexts of use. The follow up work to this study will be focussing on what support human interlocutors provide to users. This information will provide insights into which particular design interventions are likely to help users at specific points along their information seeking journeys.

6 Conclusions

In this paper we have conducted a content analysis of several channels that users of the National Archives uses for interacting with their patrons. These channels comprised both traditional, such as phone and email, as well as modern media such as livechat, and Twitter. Further, we collected user-generated data from Rootschat for comparison and contrast with TNA's data.

From the analysis of the channels, disintermediation appeared to serve as a key driver for the enquiries seen in the different channels. This disintermediation, where the user is compelled to rely solely on the search systems and help documentation online without an archivist to assist them, leads to many users feeling unable to act. They find themselves with an abundance of choice, but very quickly do not know where to turn. Users are, essentially, lost in a sea of islands of digitised data not knowing where their journey began, where they are now, or whether they have reached their destination. This lack of awareness leads to anxiety in the users, which manifests in them reaching out to real world archivists for direction and reassurance when they have found something, that they have proceeded correctly.

This has implications on how we design new search and interactive systems for archives. These systems must integrate support for navigating the archive structure so that users can grasp how to start their information seeking journeys, and direct or signal them onward to their next location. This, when combined with user education facilities to help the user learn where to find particular types of information that are relevant to them, could lead to a powerful step change in the types of interactions that would come in through the channels of enquiry. Instead of the archivist having to act like a search engine, they could instead collaborate with users in developing more sophisticated information seeking strategies.

Acknowledgments

This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC).

Research data access. Researchers wishing access to the data used in this study should visit the following URL for more information:
http://www.cs.york.ac.uk/hci/jjp513

7 References

1. Brannan, G.: Talking in the night: exploring webchats at the West Yorkshire Archive Service. In: Theimer, K.: Reference and Access: Innovative Practices for Archives and Special Collections. Rowman & Littlefield (2014).
2. Butterworth, R.: The Accessing our Archival and Manuscript Heritage project and the development of the "Helpers" website. Middlesex University, Interaction Design Centre (2006).
3. Duff, W., Johnson, C.: A Virtual Expression of Need: An Analysis of E-mail Reference Questions. American Archivist. 64, 43–60 (2001).
4. Duff, W., Johnson, C.: Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. The Library Quarterly. 72, pp. 472–496 (2002).
5. Duff, W., Johnson, C.: Where Is the List with All the Names? Information-Seeking Behavior of Genealogists. American Archivist. 66, 79–95 (2003).
6. Duff, W.M., Stoyanova, P.: Transforming the Crazy Quilt: Archival Displays from a User's Point of View. Archivaria. 1, (1998).
7. Freeman, E.: In the Eye of the Beholder: Archives Administration from the User's Point of View. American Archivist. 47, 111–123 (1984).
8. Johnson, A.: Users, use and context: supporting interaction between users and digital archives. In: Craven, D.L.: What are Archives?: Cultural and Theoretical Perspectives: a reader. Ashgate Publishing, Ltd. (2012).
9. Kirchhoff, T., Schweibenz, W., Sieglerschmidt, J.: Archives, libraries, museums and the spell of ubiquitous knowledge. Arch Sci. 8, 251–266 (2008).
10. Krippendorff, K., and Bock, M. A. (eds): *The Content Analysis Reader*. Thousand Oaks, CA: Sage (2008).
11. Malbin, S.L.: The Reference Interview in Archival Literature. Coll. res. libr. 58, 69–80 (1997).
12. Marchionini, G.: Information Seeking in Electronic Environments. Cambridge University Press (1997).
13. Martin, K.: Analysis of Remote Reference Correspondence at a Large Academic Manuscripts Collection. American Archivist. 64, 17–42 (2001).
14. Pugh, M.J.: Providing reference services for archives and manuscripts. Society of American Archivists, Chicago, Ill. (1992).
15. Strauss, A, Corbin, J.: Basics of qualitative research: Grounded theory procedures and techniques. Thousand Oaks, CA, US: Sage Publications, Inc. (1990).
16. Wilson, T.D.: Models in information behaviour research. Journal of Documentation. 55, 249–270 (1999).
17. Wineburg, S.S.: Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. Journal of Educational Psychology. 83, 73–87 (1991).
18. Yakel, E., Torres, D.: AI: Archival Intelligence and User Expertise. American Archivist. 66, 51–78 (2003).