Kelsey Rogers

LIS 60500-30

Research Proposal

Search Box or Archival Box: Non-Specialist Researchers' Discovery of Archives Based on

Format in Genealogy

Introduction

A small subset of research in archival information behavior centers on non-specialist researchers of family history and genealogy. These researchers trace their ancestors through historical documents, such as church and county records, censuses, and cemetery indexes. Directories, personal letters, wills, and other resources are used to flesh out family histories and narratives, tracing where ancestors lived, what they did for a living and other facts about their lives. These researchers usually have very little to no formal training in historical and genealogical research yet many make extensive, sophisticated use of historical records. These records are found in a variety of locales, but public libraries in particular are destinations for family history and genealogy (FHG) researchers, acting as repositories for a variety of valuable resources. With the advent of the internet, the format of these resources has expanded from microfilm and bound records, to databases, online indexes, and digital collections. In 2014, Lentstra surveyed several librarians at a workshop held around providing FHG resources at public libraries. Digital resources reported by the librarians were digital archives and databases run by the libraries, scanners for public use to digitize personal family history, local history wikis, and subscriptions to commercial databases. Many projects even crossed the

digital/physical boundary as print documents were transformed into interactive online portals and born-digital resources were published in book form.

Much of the previous research of FHG researchers has centered on the overall patterns of their information behavior (Yakel, 2004; Darby and Clough, 2013; Friday, 2014). However, in-depth patterns of resource use and discovery have not been studied despite efforts by public libraries to expand their FHG resources. Previous research has recorded the heavy use of big name genealogy websites such as *Ancestry*, which many libraries provide free subscription access to on-site (Skinner, 2010; Darby and Clough, 2013; Lucy, 2015). However, localized records are not always available through these large websites. County-level records, for example, are mostly found in physical format, and those that are being digitized are incorporated into smaller, non-profit databases and archives. These can be difficult to find as they may not be indexed by search engines (Sinn and Soares, 2014). Additionally, there may be preconceptions of the validity and availability of documents in digital formats in comparison to physical formats. These issues have been explored in terms of professional historians' conceptions of, and research behaviors towards, digital archives versus physical archives, most notably in Sinn and Soares's work (2014). Donaldson and Conway's work (2015) on digital document trustworthiness for amateur historians also sheds some light on the issue. However, non-specialist researchers' conceptions of format and the effects this has on resource discovery have not been studied extensively. This is increasingly relevant as public libraries balance budget and staffing concerns in regards to expanding their FHG collections both physically and digitally (Skinner, 2010; Lenstra, 2014).

The proposed research question is: *How do non-specialist researchers of local history and genealogy conceptualize digital archives in comparison to physical archives and does this*

affect their search for, and discovery of such archives? Answering this question is important for the creation and promotion of relevant resources for FHG researchers so that public libraries can maximize their support of these patrons, many of whom will travel 100 miles or more to utilize such resources (Skinner, 2010).

Literature Review

Family History and Genealogy Information Behavior

Darby and Clough (2013) identify a model for the information behavior of FHG researchers that focuses on the intersection of resources and activities. Through this model, types of resources that are consulted are associated with steps in the research process. 'Collect Family Information' is associated with anecdotes, photographs, and artefacts within the family's possession. After this, "Learn the Process" and "Break In" deal with introductory resources. Censuses records are cited as the resource most used to spearhead the entire endeavour. In fact, online census records were rated as the number one resource, followed closely by vital statistics indexes and *Ancestry.* Once the researchers have "broken in," tree-building commences, split into three phases of increasing difficulty. Within these phases, the "easy" phase was associated with high use of online resources, which trailed off in the later phases. This is interpreted as a preference for easily accessible resources in the beginning of the process. However, physical resources were used frequently in every phase of tree-building, suggesting that on-site visits, despite being slightly less accessible than online research, were prized.

One possible reason for seeking out physical resources despite a preference for easily accessible resources may be due to a search strategy identified by Friday (2014)'s work on online FHG researcher behavior. Researchers often combined multiple versions of the same

resource, such as comparing a census search on *Ancestry* with a census search on *FamilySearch*. Little is reported on the occurrence of cross-format searching of similar resources such as conducting a census search on *Ancestry* compared to searching through the census on microfilm or in book form. However, the occurrence of physical resource use in the midst of heavy online resource use in the first phase of tree-building may point to strategies of cross-format comparisons. In fact, easy access seems to be overshadowed by the importance of physical resources. A little over half of all respondents to Lucy's 2015 online survey of FHG researchers visited library, archives, or other repositories in person at least "sometimes" according to a Likert scale. Some of the respondents to Skinner's survey (2010) had even travelled 100 or more miles to use the library's collections. One of the librarians who responded to the survey stated that many of the researchers travel extensively for their research. Skinner points out,

> "The frequent mention made of both physical resources and digital ones suggests that users may value both face-to-face interaction and working with original documents while also valuing the convenience of Internet-based resources. This was explicitly stated by one user, who said she prefers to work directly with materials, but enjoys the ease of remote access. Most other users conducted research both within institutions and on the Internet, suggesting either that both types of resources offer different types of materials or experiences."

This attitude is reflected by professional historians; in fact several respondents to Sinn and Soares' questionnaire emphasized that the two formats were not comparable, even if digital resources were extremely useful.

Trustworthiness of Digital Documents

The emphasis on the difference between digital and physical formats of the same documents suggests that it may affect users' perceptions of the trustworthiness of the documents. Lucy's 2015 survey of non-specialist researchers found that the most common problem in genealogical research was dealing with questionable sources and information (29.8%), and a tangential problem was fraudulent pedigrees (12.2%). However, Donaldson and Conway's research (2015) suggest that this may be related to the repositories rather than the digital format of the documents. In a focus group of genealogists familiar with the Washington State Digital Archives, the top Kelton-Based Themes for trustworthiness were accuracy, validity, and coverage. This means that genealogists recognized a document as trustworthy if the information within the document is accurate, the document itself was created using the proper means, and the document is complete. For example, a census record would still be trustworthy despite having erroneous information, simply because it is a government census record. Overall, participants felt that the digital documents were authentic and looking out for "fakes" is something that participants do not consider relevant while looking through the archives. However, emphasis is put on information that is as close to the original event as possible. The more times the information has been transcribed, such as in ledgers or indexes, the less trustworthy it becomes. Another possible detractor of trustworthiness of digital documents is the quality of digitization. If the document is a "real fuzzy reproduction" the trustworthiness may be decreased. Additionally, having an idea of what the original document should look like is cited as an element in deciding the trustworthiness of the digital document.

Resource Discovery

A variety of methods were used by historians to find digital archival collections according to Sinn and Soares (2014). The top three methods were web searches, heard about

from a colleague, or recommended by a librarian/archivist/museum curator. One respondent

utilized a web search to find the University of Texas Perry-Castañeda Library Map Collection

while looking for a copyright free map. As Sinn and Soares suggest, making digital collection

databases indexable by search engines might provide broader reach to both specialists and

amateur researchers. Currently, many databases are generated dynamically and thus cannot be

reached by web crawlers used by search engines like Google. Due to this, many historians

searched through the websites of institutions that were relevant to their topic and sought out

digital collections even if they did not know whether the institution had them. Sometimes finding

out about the digital collections occurred adjacent to searching the website for the institution's

holdings in preparation for a visit. For historians who had not used digital archives in their

published work, it was not because they had not looked for them. Of the eight who did not use

digital archives, seven respondents said that they could not find digital collections for their

research topics. Another reason for the lack of use was that researchers used finding aids or

online catalogs to find relevant documents but could not access them online as they had not been

digitized. Overall, the top two factors that affected the use of digital archives were institutional

reputation and the format of the documents themselves, i.e. HTML text vs. scanned images. This

reflects the perceptions of FHG researchers of trustworthiness of digital documents found by

Donaldson and Conway (2015). The participants trusted the digital documents because they

trusted the Washington State Digital Archives, and if the documents were poorly digitized, their

trustworthiness decreased. This suggests that FHG researchers may use similar methods of

relying on institution reputation to find digital resources. However, due to the varying levels of

information literacy of FHG researchers reported by Friday (2014), there may not be a direct

comparison to be made. While most FHG researchers were skilled in reframing their searches

and employing novel search strategies, their ability to navigate unfamiliar resources was limited.

Over-reliance on a few websites, such as *Ancestry,* was documented by Friday, as well as a lack

of confidence in navigating the Internet. Beyond this, little has been reported by any previous

research on the discovery methods of FHG resources, and Darby and Clough (2013) only report

on the broad task of resource discovery rather than the specifics.

Research Methods

The primary research method for the proposed study will be interviews and diaries.

Coding of the interview notes and collected diaries will then be used to interpret the data, with an

eye towards common patterns and themes surrounding archive usage. As the proposed study

targets how researchers conceptualize archive format as well as interrogates the multi-faceted

approaches researchers use, a qualitative method best compliments the data I am attempting to

gather. Narratives in interview and diary form will more closely reflect the organic, emotional

nature of non-professional genealogical research.

Population

The population will be non-specialist FHG researchers. The sample will be drawn from

researchers who utilize public libraries' FHG departments as well as members of local historical

societies. After several regional libraries and societies are contacted and recruited to disperse a

call for participants, a recruitment message will be attached to the departments' and societies'

monthly newsletters. In addition, participants will be recruited in person at larger events that the

libraries and/or local historical societies sponsor. Sampling will seek to recruit participants who

possess a wide variety of skill levels, from beginners to seasoned researchers. Hu's review of

methods for digital archival research (2012) emphasizes the usefulness of small sample sizes as

permissible for interviews, citing Nielsen's suggestion that five user interviews may be suitable

to identify a pattern of behavior. Thus, a sample size of five to twenty non-specialist family

history and genealogy researchers is proposed. It is likely that the first stage of the research will

have a larger sample size than the second stage due to the effort involved in keeping a diary for

six months.

Procedures

Interviews will be conducted in person. As the study is targeting differences in digital and

physical archive use, and the assumed variability in the technological skills of the population,

data collection should be done with as few technological barriers of entry as possible. The

interviews will be guided by the interview guide (Appendix A) with extra questions used as

prompts for more information, and dependent on participants' responses. Interviews will be

conducted one-on-one and, after consent is given by the participant, recorded with a digital audio

recorder. The interview will be conducted in a casual, conversational manner, with an emphasis

on following the conversation where the participant would like to go while still touching on the

main questions of the interview guide. The interviewer will take unobtrusive notes during the

interview to track any thoughts or impressions.

All participants will be asked to participate in the next stage of the study, but an inability

to participate will not invalidate their participation in Stage One. Stage Two will require

participants to keep a diary detailing their research activities. In a similar process and time period

as Friday (2014), participants will be asked to record instances of research over a period of six

months. This will account for variability in research habits of the participants and their ability to

access resources. Three months into Stage Two there will be a mid-point interview to reinforce

the diary process as well as gather data about any research occurring in that time period. It will be much less structured than the initial interview, being directed by the participants' research activity and diary entries during the three months. At the end of the six month period, the diaries will be collected and a closing interview will be administered in the same style as the mid-point interview.

Diaries can be kept in any format the participants wish, either electronic or written. Tom and Duff (2002) report the variability of participant adherence to provided diary format and the difficulty of ensuring adherence when researching the usage of archives. They theorize this is due to the already complex environment of archival research and the difficulty participants find in juggling both the diary and the actual research. Despite the many difficulties they found in securing adherence to the diary format they requested, they found that by combining all of the formats received with interviews yielded more comprehensive and effective interviews. Therefore, a rough guide will be provided for the participants when preparing them for the diary stage of the study but otherwise, the format of the diary will be determined by the participants and used as a supplement to the interview.

Interview recordings, diaries, and interviewer notes will be transcribed electronically and saved on an encrypted, private server. Participants will be identified by code so as to match diaries with interviews and keep track of participants across interview sessions. Participant's contact information will saved on a seperate, encrypted system with matching codes. Once the study is complete the identifying document will be destroyed.

Data Preparation and Analysis

Qualitative coding will be implemented following open-coding processes, with no previous code schema applied. From these codes, categories and patterns will be tracked and from these themes will be identified. This will be an iterative process, with multiple reviews of the data to add and refine coding, identify new instances of categories and patterns, and compare and contrast instances across participants. Once themes have been identified, they will be applied to the research question and demographics of the participants.

Reporting

Data will be reported in a primarily narrative structure with excerpts from interviews and diaries to demonstrate the actual data gathered and connect the interpretation to the actual experiences of the participants. There will be sections describing the themes that arose from the data and their implications for the research question. A concluding section will provide suggestions for LIS institutions based on the results of the themes as applied to the research question. These will focus on the structuring and promotion of genealogical and local history resources for use by non-specialist researchers. Libraries, archives, and museums (LAMs) that manage and seek to create resources for FHG researchers require an understanding of the users' patterns of information seeking and what formats attract them and why. The proposed study will provide data that will hopefully support efforts for more effective promotion of these resources as well as improving user interfaces for more effective non-specialist utilization of both physical and online resources. Additionally, the study will hopefully provide insight into the usefulness of digital and/or physical access to local history and genealogical resources.

Timeline

FIRST STAGE

Week 1:

- Identify libraries and historical societies for potential participation

Week 2-5:

- Contact libraries and historical societies to obtain participation
- Create newsletter recruitment messages and schedule appearances at events with participating libraries and historical societies.
- Send out newsletter messages
- Recruit participants at events

Week 6-8:

- Set-up and conduct interviews
  - Keep interview notes
- Recruit participants from interview pool to participate in second stage
- Write interview reports
- Create encrypted storage and pseudonym document

SECOND STAGE

Week 6-8:

- Prepare and instruct participants in keeping research diary

Week 19 - 21/Three Months After First Interview:

- Gather diary entries to date and briefly review
- Conduct mid-point interviews
  - Keep interview notes
- Reinforce diary instructions
- Write interview reports

Week 31-33/Three Months After Second Interview:

- Gather diary entries and briefly review

- Conduct last interview
    - Keep interview notes
- Write interview reports

Week 33-35:

- Transcribe interviews and diaries

Week 35-37:

- Code data
- Analyze data

Week 37-39:

- Write report

Limitations and Research Quality

Limitations of this study will be the diary data collection method, the small sample size, variability in research undertaken by participants over the six month period, and the relative skill levels of the FHG researchers. Diaries are contested by Hu (2012) and Tom and Duff (2002) as being less than useful data collection methods. However, Tom and Duff demonstrate that by combining diaries with interviews, the data gathered becomes much more useful. Additionally, Friday (2014) shows strong support for this method. This must then be taken into account during data collection and analysis. While Hu (2012) presents diaries as a less than ideal data collection method, limited sample sizes in interviews are seen as acceptable limitations to the research. As there is an overall difficulty in recruiting a large sample size of FHG researchers, this is an acceptable limitation despite its possible effects on the appropriateness of the results for application to the wider population.

The research undertaken by LHG researchers is characterized by intense sessions followed by long periods of time without any research undertaken (Friday, 2014). This may cause variability in the frequency of research done within the six month period, as well as the intensity and quality of the research. However, Friday (2014) was able to gather valuable data from the diary method of data collection despite this limitation. Finally, the skill levels of the FHG researchers in the study may influence the outcome as more seasoned researchers may have more knowledge of the resources available to them as well as more experience in finding niche resources, as suggested by Darby and Clough (2013). This makes it all the more imperative that a variety of FHG researchers at varying skill levels are recruited for the study as well as ensuring their continued participation throughout the two stages of the study.

References

Darby, P. & Clough, P. (2013). Investigating the information seeking behaviour of genealogists

    and family historians. *Journal of Information Science, 39*(1) pp. 1-13. Retrieved from:

    doi:10.1177/016555150

Donaldson, D. & Conway, P. (2015). User conceptions of trustworthiness for digital archival

    documents. *Journal of the Association For Information Science and Technology.*

    Retrieved from: doi: 10.1002/asi.23330

Friday, K. (2014). Learning from e-family history: a model of online family historian research

    behaviour. *Information Research 19*(4). Retrieved from:

    http://www.informationr.net/ir/19-4/paper641.html#.XFXnVFzYqUl

Hu, R. (2012). Methods to tame the madness: A practioner's guide to user assessment techniques

    for online finding aid and website design. *RBM: A Journal of Rare Books, Manuscripts,*

    and Culutral Heritage, 13(2), 175-190. Retrieved from: doi:10.5860/rbm.13.2.381

Lenstra, N. (2014). Cyberorganizing everyday heritage in and around public libraries: An

    exploratory study in Illinois. *Public Library Quarterly, 33* pp. 1-22. Retrieved from:

    doi:10.1080/01616846.2014.877716

Lucy, C. (2015). Research trends & emerging technologies for genealogists. *School of*

    Information Student Research Journal, 5(2) Retrieved from:

    https://scholarworks.sjsu.edu/ischoolsrj/vol5/iss2/4/

Sinn, D. & Soares, N. (2014). Historian's use of digital archival collections: The web, historical scholarship, and archival research. *Journal of the Association For Information Science and Technology, 69*(9) pp. 1795-1809. Retrieved from: doi:10.1002/asi.23091

Skinner, J. (2010). Does greater specialization imply greater satisfaction? Amateur genealogists and resource use at the State Historical Society of Iowa Libraries. *Libri, 60* pp. 27-37. Retrieved from: doi:10.1515/libr.2010.003

Tom, E. G. & Duff, W. (2002). "I spent 1 ½ hours sifting through one large box. . . .": Diaries as information behavior of the archives user: Lessons learned. *Journal of the American Society for Information Science and Technology, 53*(14) pp. 1232-1238. Retrieved from: doi:10.1002/asi.10165

Yakel, E. (2004). Seeking information, seeking connections, seeking meaning: genealogists and family historians. *Information Research, 10*(1). Retrieved from: http://informationr.net/ir/10-1/paper205.html

Appendix A

1. (Warm Up) I've been working on my family's genealogy lately because my grandma talks about her Swedish grandmother frequently and I wanted to give her a present detailing her Swedish ancestry. How did you get started researching your genealogy/local history?

2. How did you learn how to do this type of research?

3. (Indirect) How do you think other people conduct their genealogy research?

4. (Direct) Could you tell me how you go about researching your genealogy?

5. What are some of your preferred resources to use when researching?

5. Could you tell me about what you do when you can't find answers in your preferred resources?

6. What are some of your experiences when using online resources?

      FURTHER PROMPT → Do you find online resources difficult to use?

      IF NOT → Are there any reasons why you haven't used online resources?

7. What are some of your experiences when using physical resources?

      FURTHER PROMPT → Do you find physical resources difficult to use?

      IF NOT → Are there any reasons why you haven't used physical resources?

8. IF USED BOTH →

      Do you enjoy using online resources or physical resources more?

9. How have you discovered these resources?

10. Have you ever felt out of your depth when researching? What happened?

11. (Wrap-Up) What are some of your favorite moments when researching?

Thank you for sharing your experiences with me. I wish you luck on your next search!