Finding Family Facts in the Digital Age: Family History Research and Production Literacies

A Thesis

Submitted to the Faculty

of

Drexel University

by

Heather Lynn Willever-Farr

in partial

fulfillment of the

requirements for the degree

of

Doctor of Philosophy

June 2017

Copyright

©Copyright 2017

Table of Contents

List of Tables

Table of Figures

Abstract

Finding Family Facts in the Digital Age: Family History Research and Production Literacies

Heather Willever-Farr

Andrea Forte, PhD

This study examines the online information behaviors of experienced and novice family history researchers, though the lens of accuracy and an increasingly digital research and production environment. It presents a model of the information behaviors of family history researchers, as well as a literacies framework, which visualizes the skills and knowledge needed to conduct accurate family history research and produce accurate family histories in the digital age.

CHAPTER 1: PROBLEM STATEMENT

Millions of Americans are contributing family history content to websites such as Ancestry.com, Findagrave.com and Wikitree.com ("Online Family History Trends Report," 2011; Rainie, Lenhart, Spooner, & Horrigan, 2000). The proliferation of user-contributed content on family history websites has led many family history researchers (FHRs) to view this contributed content as an important information resource (Willever-Farr & Forte, 2014). At the same time, family history websites are experiencing widespread problems with the proliferation of inaccurate and poor-quality user-contributed content on their websites (Willever-Farr & Forte, 2014). Producing accurate family histories appears to require the acquisition of certain research skills and knowledge—family history production is a learned craft (W. M. Duff & Johnson, 2003; Willever-Farr & Forte, 2014; Willever-Farr, Zach, & Forte, 2012; Yakel, 2004b; Yakel & Torres, 2007). Yet anyone with or without good research skills can easily contribute content to family history websites. Given that user-contributed family history content has become an important information resource, what can be done to address the problem of inaccurate content on family history websites? This dissertation will address the problem of inaccurate user-contributed content on family history websites by answering the following questions:

1) What are the online research practices of experienced FHRs vs. novice FHRs?

2) How might these practices influence the accuracy of the family history content they produce?

3) What new literacies are needed to produce accurate family history content in online environments?

4) How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?

5) How might family history websites be designed to better facilitate the production of more accurate content?

There are many gaps in existing literature on FHRs' online information behaviors. To date, few studies have considered online spaces where much of the family research and production is occurring. As a result, the close intertwining of family history research and production in online environments has not been fully explored. Literature on FHRs' information-seeking provides a cursory view of how FHRs work alone or with others in online contexts to find resources, determine their relevancy, interpret information from sources, manage conflicting information found in the myriad of online (and offline) sources, and then use found information to build biographical content about their ancestors on family history production sites. The information literacies needed to build accurate family histories in online contexts is largely unknown.

Furthermore, existing studies focus on the behaviors of more experienced FHRs. There is little knowledge of the information behaviors of novices compared to those of experienced FHRs, and how these information behaviors may impact content production on FHR websites. This dissertation will begin to fill those research gaps and provide data that can be employed to improve the design of family history websites so that they better support the production of accurate family histories.

CHAPTER 2: INTRODUCTION TO THE PROBLEM SPACE

2.1 Family History Research and Production: A Ubiquitous Hobby in the U.S.

In the 2010s, family history research or genealogy became "almost as popular as porn" and found its place as the second-most popular hobby in the United States (News & Farnham, 2012; Rodriguez, 2014). Ancestry.com, a for-profit purveyor of online genealogical data and family tree production tools, likely played a key role in the popularity of this hobby by lowering the bar for entry. Before Ancestry.com and Familysearch.org, family history research often required the researcher to travel to numerous brick-and-mortar archives and wade through large archival collections to find a single record pertaining to an ancestor. Knowledge of archival collections and systems, as well as knowledge of various search strategies for finding relevant material in often dispersed collections, was required in the days before the World Wide Web. To support FHRs in this often difficult process, many face-to-face genealogical groups were formed to share know how and support in the research process (Duff & Johnson, 2003; Yakel, 2004b; Yakel & Torres, 2007).

Ancestry.com, LLC (parent company of Ancestry.com) and FamilySearch, Inc. (parent company of familysearch.org) stepped into the family history scene and transformed the hobby from a niche avocation to a mass hobby. Their websites, ancestry.com and familysearch.org, are the two most popular websites for genealogical research in the U.S. ("Top 100 Genealogy Websites of 2014 (page 2)," 2014; "Top 100 Genealogy Websites of 2015," 2015; "Top 100 Genealogy Websites of 2016," 2016). Familysearch.org ("FamilySearch") is a free family history website hosted by the Church of Jesus Christ of Latter-Day Saints. Beginning in the 1960s, the Mormon Church began collecting genealogical data and records, often saving analog collections that

contain biographical information, such as defunct newspaper collections, from oblivion by microfilming them and then sharing the microfilms with libraries across the country. The church also granted access to the genealogical data and original documents through its family research center in Salt Lake City and smaller regional family history centers throughout the U.S. The emergence of the World Wide Web presented an opportunity for the church to share imaged document collections and genealogical data with a much larger audience through its free website, Familysearch.org. The church is committed to providing free access to the genealogical information, online or offline, as it is a means to expand its posthumous congregation (News & Farnham, 2012; Rodriguez, 2014). If a living Mormon can establish a familial kinship to someone deceased, then the deceased can be welcomed into the faith retroactively.

Ancestry.com, LLC ("Ancestry") has no such religious mission. It began providing access to genealogical data, and later imaged records and online production tools for a fee. The website, ancestry.com, is the biggest player in the US genealogical market, and is the starting place for many beginners. In 2015, Ancestry reported it had 2,219,000 paid subscribers and revenue totaled $683.1 million, up 10.3% from 2014. Through arrangements with archival repositories in both the U.S. and Europe, Ancestry has digitized, indexed, and transcribed records and made them available to subscribers through online databases on Ancestry.com and its sister sites in Canada (ancestry.ca), the U.K. (ancestry.co.uk), and Australia (ancestry.com.au). In addition to providing access to records and transcribed data, Ancestry provides online family tree construction tools. Ancestry has integrated tree construction tools with search tools, thus blurring the lines between production of public-facing family trees, and the often messy research process. For example, subscribers can search Ancestry's databases for deceased individuals by name, date, and other variables and then automatically populate their family

trees with the found information and images. Ancestry also integrated a "hint feature" within the tree construction tool: Once a contributor creates a portion of a family tree, the website will generate "hints" that appear as shaking leaves. These hints are automatically generated via search algorithms that attempt to find additional ancestors based on the provided data. The subscriber can simply click on the green leaf to automatically add the newly found ancestor to their tree.

Ancestry.com, LLC.'s influence extends beyond its flagship website, ancestry.com. In addition to the U.S.-based website, some of Ancestry's sister sites (ancestry.ca, ancestry.co.uk, ancestry.com.au) became dominant players in family history markets in other countries. In the U.S., Ancestry's domination of the family history market has been bolstered by its acquisition of other popular U.S. for-profit family history websites, such as fold3.com, genealogy.com, and archives.com. While owned by Ancestry, these sites have kept their own identities. Many of these acquired sites provide genealogical data from original source material, as well as images of original records, along with family history production tools.

Ancestry also has begun to purchase U.S. family history websites that are "free" to users and contain content largely created by users. These sites offer free production tools, with the expectation that users will make their content publicly accessible on those websites. The most popular such site is Findagrave.com. Like similar community-run sites, Find A Grave was largely managed by a group of volunteer administrators and the website's biographical and genealogical content was created by thousands of volunteer contributors. Find A Grave content takes the form of "memorials" rather than family trees. Memorials for deceased individuals include information such as a biographical sketch of the deceased, images of the deceased person's grave marker, images of the deceased, and immediate familial connections (such the

names of deceased person's parents, spouse and children). Ancestry purchased Find A Grave in

2013, when it had close to 13,000 visitors a day, and has maintained it as a free website as usage

has grown. ("Top 100 Genealogy Websites of 2014 (page 2)," 2014; "Top 100 Genealogy

Websites of 2015," 2015; "Top 100 Genealogy Websites of 2016," 2016). With the purchase of

Find A Grave and several other family history websites, Ancestry solidified its control and

influence in the online family history arena.

The emergence of Ancestry, Family Search, Find A Grave and several smaller family history

websites has lowered the bar for entry into the family history hobby. Millions of Americans are

trying their hand at family history and production ("Ancestry.com LLC Reports Fourth Quarter

and Full Year 2015 Financial," 2016; "Ancestry.com Traffic, Demographics and Competitors -

Alexa," 2016; News & Farnham, 2012; Rodriguez, 2014). Given the influx of novice-producers of

family history content on the web and an awareness of the need to educate newcomers about

family history research and production, many family history websites, including Ancestry and

Familysearch, have provided online didactic learning resources to its users. However, inaccurate

and poor-quality user-contributed content is an ongoing problem for many family history

production sites (Willever-Farr & Forte, 2014). It remains to be seen whether online

instructional resources are being used by many of the contributors, or, if used, whether these

resources are having any impact on the quality of user-contributed content. In addition, online

family history production systems provide few checks and balances to ensure the accuracy of

user-contributed content. For example, Ancestry does not support open-editing of user-

contributed content, as is the case with Wikipedia, so there is little opportunity for inaccurate

content to be "fixed" by more skilled researchers (Willever-Farr & Forte, 2014). Similarly, while

Find A Grave does support some open- editing of a small subset of memorials, it is not permitted

for most memorials. As a result, there are few, if any, opportunities for more skilled researchers to correct inaccurate or poorly researched content on the website (Willever-Farr & Forte, 2014).

Given the popularity of Ancestry and Find A Grave as production sites and the extent of their heavily mined user-contributed content, inaccurate content on these sites has a disproportional impact on the quality of family history content being produced on the web. If Ancestry subscribers and Find A Grave were building content and not sharing that content via the website, the impact of inaccurate content would be less problematic. This is not case: Ancestry encourages contributors to make their trees publicly viewable (to Ancestry subscribers) and accordingly, has made the access default setting "public" for all trees. As of 2015, the site provided access to 70 million user-contributed trees ("Ancestry.com LLC Reports Fourth Quarter and Full Year 2015 Financial," 2016), which are mined by millions of FHRs for their own research. Subscribers are also encouraged to connect individuals represented in their family trees to related individuals in others' trees. These links have the potential to transform isolated family tree building projects into contributions to an extensive public resource—a worldwide kinship graph. If a sizable portion of the content is inaccurate, the linking to inaccurate trees can lead to a domino effect, dispersing inaccurate information widely. Similarly, Find A Grave has upward of 157 million memorials or "grave records" and is the most heavily mined "free" family history-oriented website in the U.S. ("Find A Grave - Millions of Cemetery Records," n.d.; "Top 100 Genealogy Websites of 2016," 2016). As with Ancestry's family trees, inaccuracy in Find A Grave memorials has widespread impact on the quality of online family history content. Find A Grave's popularity and the tendency of family history researchers to copy and paste others' shared content into their own online family histories means inaccurate user-contributed Find A Grave content can proliferate across websites (Willever-Farr & Forte, 2014).

This dissertation examines the problem of inaccurate user content on family history websites, with a particular focus on Ancestry and Find A Grave, by gaining an understanding of the online research practices of family history researchers that tend to manifest in the production of more accurate family history content. Further, it will explore how family history websites may engender and better support these research practices to encourage the production of more accurate user-contributed content.

CHAPTER 3: PILOT PROJECTS

Two pilot projects were undertaken to examine the problem space – family history research and

production websites—and to gain an understanding of the dynamics of online collaboration

among family history researchers (FHRs). Many FHR websites and online family history

production tools were examined, but in the end, two websites were selected for the pilots:

Ancestry.com ("Ancestry") and Findagrave.com ("Find A Grave"). Ancestry was chosen due to its

popularity and its prominence in the U.S. Find A Grave was chosen as a research setting to

determine if a nonprofit, more community-governed family history production site presented

different experiences for FHRs. The first pilot examined the interactions of FHRs on a popular,

publicly accessible Ancestry.com message board; the second delved into the production

practices and behaviors of contributors on Ancestry and Find A Grave.

3.1 Pilot #1-- Analysis of Ancestry.com Message Board

To support the large numbers of FHRs who are online, web-based Q&A forums have cropped up

for those who are seeking and sharing family history information. While the use of genealogy

Q&A forums is widespread, relatively little is known about the interactions among users of these

forums. This knowledge gap led a research group of Lisl Zach, Andrea Forte, and I to examine

message posts on a heavily used Q&A message board on the popular genealogy website,

Ancestry.com (Willever-Farr et al., 2012). Research questions that informed the study included:

1) What kinds of social interactions emerge among askers and answerers on a message

     board for genealogists?

2) In what ways do the mediated environment and existing practices of genealogists influence those interactions?

3) Are Q&A forums serving as means to educate users about the practices of a specific community of practice, in this case, genealogists?

Method

To answer these questions, we examined interactions on Ancestry.com's United States General Message Board. I extracted and coded text from all messages that appeared on this board during the period January 1, 2010, to December 31, 2010, totaling 1,086 posts. Institutional Review Board (IRB) approval was received for use of the publicly available data. Existing typologies taken from previous Q&A studies were used to develop an initial list of coding categories. The content of an initial set of two hundred posts was analyzed using the initial list of coding categories. Code refinement was carried out iteratively, with additional codes being identified and codes that did not adequately represent posts being replaced during the coding process. Coding was done at the post level; each post was given one or more codes based on its primary content. I categorized all posts in each message thread, rather than only categorizing posts that appear to be "answers" or "questions." Doing so allowed for the examination of all interactions between the posters, not only those interactions that have been deemed useful for automated retrieval (i.e. those that can be clearly defined as questions and answers). Message threads were also analyzed in light of what is known about the genealogical community of practice, the technological context, and the emergence of family history data sources on the web. A summary of the final coding categories is shown below (Table 1).

Table 1: Askers' and Answerers' Posts

Askers' Posts	Answerer's Posts
Factual Question	Factual Answer
Instruct Question	Instruct Answer
Source Question	Source Answer
Request for Family Contact	Family Connection Answer
Unclear Question	Probes
Elaboration	Opinion
Gratitude	Encouragement

Factual answers were also examined to determine, if possible, the source of the data. In addition to coding individual posts, statistics were generated on the number and types of posts each asker and answerer generated. Other frequency data, such as the number of answers posted in response to a question, were collected. SPSS was used to generate descriptive statistics.

Findings

The content and statistical analyses revealed three major findings. First, the abundance of online genealogical data and the structure of online technological tools appear to be influencing answerer response behavior. Although Yakel and Torres (2007) found in face-to-face settings that genealogists provided assistance to other FHRs that was instructional in nature, most answerers in our study (53.6%) provided factual family history data, not instructional

information (24.8%). Access to family data online and the ease with which this data can be copied and pasted into a message is facilitating the provision of factual answers and appears to be privileging factual answers over instructional ones.

Second, easy access to family data may also be shaping other elements of the discourse between askers and answerers. The answerers' approach to "answering" questions by locating factual information about specific families for the askers may have encouraged the use of "probes." Some answerers (21.6%) and many "super sharers" (66.6%) used probes or probing questions to cull contextual data from the askers. Genealogists, particularly seasoned ones, are aware of the difficulty of finding the right data. Many families share the same surnames, so finding the right data for the right family is difficult. Contextual information about the family becomes paramount when attempting to locate correct family data, particularly when searching large online genealogical databases in which hundreds, or even thousands, of records exist for the same surnames. With such obstacles in place, many answerers wisely used probes to gain more knowledge of the family in question to improve the likelihood that the data they collected was correct. Answerer probes would be useless if the askers did not respond with further data. Almost seventy percent (69.6%) of the answerer requests for more information (probes) received asker responses (elaboration) in which additional information was offered. This indicates that askers do not always know what information to include in their questions and that answerers may benefit from using probes to extract information that was left unstated.

Third, cooperation did appear to be occurring on some of the studied threads, but it did not reflect the types of cooperation and collaboration found in face-to-face settings. In face-to-face meetings of FHRs, Yakel and Torres (2007) found collaborative group problem-solving, which involved FHRs sharing both implicit and explicit knowledge of the research process. Group

problem-solving behaviors were also found on the Ancestry message board, with over eighty percent (83.6%) of the message board threads revealing cooperative work among a group of answerers. Collaboration in which answerers jointly reached some sort of synthesis on an answer was rare on this message board. However, cooperation between answerers was common, in that many answerers employed each other's answers to find additional data for the asker. This cooperative engagement, however, rarely included cooperative instruction: The online environment and the availability of family data at one's finger tips appears to shape the types of cooperation that occurs on the message board, leading genealogists to engage in cooperative research, rather than collaborative instruction.

Discussion

The lack of instructional guidance being offered on message boards may present challenges for virtual family history communities of practice. Message boards appear to be primarily facilitating data provision, rather than instruction. Since family history research appears to involve specific types of knowledge and skills, how are family history production sites such as Ancestry.com facilitating the learning of these skills and knowledge if message boards are not serving this purpose? Ancestry.com does provide a plethora of didactic and static learning resources, available for free on its website. Yet there was no evidence in the analyzed posts that answerers are referring askers to these resources. These learning resources were not mentioned once in the message posts. The findings raise questions about how individuals are learning the family history craft in virtual environments and the implications of those learning experiences on FHRs' online interactions.

3.2 Pilot #2 -- Production Practices on Ancestry and Find A Grave

The first pilot left me with an inchoate picture of family history research and production websites. With Andrea Forte's assistance, I undertook a study to gain a better understanding of this problem space. The study enabled me to immerse myself in the family history research and production space, and gain a rich and in-depth understanding of these sites and their production tools.

Methods

The high-level question that motivated the study was, "What are the social and technical features of open collaboration systems that support family history researchers?" To answer this question, I undertook a qualitative study using multiple data sources. The primary empirical support for my analysis came from message board/forum data and interviews with FHRs. I selected two family history production websites to examine: Findagrave.com (Find A Grave) and Ancestry.com (Ancestry). These popular websites were chosen because of their prominence, the number of contributors involved, the historical value of the materials being archived, and the potential differences that may exist between family history production on a commercial website vs. a community-run website. To better understand how the sites support production and what kinds of technological constraints they impose on production activities, I spent months observing activity and trying out both Ancestry and Find A Grave. I also examined the sites' terms of service and guidelines to understand the production rules. These observations raised many questions about how people perceive their activities. Did users view their work on these sites as building a public resource or as a personal project? For what audience was this often-painstaking work intended? What kind of research experience do contributors bring to the sites?

As these questions accumulated, I developed an interview guide to explore Ancestry and Find A Grave contributors' experiences and perceptions of the sites.

I sought to interview individuals who had provided content on both websites, and used message boards and word-of-mouth on both Ancestry and Find A Grave to recruit participants. Eight individuals agreed to be interviewed: six women and two men. Participant ages ranged widely: one in his 20s, three in their 40s, two in their 50s, and two in their 60s. All eight interviewees had contributed content to both websites and had participated on both sites' message boards/forums. Interviews lasted from 40 minutes to 2.5 hours; most interviews were over an hour in length. Interviews were recorded and transcribed. I performed three iterations of coding to identify thematic patterns in the transcribed data using Atlas.ti (open coding) (Corbin & Strauss, 2008) . Then Andrea and I discussed and worked with the resulting codebook to refine and further articulate the relationships among codes (axial coding) (Corbin & Strauss, 2008).

Interviews revealed that production problems were sometimes hashed out on Ancestry's message boards, Find A Grave forums, or via private messaging. Although private message exchanges are not observable, message board/forum posts are publicly available, and I collected a strategic sample of public discussion threads to further understand the dynamics of online family history production. Similar to the method used by Forte, Humphreys, and Park (2012) , I used the codebooks developed from interview data to inform our analysis of public message board/forum posts.

Ancestry and Find A Grave message boards are different and required different sampling strategies to assemble a corpus for analysis. Although interviewees had identified the Ancestry message boards as an important communication channel for discussing community issues and

airing grievances, I knew from my previous research that much of the content on these boards involved help requests that are not clearly related to the production of artifacts on the site. Some interviewees, fortunately, mentioned that discussions regarding family tree production were often conducted on the "Ancestry" threads. This pointed me to a section of the message boards which enabled the researchers to identify specific boards that might provide additional information on production activities on the site. This yielded five message boards with 1,512 total posts between January 1, 2011, and March 1, 2013. Compared with Ancestry, Find A Grave forums are smaller in number, and I knew less about their composition. As a result, I initially read all 2,291 posts for the period January 1, 2013, to March 1, 2013, to get a sense of the content of the different forums. Forums that were overtly off-topic, such as discussions about politics or vintage cars were discarded, leaving five forums to explore. For the period January 1, 2011, to March 1, 2013, there were 763 posts on the five production-oriented Find A Grave forums. I analyzed a total message board corpus of 2,275 posts.

Finally, I used the interview codebook to categorize content from all the posts, which worked well and yielded a rich set of complementary data from both intervention-based interviews and naturally occurring online discussions.

Research Settings

As mentioned earlier, two FH production sites were chosen for the study: Findagrave.com (Find A Grave) and Ancestry.com (Ancestry). I view these websites as sociotechnical systems, each influenced by the website owners and administrators, as well as the contributors. This section will give an overview of the production systems, artifacts, and guidelines that set the stage for family history work on each site.

Findagrave.com

On Find A Grave, tens of thousands of contributors have uploaded artifacts related to deceased individuals. Each entry for a deceased person is called a memorial and typically includes a biographical sketch of the deceased, images of the grave marker, and sometimes images of the deceased. In addition, any registered user can upload virtual memorial tokens (e.g. virtual flowers) to the memorial. An example of a typical memorial is presented in Figure 1.

Figure 1: Find A Grave Memorial

Submitters retain editorial control of memorials they create. In the case of famous persons and Medal of Honor recipients, the site retains editorial control. Find A Grave is free to use and all of its content is freely accessible. There is no licensing requirement for submitted content beyond the requirement that it not infringe on intellectual property laws of any country.

The Find A Grave website is maintained by its founder, Jim Tipton, who first developed the website as a place for contributors to provide content about famous persons' graves. Over time, interviewees stated that the site has become more family-history oriented, as more and more FHRs have added content to the site. Features such as the ability to hyperlink memorials to the memorials of related individuals have been added to meet the needs of FHRs. The Find A Grave website is currently administered by Tipton and six additional volunteers, who together manage conflicts among contributors and resolve other problems and technical issues. The site also features discussion forums for socializing and help seeking.

Ancestry.com

Ancestry is a large site owned by a private equity firm and for a subscription fee provides online access to records and family tree production tools. Through arrangements with archival repositories in both the U.S. and Europe, Ancestry has digitized, indexed, and transcribed records and made them available to subscribers through online databases. In addition to providing access to records and transcribed data, Ancestry's family tree offers a complex suite of production tools that enable subscribers to enter simple information about deceased individuals and record their relationships to others. Documents, images, and videos can be uploaded and linked to individuals represented in the tree. To build family trees on the website, subscribers can search Ancestry's databases for deceased individuals by name, date, and other variables and then automatically populate their family trees with the found information and images. Ancestry

also provides a hint feature: Once a contributor creates a portion of a family tree, the website

will generate "hints" that appear as shaking leaves within the tree. These hints are automatically

generated and may lead to relevant transcribed data or digitized records, such as census

records, military records, and other user-contributed family trees. Ancestry encourages

contributors to make their trees publicly viewable (to Ancestry subscribers) and to connect

individuals represented in their family trees to related individuals in others' trees. These links

have the potential to transform isolated family tree building projects into contributions to an

extensive public resource—a worldwide kinship graph. Figure 2 features an Ancestry family tree.

Figure 2: Ancestry Family Tree

Contributors, however, have control over whether to make their tree public or private and whether to allow their trees to be connected to others. Other features of Ancestry include free, public message boards, and a private messaging function for users with accounts. Message boards are moderated by volunteer administrators.

Findings

Through interviews and analysis of contributor posts four themes were uncovered that are relevant to my dissertation study:

- representation of family;

- conflict between experienced, meticulous researchers and careless researchers;

- awareness of family history production as a public resource and the need for accuracy.

In the following paragraphs, each of these themes will be explored in more detail.

Family Matters: The Representation of Kin

All of the interviewees, as well as many Find A Grave and Ancestry forum/message board participants, spoke or wrote of the importance of controlling the content of their own family information on FHR websites. How relatives are remembered was a primary concern of family history contributors. Family trees and Find A Grave entries were viewed as means of memorializing and presenting information about the deceased for future generations. Thus, deceased relatives' identities were at stake in the creation of virtual memorials or family trees. As one interviewee explained:

> So my family, we have two Frank Greeleys and I posted a picture of one online. On Find A Grave. Well, I see it on Ancestry that that picture is attributed to both of the Frank Greeleys. And it's anymore, people just grab the wrong picture and

upload the wrong person and then they never go back and check and so I think
that's another thing that we're running into. That people just aren't checking
and not realizing that just because you have a picture of someone with this
name, it doesn't mean that it's definitely this person. You have to look and be
like, well, there's two of them. So obviously you have to look at both. So which
one could it be?...Future generations are going to be like, I found five different
family trees and they're all different and which one is correct?

Contributors' heighted sense of the role virtual family histories will play in the collective
memory of their family's past influenced their thinking about editorial control. They want to
create and control their family stories for their own families. On the other hand, Ancestry and
Find A Grave serve as historical resources for larger publics and, according to our findings, rely
on both family and non-family contributors to increase the amount of content on their websites.
This raises the question: How does one build policies, norms, and permissions on a collaborative
platform that respects the existential need of families to control the representation of kin, while
supporting the production of a historical resources for broader publics?

Closed Editing Systems

I found that Ancestry and Find A Grave take different approaches. Ancestry does not provide
community guidelines that address an individual's right to control content about his/her own
family. However, guidelines generally assume that contributors are creating family trees for
their own families, and the site's infrastructure supports editorial control of family tree content
by the original creator, who is often a family member. Yet posts on the studied message boards
suggest that some contributors are building family trees for non-related families, particularly for
famous people, on the site. Ancestry guidelines do not bar this activity outright, and Ancestry
appears to benefit from the increasing number of family trees in its online databases. The family

tree database is not only a production space for personal family histories but an information resource for its subscribers.

There are few opportunities for Ancestry contributors to work cooperatively to create content, which appears to support the notion of restricted control of family history content, presumably by family members. Subscribers can invite specific individuals (such as interested family members) to collaborate on a family tree, but cannot make a tree editable by the Ancestry public. The only other opportunities for cooperative production are linking trees with other contributors' trees (if the other contributors approve) or offering suggestions (such as corrections) or other information to other contributors by way of messaging.

Find A Grave takes a different approach. The website provides opportunities for joint construction of content by family and non-family contributors. On the site, any registered user can create a memorial for a deceased person, so long as one does not already exist for that individual. The original creator of the memorial retains editorial control of biographical information in the memorial, but any registered contributor can upload photographs to the memorial, as well as virtual memorial tokens (such as virtual flowers). This hybrid approach to editorial control (keeping images of graves or memorial tokens and textual authorship separate) facilitates cooperation among family and non-family contributors, and supports the development of Find A Grave as a public resource. The editorial structure of the site, for example, allows non-family members to provide grave images for other contributors' memorials, thereby creating richer and more complete memorials.

When it comes to biographical content of memorials, however, Find A Grave production guidelines explicitly favor family control. According to the Find A Grave's "Family First" policy, a

direct relative within four generations (siblings, parents, grandparents, and great-grandparents)

has the editorial rights to a related family member's biographical content ("the memorial") if the

original contributor is not a direct family member. If the original contributor refuses to transfer

the memorial, direct relatives can take up their case with volunteer administrators, who serve as

community mediators.

Echoing the Find A Grave production rules, interviewees and posters voiced support for familial

ownership or restricted editorial control of memorials and family trees. One might expect that

the desire to control potentially negative content about family members was a key motivating

factor in FHRs' desire for restricted editorial controls. In fact, interviewees never mentioned this

and few message board posts addressed this issue. The more prominent concern was accuracy

of biographical information, not whether the content was flattering. Contributors feared that if

anyone could modify content about their family members, lots of inaccurate information might

be added. Inaccurate information was tantamount to besmirching their relatives' reputation, as

one Ancestry message board poster declared:

> "People that play fast and loose with facts when they simply collect names on
> their trees are creating problems for those people that care about truth in their
> family histories. It's also disrespectful to the ancestors and the lives they led to
> deliberately disregard facts. Even if the lives they led were less than
> 'admirable'."

"Getting the facts right" about one's ancestors for current and future generations appeared to

be a key motivating factor for contributing and controlling content.

A Battle Over Accuracy: The Clash Between Careful Researchers and "Clickologists"

Fears of inexperienced FHRs polluting others' family trees or memorials with inaccurate

information contributed to the study participants' strong desire to control content on their

family trees and memorials. Most interviewees saw themselves as serious, experienced FHRs. Many posters on the message boards and forums we examined touted their experience level and their careful and thorough research practices in discussions. These vocal, central members of the FHR community stress the importance of careful research and the collection of evidence to back up historical claims about family members. Since the majority of records or documents about deceased individuals are not digitized, collecting evidence often entails traveling to brick and mortar archives and graveyards. Many experienced FHRs have developed artifactual literacy (the ability to analyze and interpret primary documents), subject knowledge (such as knowledge of the history of a geographical region), as well as "archival intelligence" or knowledge of archival principles and practices. These literacies enable FHRs to triangulate data from various sources to ensure that found information is relevant to their ancestors.

The FHRs we studied prided themselves on these skills; however, they also frequently lamented FHRs who were careless in their research practices and who cared more about increasing the size of their family trees or the number of memorials they contributed than about accuracy. In fact, Ancestry was seen as a provocateur of poor-quality family histories on the web. Many asserted that Ancestry had so lowered the bar for entry into family history construction that people were creating content without really knowing how to properly conduct family history research. Contributors referred to such FHRs as "clickologists" and "half researchers" as they populated their trees and memorials without attempting to verify that the information was actually about their relatives. Many asserted that Ancestry had encouraged "clickologist" behavior by suggesting that a subscription would allow people to create an accurate family tree in a weekend with the tools and data on the site.

Automation also played an insidious role. Some complained that inexperienced researchers

mindlessly accept Ancestry hints — the little shaking leaves ("poison ivy" or "poison leaves," as

two interviewees described them) offered by the sites' automatic search feature — without

verifying the data, leading to family trees full of erroneous information, or "genealojunk."

According to many FHRs, Ancestry's algorithms that identify potential kinship links do a poor job

at triangulating data from different documents and data points, leading to many false positives.

Yet inexperienced FHRs still click on them and populate their trees with the incorrect data. This

has created a domino effect as erroneous information propagates from tree to tree. An

interviewee opined:

> Well sometimes those little leaves are poison ivy. And people will ask us to
> trees that are no true facts. Like my husband, after I tried to get some people to
> make some corrections I just went ahead and made my tree private...My
> husband is a direct descendant from a woman who had a child and he was born
> in the 1830s and she never married. She never had any more children...but
> people keep wanting to marry her off and have her with two or three other
> children and I email them and I say look, Nancy never had any other children.
> She never married. Your information isn't correct. But it's still perpetrating a
> falsehood. So that's the only reason I went private but most anybody who asks
> me, I give them access or I give them what they want to know.

On Find A Grave, interviewees and posters also indicated concern about "half researchers" and

"half-fast researchers" adding or editing content on their family's memorials. Negative

experiences with careless researchers appeared to be a strong motivating factor for the studied

contributors to maintain control of their own family trees or memorials. These experiences

beget skepticism of radically open editing policies and colored the interviewees' and posters'

views of sharing content and working collaboratively with others to build family history content

for the wider public.

Family History as Public Resource: Conflicted Views of Sharing and Editorial Control

All interviewees described a similar trajectory from researching their own families and sharing

family history content on FHR-oriented websites to conducting research about non-family

members. Contributors moved from meeting an immediate personal information need to

attempting to meet the potential needs of future researchers. Among interviewees who

described contributions to Find A Grave, it was evident that they viewed the site as a historical

resource for the public and not only a resting place for their own family history. For example,

interviewees described visiting graveyards that were geographically close to them in order to

collect and share information about non-related individuals buried there. Others described

collecting information on deceased individuals at local archives and adding it to Find A Grave

because others may need that information. Coordination also took place virtually among

contributors who were not co-located. Many interviewees spoke of meeting other contributors

on one of the Find A Grave forums and deciding to partner with them to create memorials.

 Problems with inaccuracies and conflict between contributors may be even more of a stumbling

block for cooperative production on Ancestry. Inaccuracies led all but one of the interviewees

and many of the studied posters on Ancestry's message boards to make their family trees

"private," preventing others from connecting to their trees. By not making their trees public and

not allowing connections, they made a choice to not share their family histories with other

Ancestry subscribers and to limit their cooperative production activities. This choice was the

result of their concern that information and images from their carefully researched family trees

were being lifted without their permission and placed on family trees that were for families that

were not related, because the tree builder did not conduct careful research. As one forum

poster on Find A Grave concluded:

> Half-fast researchers are what poison Ancestry.com trees…I warn new researchers to RUN from those, and do their own primary research. Those sources are a thousand times more accessible than they were when I started, but sometimes you have to get off the couch and do the legwork yourself. The point-and-click researchers who think info needs to come to them are the same ones who think others are 'meanies' when actual re-search-ers do re-search and then don't share everything freely. Screw your time, effort, gas, injuries, bug bites, sun burns.

Poor research practices and inaccuracies appear to be hindering cooperative work on both Ancestry and Find A Grave.

Discussion

Websites for family history production attract millions of contributors. It is unclear, however, whether a large number of participants is sufficient to sustain the long-term cooperative production of an accurate family history resource for current and future generations. The need to not only attract participants, but enculturate them into the practices of a community has been recognized as a primary challenge for open collaboration systems (Forte & Lampe, 2013). This often entails interactions among more and less experienced contributors; however, I observed that at times conflict over accuracy has pitted experienced, careful researchers against less experienced researchers who are viewed as careless. Thus, the very individuals who could be mentoring and teaching newcomers about good research practices are angry and lose patience with the restrictions and editorial controls on both Ancestry and Find A Grave. Many committed contributors want the content to be accurate but must rely on others to address inaccuracies and have few avenues to teach newly minted contributors good research practices.

The culture of the FH community stresses the importance of helping others with their family history research (Fulton, 2009; Willever-Farr et al., 2012; Yakel & Torres, 2007). In face-to-face

environments, this often takes the form of informal instruction (Yakel & Torres, 2007). In a virtual environment with information about deceased persons at the researchers' finger tips, "help" often takes the form of sharing data instead of providing instruction (Willever-Farr et al., 2012). This study's findings suggest that less experienced contributors may need instruction on good research practices in order for them to be able to create quality content online. This raises a broader question: What is the most appropriate model for online FHR communities to bring newcomers into the community of practice?

Familial oversight is an important feature of sites that support family history production and introduces restrictions on who can edit content. These restrictions raise the bar for participation and reduce the opportunities for low-risk peripheral participation that could ease new contributors into the creation of biographical content. For example, on Ancestry, one must create a family tree in order to make an original contribution and populate it with all the requisite information. One obvious solution is to allow families to choose whether they want to open their family tree or memorial for editing, thus creating some opportunities for newbies to make small edits. I found, however, that experienced FHRs are resistant to open editing, and may not welcome this possible solution.

Ancestry has attempted to solve the problem by automating some of the family tree-building process, thus lowering the bar for newcomers. Yet I learned that this approach has not proved effective for enculturating newcomers, who tend to rely on a recommendation engine to identify family links without understanding how to verify the quality of recommendations. Instead of encouraging newcomers to learn how to construct family histories and conduct research, this approach has led to family trees and memorials littered with incorrect information and a general distrust of automated tools among experienced FHRs.

As was found in my previous study of Ancestry message boards, unstructured discussion spaces do not necessarily facilitate the exchange of instructional guidance. Yet research on FHRs in face-to-face contexts suggests that many experienced practitioners are committed to helping newcomers learn the craft. I also found that the interviewees wanted to help others but were frustrated with online spaces that did not provide good opportunities for learning to take place before contributors were producing content. Given the willingness of many FHRs to help others learn the craft, it makes sense to provide virtual spaces to better facilitate learning that are more robust than typical message boards.

I suggest that family history "learning spaces" should support both synchronous and asynchronous interactions between experienced FHRs and newcomers. These learning spaces could include resources, such as learning modules and video tutorials, to which experienced FHRs could direct newcomers, as well as "practice" spaces in which newcomers could work with experienced FHRs to build memorials or family trees. Such learning spaces may be a means of bringing newcomers into the family history community of practice and ultimately may help improve problems with content inaccuracies on the sites.

Unanswered Questions

While this study sketched out some of the dimensions of the problem space, it left many unanswered questions. Beyond automated suggestions on Ancestry, what other factors are contributing to inaccuracy issues in user-contributed content? What features can be incorporated in the research and production tools to support the production of high-quality user-contributed content?

CHAPTER 4: LITERATURE REVIEW

Literature from several domains provides both a theoretical and empirical foundation from which to build this research study. Topics and research domains that will be addressed in the literature review include:

- Information-seeking Behaviors

- Information Behaviors of Family History Researchers

- Historical Expertise

- Open Collaboration Systems

- Social and Collaborative Learning

First, literature on information behaviors, such as information-seeking and research behaviors, will be explored. While this literature does not directly examine FHRs' information behaviors, it does provide a starting point from which to explore those behaviors. Next, literature on the information behaviors of family history researchers (FHRs) will be discussed. This literature provides rich contextual information about FHRs, mostly in face-to-face environments, that serves as a beginning point for my research on FHRs' research and production behaviors in virtual participatory environments.

In addition, literature on historical expertise will be explored. From the pilot studies, it appears that more experienced FHRs learn how to avoid pitfalls that can lead to the production of inaccurate content. This study will explore FHR expertise and its impact on content production. To provide some foundation for this examination, literature pertaining to expertise in historical research, in general, will be discussed. Most of the studies on historical research expertise are

situated in physical spaces, such as face-to-face environments and brick and mortar memory institutions. Nonetheless, this literature may serve as baseline to understand the role research expertise plays in the production of family history.

To provide overarching conceptual frameworks for online participatory family history production spaces, literature on web-based open collaboration systems will be examined. Family history production systems share many characteristics, as well as important differences, with existing open collaboration systems. Nonetheless, socioeconomic and sociotechnical models of open collaboration systems may provide frameworks to understand FHRs' research and production behaviors in online spaces. The bulk of empirical studies on open collaboration systems focus on the Wikipedia and Open Source communities. For the purposes of this literature review, Wikipedia studies will be the main focus, as Wikipedia biographical content and the Wikipedia production system share some characteristics with family history content and FH production systems.

Lastly, collaborative learning theory and situated learning in communities of practice will be explored. Participatory environments such as FH production websites have the potential to support collaborative learning between experts and novices, which may engender better research and production practices among contributors to family history websites. Collaborative learning and situated learning theories inform my understanding of learning in participatory spaces. Through collaborative learning, online FH communities may have the potential to become production *and* learning spaces that enable contributors to perform tasks through peer collaboration that could not be achieved alone (Vygotsky, 1931). Since the family history research community can be considered a community of practice, theories associated with learning in such communities are relevant to my work.

4.1 Information Behaviors

This dissertation study was designed to provide a clearer and more in-depth understanding of

the online information-seeking and production behaviors of FHRs. Many existing studies on

information behaviors provide a foundation from which to understand FHRs. This is not an

exhaustive review of all literature on information behaviors; rather, it is a synopsis of literature

that is specifically relevant to the information behaviors of FHRs that were uncovered in the

pilot studies.

First, it is helpful to define foundational terms related to information and the user. The following

terms capture the main concepts related to information-related behaviors:

- *Information*: any difference one experiences in her environment or within herself

  (Bates, 2009; Case & Given, 2016).

- *Information need*: a realization that one's knowledge is not sufficient to meet a need or

  a goal, what Belkin et al. (1982) labeled as an "anomalous state of knowledge".

- *Information searching:* the act of searching for information, whether in an online or

  offline context (Bates, 2009). Includes less structured activities such as browsing.

- *Information-seeking*: a conscious effort to acquire information to fill a need or a gap in

  knowledge. Information can also be encountered via serendipity or via people sharing

  information that may meet an information need (Case & Given, 2016; Erdelez, 1997;

  Williamson, 1998).

- *Information use*: what one does with information once actively found or through

  serendipity. May include ignoring the information (Bates, 2009; Case & Given, 2016).

- *Information behavior*: the broader contexts of how people deal with or interact with

  information in their lives, which takes in account contextual factors such as time and

geography. Simply put, the ways in which people "need, seek, give, and use

information" (Pettigrew, Fidel, & Bruce, 2001, p. 44) . It encompasses active

information-seeking and information behaviors associated with non-intentional

encountering of information (Case & Given, 2016; Fisher, Erdelez, & McKechnie, 2005).

- *Information practice*: Like the concept of information behavior, information practice

  denotes discernable patterns exhibited when individuals interact with information, but

  the latter term has been preferred by many scholars who view information-related

  behaviors as socially situated (Savolainen, 2007). For these scholars, information

  practices encompass the idea that "information-seeking and use are constituted socially

  and dialogically, rather than based on the ideas and motives of individual actors"

  (Tuominen, Talja, & Savolainen, 2005, p. 328)

This dissertation study holds that FHRs' information behaviors are embedded in an established

but evolving community of practice, and that those information behaviors are socially situated

in that community of practice. Thus, the term "information practices" is often used.

4.2 User-Centric Studies of Information Behaviors

Information behaviors have been studied for decades. Earlier studies focused on users'

interactions with institutional systems and sources. In the past 30 years, the focus of this

research has expanded to include how information seekers make sense of the world around

them through information (Case & Given, 2016). These studies began to exam contextual

factors, such as time, geography, and situational contexts, to gain a holistic understanding of the

information worlds of individuals and groups. This dissertation explored information behaviors

from the user-centered perspective, a perspective supported by the work of many information

science scholars.

Foundational User Behavior Models

A large corpus of theoretical research examines information-seeking in both digital and analog

contexts (e.g. Dervin; Belkin, 1980; Ellis, 1989; Kuhlthau, 1991; Marchionini, 1995; Wilson,

1997; Chun Wei Choo, Brian Detlor and Don Turnbull). Dervin's early work shaped the discourse

surrounding information behaviors for years. She asserted that information was not simply

received by the information user; rather, the information user actively makes sense of her

surrounding reality, attaching personal meanings to information (Dervin, Foreman-Wernet, &

Lauterbach, 2003). Dervin's sense-making concept has four elements: a situation in time and

space that defines the information problem(s); a gap between the situational context and the

desired context; an outcome, which is the consequence of the sense-making process; and lastly,

the bridge, a means by which the gap between the situation and the outcome is spanned.

Dervin's work identified situational contexts as having an important place in understanding

people's information worlds. Belkin took this concept further with his ASK model, which

identified the significance of situational as well as task-related factors in the development of

"anomalous states of knowledge."

Wilson (1999) proposed a model that focused more on information seeker and their interactions

with information systems and services. Wilson suggested that information-seeking behaviors

occur as a consequence of a need perceived by an information user. His model illustrates a

process in which the information user makes demands upon formal or informal information

sources or services to address an information need, which results in success or failure to find

information that fulfills the user's need. The model concludes with the individual making use of

the information found, which may either partially or fully satisfy the perceived need. If the

found information does not meet the need, the individual reiterates the search process.

Kuhlthau's (1988, 1991) influential information search model delineates several stages in the

search process: *initiation, selection, exploration, formulation, collection,* and finally *presentation.*

Over the course of this process, user thoughts move from being general and vague to more

focused and concise. Emotional responses also change over the course of the search process,

from anxiety and uncertainty at the outset to increased confidence and satisfaction at the

successful conclusion of a search. Another influential information model developed by Ellis

(1989) has been the basis of many information-seeking studies (e.g. Bates, 1989; Choo, Detlor, &

Turnbull, 2000; Meho & Tibbo, 2003). Researching the behaviors of academic researchers in the

sciences, social sciences and the humanities, Ellis (D. Ellis, 2005; 1989) developed an information

behavior model that included six activities: *starting*, identifying key authors or studies relevant

to topic; *chaining*, following chains of citations, both forwards and backwards; *browsing* through

relevant sources or systems; *differentiating* or filtering material; *monitoring*, maintaining

awareness by regularly following particular sources; *extracting* relevant materials; and *ending*

the search. Meho and Tibbo (Meho & Tibbo, 2003) extended Ellis' model to include four

additional features: accessing, networking, verifying, and information managing. The accessing

stage encompasses behaviors associated with attempts to access the needed information, as

information seekers do not necessarily always start with raw materials or direct sources of

information, but rather may first interact with bibliographic records or search results. Verifying

encompasses activities related to checking the accuracy of information, and networking involves

interacting with others to gather information, build collections, and share information. Lastly,

information management is characterized by activities related to filing, archiving, and organizing found information. Ellis's model, along with Meho and Tibbo's enhancements, is relevant to some of the FHRs' web-based research activities uncovered in the pilot studies. As a result, this model helps frame findings of this study.

Many of these models are built on empirical evidence gathered from college students and scholars. In contrast, other influential information behavior frameworks elucidate information behaviors in the context of life outside of the workplace or school. Savolainen (1995) was one of the first to examine "everyday life information-seeking" and to highlight its importance. Most information behavior studies to that point explored workplace and academic information-seeking processes. Savolainen identified two major dimensions of everyday life information-seeking: the seeking of orientating (monitoring, day-to-day awareness) and the seeking of practical information to answer a specific question or fill a particular need. Spink and Cole (2001) suggest that the difference between occupational or academic information-seeking and everyday life information-seeking is that the former takes place in a more controlled environment, with beginning and end points, and culminates in a tangible product. In contrast, everyday life information-seeking is more fluid, variable, and dependent on the "the motivation, education, and other characteristics of the multitude of ordinary people" (Spink and Cole, 2001, p. 301).

Stebbins' (2009) examination of information behaviors associated with leisure pursuits suggests that there are similarities between work- or academic-related information behaviors and hobby-related information behaviors. Stebbins identified three types of leisure pursuits: serious leisure,

casual leisure, and project-based leisure. Serious leisure is "the systematic pursuit of an amateur, hobbyist, or volunteer core activity that captivates the participant with its challenges and complexity" (Stebbins, 2009, p. 622). Casual leisure is a short-lived activity which is immediately and intrinsically rewarding and requires little or no special knowledge. Project-based leisure is a short-term "one-shot or occasional, though infrequent, creative undertaking carried out in free time"(Stebbins, 2009, p. 622). Due to the complex nature of serious leisure, Stebbins argues that serious leisure is a ripe area for investigation by information science researchers. Stebbins also examines the "publics," or social worlds, that hobbyist tend to develop. Within these publics are different actors: tourists, regulars, and insiders. Tourists participate in a social world temporarily for momentary reasons. Regulars routinely participate in the social world. Insiders take this involvement further by showing "exceptional devotion to the social world they share, to maintaining it, to advancing it"(Stebbins, 2009, p. 627). Participants in serious leisure tend to identify strongly with their chosen hobby. However, casual leisure is too fleeting for most people to do so. Given that family history research and production requires specialized skills and knowledge, it is a particularly apt problem space in which to examine serious leisure. However, it appears that participants with varying degrees of commitment and knowledge of the family history craft may be inhabiting and contributing to online spaces for family history research and production (Willever-Farr & Forte, 2014). How might the presence of tourists, regulars, and insiders in online publics for FHRs impact the character of contributed content?

Relevance

This dissertation study is not focused on a system approach to relevance. While the algorithms that underlie search engines on family history websites can be improved, this study focuses on how FHRs make their own judgements of relevance, specifically how they decide if found information pertains to their ancestors. Search engines may aid them in this process, but they may also lead them astray (Willever-Farr & Forte, 2014). Arguably, FHRs' knowledge of how to establish "facts" about their ancestors is the most valuable tool FHRs have in determining whether found information pertains to their actual ancestors. Given this, the large corpus of literature related to developing algorithms to improve relevance of search results will not be discussed. Studies that focus on the nature of relevance from the user-perspective will be explored. These studies can be roughly categorized as: studies that explore topicality and other criteria used to judge relevance; studies that examine the changeable nature of relevance and situational relevance; and studies that explore cognitive and affective relevance (Zimmer, Arsal, Al-Marzouq, & Grover, 2010).

Relevance Criteria

An important area of relevance-related research focuses on what Saracevic calls "relevance clues" or the criteria individuals focus on when making relevance inferences (2007). Research on relevance clues abounds, with a primary focus on "topicality" relevance (Barry, 1998; Fitzgerald & Galloway, 2001; Maglaughlin & Sonnenwald, 2002; Toms, O'Brien, Kopak, & Freund, 2005; Zimmer et al., 2010). Scholars have identified many criteria used to make relevance decisions, and while they often employed different terms to describe the criteria, the criteria have many

conceptual similarities (Saracevic, 2007). According to Saracevic, seven major criteria have

emerged from relevance clue research:

Content: topic, depth, scope, currency and clarity

Object: tangible characteristics of the information object, such as type of object, format,

availability

Validity: accuracy, trustworthiness, verifiability of information

Use or situational match: information is appropriate for use situation or task at hand

Cognitive match: information is understandable, novel, matches mental effort user is willing to

put forth

Affective match: information aligns with emotional need, expectation

Belief match: information aligns with user's beliefs

As there are no existing studies that examine how FHRs make relevance judgements (i.e.

whether found information or documents are relevant to their ancestors) in online

environments, it is unclear to what extend these clues are employed by FHRs.

While there are no existing studies that specifically examine FHR's relevance judgements, a few

studies focus on topics that are particularly pertinent to FHRs' information behaviors in online

contexts. The FHR process involves examining not only representations of original sources, but

also images of the actual sources. Choi and Rasmussen (Y. Choi & Rasmussen, 2002) studied

faculty and graduate students' retrieval of historic images included in the Library of Congress

American Memory archive. The study found that the user's perception of topicality was the

most important factor across information-seeking activities. Other factors that were important

in determining relevance were image quality and clarity. Participants also judged relevance

based on title, date, subject descriptors, and notes provided in the image metadata. Dziadosz

and Chandrasekar (2002) found that thumbnails coupled with text summaries better supported

relevance decision-making than text-only summaries or thumbnails alone. Another study

indicated that textual content of webpages (e.g. scope and depth of content), coupled with

other features such as layout and authority, were the most frequently mentioned criteria for

determining relevance (Tombros, Ruthven, & Jose, 2005).

Fact-finding is also a major component of the FHR process. Kelly, Murdock, Yuan, Croft and

Belkin (2002) examined features of web documents which influence users' relevance

assessments for two types of information activities: task-oriented (how do I do something?) and

fact-finding. They found that depending on the task, different web page features, such as length

and presence of tables and figures, influenced relevance judgments.

The Changeable Nature of Relevance

Several scholars have characterized the act of judging relevance as a dynamic, cognitive process

that is dependent on the individual's perceptions and knowledge (Bates, 1989; Belkin, 1980;

Dervin, 1983, 1998; Taylor, 1986; Wilson, 1999). Relevance is a fluid concept in which an

individual's view of what is relevant may change. Wilson (Wilson, 1981) suggests that a user's

judgement of relevance may change based on her knowledge level at a given time and based on

information that has been found. Similarly, Bates' (Bates, 1989) research suggests that relevance

is an evolving, rather than static, construct, even in the context of a single search. In other

words, what constitutes "relevant" information to the researcher can change or evolve as the researcher goes through the search process. Bates asserts that a single search can involve the use of different tactics (e.g. footnote chasing, author searching, citation searching, etc.) and the use of multiple queries and search terms that are revised as the researcher is exposed to different information sources found in the search process. Bates suggests that this multifaceted search process resembles "berry-picking," with the researcher picking different bits of information from here and there to acquire a sufficient answer to her question.

Several studies of college or university faculty and students found that information seekers' inferences pertaining to found information or information objects is dependent on the stage of the research task (Anderson, 2005; Barry, 1994; Spink, Greisdorf, & Bateman, 1998; Tang & Solomon, 2001; Vakkari & Hakala, 2000; Wang & White, 1995). The criteria participants used to determine the relevance of information remained the same at different stages in the research process. However, the weight given to different criteria often changed. Information seekers also became more discerning and stringent in their relevance inferences as they became more focused in the research process.

A small number of studies examine whether users judge information objects independently or not, and whether the size and number of information objects influence relevance judgments. Katzer and Fletcher (Katzer & Fletcher, 1992) found that users' perceptions of the type and extent of information available will influence their perceptions of relevance. Information objects presented earlier are more likely to be deemed relevant than those presented later (Bar-Ilan, Keenoy, Levene, & Yaari, 2009; Eisenberg & Barry, 1988; Huang & Wang, 2004; Purgailis Parker

& Johnson, 1990; Xu & Wang, 2008). However, if there are only small number of information objects, order of presentation has little effect. The extent of the information included, such as title, author or abstract, also influences relevance judgements (James, 1996).

Other scholars have examined the impact domain knowledge has in relevance judgements. Many such studies enlist "judges" to determine to what extent relevance judgements are consistent across individuals and/or groups. Judges who possess domain knowledge or higher expertise tend to have higher agreement in relevance judgments and are more stringent in their relevance judgments (Dong, Loh, & Mondry, 2005; Kinney, Huffman, & Zhai, 2008; Wildemuth, 2004). Lesser subject knowledge or expertise results in more leniency in relevance judgments.

In 2007, Saracevic pointed out that most relevance studies focused on college/university students and to a lesser extent, faculty: "We are really getting a good understanding of student relevance" (Saracevic, 2007, p. 2141). He further asserted that, "If we are to gain a better understanding of relevance behavior and effects in diverse populations, other populations should (or even must) be studied as well" (Saracevic, 2007, p. 2141). This dissertation study helps to fill that gap by developing an understanding of how FHRs make relevance judgments.

4.3 Web-based Information-seeking Models

The ever-growing amount of online archival data and digitized archival documents along with the popularity of family history production websites has transformed family history research from a largely face-to-face, analog-based hobby, to one firmly embedded in the web environment. Further, family history research involves a number of information tasks and is tied

to an existing, but evolving, community of practice that began in the analog world and has

moved to a digital world. As Talja and Preben Hansen (2006) emphasize, information practices

are firmly embedded in work and other social practices, and these practices draw on the social

practices of a community of practitioners, a sociotechnical infrastructure, and a common

language. Accordingly, studies that examine web information behaviors from the perspective of

information-seeking tasks conducted by groups of practitioners, typically in work environments,

are the most relevant to FHRs' information behaviors and will be examined in this section of the

literature review. Everyday life information-seeking studies do not adequately capture the

information-seeking and production behaviors of FHRs, and, therefore, these studies shed little

light on FHRs' information behaviors.

Drawing heavily from the work of Wilson (1997) and Ellis (1989), Choo, Detlor, and Turnbull

(2000b) developed an integrated model of information-seeking on the web that provides a

holistic depiction of web-based information activities (Table 2). On the left axis of the model,

episodes or types of information-seeking activities were plotted according to the four

information-seeking behaviors: undirected viewing, conditioned viewing, informal search, and

formal search. On the top axis of the model, episodes were plotted according to the occurrence

of one or more of the six types of information-seeking behaviors identified by Ellis (1989, 1990):

starting, chaining, browsing, differentiating, monitoring, and extracting. The model illustrates

the connections between motivations (the strategies and reasons for viewing and searching)

and moves (the tactics used to find and use information).

Table 2: Choo, Detlor, and Turnbull's model of Behavioral Modes and Moves of Information-seeking on the Web

Behavioral Modes and Moves of Information-seeking on the Web

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	Identifying, selecting, starting pages and sites	Following links on initial pages				
Conditioned Viewing			Browsing entry pages, headings, site maps	Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	
Informal Search				Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	Using (local) search engines to extract information
Formal Search					Revisiting 'favorite' or bookmarked sites for new information	Using search engines to extract information

Empirical evidence for this model was derived from Choo, Detlor and Turnbull's (2000b) study of

the information-seeking behaviors of 34 work professionals from seven companies. Using a

survey, an application that recorded Web browser actions, and personal interviews, Choo et al.

found that their model captured the universe of web-based seeking tasks carried out by the

participants. The participants engaged in a range of complementary modes of information-

seeking, ranging from undirected viewing that does not involve the pursuit of specific

information, to more formal searching that involved focused searches for information for

decision-making or to take work-related actions. Undirected viewing involved starting and

chaining actions; conditioned viewing involved differentiating, browsing, and monitoring

actions; informal search consisted of differentiating and localized extracting; and lastly, formal

search consisted of systematic and thorough extracting.

Other scholars described similar web-based information activities, but often in different terms.

Morrison, Pirolli, and Card (2001) studied web-based information activities, or "actions," by

asking study participants to describe a recent episode in which they found information on the

web that led to a significant decision or action. Morrison et al. found that participants engaged

in four main actions: collect, find, explore, and monitor. Similarly, Sellen, Murphy and Shaw

(2002) studied the Web activities of 24 knowledge workers and found they engaged in six main

actions: finding, information gathering, browsing, transacting, communicating, and

housekeeping. Kellar, Watters, Shepherd (2007) found several distinguishing characteristics of

four types of information-seeking activities (fact-finding, information-gathering, browsing,

transactions). In particular, information-gathering was found to be the most complex task. On

average, study participants spent more time on information-gathering, and when doing so,

viewed more webpages and more heavily used web browser functions. Kellar et al. assert that

this finding suggests "more research is needed to support users in their Information gathering

tasks" (Kellar et al., 2007, p. 1017).

Rieh's (2003) research was a sharp departure from previous research with its predominant focus

on web-based information behaviors in workplace contexts. Exploring web searching behavior in

home environments, Rieh found that participants searched differently in the home than what

had been found in previous studies of workplace information-searching. In Rieh's study,

participants searched the web more frequently for shorter periods of time, and their searches

tended to be broader in nature. Several studies that examined search engine log files, which

included many types of users, revealed that the typical web users engage in short search

sessions, use few terms per query, typically peruse one result page per query, and, if they use

advanced query operators, do not use them effectively (Jansen & Spink, 2006).

Expertise and Web-based Information-seeking

Most studies of expertise and web-based information behaviors explore the differences in

experts and novices search behaviors when using the web. These studies have largely found that

expert web searchers' behaviors differ from those of novices (Aula, Jhaveri, & Käki, 2005;

Cothey, 2002; Hölscher & Strube, 2000; Hsieh-Yee, 1993; Jenkins, Corritore, & Wiedenbeck,

2003; Kellar et al., 2007). Those differences include:

- Experts employ more varied search strategies
- Experts utilize more diverse sources

- Experts employ both browsing and searching

- Experts reformulate queries if initial search is unsuccessful

- Experts persist longer with searches

- Experts are more proficient at query formulation

Cothey (2002) found that as college students became more experienced, they began to visit a more distinct set of web pages and used fewer search queries to locate needed information. This finding suggests that experienced searchers were more proficient at search-term formulation. Jenkins, Corritore, and Wiedenbeck (Jenkins et al., 2003), found that web novices tended to use a breadth-first pattern of information-seeking. This pattern was characterized by an unwillingness to stray more than one click beyond a hub webpage, as the web novices became disoriented even one link from their home base.

Concerned with the varying definitions of novices and experts and different measures of success in these studies, Aula and Nordhausen (2006) used a different approach -- the "task completion speed" or TCS model -- to assess web searching experience and skill. The model involved variables such as speed of query iteration, the length of queries, the proportion of precise queries, and the speed of evaluating result documents. They found that the increase in the years of web use was related to improvement in TCS in terms of broader search tasks. Less frequent Web use was related to a decrease in TCS in the fact-finding tasks.

Other studies explore the influence of subject domain expertise on web information-seeking behaviors. Holscher and Strube (2000) compared a group with expertise in information-seeking

with another group dubbed "double experts" or experts in information-seeking and the subject

domain of the search. They found that domain knowledge helped participants devise query

terms that were more likely to produce a result set of useful size and relevance. Jenkins,

Corritore, and Wiedenbeck (Jenkins et al., 2003) found that even individuals with web searching

experience, who possessed little domain knowledge, had difficulty locating relevant information

on the web. However, they also found that domain experts with little web searching experience

exhibited a "pattern of searching that had greater breadth than depth to minimize getting lost"

(Jenkins et al., 2003, p. 81). This same pattern was evident in non-domain experts with little web

searching experience.

Bhavnani (2002) found similar novice information behaviors. Studying the search strategies of

health care and shopping experts in their own domain, as well as the other domain, Bhavnani

found that domain experts were aware of key resources for their domain and often went

directly to those online resources, rather than employing general web search engines. He also

found that domain experts had a general strategy for performing search tasks. In contrast,

domain novices typically started with general web search engines, examined few items in the

results sets, and usually terminated their searches before relevant information was found.

Employing a slightly different use context, a full-text hypertext system, Marchionini, Lin, and

Dwiggins (1990) compared the searching performance of search specialists, domain specialists,

and novices. They found that both types of experts were more successful than novices in

completing search tasks. Zhang, Anghelescu, and Yuan (2005), and Duggan and Payne (2008)

also found that domain experts (engineering and football, respectively) were more successful in

their searches (i.e. finding relevant information) than domain novices. However, these studies

found different results in terms of length of search queries of experts. Zhang et al.'s engineering

experts used longer queries; Duggan and Payne's football experts used shorter queries than

domain novices.

White, Dumais, and Teevan (White, Dumais, & Teevan, 2009) studied the effect of domain

expertise (medicine, finance, law and computer science) on web search behavior. They found

that domain experts search differently than non-experts in terms of the sites they visit, the

search terms they use, their search patterns, and the success of their searches. White et al.

employed these finding to develop a model to predict domain expertise effect on search

behaviors. Similarly, Wildemuth (Wildemuth, 2004) and Vakkari, Pennanen, and Serola (2003)

found that as students gained more knowledge of a topic over the course of their education,

they became more skilled in search term formulation, including using more varied and more

specific vocabulary in searches.

Information Coping Strategies in Web Environments

The online environment — characterized by the ease of access to a sea of information and the

need to cope with large result sets -- may influence information-seeking behaviors and lead to

adaptive strategies in search strategies. Today, information seekers are inundated with online

information; in many cases, it is impossible for information seekers to view every potentially

relevant source. Individuals must make decisions regarding how many sources to explore.

"Satisficing," a concept defined by Simon (1972), is an important coping mechanism in managing

large quantities of information. Satisficing is the act of individuals making choices without

contemplating all of the possible rational options. Several scholars have found that different

types of users employ satisficing behavior in online contexts. Warwick, Rimmer, Blandford,

Gow, and Buchanan (2009) introduced the concept of a "strategic satisficing stage" in which

undergraduate students "used the expertise that they had, by now, gained in information-

seeking to create time-saving strategies to complete the coursework with minimum effort" (p.

2409). Further, the ease with which one can access information online may have conditioned

individuals to expect easy access to information in general, and this may influence information-

seeking (Nicholas, Rowlands, Huntington, Jamali, & Hernández Salazar, 2010). Connaway,

Dickey, and Radford (2011) found that convenience is a criterion in college students' choices and

actions during all stages of the information-seeking process. The notion of "convenience"

influences users' selection of sources, their satisfaction with the source and its ease of use, and

their time horizon in information-seeking. Participants in the study satisficed their information

needs to quickly select sources whose convenience made them "good enough."

Rimmer, Blandford, Gow, and Buchanan (2009) suggest that satisficing behavior is likely to be

found in non-academic settings. Choo (2007) found evidence of similar online satisficing

behaviors in corporate settings. He found that the organization accepted a satisficing approach

to information-seeking; in other words, it was considered acceptable for staff to stop when they

had found adequate information to inform the current decision rather than seeking all the

information that might be pertinent to that decision. Agosto (2002) found that several factors

influenced young web surfers' decisions to stop searching. Factors such as limitations of time,

cognitive capacity, boredom, and physical discomfort influenced participants to stop searching.

However, in contrast to Simon's concept of satisficing, these factors led participants to stop

searching even when a good enough website had not been found. Agosto also discovered that

participants employed "reduction" to manage information overload on the web. The most frequently used reduction method was to return to known sites to find information.

Mansourian's (2008) research sheds further light on how individuals cope with the inability to "immediately" find relevant information in the web environment. Studying the information behaviors of academic biological researchers, he found that they employed both passive (giving up; goal modification) and active strategies to cope with the lack of immediate access to desired information. Active strategies included: revising strategies (altering search queries or information sources); seeking help (from colleagues and information intermediaries); or postponing the search until another time. His research suggests that the perceived importance and interest level in the search will impact the level of effort made, as well as the number of coping strategies employed.

4.4 Information Behaviors of Family History Researchers

There is a dearth of literature on FHRs' online information-seeking and production behaviors in web environments. Nonetheless, existing research on FHRs' information-seeking and sharing behaviors may serve as a baseline for understanding FHRs' online production behaviors. In this section, information-seeking and sharing behaviors will be explored, as these appear to be closely intertwined behaviors in the production process that undergirds the creation of family histories.

Information-seeking Behaviors of FHRs

Information-seeking studies suggest that FHRs follow common research practices. Duff and

Johnson (2003) studied amateur genealogists' use of in-house archival finding tools and their

face-to-face interactions with information professionals. They identified three different stages in

the genealogical research process: 1) collecting names of family members; 2) gathering detailed

information on family members; and 3) contextualizing the detailed information by learning

about broader history. Duff and Johnson also found that information-sharing was important in

the information-seeking process, because finding aids failed to serve the needs of FHRs,

prompting FHRs to seek help from one another.

Building on Duff's and Johnson's seminal paper, Darby (2011) and Darby and Clough (2013a)

studied the information-seeking behaviors of FHRs in both online and brick-and-mortar archival

contexts. Based on interviews with FHRs, Darby created a model of FHR information-seeking

that has eight distinct stages: 1) trigger event (starting reason for the research); 2) collect items

(collect family anecdotes, documents, artifacts); 3) learn the process (learn basic approach to

research); 4) getting started (initial step leading to "breaking into" the records); 5) fill out tree --

easy (add easy individuals to tree); 6) fill out tree –medium (add more difficult individuals); 7)

fill out tree – hard (fill in hard gaps and sort out ambiguities); 8) push back (finding more

obscure family members and sources). While Duff's and Johnson's model is sufficiently broad to

cover the many information-seeking and production activities of FHRs, Darby's model focuses on

family tree creation, which is only one of the many production activities FHRs engage in (e.g. the

production of family history narratives, memorials, scrapbooks, image and document archives).

Darby's model also does not account for social information-seeking, which is common on FHR

websites (Willever-Farr et al., 2012). It seems that Darby's model exists in a vacuum in which

there is no interplay between FHRs and others, and FHRs and online-production communities. In addition, Darby and Darby and Clough do not explain how FHRs "learn the process." Nor do they address how the prevalence of online genealogical data may be impacting the research process. For instance, how do FHRs decide which records are relevant in a sea of online genealogical data?

Lucas (2008) began to answer some of these questions by employing interviews, observations, and talk-in-action studies to capture the FHRs' information-seeking behaviors at brick-and-mortar memory institutions. He found that participants followed two distinct information-seeking processes. The first centered on locating and selecting sources in the library. The second was the information-selection process participants utilized to search for information within the sources. Lucas's findings suggest that FHRs develop skills and knowledge related to library and archival information systems and other reference sources. These two areas of knowledge – "archival intelligence" (Yakel & Torres, 2003) and reference source literacy – likely play role in the production of online family histories, but this study does not explore information-seeking and production of family histories in online contexts. Is the prevalence of online genealogical resources and user contributed historical content shaping FHRs information-seeking behaviors? Another limitation of this study, is that the research setting (libraries/archives) provides little insight on social information-seeking and cooperative production behaviors, which are common in the FHR community (Fulton, 2009; Willever-Farr & Forte, 2014; Yakel, 2004b; Yakel & Torres, 2007).

Some scholars have begun to explore FHRs' information behaviors in online contexts, but there continue to be significant gaps in this area of research. While Yakel's 2004 study did not drill deeply into the online research practices of FHRs, she did describe these practices broadly:

looking for "information in chat rooms or listservs, finding contact information for libraries or archives...planning research trips, constructing websites which record family trees and other research results" (Yakel 2004). In her master's thesis, Garrett (2010) surveyed thirty FHRs and found that study participants lauded many of the features of Ancestry.com, such as ease of use, speed, and numerous collections; almost half of the participants were able to locate information on difficult-to-find ancestors by using Ancestry.com databases. However, despite the acknowledged benefits of the website, the majority of participants' responses suggested that it would not eliminate the need to visit physical repositories, as Ancestry.com was a starting point, not an end point. All of the survey participants were regular users of the Alabama Department of Archives and History, and this may have skewed the results pertaining to archives use. Garret does not include any questions on her survey about the use of Ancestry.com for peer production of family trees or other resources – she appears to view Ancestry.com as a virtual space for information consumption, but not for information production. Her study does not examine, in depth, web-based information-seeking patterns of FHRs. In exploring how archives can provide better online access to local records, Friday (2014) began to address this gap. Friday found that the online family history research process was highly iterative and circular in nature. FHRs in her study demonstrated information needs that were both informational and affective, with a high degree of personal involvement and commitment. She also found that participants displayed well-developed information literacy skills, but discussion of these literacies was not well-developed, as this was not a focus of the study.

While existing literature does fill in some knowledge gaps regarding FHRs' online information behaviors, these studies do not focus on the actual online spaces where much of the family

research and production is occurring. As a result, the close intertwining of production and

research have not been fully explored. Nor do we have an understanding of how people are

learning the FHR craft and how that may influence their online research and production

behaviors. Additionally, these studies focus exclusively on the behaviors of more experienced

FHRs. There is little knowledge of the information behaviors of novices compared to those of

experienced FHRs, and how these information behaviors may be impacting content production

on FHR websites. Further, literature on FHRs information-seeking provides a cursory view of

how FHRs work alone or with others in online contexts to find resources, determine their

relevancy, interpret information from sources, manage conflicting information found in the

myriad of online (and offline) sources (a common problem FHRs face, e.g. Yakel, 2004) and then

use the data to build biographical content on families and individuals on production sites such as

Ancestry.com or Findagrave.com. The information literacies needed to build accurate family

histories in online contexts is largely unknown.

Information-Sharing and Cooperative Production Behaviors of FHRs

Studies of FHRs' information-sharing behaviors primarily focus on informal information-sharing

among FHRs. Duff's and Johnson's (2003) seminal paper on FHRs is one of the few papers on

FHRs' information behaviors that refers to the idea of user production. Duff and Johnson (2003)

found that genealogists often worked around archival access systems and relied more on their

own social networks than on information professionals to obtain information:

> Many genealogists develop a parallel system to help them retrieve records
>
> because the archival information system fails them. This parallel system

includes finding aids organized according to the genealogist's point of view,

along with a strong network of colleagues and courses. With this parallel system

and personal expertise, genealogists are able to circumvent the archival

retrieval system and find the information they need (p. 92).

This suggests that FHRs began building their own research aids or guides in response to the

limitations of archival information systems that were not designed to meet the information

needs of FHRs. Duff and Johnson studied family history researchers before the web became a

ubiquitous tool for family history research and production. The web has become a successful

venue for FHRs to create and make accessible their own research aids, but it also enabled a new

type of production: collaboratively built family history archives filled with primary materials such

as digitized photographic images of people, places, and documents. While this has become a

widespread phenomenon, no empirical study explores this online production activity of FHRs.

Building on Duff's and Johnson's findings about information-sharing among FHRs, Yakel (2004b)

and Yakel and Torres (2007) interviewed FHRs and observed their interactions during genealogy

society meetings and found that FHRs' information-seeking involves the gathering of factual

information, which leads FHRs to connect and build a common identity with other FHRs and

relatives. Part of connecting with other FHRs is sharing information, advice, and instructional

guidance -- either virtually or in face-to-face meetings (Tucker, 2009; Yakel, 2004b; Yakel &

Torres, 2007). Sharing information was not only a means of connecting but a means of

advancing individual research interests: by sharing information, FHRs were creating a reciprocal

information-sharing culture that could be drawn upon to meet individual information needs

(Fulton, 2009; Tucker, 2009). Yakel (2004b), Fulton (2009), and Tucker (2009) found that web-

based chat rooms and forums were important venues for such informal information-sharing among FHRs.

While the web is seen as an important information-sharing venue by Yakel (2004b) and Fulton (2009), they do not explore other uses of Web 2.0 tools, such as for the collective production of family history archives. Tucker (2009), however, envisions the possibility of shared spaces for the production of family history archives when she writes, "family historians already work in such spaces as various photo-sharing sites. Archives could offer similar preservation and access, with custody shared or specifically allotted" (p. 268). It is unclear, though, whether she views such spaces as collaborative spaces, in which many contributors work collaboratively to build family history archives. Nonetheless, Tucker is one of the few scholars studying FHRs who views the web as a production space for shared family history content.

Three additional studies on FHR's use online artifacts to explore the online information-sharing behaviors of FHRs. Using a quantitative approach, Veale (2004) studied a FHR newsgroup, soc.genealogy.britain, to determine whether the newsgroup constitutes an online community based on two models of virtual communities: the OC Model (Schoberth, Preece, & Heinzl, 2003) and the Common Ground Model (Whittaker & Terveen, 1998). Using metrics derived from these models, Veale found that participants in the newsgroup were increasingly familiar community members, were largely interactive in their communication activities, and conducted rich conversations as indicated by long threads. Long threads, however, may not necessarily represent rich conversations.

Using both qualitative thread analysis and descriptive statistics, Willever-Farr, Zach and Forte (2012) studied the exchanges of FHRs on a popular message board on Ancestry.com. Through

qualitative analysis of the message threads, the researchers were able to understand the meaning and purpose of the exchanges between users. Previous research on the face-to-face interactions of genealogists found that genealogists tend to help other genealogists by providing instructional guidance both on a one-to-one and a many-to-one basis (Yakel & Torres, 2007). In contrast, this study found that the presence of online genealogical data and the affordances of interactive computer technologies appear to be pushing message board answerers away from providing instruction on how to find family history data and toward providing family data outright. Additionally, answerers on the message board worked cooperatively to provide family data, suggesting that the web context is leading many genealogists to engage in cooperative research, not collaborative instruction. This study raises concerns about how FH virtual communities can help novices learn the craft of family history production. Message boards do not appear to facilitate many exchanges in which novices learn about how to conduct FH research. What other types of learning spaces might be more effective in enculturating newcomers into the FH community of practice?

A study of contributors' production experiences on Ancestry.com (Ancestry) and Findagrave.com (Find A Grave), a community-based website, yielded findings that suggest that not enculturating novices into the community of practice is creating conflict and leading to inaccurate biographical information on these sites (Willever-Farr & Forte, 2014). Through interviews of content contributors and content analysis of message board posts, Willever-Farr and Forte found "getting the facts right" about one's ancestors for current and future generations was a key motivating factor for contributing and controlling content about their ancestors. Ancestry's and Find A Grave's built-in editing controls allow the first contributor to retain editorial control over the content they submitted; these controls dovetailed with the

interviewees' and forum/message board posters' expressed desire to control the content they

submitted. However, while the interviewees and many posters' wanted to control content

about their relatives, they were frustrated that they could not directly "fix" inaccurate or poor

quality content being produced by careless and inexperienced contributors or "half-

researchers." Due to the editorial controls in place on both websites, experienced,

knowledgeable contributors could only suggest to the original contributors that corrections are

needed, but original contributors are under no obligation to take heed of these suggested

corrections. In addition, there are few avenues on the website for experienced contributors to

teach newly minted contributors research skills. Yet many research studies have found that

FHRs want to help novices learn the craft of family history research (W. M. Duff & Johnson,

2003; Fulton, 2009; Willever-Farr & Forte, 2014; Willever-Farr et al., 2012; Yakel & Torres, 2007).

Yakel and Torres (2007), on the other hand, asserted that FHRs "are seekers of meaning….As

seekers of meaning, they are less invested in proving the truth of stories and records, but more

in uncovering coherent narratives" (p. 111). Willever-Farr and Forte (2014) found, however, that

in the context of the online cooperative production of family histories, accuracy matters to

experienced contributors due to the public nature of user contributions – user contributed

content on these sites is a public resource and future generations may base their understanding

of deceased individuals and families on that content. As one interviewee stated, "Future

generations are going to be like, `I found five different family trees and they're all different and

which one is correct?'" If accuracy matters, how might the online FH community facilitate the

learning of good research skills so that user content is more likely to be accurate? How might

FHR websites be designed to support the production and sharing of more accurate user-

contributed content?

While not directly related to the building of histories about deceased individuals and their families, a study about peer indexing of census records on a FH website explores ways to improve user-contributed content. A study conducted by Hansen et al. looks at how to make peer indexing of documents, such as census records, more efficient while retaining accuracy in user transcriptions (Hansen, Schone, Corey, Reid, & Gehring, 2013). Peer indexing involves volunteers extracting data from digitized government documents, such as census records, and uploading transcribed data to family history websites. Hansen et al. compared two different types of quality control methods — arbitration and peer review — to determine which might prove to be more efficient and provide more accurate peer indexing on FamilySearch.org. They found that the peer review method took considerably less time to resolve transcription errors than arbitration, but with slightly less accuracy. While both appear to be viable options for resolving problems with the accuracy of questionable user-submitted transcriptions, these quality control measures are unlikely to mesh with the production rules on most FH sites. The restricted editorial controls built into most FH production sites and embraced by many FHRs do not allow for peer review or arbitration of user content (Willever-Farr & Forte, 2014). Collaborative learning of good research skills in virtual spaces on FH production sites may be a more viable approach, given the strong desire of many contributors to control the content related to their ancestors.

The dearth of literature on the online cooperative production of histories of deceased peoples and their families leaves an inchoate view of FHRs' research and production behaviors. Such knowledge is foundational to building effective online collaborative learning opportunities for FHRs.

4.5 Historical and Archival Expertise

How might the increasing numbers of digitized primary sources, search tools, and social and collaborative spaces on the web shape the skills and knowledge needed to become an expert at historical research? There is a dearth of studies that address this question. A few scholars have explored the trajectory from being a novice to an expert in history. These studies focus mainly on professional historians, not amateur family history researchers. However, skills related to historical research may be similar to the skills that need to be mastered to create accurate family history content. Cole (2000) found that historians move from the novice to expert stage due to their ability to use "name collection" — the identification and tracing of personal and organizational names relevant to their research project — to further their research goals. Name collection enables the expert historian to quickly scan large qualities of materials to determine their relevancy. Cole also asserts that name collection signifies the expert's ability to recognize patterns.

Wineburg (1991, 1998) found that historians had superior abilities at organizing and applying subject knowledge. In his first study, a group of high-achieving high school students and a group of historians with different subject expertise were given a test on American history facts (Wineburg, 1991). Some of the students performed better on the factual test than some of the historians; however, differences surfaced between all of the historians and all of the students when Wineburg studied how these two groups made sense of historic documents. Unlike the students, all of the studied historians, no matter the differences in their areas of concentration, were able to provide more elaborate explanations of events, provide alternative interpretations of those events, and use corroborating evidence more effectively to support their interpretations. In his second study, Wineburg studied two history experts —one a Lincoln/Civil

War expert and the other with expertise in American history, but not in 19th-century American history. The historians were asked to read and interpret a group of documents pertaining to Lincoln's complex view of slavery and equality. The Lincoln expert had no difficulty making sense of the documents; he brought his content knowledge to bear in when interpreting complicated and sometimes contradictory documents. The other historian quickly became confused by the details and the contradictions found in the documents, and initially performed no better than a group of student teachers that performed the same task in a parallel study conducted by Wineburg and Fournier (1994). However, the second historian, went further than his initial analysis, and developed a working hypothesis suggesting that the found contradictions were less a result of Lincoln's duplicity and the result of the historian's lack of knowledge of 19th-century America. After considerable effort and time, the second historian delved deeper into the provided documents and came to conclusions that echoed the Lincoln expert's conclusions. The student teachers never went beyond their initial interpretations. The second historian was able to monitor his current knowledge level, identify deficiencies in his knowledge, and use that understanding to develop strategies to fill in those gaps.

Other scholars explore historical research expertise from the perspective of archival and information science. Based on this literature, knowledge areas that have been identified as important to historical research include: subject or domain knowledge, artifactual literacy, and archival intelligence (Yakel & Torres, 2003). Domain or subject knowledge is an understanding of the topic being researched; in the case of genealogical research, this could be, for example, knowledge of immigration history or the history of a specific geographical region. Artifactual literacy is the ability to interpret records and assess their value as evidence. Artifacts used in historical research are often referred to as primary sources (Robyns, 2001). Primary sources are

subjective in nature and take many forms (e.g. documents and photographs). According to

manuals on the historical research process, interpreting primary sources requires conducting

external and internal criticism of the sources (e.g. Furay & Salevouris, 1988; Hockett, 1955).

External criticism involves the process of authentication and verification of authorship or

determining the "where, when, why, and by whom" a document was written. Internal criticism

is the process of evaluation and interpretation of the content after the researcher has

determined the provenance and authenticity of the source. At this point, the researcher must

determine the meaning of the information or evidence contained within the source. Only one

existing study explores primary source literacies in online contexts (Archer, Hanlon, & Levine,

2009). This study explored the effectiveness of a web guide to teach basic concepts pertaining to

primary source literacy. The study also explored the strengths and weaknesses of college

students' approaches to primary source research. Among their findings, two are particularly

pertinent to family history research:

- "Students relied on familiar tools without a clear understanding of whether those tools
  would produce what they were looking for. The lack of knowledge about the what
  attributes constitute a primary source, and how materials are made available, as well as
  unfamiliar terminology, meant students did not know how or where to look" (Archer et
  al., 2009, p. 419).
- Students also struggled with making distinctions between primary sources and other
  types of content in part due to the fact that the "Web blurs these distinctions by
  combining access to all research materials, primary and secondary, into a variety of
  different tools that are widely available" (Archer et al., 2009, p. 419).

They conclude by identifying specific types of knowledge that students need to possess to be successful at locating primary material that meets their research needs, including: an understanding of the scope and definition of primary sources; knowledge of key terminology and specialized tools for primary source research; and an understanding of how tools they are already familiar with can be exploited more effectively to locate primary sources.

Another literacy identified as being critical to historical research is "archival intelligence," or "a researcher's knowledge of archival principles, practices, and institutions, such as the reasons underlying archival rules and procedures, how to develop search strategies to explore research questions, and an understanding of the relationship between primary sources and their surrogates" (Yakel & Torres, 2003). Yakel and Torres interviewed 28 individuals about their research practices in brick-and-mortar archival settings and found that the main dimensions of archival intelligence are: 1) knowledge of archival theory, practices, and procedures; 2) strategies for reducing uncertainty and ambiguity when unstructured problems and ill-defined solutions are the norm; and 3) intellective skills, or the ability to understand the connection between representations of documents, activities, and processes and the actual object or process being represented.

Similarly, Duff and Johnson (2002) identified four major information-seeking activities historians undertook in archives: orientation (to a new archive, finding aid, source, or collection); searching for a known material (item, collection or form); accumulating contextual knowledge; and identifying relevant material. These activities often occurred "simultaneously and in no particular order" (W. M. Duff & Johnson, 2002, p. 492). Names were commonly used as an access point into materials. Tibbo (2003) surveyed 258 U.S. historians and found that they use a wide variety of primary sources and an equally wide variety of methods to locate these sources.

Methods included footnote-chasing and web-searching; web-searching most often involved

visiting a memory institution or archival repository website, but using web search engines (e.g.

Google) was not a commonly employed search method. Duff, Craig, and Cherry (2004) found

that many historians, even in the digital age, prefer originals to digital surrogates, which I would

suggest may be related to the poor quality of many digitized records (low dpi, blurry images)

and the fact that digitized records may lack important contextual and provenance-related

information. Their findings may suggest that a lack of contextual information makes

interrogation of digital primary sources more challenging, and this may impact how historical

research is carried out in web environments. However, more research is needed to validate my

assertions and the other ways online primary materials may be impacting historical research.

The dearth of studies on historical research in web environments leaves many unanswered

questions regarding what skills and knowledge are needed to become an expert at historical

research in the digital age. Few studies deeply examine primary-source or artifactual literacies, a

literacy at the heart of historical research; and even fewer examine these literacies in the web

context. As Archer, Hanlon and Levine opine, "research on 'primary source literacy' is still in its

infancy" (2009, p. 411). My dissertation will address some of these literature gaps by examining

literacies that are needed to research and produce family histories in the web context.

4.6 Open Collaboration Systems

This dissertation asserts that FHRs' information practices, particular those of experienced FHRs,

are socially situated, and that those information behaviors are embedded in an existing

community of practice. Further, online production of family histories has a cooperative

dimension, as production is occurring in the context of online sociotechnical systems that support cooperation, albeit a limited cooperation. While little research has been conducted on cooperative history production communities, other communities that collaboratively produce online content or products have been extensively explored. Open collaboration literature can be divided into theoretical and empirical studies. Unlike participatory archives literature, theoretical works on open collaboration systems do not conceptualize collaborative production activities in terms of professional practice, but rather draw upon socioeconomic and sociotechnical models. The bulk of empirical studies on open collaboration systems focus on the Wikipedia and Open Source communities. For the purposes of this literature review, Wikipedia studies will be the main focus, as Wikipedia biographical content and the Wikipedia production system shares some characteristics with family history content and FH production systems with a few notable differences:

- Wikipedia is an open editing sociotechnical system; most FH production sociotechnical systems are closed editing systems
- Wikipedia bars the inclusion of original content or content that stems from original research; FH production websites rely on content contributions that are a product of contributors' original research
- Wikipedia contributors have some control over the development or modification of production rules; many FH sites do not give contributors direct control of production rules – they can make suggestions but they must rely on volunteer administrators or website hosts to make the final production decisions.

Nonetheless, empirical studies that focus on Wikipedia biographical content, enculturating

Wikipedia contributors, and conflict and resolution within the Wikipedia community may shed

light on the challenges facing online FH production communities and potential solutions to these

challenges.

Empirical Studies of Wikipedia

While there are only a small number of empirical studies on co-production, there is a plethora of

studies on Wikipedia contributors or "editors." This literature review will explore "editor"

studies that pertain to the following research areas:

- Editor motivations

- Attracting, enculturating, and retaining editors

- Editorial collaboration and content quality

- Editorial ownership, conflict, and resolution

Contributor or Editor Motivations

Researchers have found that Wikipedians contribute for three main reasons: altruism or the

desire to share knowledge with users (Baytiyeh & Pfaffman, 2010; Cho, Chen, & Chung, 2010;

Kuznetsov, 2006), generalized reciprocity or collectivism (Batson, Ahmad, & Tsang, 2002; Cho et

al., 2010; Prasarnphanich & Wagner, 2011), and satisfaction or enjoyment (Kuznetsov, 2006).

Adding to this line of scholarship, Forte and Bruckman (2005) analyzed the motivations of

Wikipedians through the lens of Latour's and Woolgar's sociological work on scientific

publishing. Forte and Bruckman (2005) found some similarities in the incentive structure for

Wikipedians and scientists: authorship is important in both communities (for some Wikipedians)

and authors claim and receive credit from the larger community. Differences are also present. For example, there is a general belief in a more egalitarian form of knowledge production among Wikipedians than what Latour and Woolgar found in the scientific community. This belief, however, creates some confusion among Wikipedians regarding the relationship between author credibility and the reliability of Wikipedia.

Attracting, Enculturating, Retaining Editors

The need to not only attract participants, but enculturate them into the practices of a community has been recognized as a primary challenge for Wikipedia and other open-collaboration systems (Forte & Lampe, 2013). There are several studies that explore the sociotechnical features of Wikipedia that support new participants. One study explored how social interventions such as welcome messages and offers of assistance (B. Choi, Alexander, Kraut, & Levine, 2010) may help enculturate newcomers into the Wikipedia community of practice. Choi et al. studied the impact of seven socialization tactics used most frequently in Wikipedia – invitations to join, welcome messages, requests to work on project-related tasks, offers of assistance, positive feedback on a new member's work, constructive criticism, and personal comments – to determine their impact on newcomer's commitment to Wikipedia. Most newcomers contributed fewer edits over time; however, these declines were slowed or reversed for those who received welcome messages, assistance, and constructive criticism. In contrast, invitations to join led to steeper declines in edits. Other studies detail extensive education efforts like the Teahouse (Morgan, Bouterse, Walls, & Stierch, 2013), which provides a nurturing social space for newcomers in order to retain them as Wikipedia editors. Their findings suggest that several interventions/features aid retention of new editors, including:

intervening early in the editor lifecycle, creating safe and easy-to-use virtual spaces for newcomers, and facilitating positive interactions between newcomers and established community members.

Another area of scholarship focuses on the movement from information consumer to information producer within Wikipedia. Repeated visits to Wikipedia may lead a small number of individuals to begin to make small contributions to the website, and some of these individuals eventually will make larger contributions (Bryant, Forte, & Bruckman, 2005; Preece & Shneiderman, 2009; Yeow, Johnson, & Faraj, 2006). Preece and Shneiderman (2009) describe different levels of producer contribution and the trajectory of those contributions in a web-based peer-production environment. Only a fraction of users become contributors of user-generated content, such as uploading a photograph or writing a product review. Some users move beyond such individual contributions to become collaborators with others, forming tightly connected groups that may produce a Wikipedia article or an edited YouTube video. A small number of users may move on to become leaders, who participate in governance of the virtual production space. Governance activities include developing and enforcing policies, repairing vandalized materials and/or mentoring novices. Preece and Shneiderman frame this trajectory from individual contribution to collaborative work to leadership roles in their "reader to leader" model. The model illustrates the successive levels of social participation that includes reading, contributing, collaborating, and leading.

The participation trajectory from consumer to producer was also explored by Bryant, Forte, and Bruckman (2005). Borrowing from Lave and Wegner's community of practice paradigm and the notion of legitimate peripheral participation, Bryant, Forte, and Bruckman examine how the character of participation in Wikipedia changes over time and how Wikipedia's technical and

social structures mediate user activity. As participants became more involved in the Wikipedia enterprise, they moved from encyclopedia consumers to encyclopedia creators. Novices engaged in legitimate peripheral participation by gathering information from Wikipedia for their own use and by correcting mistakes on Wikipedia pages based on their personal expertise. Eventually, novices may develop article content and become experts. The designed functions of Wikipedia support peripheral participation, such as the "edit this page" feature on most Wikipedia pages. However, the majority of tools, such as the Watch lists, support experts as they take responsibility for larger chunks of Wikipedia. Legitimate peripheral participation does not exist on Find A Grave or Ancestry; for example, contributors are not given opportunities to make small edits to existing memorials or family trees due to the restricted editorial controls embedded into these two sociotechnical systems. How does a novice, then, learn the craft of FH research on these websites before being thrown into making major contributions?

Beyond the work of Bryant, Forte, and Bruckman (2005), there are no studies of how contributors learn to produce informational content for open collaboration systems. The focus primarily has been on how to attract and retain contributors, not on the connections between learning and continued participation or between learning and the production of quality content. This leaves many unanswered questions about how the facilitation of learning on open collaboration websites may impact content quality.

Editorial Collaboration and Content Quality

Several scholars have investigated the relationship between collaboration and data quality in Wikipedia. Many scholars have found that two indicators — number of edits and number of

unique editors — are associated with an article's quality (Arazy, Morgan, & Patterson, 2006; Kittur & Kraut, 2008; Lih, 2004; Stvilia, Twidale, Gasser, & Smith, 2005; Wilkinson & Huberman, 2007). In addition, high-quality pages (i.e., those selected as Featured Articles) have larger discussion pages (Stvilia, Twidale, Smith, & Gasser, 2008). Wilkinson and Huberman (2007) also demonstrated that there are more intense patterns of cooperation in the high-quality articles than in other articles.

Other scholars have found that the types of contributors and collaboration patterns they engage in also impact the quality of articles. Anthony, Smith, and Williamson (2009) found that high-quality content comes from two types of users: zealots (registered users with a strong interest in reputation and high level of participation) and good Samaritans (unregistered, anonymous and occasional contributors). Liu and Ram (2009) found that different collaboration patterns influence content quality: articles developed using patterns for which all-round editors (those who engage in many types of actions, such as sentence creation, deletions, source linking) played a dominant role are often of high quality; while patterns for which starters (who created sentences but seldom engaged in other actions) and casual contributors (who took fewer than four actions on a given article) dominate are often associated with low data quality. Kittur and Kraut (2008) examined how the coordination methods editors use affect article quality. They found two types of methods: explicit coordination, in which editors plan the article through communication, and implicit coordination, in which a subset of editors set direction by doing the majority of the work. They found that adding more editors to an article improved article quality only when they used appropriate coordination techniques and was harmful when they did not. Implicit coordination through concentrating the work was more helpful when many editors contributed, but explicit coordination through communication was not. Both types of

coordination improved quality more when an article was in a formative stage. While these studies provide a foundation for improving content quality, they do not address how learning the encyclopedic craft may increase the quality of user-contributed content.

Editorial Ownership, Conflict, and Resolution

Conflict is present on many open collaboration websites, including FH production websites and Wikipedia. Conflict can hinder the production of content, and thus, the management of conflict in open collaboration systems has been an area of active research. Researchers have studied the conflict and resolution process on Wikipedia extensively, viewing the technical environment as an influencing factor. Conflict on Wikipedia is typically resolved through various avenues that ideally involve rational discourse and consensus building. For example, "talk" pages that are separate from the encyclopedia content are spaces for contributors to discuss content changes and resolve disputes related to a specific article (Viegas, Wattenberg, Kriss, & Van Ham, 2007). Additionally, a revision comment field -- a free-form text field supplied by each contributor when submitting an edit -- is available for the contributor to describe what kind of changes were made and why (Kittur, Suh, Pendleton, & Chi, 2007). As a last resort, the Arbitration Committee can be called upon to resolve a dispute among contributors (Forte, Larco, & Bruckman, 2009).

Avenues for resolving disputes are a critical feature of Wikipedia, as conflict resolution is becoming a mainstay in the development of many Wikipedia articles. Kittur, Suh, Pendleton, and Chi (2007) found conflict impacts activity at the article level: The number of article edits is decreasing, while the overhead (the number of edits intended for communication and policy making) is increasing for individual articles. Later, Suh, Chi, Pendleton, and Kittur developed a

user-conflict model for Wikipedia ("Revert Graph") based on revert activity (revisions that void previous edits) and found four social patterns of conflict. The authors observed that conflict was not always tied to a specific piece of content — disagreements among user factions often propagated from one article into others (Suh, Chi, Pendleton, & Kittur, 2007).

Biographical content appears to be a particular source of ongoing conflict for the Wikipedia community. Like any Wikipedia article, biographies are expected to follow Wikipedia's established content policies, including Neutral Point of View (NPOV), Verifiability (V), and No Original Research (NOR). Biographical articles, particularly those about living people, have captured researchers' attention in part because they have necessitated a lot of policy intervention (Forte et al., 2009; Joyce, Butler, & Pike, 2011). In particular, "the impact on biographies of living persons (BLPs) in Wikipedia is that what used to be a tightly controlled artifact, created for a very small number of subjects, can now be created for and about anyone" (Joyce et al., 2011, p. 26). Unsurprisingly, BLPs and attendant policies have created conflict among contributors. Joyce, Butler, and Pike assert that conflict arises largely through differences in goals between and within groups on Wikipedia and that these differences are embodied in artifacts they refer to as "contentious objects." Contentious objects are elements in a social organization that serve as a common object of activity, but these activities contain elements of unrealized conflict. For the Wikipedia community, BLPs are particularly contentious objects.

Territorialism, the expression of ownership toward an object, is another source of contention among Wikipedia contributors. Territorial behaviors, if left unchecked, may result in conflicts over ownership and hinder cooperation. In Wikipedia, such territorialism is discouraged, although Thom-Santelli, Cosley and Gay (2009) found in an interview-based study that some Wikipedia editors develop the means to take "ownership" of articles by being the primary

contributor, initial content creator, main reviser of content, and/or using an editing style that

bolsters their primary contributor role. Reasons for this behavior included commitment to

producing quality articles that would stand up to heavy critique, rectifying vandalism, and

removing erroneous information or unsupported assertions.

Conflict spurred by territorialism may become apparent in such systems particularly when

contributors perceive themselves to be experts (Thom-Santelli et al., 2009). As contributors gain

expertise, they may become possessive of their contributions and the position of perceived

authority being an expert often brings. When others threaten that position, contributors'

perceived role as an expert can itself become a territory that they must defend by appropriating

features of the Wikipedia system, such as using watchlists to carefully monitor an article the

contributor feels stewardship over. Thom-Santelli, Cosley, and Gayet (2010) explore solutions to

the problem of expert territorialism in open collaboration systems by studying a mobile social

tagging system for a museum collection, rather than the Wikipedia system. The researchers

found that experts believed that certain tags were suitable and devalued contributions made by

novices by negatively rating them. As a result, Thom-Santelli et al suggest that novices may

require additional validation to encourage participation and engender feelings of ownership

toward the collaborative community. To accomplish this, they propose that tag markers, a space

for others to offer positive feedback to novices, may provide such encouragement. They also

suggest that experts' strong feelings of ownership of the community could be directed away

from harmful territorialism toward governance of the social tagging community. The idea that

experts' desire for high quality, accurate, and appropriate tags for museum collections should be

sublimated into governing the space is essentially undermining the potential role experts could

have in teaching novices about how they develop tags. Even the authors admit that novices may

need to learn more jargon to improve their tags -- why shouldn't the experts work cooperatively with novices to teach them about tagging?

Other researchers have examined tools that may sustain or improve the quality of Wikipedia content. One such tool, Huggle, is a popular desktop application that presents a list of before-and-after edits to review that have been identified as being suspiciousness (Geiger & Ribes, 2010). With a click, editors can easily reject an edit, send its author a canned warning, and mark the author as a potential vandal. Such tools are critical in Wikipedia's efforts to handle the ever-increasing vandalism on the site. While necessary, some scholars have noted the downsides of tools like Huggle:

> Situate users as police, not mentors, affording rejection and punishment. Newcomers who make low-quality edits are situated as potential vandals or spammers to be thwarted, instead of individuals who may need assistance and guidance in navigating Wikipedia's labyrinth of policies, rules, and norms. These highly automated tools have become the dominant way in which Wikipedians – established editors with well-defined social roles who make hundreds or even thousands of edits a month – interact with non-Wikipedians (Halfaker, Geiger, & Terveen, 2014, p. 311).

Halfaker, Geiger, and Terveen further opine that these tools are causing problems with newcomer socialization and retention issues with newcomers. While such counter-vandalism tools are critical to how Wikipedians maintain a massive decentralized and open sociotechnical system, efforts to socialize newcomers has been less successful. More traditional socialization practices, such as mentoring, have received less attention and without adequate socialization support, newcomers, often on their own, are expected to learn the myriad of policies,

procedures, and guidelines that govern Wikipedia production. In an attempt to address this
shortcoming, Halfaker, Geiger, and Terveen built a socialization tool, Snuggle, that includes a
set of features enabling more experienced editors to identify newcomers in need of help, share
assessments of newcomer activities with other gatekeepers, and intervene when "good-faith
newcomers" experience inhospitable treatment. It remains to be seen where Snuggle will
alleviate some of the problems Wikipedia has experienced with newcomers.

While Wikipedia remains a successful collaboratively produced information resource for millions
of web users, it faces many challenges. Some of those same challenges are also present on
family history websites, but fewer tools are available to FHRs to solve those problems. No
research exists, except for my limited 2014 study, on how to improve socialization of FHR
newcomers or support them in the production of accurate content on websites such as
Ancestry, Find A Grave, or Wikitree.

4.7 Social and Collaborative Learning

If participatory spaces for family history production have the potential to facilitate collaborative
learning interactions between contributors, what theories might help in building an
understanding of that potential? Social learning and collaborative learning theories are
illustrative of the potential for learning in groups or communities. Many scholars have asserted
that learning takes place in social contexts and is essentially a social process. Social learning
theory, introduced by Bandura in 1977, is premised on the idea that people can learn by
watching others, or observational learning (Bandura, 1977). Further, this theory is based on the
notion of reciprocal determinism — the idea that an individual's behavior, environment, and

personal traits all reciprocally influence each other. Bandura identified three basic models of observational learning:

- A live model — an individual demonstrates a behavior

- A verbal instructional model — an individual describes or explains a behavior

- A symbolic model — real or fictional characters displaying behaviors in various media, such as books, television programs, or digital media

The theories of Vygotsky (1931), Lave (1991), and Lave and Wegner (1991) also emphasize the central role of social learning in knowledge acquisition. According to Vygotsky's theory, learning takes place through the interactions individuals have with peers, teachers, and other experts. Vygotsky further asserted that individual learners have different developmental capabilities in collaborative situations than when they are learning alone. He labeled the difference between these developmental capabilities (alone and collaborative) as the "zone of proximal development." His theory suggests that to understand what occurs during collaborative learning, one should not focus on constructing mental models of what is going on in the heads of individuals; rather, one should develop an understanding of the shared meaning-making that occurs during collaborative interactions. This further suggests that environments that provide many opportunities for learners to interact with each other through discussion, collaboration, and feedback are an effective means of supporting learning. Vygotsky also suggested that culture is a primary determining factor in knowledge construction that individuals learn by interacting with others and following the rules, skills, and abilities shaped by a culture.

Running in a theoretical parallel, Lave and Wenger (1991) developed the idea that community is the agent of situated learning. Rather than looking at learning as the acquisition of certain forms

of knowledge, Lave and Wenger view learning as a process of social participation. In their model of situated learning, learners participate in communities of practice and, through their participation, master the knowledge and skills required to move from newcomers to full participants in the sociocultural practices of the community. Individuals are able to move from being newcomers to experienced practitioners through "legitimate peripheral participation" (LLP). According to LPP, newcomers initially become members of a community by participating in simple and low-risk tasks or peripheral activities. Through this low-risk participation, novices become aware of the tasks, vocabulary, and organizing principles of the community. Over time, those who are newer to community are given the opportunity to participate in tasks that are more central to the functioning of the community. To facilitate this process, it is helpful if newcomers can directly observe the practices of experts, so that they can develop an understanding of the broader context in which to understand their own contributions. If newcomers are separated from experts and the larger community, they have limited access to tools and knowledge needed to grow within the community. As participation levels increase, situations arise that allow participants to assess their own contributions.

Situated learning is a not simply "learning by doing." As Tennant (1997) has pointed out, Lave's and Wenger's concept of situatedness involves people being full participants in the world and in generating meaning within a particular context. Situated learning is based on two claims:

• Knowledge is and needs to be contextualized

• New knowledge and learning occurs in communities of practice

Lave and Wenger (1991) viewed learning as central to human identity, and learning as social participation – that is, as an active participant in the practices of social communities, one is

constructing his or her identity through these communities. People continuously create their

shared identity through engaging in and contributing to the practices of their communities. For a

community of practice to function, it needs to generate and appropriate a shared set of ideas

and commitments. It also needs to develop various resources such as tools, documents,

routines, vocabulary and symbols that carry the accumulated knowledge of the community.

Designing for virtual collaborative spaces for communities of practice often involves creating

artifacts, activities, and environments that enhance the practices of group meaning-making

(Stahl, Koschmann, & Suthers, 2006). Meanings are shaped by an individual's past experiences

(Piaget, 1976) and are open to ongoing negotiation and evaluation by members of a CoP. While

participants may not understand others' meanings fully, they still engage in coordinated activity

and function as if shared understandings were possible and being realized. This suggests that

the design of technologies to support collaborative learning within CoPs must be "coupled with

analysis of the meanings constructed within emergent practice" (Stahl et al., 2006, p. 415). Stahl

et al. further argue that to design technology to support collaborative learning, "we must

understand in more detail how small groups of learners construct shared meaning using various

artifacts and media" (p. 415).

CHAPTER 5: METHODS

I selected a qualitative, grounded theory-inspired (GT) approach to study family history research and production behaviors (Glaser & Straus, 1967; Straus & Corbin, 1998). A qualitative approach is particularly appropriate for research that seeks to answer how and why questions (Yin, 2015), and this study poses several questions that explore the "hows" of family history research and production. In addition, family history research and production are process-oriented activities, and a GT approach provides an effective means for generating understandings of various processes (Gasson & Waters, 2011). Furthermore, due to the fact that we know little about the online research and production processes in family history communities, a more exploratory approach is warranted and GT is well-suited for such inquiries. GT's focus on finding patterns in complex textual data allows the data to reveal its own structure, rather than the researcher deductively imposing structure upon the data (Glaser & Straus, 1967; Straus & Corbin, 1998).

Additionally, the GT approach dovetails with my belief that family history production spaces are socially constructed through human interactions and thoughts. Further, historical knowledge, like other forms of knowledge, is not merely transmitted, but received and interpreted in the context of community and social expectations (Bransford, Brown, & Cocking, 2000; Lave, 1991). A GT approach allows me to investigate FH research and production practices from a socio-constructivist perspective and to develop understandings of the social dimensions of online research and production environments.

5.1 Overall Study Design

Initially, I chose two qualitative research approaches, interviews and the think-aloud protocol

("think-alouds") that would allow me to develop a rich and multifaceted understanding of

virtual family history production environments. My pilot projects left many unanswered

questions regarding accuracy of content, research practices, and how individuals learn how to

conduct family history research in online environments. In my dissertation study, I decided to

further explore these areas of interest by interviewing FHRs more specifically about their

learning experiences and research practices in online family history production environments.

While I expected these interviews to yield rich data about FHRs' research practices and what it

takes to be accurate in family history production, I knew from the second pilot project that

interviewees found it difficult to remember particulars about how they searched for and found

relevant data and how they determined the relevancy of found information – research practices

that appeared to be important to accuracy. To address this potential shortcoming in the

interview data, I included think-alouds as an additional data-collection method for my

dissertation study. Think-aloud sessions allowed me to directly observe participants as they

searched for family history information on the web, and as they assessed the relevancy of found

information. The think-aloud approach will be explained in detail in the section, "Think-alouds".

In addition to the interviews and think-alouds, a survey was developed to provide an impartial

means to determine the experience level of the interview and think-aloud participants. After

using the survey to classify the initial interviewees, I became aware of the utility of the survey

beyond its classification capability: the survey was producing interesting data that were well

worth analyzing. I decided to expand the survey to include some of the questions posed in the

interviews and to distribute the survey to a larger swath of FHRs. In the end, the survey

contributed an important data set that could be triangulated with my findings from the

interviews and think-alouds. Methodological triangulation was thus achieved by including three

different data collection methods in the study – interviews, think-alouds, and a survey (Patton,

1990).
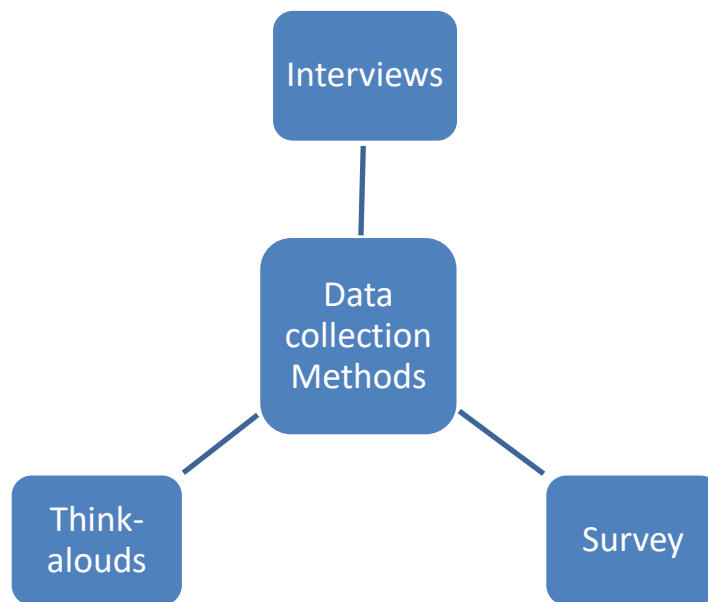
Figure 1: Data collection

The following sections will detail each research method used in this study.

5.2 Initial Interviews

The pilot projects provided a preliminary sketch of problems arising in online FHR communities

due to the proliferation of inaccurate information and what appeared to be an overabundance

of contributors who were engaging in poor research practices. To further my understanding of

contributors' research practices and this inaccuracy problem, I decided that for my dissertation

additional semi-structured interviews were needed. Another motivation for the interviews was

to delve more deeply into the kinds of obstacles FHRs face in finding relevant information about

their ancestors, which was not a main focus of the pilot interviews. Knowledge of the kinds of

common obstacles faced in the genealogical research process was then used to inform the

design of the think-aloud protocols.

My pilot study findings, as well as literature on archival intelligence and historical expertise,

informed the development of the interview questions. The pilot studies and relevant literature

(Darby & Clough, 2013b; W. M. Duff & Johnson, 2003; Garrett, 2010; Lucas, 2008; Willever-Farr

& Forte, 2014; Yakel, 2004a, 2004b, Yakel & Torres, 2003, 2007) revealed several factors that

may be influencing research and production behaviors of FHRs in online production

environments. These factors were then used to construct the interview questions:

1) Familiarity with archival descriptive systems, interactions with archivists/librarians, and

   hands-on experience with analog archival collections.

2) Collaboration with other FHRs.

3) Document literacy.

4) Skepticism of original sources and user-contributed content.

5) Knowledge of how online search tools/engines work.

6) Knowledge of contextual information, such as the regional history, history of immigrant

   groups, etc.

How and why these factors, as well as possibly additional yet unidentified factors, were influencing the accuracy of content on FHR websites was worthy of further study, as the pilot studies only provided preliminary and limited data.

The interview protocol included questions regarding how the participants began constructing family histories and their research experiences: how they learned the craft – informally or formally, face-to-face or online; the obstacles they faced both in conducting research online and in brick-and-mortar institutions; how they determined what information was relevant to their ancestors; their experiences with primary sources; and their ideas about how to facilitate learning on the FH production websites. The research questions for this study also shaped the interview questions. A visual mapping of interview questions and the study's main research questions can be found below (Table 3). The interview guide can be found in Appendix E.

Table 3: Visual mapping of interview questions and the study's research questions

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
Did you learn about how to research before you started searching for information about your ancestors or deceased individuals?		X	X		
What resources did you use to learn how to research?		X	X		
For experts: do you research any differently now that you have more experience with family history or research on deceased people?		X			

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
What sorts of websites do you use to find information about your ancestors or deceased individuals?	X	X			
Do you use websites that are for historical institutions or organizations? What were your experiences like with those websites?	X	X			
Do you use websites that managed by for-profit companies? What were your experiences like with those websites?	X	X			
Please tell me about your most recent experience with online searching for family history information or information about deceased individuals.	X	X	X		X

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
What did you need to know to be successful at this? Please describe any obstacles you may have faced.					
What do you do with the information that you are not sure is relevant to your family or the deceased person you are researching?	X	X	X		X
Can you remember one of your first experiences searching for family history information on the Web? Please describe this experience. What sorts of obstacles, if any, did you experience?	X	X	X		X

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
If you have asked other researchers to help you find information, please describe your experiences.	X	X	X	X	X
If you have asked helped other researchers find information, please describe your experiences.	X	X	X	X	X
Please tell me about the first time you contributed information, images, family trees, etc. to a family history websites, considering the following questions: *What kind of contribution did you make? *What inspired you to contribute?	X	X	X	X	X

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
*How did you learn how to create the family tree, memorial, etc.? *What was it like contributing? *Do you remember any problems you ran into? *How do you make sure the content you contributed is right or correct for the family or deceased individual? *Did you to converse with other contributors in the process of contributing? *Why did you do so?					
Please tell me about your most recent contribution, considering the following questions:	X	X	X	X	X

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
(same questions as listed in proceeding question)					
Please tell me about your most recent use of a brick-and-mortar institution or organization that has historic materials, such as an archive, historical society, etc.?		X	X		
Please tell me about your first use of a brick-and-mortar archive, historical society, etc.?		X	X		
What did you need to know to find records or other materials in these brick-and-mortar institutions/organizations?		X	X		

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
What sorts of difficulties, if any, did you encounter?		X	X		X
Please tell me about your most recent experiences using historical records, such as census records, military records, photographs, etc. Please describe any difficulties you encountered while interpreting these records.	X	X	X	X	X
Please tell me about the first time you used historical records, such as census records, military records, photographs, etc.? Please describe any difficulties you encountered while interpreting these records.	X	X	X	X	X

	How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	What are the online research practices of experienced FHRs vs. novice FHRs?	How might these practices influence the accuracy of the family history content they produce?	What new literacies are needed to produce accurate family history content in online environments?	How might family history websites be designed to better facilitate the production of more accurate content?
How important is it to look at the original records?		X	X	X	X
What do you think is the best way to learn how to conduct family history research?		X	X		X
How do you think websites for family history can support learning the craft of family history research?	X	X	X		X
How can family history websites better support you in the research process?	X	X	X		X

Due to my adherence to a grounded theory inspired approach, I wanted the interviews to be
organic – the questions in the interview guide were only starting points from which the

participants could explore their experiences with family history research and production without the interviewer placing undue constraints on those discussions. The guide includes consent language that was verbally stated before interviews commenced, and participants' consent was captured in the audio recordings of the interviews.

To recruit interviewees, a solicitation message (see Appendix A) was posted on several family history websites (Ancestry.com, Findagrave.com, Wikitree.com, and Geni.com). Participants were asked to fill out an online survey and were asked to include their email address or contact number if they were interested in being interviewed. I then emailed the potential participant to set up a date and time for the interviews that was convenient for the participant. Skype was used to conduct the interviews and a recording application for Skype (Pamela) was employed to record the phone calls. Interview recordings were transcribed and placed in the qualitative software, Atlas.ti, to support analysis of the transcriptions.

I continued to interview individuals for this portion of the study until "theoretical saturation" was reached or when additional data analysis revealed no new substantive concepts (Glaser & Straus, 1967). In the end, I interviewed 14 FHRs, most whom had considerable experience with family history research.

5.3 Survey

I wanted an impartial means of determining the experience level of interviewees and think-aloud participants, and, to achieve this, I designed a survey. Drawing from the literature on historical expertise, FHR information behaviors, and my pilot studies, I developed a survey that

measures family history expertise from several dimensions: length of time conducting family history research; breadth of information searching experience; breadth of family history production experience; breadth of experience with primary records; as well as participants' own assessments of their research experience. The survey initially contained 20 multiple choice and six short answer questions, and took participants approximately 15 minutes to complete (see Appendix C for the first version of the survey).

To use the survey as a measure of experience and knowledge of family history research, I developed criteria to segment the participants into three groups: beginners, intermediates, and experts. The following criteria were used to group interview and think-aloud participants.

To be considered beginners, participants must meet the following criteria:

1) they have fewer than two years' experience doing family history research;

2) they consider themselves to be at a beginner level in information-searching, selection of relevant information, interpretation of primary sources , and production of family histories;

3) they have created no more than one family tree;

4) they have used two or fewer family history websites to locate information about ancestors/deceased individuals;

5) they have not visited a brick-and-mortar archive/memory institution;

6) they have used two or fewer types of primary records in their research.

To be considered intermediates, participants must meet the following criteria:

1) they have three to five years' experience doing family history research;

2) they consider themselves to be at an intermediate level in information-searching, selection of relevant information, interpretation of primary sources, and production of family histories;

3) they have created two or more family trees (this ends up not being helpful – many beginners have created more than one family tree, probably based on some family tree or history a family member has given them, so they are basically copying it into an ancestry tree);

4) they have used three to five family history websites to locate information about ancestors;

5) they have visited one brick-and-mortar archive/memory institution;

6) they have used three to five types of primary records in their research.

To be considered "experts," participants must meet the following criteria:

1) they have at least six years' experience doing family history research;

2) they consider themselves to be advanced in information-searching, selection of relevant information, interpretation of primary sources, and production of family history/history of deceased individuals (even very experienced FHRs I have interviewed are hesitant to call themselves advanced or expert; intermediate seems a label they are more comfortable with);

3) they have created at least three family trees;

4) they have used six or more family history websites to locate information about ancestors/deceased individuals;

5) they have visited at least two brick-and-mortar archives/memory institutions;

6) they have used more than six types of primary records in their research.

I built the survey using the website Survey Monkey, which allowed me to develop a web-based survey. Participants were required to answer "yes" to a consent question at the beginning of the survey.

If they neglected to answer the consent question or answered "no," Survey Monkey would not accept their survey responses at the submittal stage. Survey Monkey also provides various survey analysis and aggregation tools.

The results of the survey for the initial interviewees revealed problems with the sorting criteria: For the most part, the sorting criteria did a good job at separating participants into beginner vs. experienced camps, but it did a poor job at determining more granular distinctions between the participants (such as dividing the participants into intermediates and expert groups). As a result, I created an "experienced" category that used the intermediate criteria thresholds. I discarded the "expert" criteria. I also found that many beginners had created more than one family tree. From what I could glean from the think-alouds and short answers, this was the result of beginners creating small family trees for different lines, such as their own family and their spouse's family. The ease with which one can throw together a family tree on Ancestry may contribute to this. Given that this criterion was not helpful in determining experience level, I omitted it.

While there were some snags with the criteria, the survey results from the initial interviews

produced interesting results that were worth exploring on a large scale. The interviewees'

survey responses revealed some common experiences that appeared to be of importance in

understanding the inaccuracy problem on family history websites. One interesting finding from

the survey was that many of the interviewees had jumped into researching and building family

histories without instruction from online or offline sources, whether human or inanimate. Was

this a widespread phenomenon? If so, could this way of learning be influencing the quality of

family history content being contributed by novices? In addition, some pertinent patterns were

percolating to the surface from the initial interviews. If these patterns were widespread, they

could shed light on the inaccuracy problem. Interview participants spoke of how their research

practices changed and how there was a corresponding change in the quality of the family

histories they produced. Were these research and production patterns more widespread than

the 14 people I interviewed? Are there other research and production trajectories I had not yet

uncovered? To explore these questions with a larger group of FHRs, I added an additional set of

questions, seven in total, to the survey. Appendix D contains the final version of the survey; new

additions are in italics. The new survey was submitted as an amendment to the original

protocol, and Drexel's IRB approved the amendment in November 2014.

The revised survey was posted to additional family history websites and several Facebook family

history groups, resulting in 560 survey results in the first three days of posting the survey. Based

on my advisor's experiences with surveying Wikipedians, I had not anticipated such a large

number of responses and did not think it necessary to implement a cutoff mechanism after a

predetermined number of responses was reached. I quickly closed the survey after realizing I

had exceeded the number of responses stated in my research protocol (500). I consulted my

advisor and submitted a HRP-214 form to report the problem, in which we proposed to delete

the last 60 responses. The Drexel IRB accepted this solution, and I proceeded to delete the last

60 responses from Survey Monkey. To avoid this problem in future research studies, I will

implement a cutoff mechanism for any survey that I employ.

Quantitative survey data were analyzed using Survey Monkey and Microsoft Excel. The survey

short answer questions generated qualitative data. To analyze this data, I first attempted to

group the respondents into beginner, intermediate and expert groups based on their survey

responses, but because respondents did not always answer each question for determining their

group placement, I ended up sorting the survey respondents into beginners and experienced

FHRs. The experienced group included a range of FHR expertise, from intermediate to expert,

based on the criteria outlined earlier in the methods chapter. After sorting the surveys into

beginner and experienced groups, I then compiled responses for six of the short answer

questions using Microsoft Word, and then ingested the compiled responses into individual

"documents" in ATLAS.ti for a total of 12 documents (six beginners and six experienced

documents). I then coded the documents using the existing codebook (from interviews and

think-alouds). I found that I needed to add additional codes to the codebook based on new

concepts that emerged from the short answers.

5.4 Think-Alouds

The Think-Aloud Approach

My pilot study revealed that FHRs often struggled to recall specifics about their online search and production experiences, and it seemed unlikely they were providing accurate descriptions of these processes. This struggle is likely the result of an inherent problem associated with humans recalling cognitive processes: the more time that elapses between the occurrence of cognitive processes and people's reports about these processes, the more likely the reports are to be distorted or fabricated (Ericsson & Simon, 1985; Nisbett & Wilson, 1977). To address this potential shortcoming in the interview data, I employed think-alouds as an additional data-collection tool for my dissertation study.

Think-alouds involve study participants thinking aloud as they perform a set of specified tasks. Participants are asked to talk about whatever they are looking at, thinking, doing, and feeling as they perform tasks, and the researchers objectively take notes on what is said. This method enables researchers to gain an understanding of the process that undergirds task completion. The think-aloud method has also been proven to be an effective means of arriving at an understanding of specialized learning, which family history production appears to involve (Ericsson & Simon, 1985).

Design of Think-Alouds

For this study, participants were asked to use the "think-aloud" method to explain their

cognitive process as they searched for information on the web and constructed online family

history content based on a provided research scenario. Thus, the think-aloud involved two

major tasks:

1) Searching for and locating information pertaining to specific deceased people on the

   web;

2) Constructing biographical content for a deceased person on a website.

Participants were confined to using online sources to carry out these tasks, as this dissertation

focuses on information-seeking and information production in web environments.

The think-aloud study also explored participants' experiences with content production on one

family history-oriented website. Selecting what type of content participants would be asked to

produce was an important design decision. Through my pilots, in which I explored several FHR

websites and their production tools, as well as findings from the initial interviews, I developed

an understanding of some of the problems inherent in family tree construction tools and the

difficulties many users experienced with these tools. Many interviewees who participated in the

pilot and the initial interviews for this study engaged in lengthy diatribes about the limitations

and problems with popular family tree production tools. Given that family tree tools can be

challenging to use for novices as well as more experienced FHRs, the limited time I would have

with each participant, and the potential for family tree production tools to become the main

and only focus of the think-alouds, I considered other forms of family history content production

for the think-alouds. Findings from the pilot projects indicated that Find A Grave memorials

were a form of family history content being produced by tens of thousands of contributors

(Willever-Farr & Forte, 2014). The survey provided further proof of the popularity of producing

Find A Grave memorials: 488 of the 500 respondents indicated that they had created memorials.

Find A Grave memorials possessed other advantages for the think-aloud study: they are a

simpler and more flexible form of family history content that allows participants to collect and

record basic information, as well as more rich data about a deceased person. Family trees often

only display basic data about a person, such as birth and death dates, and I didn't want the

think-aloud participants to be confined by this — I wanted them to search for any kind of

information, data, etc. about the family they were asked to research. Given that memorials can

contain narratives and multiple photographs, as well as basic life facts about a person, it seemed

to be a more flexible product for the participants to construct. An entry for a deceased person,

or memorial, may include a biographical sketch of the deceased, names of immediate relatives

(parents, spouse, and children), images of the grave marker, and sometimes images of the

deceased. Memorial content has obvious utility for family history research, so it is heavily mined

by many FHRs (Willever-Farr & Forte, 2014).

The design of the family history research scenario for the think-alouds was informed by the

initial interviews. To build the research scenario, commonly mentioned research challenges that

surfaced in the initial interviews were identified and then incorporated into the research

activity. Literature pertaining to archival literacy and historical expertise was re-consulted to

determine whether additional challenges might be present in FHR processes that were not

mentioned by the interview participants. Existing literature did not detail any additional

challenges that may be present in the historical research process. This suggested that the initial

interviews had provided adequate fodder to design a research scenario that would mimic many

aspects of the realities of family history research in the web environment.

The research challenges that were built into the research scenario include:

1) Lots of people share the same given and surnames, making it difficult to determine

   which data is relevant to the researcher's ancestor.

2) Variant spellings of names, which are commonplace in families over time, make it

   difficult for the researchers to determine if the found data is relevant to their ancestors.

3) Incomplete names (such as not knowing the ancestor's given, middle, or maiden

   names), also make it difficult to assess the relevancy of found data.

4) Researchers frequently could only locate very limited data on an individual, making it

   difficult to determine familial connections.

5) Data/information pertaining to an individual can be difficult to uncover when it's

   located in databases and webpages on many different sites, many of which are

   unknown to the researchers.

6) Many, if not most, archival records and other sorts of materials are not digitized or indexed. Such records can be scattered across repositories, sometimes geographically distant from one another. Researchers often lack knowledge of the analog collections housed at various repositories.

7) Researchers must contend with inaccurate user-contributed content and the copying and linking of that inaccurate information in online family history production spaces.

8) Researchers sometimes struggled to read and interpret primary sources.

Additionally, the strategies for overcoming these challenges as identified by initial interview participants were considered in the design of the research scenario. Many of the interviewees spoke about using multiple data points and more trustworthy data about more recently departed ancestors as a means to find relevant data on lesser-known ancestors. Interviewees who were experienced practitioners also asserted that accurate family history content can only be created if FHRs collect and examine original sources. Similarly, Yakel and Torres (2003) posit that to be successful at historical research, one must be skilled at interpreting documents and other sorts of primary sources.

Employing both the identified FHR challenges and some of the strategies employed by the initial interviewees to overcome those challenges, I built a research scenario using my own deceased relatives as the subject matter. The scenario had the following features:

1) **Multiple data points**: scenario included information, albeit incomplete information, on the main person to be researched as well as some of her family members. This gave the researchers the option of triangulating data points from different relatives if they knew to do so. It also supported "name collection" (Cole, 2000) if the researchers chose to engage in this activity.

2) **Names are common**: the main person to be researched had a common surname and given name, which meant there were **several** online records for deceased people with that same name. Researchers had to determine which of these records (if found) were relevant to person they were researching.

3) **Variant name spellings**: the main person to be researched was presented as having two possible spellings of her surname. There are several online records for both of those spellings.

4) **Incomplete names**: some of ancillary family members did not have complete names.

5) **Inaccurate or incorrect user-contributed content**: two of the ancillary family members have incorrect data listed about them on user-contributed family trees on Ancestry.com and Wikitree.com.

6) **Primary/original sources are important to the creation of accurate family content**: at least five different types of primary or original sources exist online for deceased individuals featured in the scenario.

Using the provided information on the main individual to be researched, Pauline Mai or May, and some of her relatives, participants were asked to locate information on Pauline and build a

memorial on Find A Grave for her. See Appendix for the research and production instructions

and the data provided to participants.

To give the participants a sense of the kind of information needed for a memorial, a printout of

an existing Find A Grave memorial was given to each participant and the Find A Grave URL was

provided on the printout. Given that this study focused on the information behaviors of FHRs in

the web environment, participants were confined to using web-accessible resources to carry out

the task. One subscriber-based website, Ancestry.com, was made available to participants – I

logged onto the site before each think-aloud session so that participants could access materials

(such as primary documents) that are behind a paywall. Although this eliminated the

subscription barrier, I felt this compromise to the ecological validity of tasks was less of a threat

to my research than eliminating Ancestry.com resources from the study, given its popularity as a

major source for genealogical data for millions of FHRs (Falconer, 2012b). By removing the

paywall barrier, I was able to observe how Ancestry materials were found, assessed, and used

by FHRs. Five participants chose to use their own Ancestry accounts.

To build a memorial on Find A Grave, one must create a username and password on the

website. I created a Find A Grave username and password for the study. Prior to each think-

aloud, I logged on to Find A Grave using the username and password through all the browsers

available on my laptop. This enabled the participant to perform the production tasks for the

protocol without spending study time navigating account creation. To ensure that participant-

created memorials for Pauline Mai were not accessible to other participants, after each think-

aloud session, I deleted the memorial created by participant. The participant-created memorials were captured using screen-capture and audio-video recording software.

My pilot studies, as well as literature on FHRs (Yakel, 2004b; Yakel & Torres, 2007), suggest that collaborating with other FHRs may play a key role in learning the craft of family history research and overcoming "brick walls" or obstacles that are encountered in the research process. Initially, I had hoped to pair an experienced FHR with a novice in the think-alouds: novices would be asked search and create the memorial on their own at first, and in the second half of the think-aloud an experienced FHR would assist them in the research and production process. Unfortunately, scheduling novice-expert teams proved impossible due to the conflicting schedules of participants.

Given that the scenario was designed to create a challenging research environment for participants, and because more experienced FHRs would not be paired with novices, I was prepared to provide minimal assistance to the novice to enable them to carry out the assigned tasks. Providing assistance to participants, often referred to as "coaching," is commonly employed in think-loud protocols (Dumas & Redish, 1999; Olmsted-Hawala, Murphy, Hawala, & Ashenfelter, 2010a, 2010b). If novice participants struggled for 20 minutes with a task, I offered assistance to allow them to continue. By noting points at which intervention was required, I hoped to illuminate which aspects of tasks might be better learned through collaboration with more experienced peers.

After the participants searched for information and created a memorial, I interviewed them about their perceptions of the experience. I also asked them questions about any previous family history research experiences they may have had. (See Appendix G for the exit interview guide).

Capturing the Think-Alouds

Both the think-aloud sessions and exit interviews were captured using Apowersoft's Screen Recorder Pro software, which supports simultaneous video, audio, and screen activity capture. For the think-aloud portion of the sessions, I transcribed the verbalizations of the participants and took notes on their screen activity and facial expressions captured in the video recording. I added the transcripts and notes to ATLAS.ti. For the interviews, I transcribed the audio portion of the recording, and made notes on the participants' facial expressions in the video recordings. Both the interview transcriptions and notes were added to ATLAS.ti.

Think-Alouds Recruitment and Number of Participants

Think-aloud participants were recruited through the survey, which was posted to several local genealogy Facebook groups and through snowballing – six participants suggested friends and family for the study, and four of those individuals participated in the study. As an incentive to participate, $20 was offered to each participant who took part in a think-aloud session. The think-aloud sessions were held at locations convenient to the participants. Those locations included libraries near the participants' home, the homes of participants, and the usability lab in the basement of Drexel University's Rush building. The length of the think-aloud sessions ranged from one hour to two and a half hours, with the median being an hour and half. A total of 14

individuals participated in the think-alouds: seven experienced FHRs and seven novices. I found

that by the sixth novice participant and fifth experienced participant that no new evidence of

import was being collected. At that point, it appeared that I had captured the range of recurring

patterns in information-seeking and production behaviors among participants with different

levels of experience. As I had already scheduled three more think-alouds at that point, I decided

to go forward with the sessions and to include data collected from those sessions in my final

data set. The last three think-alouds mirrored behaviors collected in earlier think-alouds, which

lent more validity to my findings.

5.5 Analysis

Textual Data

Having selected a grounded theory-inspired approach for my research study, I followed an

inductive approach to analyzing the collected textual data. Grounded theory method requires

the researcher to make theoretical sense of diversity in the collected data by developing "ideas

on a level of generality higher in conceptual abstraction than the qualitative material being

analyzed" (Glaser & Straus, 1967, p. 114). Accordingly, several iterations of coding were

performed to identify thematic patterns in the transcribed text, including interview and think-

aloud transcriptions, as well as survey short answer responses. Constant comparison of

indicators, concepts, and categories was employed throughout the coding process to increase

the internal validity of my findings (Glaser & Straus, 1967). I diligently looked for any divergent

behaviors and opinions, and made certain all patterns were identified. This process is

considered to be "open-coding" and culminates in the creation of a "codebook" (Corbin &

Strauss, 2008). I further refined the codebook, combining some of the more idiosyncratic or

overly granular codes and replacing them with more generic terms, as well as combining codes

that represented the same concept but were worded differently. I then established relationships

among codes (axial coding) (Corbin & Strauss, 2008) and created visual groupings of related

codes. This was also an iterative process: codes were organized and reorganized into groupings

until meaningful explanations could be built from the groupings.

ATLAS.ti provided support in the coding process by enabling me to:

- Ingest transcripts/notes from each interview/think-aloud into separate "documents."
  (Responses were combined into one document for each short answer question on the
  survey).

- Mark passages of text with initial comments describing the text.

- Review these comments across "documents," which aided me in creating the first set of
  codes.

- Mark passages of text with the first set of codes.

- View initial codes and occurrence of those codes across documents and within
  documents. This tool supported my efforts to refine the code book, which included
  consolidating codes that represented duplicate or very similar behaviors, opinions, etc.,
  and redrafting codes that were overly granular into more generalizable codes.

- Efficiently mark documents with the final set of codes.

- Group codes into families and then visualize those code families.

The codebook consisted of 134 codes, which were grouped into 26 code families. The following

is an example of the code families I generated using ATLAS.ti (Figure 3).
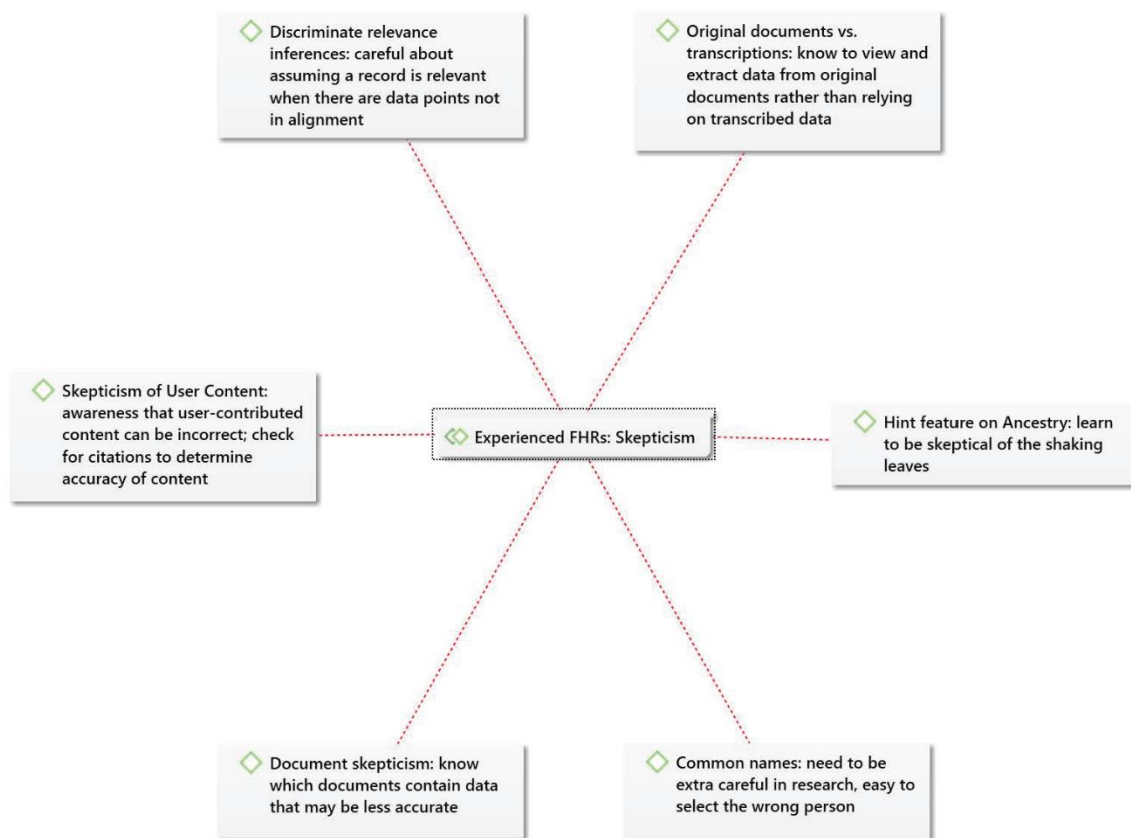
Figure 3: Example of a Code Family Generated with ATLAS.ti

Quantitative Data

Survey Monkey provides aggregation tools and visualization tools for quantifiable survey

responses. I used these tools as well as Microsoft Excel to create visuals of the collected

quantitative data, which aided me in comparing textual and quantitative data. The quantitative

data were largely convergent with my findings from the textual analysis, which gave my findings

further validity.

CHAPTER 6: FINDINGS

In this chapter, salient findings from the study will be presented. First, to give the reader a sense

of the makeup of the study's participants, the demographics and family history experience level

of the survey respondents, initial interviewees, and think-aloud participants will be discussed.

Next, the dimensions of the problem space will be explored, with an eye to understanding how

the design of existing family history websites and other online factors may be contributing to

the proliferation of inaccurate family history content. This will be followed by an examination of

the online information-searching behaviors of novice and experienced FHRs. This examination

gives the reader a glimpse of the difficulties FHRs face in finding relevant information about

ancestors in the online environment and begins to point to the many literacies needed to be

successful at family history research. Next, the literacies that undergird the successful searching

activities of experienced FHRs will be explored. The following table maps the study's directive

questions to their relevant section(s) within the Findings Chapter.

Table 4: Research Questions and Findings Sections

Research Questions	Findings Section
How are the designs of family history websites, including their search tools, production tools, and editorial control systems, impacting the accuracy of user-contributed content?	Characteristics of the Problem Space/ Setting
What are the online research practices of experienced FHRs vs. novice FHRs?	Searching Behaviors and Literacies Other Family History Research and Production Literacies
How might these practices influence the accuracy of the family history content they produce?	Searching Behaviors and Literacies Other Family History Research and Production Literacies
What new literacies are needed to be accurate in family history production in online environments?	Searching Behaviors and Literacies Other Family History Research and Production Literacies

6.1 Participant Demographics

Demographics of Survey Respondents

More women than men responded to the survey: 308 women participated in the survey

compared with 192 men. Most participants were Caucasian (469), with only a small number of

respondents identifying as Black, Hispanic/Latino, Native American or Asian (Figure 4). The

overwhelming majority (90%) of respondents were experienced FHRs; only 10% were beginners.

Survey Respondent Demographics

	Caucasian	Black	Hispanic or Latino	Native American	Asian
■ Female	287	2	5	13	1
■ Male	182	2	1	7	
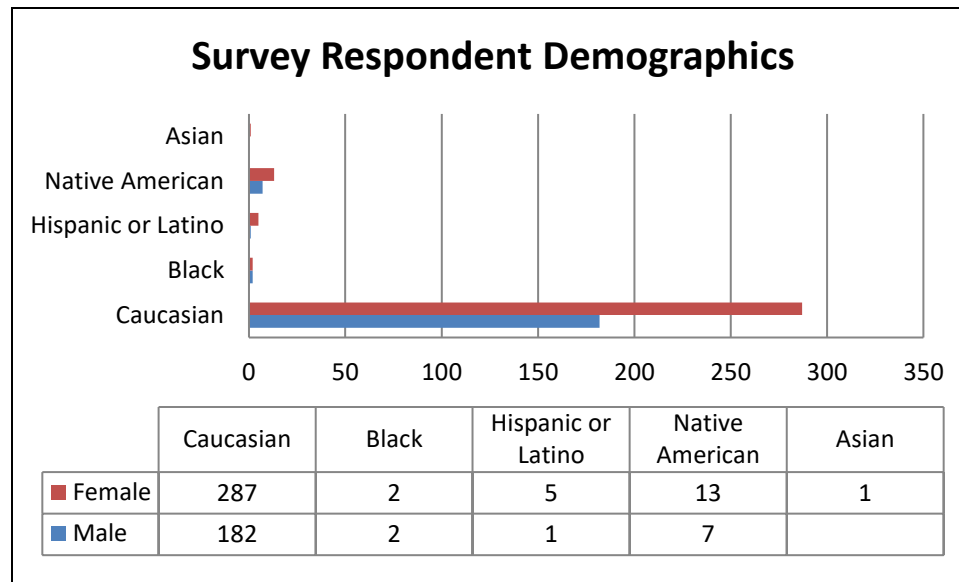
Figure 4: Race and Gender of Survey Respondents

Respondents' current occupations or occupations before retirement varied greatly, with 217

different types of occupations represented. Figure 5 indicates the 10 most common answers to

the occupation question. The terms used to describe the occupations were drawn from the

survey responses. For example, "homemaker" and "stay-at-home mom" were the terms used by

survey respondents.
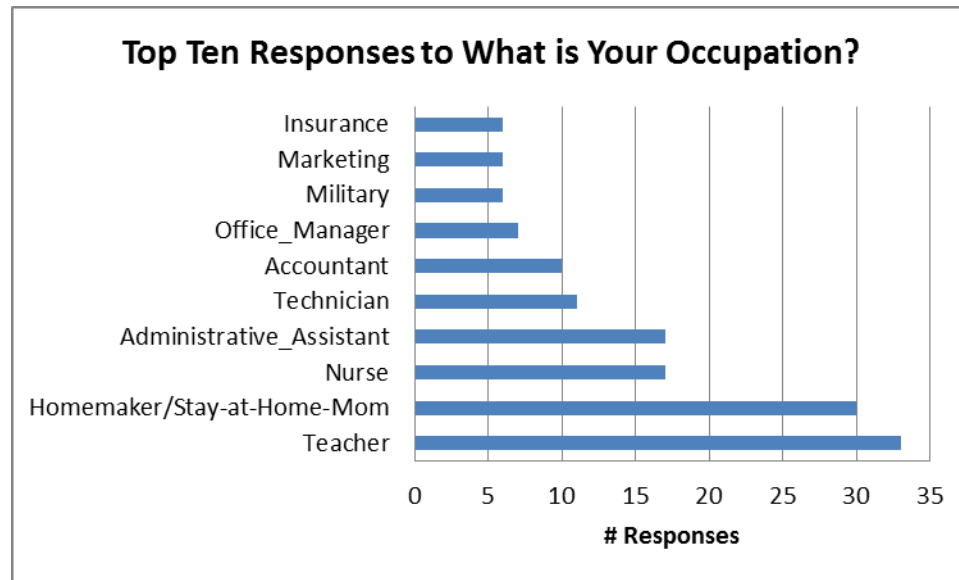
Figure 5: Top Ten Responses to Occupation Question

Other occupations that were listed less frequently (6<) include (but are not limited to): sales associate, information technologist, minister, factory worker, general contractor, archeologist, baker, waitress, restaurant manager, police officer, usher, truck driver, retail worker, reporter, graphic designer, photographer, conservator, physician, university administrator, office manager, and librarian. A word cloud (Figure 6) provides a visual representation of all 217 different occupation types listed in the survey responses.

Figure 6: Word Cloud of Survey Respondents' Occupations

While there appear to be a few core demographic groups that FHR hobbyists likely hail from, the

range of careers indicated in the survey suggests that there is a "long tail" of FHRs that possess

diverse educational backgrounds, professions, and life experiences. Given this occupational and

educational heterogeneity, there may a range of research and analytical skills that individuals

are bringing to bear in their family history research. There was also a wide range of ages

represented in the survey results. Survey respondents' ages ranged from 25 to 83, with the

majority of participants in their 50s and 60s (Figure 7). These findings suggest that FHR

production sites are attracting a demographically different group of contributors than many

other types of user-contributed websites have attracted. The majority of contributors to

websites, such as Wikipedia, code-sharing sites such as Stackflow, and cartographical sites such

as OpenStreetMap, are males in their younger adult years (Eckert & Steiner, 2013; Hill & Shaw,

2013; Lin & Serebrenik, 2016; Stephens, 2013; Vasilescu et al., 2015; Vasilescu, Capiluppi, &

Serebrenik, 2012). Women and older adults are not well represented. This appears to not be

true for family history production sites. The demographic differences of contributors to family

history production sites makes it a novel problem space to explore.
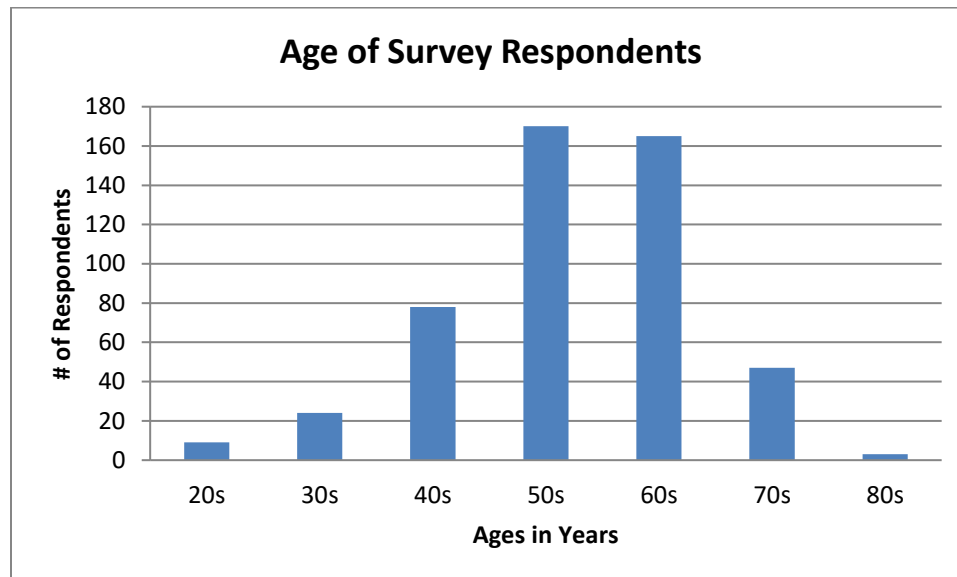
Figure 7: Ages of Survey Respondents

Demographics: Initial Interviews

A total of 9 women and 5 men participated in the initial interviews. Their ages ranged from 25 to 63, with the majority of participants (8 participants) in their 40s and 50s. There was little racial diversity among the participants as 13 participants were white and 1 was black. However, there was multiplicity in their occupations and in their places of residence. Most participants hailed from places throughout the U.S., but one participant was a Swiss citizen residing in Switzerland and the other an American who was working in Brazil at the time of the study. As with the survey, more experienced FHRs (10) participated in the interviews than novices (4). The table displays demographic information for all of the initial interview participants.

Table 5: Demographics of Initial Interview Participants

Gender	Race	Age	Occupation	Residence	Experience-level
Female	White	52	Behavioral researcher for Disney	CA	Experienced
Female	White	44	College faculty - digital humanities	Michigan	Experienced
Male	White	25	Master's student- library science	District of Columbia	Experienced
Female	White	63	Human resources	Massachusetts	Experienced
Female	White	50	Caregiver for her father	Washington	Experienced
Male	White	46	Computer scientist	Illinois/Brazil	Experienced
Female	White	25	Graduate student – sociology	Mississippi	Novice
Female	White	46	Federal State Department administrator	Virginia	Experienced
Female	White	55	Secretary	Florida	Experienced
Male	Black	60	Retired/Navy enlisted	Tennessee	Experienced
Female	White	53	Owns insurance business	Pennsylvania	Novice
Male	White	43	Adjunct Instructor - public history	Pennsylvania	Experienced
Female	White	35	Web designer	Florida	Novice
Male	White	53	Engineer	Switzerland	Novice

Demographics: Think-Aloud Participants

A total of 11 women and 3 men participated in the think-aloud sessions. The participants' ages

ranged widely from 35 to 82 years of age. There was little racial diversity among the

participants: 12 participants were white and 2 were Native American. All of the participants

resided in Pennsylvania, due to the requirement that participants attend think-aloud sessions in

person. There was diversity in participants' occupations, including artist, university

administrator, administrative assistant, school teacher, and gunsmith shop manager. An even

number of novices and experienced participated in the think-aloud sessions. Table 6 details

demographic information pertaining to the think-aloud participants.

Table 6: Demographics of Think-Aloud Participants

Gender	Race	Age	Occupation	Residence	Experience Level
Female	White	73	Retired, project manager for healthcare system	PA	Novice
Female	White	47	Manager, gunsmith business & Air Force Reserves (enlisted)	PA	Experienced
Female	White/Native American	63	Retired, physical therapist	PA	Novice
Female	White	72	Retired, phone company (worked for phone company for 30 years, held various positions)	PA	Novice
Female	White	51	Office manager	PA	Experienced
Female	White	69	Artist	PA	Novice
Female	White	72	Retired, administrative assistant	PA	Experienced
Male	White	77	Retired, insurance agent	PA	Experienced
Female	White	82	Retired, school teacher	PA	Novice
Female	White	46	Account manager, advertising agency	PA	Experienced
Male	White	74	Academic administrator & college professor	PA	Experienced
Female	White	75	Retired, college instructor (writing)	PA	Novice
Female	White	47	University administrator	PA	Novice
Male	White/Native American	35	University database administrator	PA	Experienced

Demographic Limitations of Collected Data

Without reliable statistics about the age, gender and race of FHRs in the U.S., it is difficult to determine whether this study's participants are representative of the larger FHR population. Given that people of color (in this case, Black, Hispanic/Latino, and Asian peoples) make up a much higher percentage of the population than are represented in this study (U.S. Census 2015, https://www.census.gov/quickfacts/), and even accounting for the possibility that fewer people of color are interested in the hobby, it seems likely that they underrepresented in this study. Native Americans, however, may be overrepresented in this study: this group makes up 4% of participants, but only account for 1.2% of the U.S. population (U.S. Census Bureau, 2015). It is possible that Native Americans make up a higher percentage of FHRs, and thus, the 4% is representative.

While the demographics of the study participants may not adequately represent the racial makeup of the current FHR population in the U.S., the high numbers of women and women over 45 who participated in this study may reflect a trend in family history research in the U.S. While publicly accessible data about FHRs is limited, some sources suggest that users of genealogy websites tend to be female aged 45 or older ("Ancestry.com Traffic, Demographics and Competitors - Alexa," 2016; Falconer, 2012a; "Online Family History Trends Report," 2011, "The Ancestry Insider," 2016).

Another potential limitation of this study is the small number of novices (59) who participated. Given the small number of novices represented in this study, it is possible that additional information behaviors and patterns are present in this group of FHRs.

6.2 Characteristics of the Problem Space/Setting

To anchor my findings within the problem space as it is experienced and interpreted by my participants, I draw data from all three research instruments and my pilot studies to establish the features and the challenges of family history research in the web environment. Where possible, I link these features and challenges to the relevant literature.

Data collected from all three sources -- the initial interviews, the survey, and the think-alouds -- point to five key dimensions of online family history research and production environment that may be contributing to the inaccuracy problem with user-contributed content:

1) Information about deceased people abounds on the web.

2) There is a plethora of inaccurate user-contributed content on FHR sites.

3) Popular FHR websites encourage the sharing of user-created content regardless of its accuracy.

4) The design of online family history search and production tools has blurred the lines between the research process and the creation of public-facing family history content.

5) The design of online family history production tools is limiting the ability of FHRs to collaborate on content production, and in turn, places the onus for accurate content production on the individual, rather than supporting collaboration in content production between more-experienced and less-experienced FHRs.

Information About Deceased People Abounds on the Web

One of the prominent features of the problem space is the easy access to troves of information about deceased people. There is an increasing amount of historic data and digitized original source material on the web, which has made family history research a more accessible activity

for those who do not have the funds, the time, or the inclination to travel to brick-and-mortar

repositories to conduct such research. While the explosion of online information about

deceased people makes it easier to access such information from your laptop or phone, it also

presents challenges. Producing accurate family history even in online environments is not a

simple process. As one participant opined, "Even with the internet, you need to realize that

[FHR] takes lots of time, effort, money, determination, and planning to uncover accurate

information about one's ancestors" (survey participant #377).

In the web environment, FHRs are confronted with a wealth of online information for any given

name, which adds to the challenge of finding the right information for an ancestor. Determining

which one of those people bearing the same name is your ancestor can be difficult if there are

many to pick from. That, in turn, makes it easy to go down the wrong paths. This study and my

two pilot studies suggest that the main cause of such diversions is the fact that many people

share the same names, many of them living in similar time periods and in similar geographical

areas. Figuring out which one of those people is actually your ancestor is challenging. As one

participant exclaimed, "It's easy to go down the rabbit hole of not actually related information

on the Internet" (initial interview participant #1). Many study participants, including the

interview, think-aloud, and survey participants, attested to the challenging nature of conducting

family history on the web. As one stated:

> The wealth of information available makes it difficult to sort. It's a double-edged
>
> sword – you can google someone's name and get thousands of results in seconds and
>
> then a lot of that stuff is junk, and then you have to really clean the search terms, and
>
> then sort stuff. It can be overwhelming. (initial interview participant, #7)

Another participant asserted that, "You need to know how to discern whether an individual is who he appears to be - -far more difficult than it sounds because of the amount of information out there!" (survey respondent, #10). One interviewee further remarked:

> You know a lot of people watch those shows and say `look who they found' but those people are professional genealogists doing that research. It's not just as easy as going on Ancestry. And if you have a really common name, it's difficult. There are some common names on my mom's side and you have to be really careful – is that really my ancestor? Are they from the right town? Are they from the right county? (initial interview participant, #9).

The abundance of information on the web about deceased people creates a challenging research environment for FHRs.

Abundance of User-Contributed Content

Family History Content

A sizable portion of the online information about deceased people is contributed by FHRs in the form of family history content. This study found that the two most popular types of content FHRs are producing are family trees and memorials. Family trees are focused on documenting a group of related people, with space to record life event data, relationship data, and images for each person on the tree; memorials are more focused on documenting one individual, with space for recording life event data, short biographies, images and basic relationship data (i.e. mother, father, spouse, and children of deceased individual). This study uncovered data primarily related to family trees and memorials, which appear to be the most commonly shared content on the web, according to the survey data: 497 survey respondents have created family trees, 488 have created memorials, and 71% of those who created family trees and memorials have shared them online.

On the one hand, it is encouraging that access to family history research is expanding — anyone

with a computer or smartphone and internet access can tinker with family history research and

share their findings online in the form of family trees or grave memorials. Yet the ease with

which one can build and share family content online is a double-edge sword. Web 2.0

technologies enable FHRs of any skill level to more easily create and share content with large

audiences. When accurate, user-provided content shared online can be of great help to other

FHRs, particularly if the content was the result of contributors delving into analog materials held

by various repositories that may not be easily accessed by others (Willever-Farr & Forte, 2014;

Willever-Farr et al., 2012). However, there is a growing corpus of inaccurate user-contributed

family history content on the web (Willever-Farr & Forte, 2014; Willever-Farr et al., 2012). The

ease with which one can copy and paste content means that inaccurate user-contributed

content can proliferate. As one survey respondent said:

> Honestly, I did not realize the impact of having a public tree with an error in it being
> added to databases and copied by others until I had more experience with the process.
> Now, I feel it is on the individual researchers to be careful about where their
> information is coming from and trying to be accurate. (survey respondent #499)

Another participant specifically complained about the inaccurate user-contributed trees on

Ancestry:

> Not to say they are all wrong, but there has been a huge number of instances when
>
> they were. I have no idea whether those trees are accurate. So I stopped using that
>
> feature on Ancestry (the hints that led to user-contributed family trees). There is no
>
> evidence to prove what they are saying. (initial interview participant #9)

Yet another participant spoke of this problem with great zeal and with specific concern about

content produced by novices:

What I see, is that many people doing genealogy who don't know what they are doing, who I run into on a daily basis, are the people who tell you I got my tree back to Adam (Adam and Eve) last weekend. They went to Ancestry or some other tree site and just copied a tree that may be or may not be, is probably incorrect. (think-aloud participant #14)

Several participants noted that as novices, they were particularly "copy happy" or were less

discriminate in selecting and then copying and pasting data into their own family trees or Find A

Grave memorials. One participant asserted that novices:

They are not doing real research, they are just copying information and they have no idea how to do the research. (think-aloud participant #14)

The problem of widespread sharing of inaccurate user-contributed content is exacerbated

by the fact that family trees and memorials are often linked to other user-contributed

content on the web, resulting in a web of linked family history content (Willever-Farr &

Forte, 2014). This means that inaccurate user-contributed content can impact not only an

individual piece of content but also other content it is linked to. One participant attested to

this problem:

In a way, Ancestry, they almost promote the taking of information since you can link into other trees, which is kinda like taking others' work, which is why the same untruths are perpetuated. Researchers, if they see five people who link back to the same ancestor, they say ok this is five different people, well this must be true. The problem is all five people linked into the same untruth. (initial interview participant #9)

The abundance of often inaccurate family history content on the web presents a significant challenge to FHRs as they attempt to construct family trees, memorials and other types of family histories.

Unfortunately, it appears that Ancestry, through its marketing campaigns, is encouraging the production of family history content, regardless of the quality of that content. Ancestry benefits from encouraging users to produce more content for its website, as the more information content it has on the site, the more attractive it is as an information resource for FHRs. But the accuracy of user-content seems to be less of a concern to Ancestry then landing subscribers. To appeal to potential subscribers, Ancestry's marketing materials emphasize that its website makes FHR easy for anyone. Its commercials and ads suggest that one can build a family tree with as little as a family name and no knowledge of the family history craft. Ancestry commercials feature taglines such as "you don't have to know what you are looking for, you just have to start looking" (https://www.ispot.tv/ad/7drx/ancestry-com-born-a-slave-died-a-business-man); and "all it takes is a name" (https://vimeo.com/47916396). In one commercial the narrator states, "I typed my name and Ancestry opened the door to my past. Before my eyes my family's history was revealed" (https://www.ispot.tv/ad/7pVw/ancestry-com-guide-throughout-the-past). In another, the narrator exclaims, "In a few days I went from knowing almost nothing to wholly crow I am related to George Washington" (https://www.ispot.tv/ad/AZbe/ancestry-com-emily).

These commercials are suggesting to the public that producing family histories does not require any particular knowledge or skills, but simply a name or two. Given that these commercials are on many TV channels and also appear on websites, it's quite possible that the commercials are influencing FHR expectations and behaviors on many FHR and memorial sites.

Transcriptions of Historic Sources

In addition to family history content, FHRs are also contributing transcriptions of original,

historic sources. Such transcriptions are critical for FHR. On all family history websites that host

search tools, transcribed data from historic records (e.g. government records, court documents,

photos, newspaper obituaries, etc.), along with data extracted from user-contributed family

histories, are employed for search. Family history websites use transcribed data from historic

sources to create searchable indexes that search tools mine when users conduct searches on

the sites. This extracted data is also presented to the user in search results and is often

presented as THE historic "record," with or without a link to an image of the original source. This

would not be problematic except for the fact that transcriptions of historic records, often

crowd-sourced through online volunteer projects, can be riddled with errors. Time and time

again study participants complained about incorrect transcriptions. As one survey respondent

stated, "There is always the possibility of transcription errors for the records you come across"

(survey respondent #173). One of the reasons for the inaccuracies found in transcriptions is

handwriting: many original documents are handwritten and volunteer transcribers can easily

misinterpret handwritten words. The challenge of disambiguating handwriting is so great that a

few family history websites have an arbitration process in place to manage the transcription of

challenging handwriting (Hansen et al., 2013).

FHRs need to know to be wary of transcriptions, but this may not be obvious to newcomers.

Novices who participated in the think-alouds visited many sites which did not provide access to

images of the original sources. Even on sites that did provide access to imaged records, the

transcribed data was prominently displayed at the expense of the images. Participants needed

to know to click on a link to view the images, and these links were overlooked by most of the

novices, as the links were not prominently placed.

The Design of Family History Websites

Combined Search and Production Tools

As discussed in the introduction, Ancestry is the dominant player in the online family history

market. Millions of people use Ancestry's website to create public-facing content for other

Ancestry subscribers. The design of Ancestry's search and production tools, however, appears

to be contributing to the proliferation of inaccurate user-contributed content. Its tools blur the

lines between research and a shareable product, which may result in users sharing content in

the early stages of research, before sufficient evidence has been collected. In other words, this

tool does not make it easy for users to distinguish between family history research, an often

messy process, and the production of public-facing, user-created content. At the time of data

collection (2014-2015), the first "search" users encountered in the body of Ancestry's homepage

was not only a search tool but a tool that combines search and family tree production into one

seamless function. As users search and fill their family trees with found data via this tool, they

may be unwittingly sharing that data with others. This is because the access or view default for

family trees produced via the Ancestry website is "public;" as soon as a user saves a tree, it's

viewable by Ancestry subscribers. Ancestry does not prominently display the private/public view

option to users. Users need to review their account settings to discover that they can make their

trees private. The opaque nature of the privacy option for trees was born out in the interviews

and think-alouds. Many of the FHRs with extensive FHR experience who participated in the

study were uncertain about the access status of their Ancestry trees. When asked if their

Ancestry trees were public, nine study participants indicated they were not certain, although

some thought they were likely public because other FHRs had contacted them about

information they had included on their trees. Other websites, such as Geni.com and

Wikitree.com, also make "public" the default view for user-contributed content. Yet their

private/public view options are slightly more prominent in the main production area of the sites.

Regardless, by making the default "public" for user-contributed content, family history websites

may be encouraging content to be "published," or accessible to others, before it's ready to be

mined by other FHRs. The public default is problematic as there are few checks or editorial

processes in place to catch inaccurate content before it's made accessible.

Closed-Editing Production Systems

Unlike Wikipedia, which is based on an open-editing model, most family history production

websites do not support open editing of user-contributed biographical content. Contributors

retain control over the content they created and shared. As a result, there are no checks and

balances that a crowd of editors armed with content and editing guidelines can provide to

ensure quality and correct family history content, as is the case with Wikipedia (Forte & Lampe,

2013). Studies of Wikipedia have shown how open-editing systems that support multiple

contributors can lead to the production of accurate content (Anthony et al., 2009; Arazy et al.,

2006; Kittur & Kraut, 2008; Lih, 2004; Liu & Ram, 2009; Stvilia et al., 2005; Wilkinson &

Huberman, 2007) that often rivals more traditional encyclopedic content (Giles, 2005). In

contrast, the closed-editing systems present on most family history production sites place the

onus for accurate content production on the individual, as these systems inadequately support

larger community involvement in content production (Willever-Farr & Forte, 2014). There is no

easy way for FHRs to obtain help with constructing their trees or memorials, besides posting to a

message board. However, message boards were found to provide inadequate support for

learning the family history craft (Willever-Farr et al., 2012)

Automated Suggestions/Hints

Another design feature of the Ancestry site also contributes to the proliferation of inaccurate

content. Ancestry has incorporated a hint feature in its search/production tool: once a

contributor creates a portion of a family tree, the website will generate "hints" that appear as

shaking leaves on a family tree. These hints are automatically generated and may lead to

relevant transcribed data from original sources and/or user-contributed family trees on the site.

With a simple click, a user can populate their trees with this suggested data. According to the

many of the experienced FHRs who participated in the interviews and think-alouds, these

automatically generated "suggestions," or leaves, are notoriously wrong. It appears that

Ancestry's algorithms that identify potential kinship links often do a poor job at triangulating

data from different documents and data points, leading to many false positives. As one

participant exclaimed:

> There is really no reason for me to presume that the little green leaves are of actual
> value. Lots of the tempting data is wrong -- wrong part of the country, wrong time
> frame. I already know some information on my relatives and it doesn't match what the
> leaves are telling me. (initial interview participant, #1)

Experienced FHRs suggested that novices were not skeptical enough of these suggestions and

often populated their trees with incorrect data offered up by Ancestry's search engines.

Automation appears to be contributing to the proliferation of inaccurate family trees.

Lack of Assessment Tools

Compounding the problem of inaccurate content being made public is the fact that there are

few online tools to help users assess the accuracy of the shared user-contributed content on

family history production websites examined for this study. For example, there is no easy way

for FHRs to know how green the FHR is who contributed the content. Moreover, there exists no

clear reputation systems that could help users of family history websites to understand the

expertise level of those whose contributions they wish to use or link to.

The features of the online family history research environment make the hobby more

accessible, but also present many challenges. Family history researchers are faced with a

plethora of online information about the deceased, a sea of inaccurate, often linked, user-

contributed content, and production systems that are helping to contribute to the proliferation

of inaccurate content.

6.3 Coping with the Online Family History Research and Production Environment:

Literacies and Strategies

This study's findings strongly suggest that family history research and production is a serious

leisure activity as defined by Stebbins (Stebbins, 2009), in that FHR requires the acquisition of an

array of specialized skills and knowledge. This was born out in the study participants'

descriptions of their trajectories from novice to experienced FHR. Findings pertaining to those

trajectories, obtained via all three of this study's collection instruments, convincingly establish

that many novices do not have the skills and knowledge to produce good-quality, accurate

family history content. It is only through experience, learning from more experienced FHRs, and

other instructional sources that FHRs learn how to produce accurate content. A whopping 494

of the 500 survey respondents indicated that they changed their research and production

practices for the better as they became more experienced. All of the experienced FHRs, who

participated in the interviews and think-alouds, also said that they changed their research and

production practices as they became more experienced.

Many participants recalled that as novices they made lots of errors. Said one participant: "I

joined Ancestry and made lots of mistakes. Later I learned a lot of good

information/skills/strategies by joining about a dozen Facebook groups and asking questions,

reading posts" (survey participant #212). Another participant stated: "I made a lot of mistakes at

first!" (survey respondent #291). It was clear from participants' descriptions of their novice

research practices that they had produced inaccurate content. One participant wrote:

> The most significant change has been to slow down and more carefully consider
>
> evidence before recording it, and then complete sourcing it. (survey participant #491).

Another participant suggested a similar change in practice:

> I became more thorough and concentrated more on evidence. At the outset a similar
> name in the same county was enough for me to append them to my tree. After a time,
> I rewrote my tree, making sure that only people that truly belonged there were on
> there. (survey participant #473).

One participant exclaimed: "I think the difference would be I would take the stuff that did not

match up well, but now, I would be less likely to include it on my tree" (initial interview

participant #11). These and similar findings obtained through all three collection instruments

suggest that FHRs must gain specific knowledge and skills to create accurate family history

content.

The specific knowledge or literacies that appear to be critical to building accurate family history content include:

- Online search literacies

- Data and document[1] collection and management literacy

- Interpretive document literacy

- Transcription literacy

- User-contributed content literacy

- Production and publishing literacies

- Data triangulation/alignment literacy

Each of these literacies will be examined, starting with an exploration of the online information-seeking behaviors of experienced FHRs compared to those of novices. This examination will give the reader a concrete sense of the difficulty of searching for information about deceased people in the web environment, and how novices and experienced FHRs tackle this task using starkly different approaches. The varied and adaptive search strategies exhibited in the think-alouds, as well as those described by interview and survey participants, indicate that experienced FHRs possess a set of underlying literacies that support them in search activities that go beyond knowledge of online search tools and strategies. Those literacies — data/document collection, interpretive document, user-contributed content, production and publishing, and data triangulation literacies — enable experienced FHRs to assess the *accuracy* and *relevance* of found data and documents, two important aspects of building quality family history content.

---

[1] For the purposes of this study, "documents" are any original source that contains biographical information or historical information that helps document or contextualize the lives of ancestors. Examples of documents that are employed by FHRs include (but are not limited to): government records, certificates, licenses, court documents, maps, photographs, diaries, and military records.

Drawing from findings related to literacies, a model of FHR's information-seeking and

production behaviors was developed and will be presented. The model highlights the integrated

nature of FHRs' information-seeking and production behaviors in the web environment, and the

integral role that assessments of **data alignment** (does the newly found data align with what is

known?) and the **accuracy** of found data (is the data from trustworthy sources?) play in the

family history research and production process.

Searching Behaviors and Literacies

To begin, an exploration of the searching behaviors that were exhibited by novices and

experienced FHRs in the think-aloud sessions will be presented. The think-aloud findings provide

a concrete glimpse into the many challenges faced by FHRs as they conduct FHR in an online

environment. These findings provide a context for later discussions of the other literacies that

are needed to build accurate family history content in online environments. While the think-

aloud sessions primarily inform the findings presented in this section, interviews and the survey

provide supporting evidence. During the think-alouds, experienced and novice FHRs

demonstrated divergent search behaviors, including: disparate approaches to using online

search tools, different levels of adaptability in search, and differing tactics to narrow search

results. These divergent behaviors, as well as the online search environment, will be discussed in

the following paragraphs. Experienced FHRs search strategies and knowledge of complex online

search tools allowed them to find more relevant data and documents than their novice

counterparts.

Online Search Environment

Literature pertaining to online information-searching behaviors focuses on the major search

engines such as Google or Excite (Hölscher & Strube, 2000; Jansen & Spink, 2006). Through the

interviews, think-alouds, as well as my own experiences searching for genealogical data on nine

family history websites most commonly mentioned in the survey responses (ancestry.com,

archives.gov, familysearch.com, findagrave.com, findmypast.com, fold3.com,

genealogybank.com, geneaology.com, rootsweb.com), I discovered that FHRs are not only

confronted with all-purpose search engines such as Google, but also a myriad of family history-

focused search engines provided on various family history websites. These search tools have

common elements, but also differences. In particular, the search tools possess different search

forms and functionalities. There are two main types of search tools on family history websites:

ones that search across such websites and ones that allow users to search for data or

documents on one website. The cross-website search tools tend to be simple searches (single

search boxes, like Google) and the underlying search engines specifically crawl family history

sites or sites that contain biographical information (such as newspapers' websites). Search tools

that allow the researcher to mine a single site, such as Ancestry.com, come in two forms: simple

search, with one or a few search boxes, and advanced search, with several search boxes for

different types of data and facets (or filters).

Given that genealogical data is spread across many websites and that these sites have their own

flavors of search tools, FHRs may need to master the use of several search tools and IR systems

to be effective and efficient researchers. To exploit the many features family history search tools

offer requires "web expertise," as well as domain knowledge or knowledge of the FHR process.

Hölscher and Strube (2000) define "web expertise as a type of media competence" or "the

knowledge and skills necessary to utilize the WWW and other Internet resources successfully to solve information problems." This was born out in the think-alouds: Expert participants demonstrated a mastery of search tools that yielded a higher numbers of relevant results; novices were reliant on simple searches, which generated far more irrelevant results.

Novice Searching Behaviors

Reliance on Inter-Website Search Engines and Simple Search Tools

Novices who participated in the think-loud sessions tended to use simple searches, rather than advanced search forms, on general-purpose search sites (such as Google), cross-website family history search sites (such as Mooseroots.com), or family history websites with genealogical records such as Ancestry.com. This reliance on simple search tools had significant consequences in terms of their ability to find relevant results. Simple searches involve a single box form or a search form with a few boxes (in this case, first name, last name and date boxes), with no additional prompts to add other contextual information or to filter out specific information. Given the large result sets that a search may yield for most names, particularly when there is more than one spelling of a surname, FHRs may need to add additional terms or filter out unwanted results to successfully surface relevant results to the top pages of search results. In other words, successful searching with simple searches may require more advanced knowledge of the terms to employ and the use of Boolean operators to filter out unwanted results.

Novices appeared to lack knowledge of what terms to use in simple search forms, and not one novice used Boolean operators. This lack of knowledge was apparent when four of the novices employed Google or Yahoo to conduct their first searches. Many simply added the search terms,

"Pauline Mai" or "Pauline May," which resulted in pages and pages of hits for living Pauline Mais and Pauline Mays the world over. Novices struggled to determine whether any of these living Pauline Mais were related to the Pauline Mai in question. They were unable to identify any search results that were relevant to Pauline Mai. This finding complements the results of Bhavnani's (2002) study of domain experts' and domain novices' web-searching behaviors. Bhavnani found that novices typically started with general web search engines, examined few items in the results sets, and usually terminated their searches before relevant information was found. So, too, for novice FHRs. These findings support Jenkins', Corritore's, and Wiedenbeck's assertion that "domain novices, even with good search and browsing tools, may simply lack the knowledge necessary to make good information choices. This suggests that system support for making choices also needs to be integrated into Web tools" (Jenkins et al., 2003, p. 84).

In addition to using all-purpose search engines, many of the novices attempted to employ cross-website search sites to locate information on Pauline Mai and her family members. As mentioned earlier, these sites mine a myriad of family history websites and other sites that contain biographical information. Cross-website search engines presented a number of challenges for the novices. The first barrier novices encountered was that these search engines seemed to give preference to pay-for-access sites; as result non-free sites tended to dominate the search results. For example, novices found that such sites would offer links to websites with newspaper obituaries, all of which required paid subscriptions. One expert think-loud participant complained about these search sites: "They really don't help you.…They send you to websites that don't have the actual information about the person, and then to another that doesn't have the information either. And Ancestry ads are all over the place." Novices experienced the circular nature of the search results provided by these cross-website FHR

search engines. They would click on links to sites that supposedly had information on "Pauline

Mai" or "Pauline May," but many of these sites lacked biographical information and were also

cross-website FHR search engines. This was maddening to many of the novice think-aloud

participants, who felt they were "going in circles" (think-aloud participant #1). The five novices

that tried their hand at using cross-website search engines for family history eventually gave up

on them and moved directly to searching on Ancestry.

All seven novices either initially or eventually visited family history websites that provided

access to genealogical and documents; most focused on Ancestry, but one targeted

FamilySearch initially and then moved to Ancestry. For all seven novices, Ancestry, once found,

played the dominant role in their search for Pauline Mai. While Ancestry and FamilySearch

offered advanced search forms and/or functions, the novices did not employ advanced search

functions, such as facets or filters, to unearth data on Pauline Mai. (Novice use of the advanced

search functions did not occur until I recommended using these search features later in the

think-aloud session). When interviewed, many of the novices spoke of using Google, Bing and

Yahoo when they searched for information, in general, on the web. These search engines all

feature simple search forms – a single search box. Given their heavy use of simple search forms

when searching on the web, it is possible that that they used the same sort of search form on

family history websites because that is the type of search they are accustomed to.

While many of the novices seemed most comfortable with simple search forms, they may have

not employed Ancestry's advance search option, because they did not know it existed.

Ancestry's homepage was so cluttered with content, that six of the novices couldn't even find

the simple search on the homepage, let alone the advanced search option, which was a few

clicks away. Instead, the first search box they found was for Ancestry's family tree search and

production tool, which at that time was prominently placed on Ancestry's homepage. As discussed earlier, this tool combines searching and building a family tree into one function. Novices found this combination of simple search inserted into a production tool very confusing: "I don't want to make a family tree so why is it doing that? I thought it was going to search" (think-aloud participant #3).

After giving up on the search and production tool, many of the novices located the simple search on the Ancestry homepage (at that time, it was just below the search/production tool box). In the simple search, novices initially input "Pauline Mai" or "Pauline May;" some also added an approximate birth year (which was a guess, since the provided information sheet included a date range for the birth date for Pauline). Those who searched on Ancestry using the simple search (search terms used: first and last name, and some added a supposed birth date) fared poorly. Given that Pauline Mai and Pauline May are fairly common names and many lived in the late to the mid-20th centuries, novices' decision to use the simple search and only input a name or a name and a potential birth year as search terms meant that relevant Pauline Mai records were pushed down in the search results. Due to the lower placement of the relevant results, all of the novices overlooked them.

If they had initially used Ancestry's advanced search form, which contained many other search boxes in which to input specific types of data, such as location of birth, children and spouse names, etc., the novices may have been prompted to include additional data in their search, based on the information provided to them about Pauline Mai and her family. Doing so would have aided them in surfacing more relevant results. In contrast, their searching on "Pauline May" and "Pauline Mai yielded no relevant hits in the first three pages of results. In general, the novice participants seemed only focused on only pursuing the very top hits in the search results,

which were not relevant due to the lack of specificity in their searches. In turn, they went down many wrong search paths. This behavior is similar to a behavior found in several relevance studies: Users tend to deem information objects presented earlier as being more relevant than those presented later (Bar-Ilan et al., 2009; Eisenberg & Barry, 1988; Huang & Wang, 2004; Purgailis Parker & Johnson, 1990; Xu & Wang, 2008).

Following a similar, fruitless search path, one novice first used the FamilySearch search tool. Though this tool includes an advanced search form (with faceted search functions that need to be "turned on"), the novice only input a name. Her FamilySearch results were slightly better than the Ancestry searches conducted by the other novices, in that one relevant Pauline Mai record was in the top five results. Unfortunately, the participant deemed it irrelevant, as it was a census record for a Pauline Mai who lived in New Jersey and the provided information did not explicitly state she lived there. Had she more closely examined the transcribed data and the image of the census document, she may have realized that much of the data aligned with the provided information sheet. After examining the first result and finding it irrelevant, she decided to move on to Ancestry, rather than examining the other results.

Lack of Adaptability in Searching

Novices who participated in the think-aloud sessions exhibited far less flexibility in their searching behaviors than experienced FHRs, often staying fixated on the same search path, and repeating the same terms, even when that path was leading to irrelevant results. Similarly, Wildemuth (Wildemuth, 2004) and Vakkari, Pennanen, and Serola (2003) found that students with little knowledge of a topic used less variation in their search strategies than after they gained more knowledge of the topic.

Even though some of the key data in the top results did not align with the provided data on

Pauline Mai and her family, novices tended to only examine the top results and were inclined to

accept data from those results. Four of the novices eventually added a date to their searches

(via the simple search form) by extracting dates from their top search results, such as incorrect

birth dates, which resulted in more irrelevant results. They also used that incorrect data to

evaluate existing search results, which led them down more wrong paths. It seemed from the

joy on many of their faces and their enthusiastic exclamations, such as "I found her!" (think-

aloud participant #13), that they were so happy to find information that looked somewhat right,

they were willing to accept data that did not line up with some of the key facts provided to them

about Pauline and her family. The tendency of domain novices to be more lenient in their

relevance inferences than domain experts also has been observed in studies of students and

scholars (Saracevic, 2007).

The novices who participated in the think-aloud sessions provided many apt examples of this

behavior. Two novices landed on the same census record (a transcript of the census) for Pauline

Mai via an Ancestry search, and stated they thought this was right for Pauline Mai, even though,

there was significant incongruence between the children's names that were listed and those

that were provided. Another was convinced she had found the right Pauline Mai, even though

the city directory (a transcript of the city directory) was for a city that was far from the

residences listed on the sheet for Pauline and the directory dated from the decade before

Pauline arrived in the USA. A few other novices found records (transcriptions of records on

Ancestry) for a Pauline Mai who resided in Wisconsin, and were certain they had found the right

Pauline Mai, even though the found data did not align with the approximate birth date that was

provided or the provided names of her children. In such circumstances, the novices seemed to

get stuck in a rut, pursuing the same search trajectory based on incorrect data. Once incorrect

data was accepted it had a snowballing effect on their search and relevance judgements:

Accepting incorrect data led them to examine other irrelevant records and then accept those

records, and so on. Their persistence in following the wrong search paths made them more

committed to those paths and to the accuracy of their search results. This snowballing of

irrelevant data caused by committing to irrelevant results early, then diligently pursuing more

data based on the irrelevant data, and the resulting confidence in the found data appears to be

a novel finding. There is no known study that examines this phenomenon. More research is

needed to determine if this behavior is unique to novice FHRs or whether this behavior is

common among novices in other domains.

The Impact of Helpful Guidance

To encourage novices to abandon unsuccessful search trajectories and the collection of

irrelevant data, helpful suggestions were offered. After several minutes (from 15-20 minutes) of

exploring and committing to irrelevant results, I offered suggestions to the novices, such as,

"Hmm, those children are different than the ones on the sheet," or, "From what it says on the

sheet, it appears that Pauline Mai also lived in Idaho." I also recommended that they use the

advanced search option on Ancestry, as this was the overwhelmingly popular family history

website for all seven novices. Such elementary suggestions often led them to try different

search terms in new searches and/or to reevaluate some of their results from previous searches.

Including additional search terms helped all but one novice find relevant records. Six of the

novices then extrapolated a few data points from these results and added one or two of these

data points (such as birth date) to their search terms (along with "Pauline Mai"). This led to

relevant results surfacing in their searches.

Two novices who participated in the think-aloud session initially exhibited a slightly different

searching approach, best described as paralysis, followed by schizophrenic searching. After

inputting "Pauline Mai" into Google or Yahoo (one novice chose Google, the other Yahoo), they

were overwhelmed with the number of irrelevant results for living Pauline Mais. One of the

novices exclaimed, "I don't know which one to pick, these all are wrong" (think-aloud participant

#9). As both these novices seemed paralyzed by the overwhelming number of hits for living

Pauline Mais, I suggested they add "deceased" to their searches, which helped surface more

results for family history websites. At that point -- when they had a result set that looked more

promising -- they began to exhibit a schizophrenic searching pattern: they would visit one of the

sites in the search results, select one or two of the listed results for Pauline Mai on that site,

quickly examine the records with looks of uncertainty, then give up on that site and return to

the Google or Yahoo search results. They continued this dizzying back-and-forth behavior from

the Google/Yahoo search results to promising websites in quick succession for 10-15 minutes.

While this searching pattern has not been found as such in other studies on web searching

behaviors, it resembles novice search behaviors that were uncovered in a study conducted by

Jenkins, Corritore, and Wiedenbeck (Jenkins et al., 2003). They found that web novices tended

to use a breadth-first pattern of information-seeking. This pattern was characterized by an

unwillingness to stray more than one click beyond a hub webpage, as the web novices became

disoriented with further clicks.

As might be expected, these two novices became very frustrated. As a result, I recommended

they search a website that contain extensive biographical information, such as

FamilySearch.com or Ancestry.com. Both opted for Ancestry and began searching for Pauline

Mai using the simple search. At that point, they exhibited less anxiety and their searching

behaviors began to resemble those of the other five novices. These two novices experienced the same challenges in selecting search terms and relevant results on Ancestry as the other novices. As with the other novices, I interjected helpful comments regarding the provided information about Pauline and her family members. As with the other novices, my comments led them to pursue more fruitful search paths.

While all the novices successfully used the advanced search to find some relevant data, they did not fully exploit the many search options available through this tool. For example, not one novice employed document/record type filters that were available in the advanced searches. They seemed to be completely unaware of their existence. Document filters enable users to narrow their searches to a defined set of "document types," such as searching for immigration/naturalization records. Using the document type filters would have helped the novices uncover additional information from documents other than census and death records. Other pitfalls were not avoided. Since some of the unearthed results were transcriptions of census data from original records, novices unwittingly encountered another challenge – determining what data from various records is trustworthy enough to include in a search. Unaware of how faulty census data can be, novices tended to assume that all transcribed data they unearthed was correct; for example, four of the novices trusted age information from a census record that listed an incorrect age for Pauline Mai. Based on this faulty data, they calculated a birth year for Pauline Mai and added this to their search terms, which led to numerous irrelevant results being retrieved. Instead of dropping what was likely incorrect data, particularly from census records, two of the novices moved to other family history websites and re-input incorrect dates in their searches. This also led to irrelevant results surfacing to the top of search results.

Regardless of some initial differences in the novices' searching behaviors, all of the novices

demonstrated a lack of adaptability in their search term selection: When search results seemed

to be leading to irrelevant results, they tended to continue to use those terms. All profited from

hearing and acting upon suggestions I offered that led them to refocus their search efforts by

better exploiting the provided information about Pauline Mai, including adding additional data

in their searches and comparing found data with that provided about Pauline Mai. All seven

novices were able to find at least one relevant transcribed record for Pauline Mai, and to have

some confidence that it was indeed relevant to the Pauline Mai in question. However, many

more relevant records might have been found had the novices been skilled at using other search

options, such as document filters, that are available on sites like Ancestry and FamilySearch. This

suggests that additional help may be needed to encourage novices to use more varied search

strategies.

Experienced FHRs' Searching Behaviors

Given that one of the main barriers to finding relevant genealogical data is the large results sets

for people with the same names or similar names, it is paramount for FHRs to learn how to

surface more relevant results toward the top pages of results. In this section, the ways

experienced FHRs coped with a plethora of search results and how they surfaced relevant

results will be discussed. Their search approaches were more varied than those of novices, and

they demonstrated more adaptability in searching than novices. They also had mastered

advanced search functions on websites such as Ancestry, which enabled them to find relevant

data and documents.

Adept Use of Advanced Search and Multiple Search Terms

In sharp contrast to the novices, none of the experienced think-aloud participants used cross-website family history search tools. Rather, most expert participants visited trusted family history websites that provide access to actual genealogical data (both user-contributed and from historic records). The experts tended to either conduct their first searches on the trusted sites or select those sites from a Google search. Similarly, Bhavnani (2002) found that domain experts were aware of key resources for their domain and often went directly to those online resources, rather than employing general web search engines.

Participants who visited trusted, data-rich websites used their advanced search forms/functions, which allowed them to surface more relevant results than the novices, who tended to use simple searches. Experienced FHRs filled in multiple fields in advanced search forms, which enabled the Ancestry or FamilySearch search engine to triangulate the entered data and retrieve more relevant results. As one think-aloud participant explained as she was filling in the advanced search form on Ancestry: "Let's get systematic about this. I am going to add her name, and children's names, and a few of the places I know she lived," and, after she examined one of the search results, "I think this is promising because the kids line up, and she's in Idaho and this is in Idaho" (think-aloud participant #4). Only a couple of the experts used all-purpose search engines, such as Google, but when they did, they combined Pauline Mai with various search terms that were effective in surfacing relevant results to the top pages of the results. While they did not demonstrate knowledge of Boolean operators, they had a good sense of what combination of data and/or search terms to include in a search to surface relevant results.

By including several data points (e.g. name, birth location, approximate life dates) in their search queries, experienced FHRs were able to unearth more relevant data and documents for Pauline

Mai. Novices, on the other hand, used few data points in their queries, which resulted in

irrelevant results appearing at the top of their search results. These findings are contradictory to

those found in Hölscher's and Strube's (2000) study, in which participants with little domain

knowledge (little knowledge of economics) made significantly longer queries than those

participants with domain knowledge (economists). In a digital world where thousands of search

results are possible for any given name, more search terms are critical in narrowing the results

and surfacing relevant results. Experienced FHRs know this and use advanced search forms that

allow for the entry of multiple data points about a person. One experienced FHR did not use

advanced search forms, but he used several search terms, including what he deemed to be more

unique terms, in his Google searches. For example, he input two full names (Pauline Mai and

Herman Pfenninger), the term "deceased," and a date for his first successful query using Google.

In their use of Google searches or advanced search forms, experienced FHRs effectively used

multiple data points to unearth relevant data. This finding suggests that more research is

needed to determine if differences in query formulation exist in other domains.

Many of the experienced FHRs who participated in the think-aloud sessions also were adept at

using document-type filters provided in the advanced search on websites such as Ancestry and

FamilySearch to narrow and refine their search-result sets. The number of document-type filters

can be extensive: In 2013 and 2014, Ancestry provided its users with over 50 different types of

document filters on its advanced search. Document filters enable the user to narrow search by

document type. For example, one can narrow a search to military records only. Participants used

approximately 10 different types of document filters during the think-aloud sessions.

All but two experienced FHRs who participated in the think-alouds used document-type filters included in the Ancestry advanced search and FamilySearch search forms to narrow results and find additional documents and data pertaining to Pauline. Five of the experienced FHR think-loud participants cleverly employed filtered document searches, starting with one type of document and moving to others. As they moved from one document type to another, they used newly found data to refine their searches and to locate additional documents. Many turned to census records first. As one think-aloud participant explained, "Census records are great starting place because they have a lot of information, but you need to check information you get in census with other sources" (think-loud participant #10). Others chose other information-rich document types that would provide them with more data points to work with.

For example, one think-aloud participant (think-aloud participant #2) with years of FHR experience conducted a search using the advanced form on Ancestry. Extracting data from the provided information about Pauline Mai, she used the following search criteria: Mai as the surname; Pauline as the given name; Arthur as her child; Germany as her birth place; and a date range for her birth date. The result set was large, with some promising results, but she felt she first needed to establish when Pauline came to the U.S., so that she might be able to trace her movements there. Accordingly, she employed the advanced search's collection-type filter for immigration records because she wanted to "check to see where people in these records emigrated from" and "to narrow down the search." By doing so, passenger list records for the right Pauline Mai were pushed to the top of the search results. Uncovering Pauline's immigration information set the participant on a fruitful search trajectory which led her to find more relevant data on Pauline than any other think-aloud participant.

Specificity vs. Fuzzy Searching and Probability Calculating for Search

Experienced FHRs knew more than to use advanced search forms, document filters, and to add multiple data points to these forms. They also knew when to conduct fuzzy or more specific searches in certain fields based on the trustworthiness of found data. This search adaptability came, in part, from their knowledge of the trustworthiness of data contained in original documents. For example, four experienced FHRs calculated a birth date for Pauline Mai from a census record. They all voiced concerns about the accuracy of dates in census records. As one participant explained:

> Census takers often did not speak the language of those being documented, or the person giving the information could change, or could even be a neighbor if the family wasn't home, so you got to be a little skeptical about census records. (think-aloud participant #10)

Given their lack of confidence in census data, four of the experienced FHRs opted to use a fuzzier search, using the date range option, for the birth date calculated from the census record. This was important as one census record for the Mai household indicated that Pauline was three years younger than she was. Experienced FHRs also spoke of being more confident with dates and other information from documents that were generated close to a life event, such as an official marriage certificate, and some think-aloud participants found such documents for two of Pauline's family members. When using dates extracted from trustworthy documents to search for more information about these two family members, the think-aloud participants used specific date searches, rather than a date range. Such search strategies directed these participants to documents I had not uncovered myself after conducting extensive searches on Pauline Mai and her family.

Experienced FHRs who participated in the think-alouds also engaged in "probability calculating"

to determine life dates that were not explicitly stated in documents. Using a series of found

documents, some of the experienced participants calculated probable dates for other life

events, based on the other dates and contextual information contained in those documents. For

example, one of the experienced FHRs (think-aloud participant #5) derived a possible range of

dates for Pauline Mai and her husband's marriage from three census records (from three

different years). Based on these census records, she also gained a sense of when the couple may

have moved to Idaho. Those who used probability calculating were careful to conduct more

fuzzy searches with the dates they calculated in this manner. They also used such calculated

dates to determine if a result was relevant: if results contained dates that were too divergent

from the calculated dates, they discarded them.

Adaptability in Searching

Seasoned FHRs who participated in the survey, initial interviews, and the think-aloud sessions

described how they learned to be more skillful in searching for data and documents. One such

skill that FHRs frequently alluded to was the ability to adapt or change their search strategies

and trajectories when a search was leading to irrelevant information. For example, a think-aloud

participant with considerable FHR experience decided to alter her research path because the

data she found was not lining up with the data she had been given about Pauline Mai. She

explained her change in search direction:

> Participant: "Now that we are probably 15 minutes into this I am going to rearrange
> my thoughts because I am going to search her kids instead of her."
> Interviewer: "Why are you doing that?"

Participant: "Because I don't have concrete evidence to go forth with to confirm what I think I have right for her." (think-aloud participant #16).

All of the think-aloud participants except one (who had constructed a nearly perfect search by combining a set of terms in a Google Search that gave him a large number of relevant results) changed their research direction by abandoning search paths that were not yielding results that aligned with the provided data and embarking on other trajectories. Some examples of the ways in which think-aloud participants changed their search directions include:

1) Change search to variant surnames: abandon search on Pauline May, and begin a search on Pauline Mai (the alternate spelling provided to participants).

2) Change search from primary name to children's names: abandon search on Pauline Mai, and begin search on her children with less common names.

3) Change search location: abandon search on Pauline Mais who resided in Wisconsin, and begin search on Pauline Mais who resided in Idaho.

By being nimble in their search trajectories, think-aloud participants were able to stop themselves from going down dead-end paths and move to more promising paths that led them to relevant results.

Take, for example, the adept handling of variant spellings of surnames by experienced think-aloud participants. Variant spellings of surnames are commonly found in family history research. Some variant spellings are due to misspelled names that appear in user-contributed transcriptions (with those misspelled names then included in the search index); others are misspelled in the original documents themselves. Experienced think-aloud participants were

adept at managing the variant spellings of Mai/May that were included in the provided

information sheet on Pauline and her family. Unlike the novices, they quickly realized that May

was not the right spelling, as "Pauline May" searches generated results that did not align well

with the provided genealogical information. Several survey respondents also indicated that as

they became more experienced, they became more adept at handling variant names in the

searching process. As one survey respondent (#377) stated: "I learned to search with a variety of

surname spellings and nicknames."

Another adaptive search strategy -- demonstrated by think-aloud participants and mentioned by

five initial interview participants -- was searching on less-common names in a family to aid in

finding information on other family members with more-common names. All of the think-aloud

participants encountered large result sets for Pauline Mai. To surface more relevant results, five

experienced FHRs conducted searches on Pauline's relatives who appeared to have less

common names (e.g. Herman Pfenninger). Doing so enabled them to track down an obituary for

Ella Pfenninger (who they discovered was Pauline's daughter), which provided important data

about Pauline. In turn, the participants used this data to conduct more narrow searches on

Pauline.

Browsing as a Search Strategy

Due to incorrectly transcribed data and because of incomplete transcriptions of documents, it

can sometimes be challenging to locate original sources using search engines on family history

websites. One participant explained the problem: "Sometimes the person doing the

transcriptions for the key word searches writes them down wrong. So you can't find it using the

search" (initial interview participant #9). To address this challenge, several experienced FHRs,

who participated in the think-aloud sessions and the initial interview, spoke of browsing imaged

documents to locate information that was not accessible through keyword searches. One

participant described how he managed this challenge when searching for information about his

great-grandmother:

> I think it's important to look at the original because there are usually transcription errors. And it may be only half the family. Like, I have googled and look for my grandfather's mother, and the census comes up and it says oh look, here is Francis Tolieu from Ancestry. It will tell you who else was at that residence. So they pulled up my grandfather's mother, father and uncle, but instead of their names being Tolieu, it was like Wallace. First names were also misspelled by transcriber. You really have to look at the census because you may find other people who are related but they don't come in the search but are on nearby grouping on the census that you are looking for. (initial interview participant #9).

Another participant also spoke of looking at other census entries that are in the "vicinity":

> You need to understand how to look at the census records. When you do the search, and sometimes you look at the name below, and it's like oh that's the person I am really looking for there. (initial interview participant #8)

Unfortunately, the online tools for browsing documents are not particularly user-friendly. Think-

aloud participants struggled to move from one page to another page in online census records on

both FamilySearch and Ancestry. On some sites (e.g. http://www.libertyellisfoundation.org), it

was nearly impossible for the participants to browse or move from one page to another page in

a record group.

Adaptability in searching and the use of varied search strategies has also been found in the

online behaviors of web search experts and individuals with various types of advanced domain

knowledge from healthcare to economics (Aula, Jhaveri, & Käki, 2005; Cothey, 2002; Hölscher &

Strube, 2000; Hsieh-Yee, 1993; Jenkins, Corritore, & Wiedenbeck, 2003; Kellar et al., 2007).

Similarly, experienced FHRs' domain knowledge and knowledge of online search tools enabled

them to be nimble searchers who found more relevant data than the novices who participated

in the think-alouds. Experienced FHRs knew when found data and documents were likely to be

irrelevant, and this moved them to try different search strategies, such as reformulating their

search queries or browsing documents. They also demonstrated an awareness of search

techniques that allowed them to uncover different relevant data than initially had been found.

This translated into the experienced think-aloud participants creating richer, more extensive

Find A Grave memorials for Pauline Mai than those created by novice think-aloud participants.

6.4 Other Family History Research and Production Literacies

As mentioned earlier, there are additional literacies, beyond search literacies, that are critical to

producing accurate family history. These literacies enable FHRs to collect a myriad of data and

documents that are both relevant and accurate. Those literacies include:

- Data/document collection and management literacy

- Interpretive document literacy

- Transcription literacy

- User-contributed content literacy

- Production and sharing literacies

- Data triangulation/alignment literacy

Each of these literacies will be delineated in the following paragraphs.

Collecting and Managing Data and Documents

The collecting of data as well as documents is an essential part of the FHR process. Collecting

and managing data and historic documents enables FHRs to build family history content and

provides them with "evidence" to support the claims they make about their ancestors. This

collecting behavior resembles the "information managing" step in Meho's and Tibbo's (Meho &

Tibbo, 2003) enhanced version of Ellis' information-seeking model. Meho and Tibbo characterize

"information managing" as activities related to filing, archiving, and organizing found

information. Information managing is a learned literacy; evidence from all three collection

instruments suggest that FHR novices typically do a poor job of collecting the right data and

appear to be unaware of the importance of collecting documents. Many of the study

participants indicated that as they became more experienced they learned what to collect, and

while online tools to manage family history information are limited, many devised methods to

organize found documents and data.

Collecting and Managing Data

FHRs need to know what data to collect to create family history content. Given that about half

of survey participants (51%) jumped right into researching their families on the web without

learning how to conduct FHR from any source (how-to sources on the web or in analog form,

other FHRs, or family), it is likely that accessible, online family history production tools (such as

family tree construction tools on sites like Ancestry and Wikitree) are informing the data

collection choices novices are making. As one participant stated:

I learned how to research with family tree software on Ancestry, which guided me in

what information to get and how to record it. (survey respondent #491)

Another participant simply stated: "I joined Ancestry.com and began building my tree" (survey

respondent #209). Family tree production tools focus on collecting and displaying basic, but

critical, data about ancestors:

1) Names of ancestors (given/first name, middle name, surname)

2) Life dates of ancestors (e.g. birth and death dates)

3) Places of birth and death

4) Names of immediate ancestors [e.g. spouses' name(s), children's name(s)]

Some family tree production tools allow for additional data to be collected and organized, but

these four types of data are the main focus of the tree tools, as it is this data that is graphically

displayed in many digital family trees on sites such as Ancestry, Wikitree.com, and Geni.com.

The following screen shot of the Farr family tree (created via Ancestry) illustrates that only basic

information about the person is displayed (Figure 8):
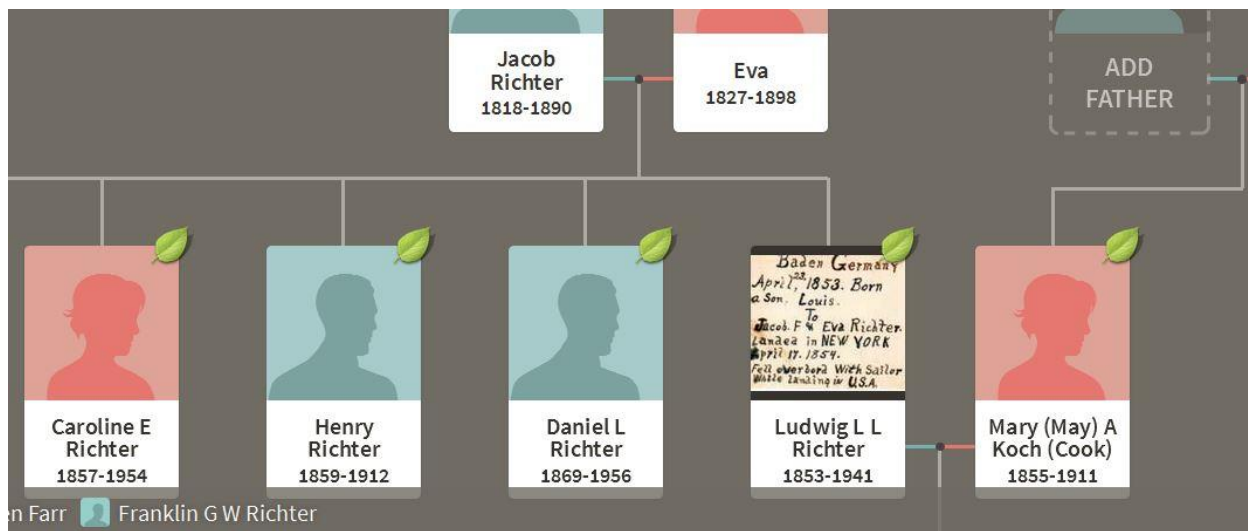
Figure 8: Farr Family Tree Created with Ancestry.com Production Tool

As the Farr family tree illustrates, most online trees display only basic life and residence data.

Several study participants mentioned that as novices they initially got their start in the FHR craft

by using online family tree production tools; as a result, they were collecting basic data about

their ancestors and little else. As one participant opined:

> At first, I was not recording enough information. I thought I just needed, you know,
> birth, death and marriage dates, that kind of stuff to fill my tree. But now I record lots
> of other information. Every bit helps. (initial interview participant #12)

This finding -- novices only collecting basic data about their ancestors -- echoes Duff's and

Johnson's (2003) observation that many genealogists start by collecting names of family

members. Later, they may move to gathering detailed information on family members.

In contrast to novices, experienced FHRs appear to collect many more types of data, even data

that at first glance may seem inconsequential. For some experienced FHRs, data collecting is a

"data snowballing" activity in which one bit of collected data leads to another and then to

another. As one participant explained:

> One cannot rely on just one source of information. I used to not write down
> professions, but now I do because other sources of information may be gained from
> that information -- military records, if they were a doctor, if they were a city official,
> etc. (survey respondent #116)

Even what might appear at first to be less important data can be useful in FHR: "I found that so

many minor details, can make a huge difference in the research" (initial interview participant

#10). Some of the additional important data types that experienced FHRs identified as being

important include:

- Variant spellings of their ancestors' surnames

- All places of residence of ancestors

- Occupations of ancestors

- Military service data about ancestors

- Ancestors' immigration dates, ports of departure, ports of entry

- Religion and church membership of ancestors

- Other membership data about ancestors, such as membership in professional or

  benevolent societies.

- Other life dates (e.g. baptismal date, marriage date, engagement date)

Such varied data is gathered and stored as more data helps support FHRs in the data

triangulation process (which will be discussed later in the chapter) and because such data

provide richer pictures of ancestors' lives. Family tree production tools do not support nor

encourage users to collect such varied data, which may lead novices to collect less data then is

needed to make good relevance inferences. This study found that for even less-common

names, FHRs are confronted with multiple records for different people with the same names on

the web. Given this, the collection of more data may be critical in locating information that is

relevant. However, many FHRs indicated that it was challenging to manage these additional data

points, as most online family tree and memorial production tools do not provide structured

spaces for such data. Participants described using complicated paper filing systems, lengthy

word processing documents, spreadsheets, PowerPoint presentations, and even Microsoft

Access databases to manage FHR data that did not fit neatly on family trees.

Collecting and Managing Documents

Experienced FHRs collect not only data, but also original documents, as original sources

represent important evidence that supports their claims. One participant described the

importance of documents in creating accurate history:

> Seeing the original itself, including a digital copy, is the closest you're going to get to
> the right information. If you can copy the original, you know by saving it if it's digital or
> photocopying it, that is the best thing to do. You need sources to prove you're right.
> (think-aloud participant #9)

Document collecting allows FHRs to "source" their claims — a key component in producing

verifiable and accurate family history. Many study participants indicated that as they became

more experienced they began to focus on sourcing. One participant described this change in

direction: "The biggest change [in her research and production practices] was learning to source

everything! Nothing is `fact' without sources!" (survey respondent #308).

Experienced FHRs emphasized that accuracy in family histories was founded on finding multiple

documents that contained the same "facts" about their ancestors. As one participant

emphasized:

> It's important to find out what record are available for relevant time periods, i.e. Civil
> Registration, Census records, Tax Records, Property Records, Parish Records, Wills and
> probate records, etc. Keep records of research done, including that which returns no
> results. Record information from documents accurately and note the source. Sources
> not only enable you to tell where you got the information from but help with the
> voracity [sic] of the information, as well as helping others to find information. (survey
> respondent #476)

Another participant indicated that as she gained experience in the FHR craft, she "learned that

unless I go on, there are like six or seven original sources that refer to that person, I don't add

them to my tree" (initial interview participant #8). Another participant suggested that even

more documents are needed to establish the facts about an ancestor: "If I can get 10-15 sources

for people I am studying a lot, that's great and helps me establish the facts!" (initial interview

participant #12). One participant described herself as "a big primary document person" (initial

interview participant #13).

The myriad of documents that study participants indicated they used to confirm facts about

their ancestors was impressive: 43 different types of documents were mentioned in the survey

responses, the interviews, and the think-aloud sessions. Documents that participants employed

for FHR included: apprentice and indenture records, church records of many types (e.g.

baptismal and christening records), birth records, census records, court records, death records,

deeds, directories (city, phone, etc.), family bibles, family histories and local histories

("published" histories), grave/graveyard records, immigration/naturalization records,

inquest/post mortem records, land/deed records, manumission records, maps, marriage

records, military service records, newspaper articles, member-organization records (e.g.

beneficiary societies, brotherhoods, professional societies, etc.), obituaries, pension records,

probate records, school records, ship manifests/passenger lists, Social Security records, tax

records, voter lists, wills and workhouse/poorhouse records.

As FHRs become more experienced and exhausted easily found documents (ones available on a

handful of sites, such as Ancestry), they learned of more obscure records, some of which are

online but many of which are offline, housed at archival repositories across the country.

Knowing where to find lesser-known and harder-to-access records requires archival intelligence,

or "knowledge of archival principles, practices, and institutions" (Yakel & Torres, 2003).

Regardless of whether the documents are analog or digital, many experienced FHRs who

participated in this study indicated they would collect such documents in whatever way they

could -- be it by photocopying, digital downloading, or having the documents digitally imaged. If

documents could not be copied, FHRs made clear that it was critical that citation information

pertaining to those documents was collected.

Most experienced FHRs who participated in the think-aloud sessions exhibited this knowledge

by attempting to collect more unique documents on archival repository websites, such as the

U.S. National Archives and New Jersey Historical Society websites. The survey responses also

indicate that experienced FHRs employ many websites to locate information about their

ancestors: close to 67% of the experienced FHRs surveyed said they used 11 or more websites to

find such information. The majority (65%) of the novice FHR survey respondents employed six

or fewer websites to conduct research on their families. It appears that as FHRs become more

experienced, they gain awareness of the myriad of websites and brick-and-mortar repositories that contain genealogical data and documents, not just the usual suspects such as Ancestry and FamilySearch.

Not only do experienced FHRs visit many websites to glean information about their ancestors, they also tend to visit brick-and-mortar repositories to locate relevant sources. As one participant explained:

> Many records aren't online, even if you have all the right information to find the right records. You actually have to go to the place where the actual records are -- the town hall or community center of some sort, churches. (initial interview participant #1)

All of the experienced FHRs who participated in the initial interviews and think-aloud sessions had visited multiple repositories to locate source materials on their ancestors. One experienced FHR explained her transition from pursuing only online records to also seeking harder-to-access offline records:

> I expanded my search range; I rely far less on websites such as familysearch and more on documents available through archive offices. (survey respondent #216).

Another participant (think-aloud participant #8) spoke at length about journeying to courthouses, county offices, and local historical societies in Pennsylvania to locate land records pertaining to one of his ancestors. The survey yielded similar results. Of the experienced FHRs who responded to the survey, 89% had visited brick-and-mortar repositories to research their ancestors, with 68% having visited repositories more than five times and 57% having visited 11 or more times. While increasing amounts of transcribed data and images of original documents are online, these findings suggest that more experienced FHRs continue to travel to brick-and-

mortar archival repositories to conduct research. These findings conflict with Veale's (2005)

assertion that the web environment is enabling FHRs to collaborate online, thus making

journeys to archival repositories less needed.

As with managing data, experienced FHRs mentioned the challenge of managing collected

documents, as well as their "research paths," which involved keeping track of what and where

they found data and documents. Several participants described this challenge:

> I am trying to keep better records of what I already have vs what I still need to avoid
> duplicate effort; this has been a real challenge. Trying to be better organized is hard.
> (survey respondent #302)

> You need to document every step, even keeping track of where you looked, even it
> (sic) you didn't find anything. But its (sic) hard to do this. (survey respondent #195)

> Documenting sources! A mistake I made when first doing research was thinking I
> would remember where I found a piece of information" (survey respondent # 398).

> As you find new things in new places, you have to keep it in your brain. It's not easy,
> you know. (think-aloud participant #4).

These comments suggest that better online tools are needed to support FHRs as they collect

data and documents from multiple websites and even offline sites.

Novices and Document Collecting

In contrast to experienced FHRS, novices who participated in the survey tended to pursue

evidence from documents readily available on the major family history websites, such as census

records. Close to 97% of the novices who participated in this study indicated they have mined

census records. Other types of records were used far less by novices. For example, less than 1%

of the novices indicated that they have examined church records, maps, or wills. And only 29% of novices indicated that they had visited archival repositories to examine analog records.

These survey findings are in-line with those from think-aloud sessions. Based on their comments and the search results they choose to pursue, it appeared that novice think-aloud participants had knowledge of only commonly known sources of biographical information, such as obituaries, birth and death records, and census records. One novice (think-aloud participant #1) became so obsessed with finding an obituary for Pauline Mai that she spent 20 minutes attempting to find one online, even though some of her searches were surfacing relevant data from other sources.

In addition to having limited knowledge of the large array of original documents used in FHR, novices may be bypassing even documents they are familiar with. Data from the think-aloud sessions suggest that novices tend to rely on transcriptions of original records, rather than the records themselves: six of the seven novices examined transcriptions they found, but completely overlooked images of the original documents that were linked to the transcribed data. Only one novice (think-aloud participant #13) viewed an original document, and in her case, she only viewed one document, although she had encountered links to many others. Several experienced FHRs who participated in the study indicated that as novices they relied on transcriptions of documents, but as they discovered that transcribed data are often incorrect, they began to peruse the documents directly — they no longer accepted the veracity of transcriptions. One participant made this clear:

> Once you locate a record consult the original, not a derivative (indexed list, transcription, etc.). I didn't do this when I started and put incorrect information in my tree all over the place. (survey respondent #377)

This quote points to the importance of examining original documents and extracting data from those documents for family histories. This is not always easy: Specialized knowledge is often needed to make sense of historic documents.

Interpretive Document Literacy

Study findings strongly suggest that FHRs not only need to know what types of documents to collect and where to find them, but they also must learn how to interpret those documents. To extract and make sense of data found in documents requires additional skills and knowledge, in particular, "artifactual literacy," or the ability to interpret records and assess their value as evidence (Yakel & Torres, 2003). There is a constellation of skills and knowledge that study participants employed to make sense and assess the veracity of data found in documents, including: image viewer literacy, handwriting and vernacular literacies, knowledge of document forms and recordkeeping processes, and other historical contextual knowledge. In addition, experienced FHRs in this study were knowledgeable about the trustworthiness of different types of documents, and were adept at managing conflicting information found in documents.

Image Viewer Literacy

To access data contained in online documents, FHRs need to know how to employ image manipulation tools or image viewers. Among other functions, image viewers allow one to focus on and magnify various portions of a document. While these tools would appear to be simple to manipulate, one novice in the think-aloud (think-aloud participant #13) who attempted to use a viewer struggled with it and eventually gave up on using it. Many of the experienced FHRs in the

think-alouds adeptly employed viewers on various websites, although a few complained about

how difficult they were to use. One such participant had very poor eyesight and found the

magnification to be insufficient. Another had arthritis in her wrist. Because the viewer required

her to move the cursor to various portions of the document to make them viewable, she

struggled, given how difficult it was for her to make slight movements with the provided touch

pad and mouse. The viewers employed by think-aloud participants were all slightly different

from each other, with similar functionality but interfaces that differed visually. This suggests

FHRs must learn how to use several types of document viewers in order to mine data from

digital documents on various family history websites.

Handwriting and Vernacular Literacies

Other artifactual competencies are needed. Two such literacies are the ability to disambiguate

handwriting and knowledge of regional vernaculars and the specialized lingo that often appears

on earlier documents. Many participants mentioned that they became better at decoding older

forms of handwriting and different vernaculars as they became more experienced. One

participant mentioned handwriting literacy as one of the main skills acquired as she gained more

experience: "I've become far more adept at reading archaic handwriting" (survey respondent

#286). In response to the question, "What does it take to be accurate in FHR?" several survey

respondents stated that acquiring handwriting and linguistic skills, or "linguistic and

handwriting-deciphering ability" (survey respondent #180), was key. Participants also

mentioned several different types of knowledge and tactics that they brought to bear in their

efforts to make sense of handwritten documents. One experienced FHR participant explained

his approach to disambiguating handwriting and the lingo used in a set of documents:

> Handwriting can be hard to read. Now with records online, like probate records, you can read 10 before it and 10 after it. I have found even with chancery and court records, if I read that record set, if I had to read 100 pages, I got where I needed to be to understand the writing, what the words were for. (initial interview participant #13)

Another described some of the peculiarities of the Pennsylvania Dutch style of handwriting and vernacular, which she had mastered (initial interview participant #2). She found a book on handschrift to be particularly helpful in making sense of these "dutchy" documents related to her ancestors.

Participants also spoke of the need to understand acronyms that appear in handwritten documents. For example, older census records often contain an array of acronyms that FHRs need to know if they want to understand the data contained in such records. Five of the think-aloud participants described the acronyms used in the immigration status column of the census records as they were examining census documents for Pauline Mai and her relations. These acronyms and others (that appeared in other documents uncovered during the think-alouds) were not always intuitive; many of the participants mentioned that human resources, such as librarians, archivists and other FHRs, and how-to books and magazine articles were the means by which they learned how to disambiguate complex documents.

Record Forms and Recordkeeping Knowledge

Understanding the structure of documents and where different types of data are placed on those documents is often essential for understanding the context and therefore the meaning of the data found on the documents. For example, with census forms, having an understanding of what each column header means helps the FHR make sense of the data listed in each column. Many experienced FHRs who participated in the think-aloud sessions took the time to describe

each column in census documents and how to interpret data in those columns. The time and

effort they took to explain the census data suggests the importance of such knowledge.

Often, understanding the purpose of a record and the recordkeeping process that went into

creating it is needed to make sense of the information the record contains. One participant

described how his knowledge of recordkeeping in the 1700s influenced the way he interpreted

records from the period:

> I know that ministers were the record keepers in the old days and they didn't always
> do it as soon as someone was born. Did it down the road when they had a chance. So
> sometimes the birth dates are off in these records. (think-aloud participant #2)

Another participant explained why understanding the census data recording process is critical

for assessing the accuracy of census data:

> With the census, sometimes it was the neighbor giving the information about a family
> not the family. And people need to understand that just because it is on the census
> doesn't mean the census taker actually talked to the family; they may have talked to a
> neighbor or a boarder living in the house. (initial interview participant #9)

This participant made it clear that the way census data was once obtained, often in less than

optimal ways, impacted the accuracy of that data. This suggests that without knowledge of

records and the recording processes, FHRs may have difficulty making sense of data found in

original documents and assessing its accuracy. As one participant responded, when asked "What

does it take to be accurate in FHR?": "Be aware why the record was created and what it actually

records" (survey respondent #453).

Other Historical Contextual Knowledge

In addition to understanding recordkeeping processes, many experienced FHRs in this study referred to their use of other contextual knowledge to make sense of original documents. This finding dovetails neatly with Duff's and Johnson's (2003) formative study on genealogists, in which they found that broad historical knowledge was important in identifying relevant records. This study extends this finding by providing rich descriptions of how historical knowledge is used by FHRs to make judgments about the accuracy of data found in documents and by explaining how this knowledge assists FHRs in managing conflicting and inaccurate data found in sources.

Experienced FHRs who participated in this study indicated that they use many types of knowledge to make sense of data found in original sources, including:

- Knowledge of the naturalization process

- Knowledge of the land purchasing processes at different times and in different places

- Knowledge of immigrant migration patterns for different ethnic groups

- Knowledge of municipal, county, and state court systems over time

- Knowledge of other historic legal processes (e.g. legal process for marriage in different states)

- Knowledge of local and regional history

- Knowledge of historic social practices for different groups

- Knowledge of professions and trades over time

Participants explained how they used historical contextual knowledge to gain a deeper understanding of the data they found:

I've also learned to look at history to put facts into perspective. Geopolitical factors influenced immigration. The more I learn about world, regional, and even local history, the better I can surmise the "Why's." (survey respondent #217)

It's important to understand their migration pattern. That's hard. They emigrated to Massachusetts and then one married someone for Connecticut and then they moved. Then maybe later they are in the more outskirts of Connecticut. You can track that, and you know logically they are probably not going to be in the South. You have to look at the timeframes. You know if someone ends up down in Atlanta during the Civil War – you know that's wrong. You know they came from Connecticut and they're not in a prison camp but living there – you know it's not logical. (initial interview participant #8)

Another study participant used her knowledge of the naturalization practices to make sense of

citizenship data she found in census records. The participant clearly knew the meanings of

acronyms pertaining to citizenship that were commonly used in older census records. She

explained why the naturalization data she had found made sense:

Otto was naturalized, and Pauline was not on this census, which makes sense, the other census said she wasn't naturalized, which is typical. Normally the wife became a citizen when she married if he was naturalized." (think-aloud participant #5)

Additionally, a think-aloud participant (participant #8) described in great detail the land

purchasing process and the kinds of records generated in that process. He further explained

how this knowledge helped him locate a hard-to-find land record at a local government office in

Pennsylvania, and how his knowledge helped him interpret what he found in the record.

Another participant (think-aloud participant #10) was able to make sense of complicated Ellis

Island arrival records in which Pauline had two entries that were years apart. Using a lead from a

user-contributed family tree on Ancestry that included Pauline, the participant was able to track

down the Ellis Island arrival records for Pauline. The tree had incorrect information about

Pauline's first arrival in the U.S., but the study participant, utilizing her knowledge of

immigration sources and associated recordkeeping processes, was able to sort out the incorrect

information on the tree:

> The information seems to be wrong about Pauline in the tree, someone has put the
> wrong first arrival date from a passenger list. …From what I can see from the actual
> passenger list, it looks like she traveled back to Germany for a visit and they have
> wrongly attributed this date to her first arrival. The earlier date on this passenger list is
> the first arrival date in the U.S. So I am going to disregard that, but otherwise she looks
> like the correct person. (think-aloud participant #10)

Other participants discussed how they employed their knowledge of historic social practices to

determine the accuracy of information found in original sources or documents. One participant

discussed the practice of recording "marriage intentions" in the 1700s:

> Back in the 1700s, you had couples recording that they intended to marry. But just
> because someone had a marriage intention date, you can't presume they married.
> That's something people may not know – just because the intention date was
> published doesn't mean they went through with it, and from my own family and some
> history books I read, I can tell you couples didn't always go through with it. If you want
> to know if they married, you need to find the marriage certificate or some reliable
> record that has the marriage date listed. (initial interview participant #9)

This quote alludes to another challenge FHRs face: making sense of conflicting information

found in original documents. The majority of experienced FHRs who participated in this study

emphasized the importance of making determinations about whether found data in original

sources was correct. Many participants indicated that they had found conflicting data in original

sources about ancestors which compelled them to make judgements about the trustworthiness

of that data. Many survey respondents, initial interviewees, and think-aloud participants

indicated that as novices, they were less skeptical of data found in original documents. But as

they became more experienced they began to contrast and compare data found in documents

and to make assessments about the accuracy of that data. As one participant stated: "I became

skeptical of some types of records. Now, I take lots of the dates you find on census records with

a grain of salt" (survey respondent #314). One survey respondent wrote at length about the

importance of assessing the accuracy of data found in original documents:

> Even original records can contain errors, both deliberate and not. Spelling wasn't
> standardized until comparatively recently. People weren't always literate and often
> relied on others to record information or draw up documents and the people doing
> that may not understand local dialect or may not have been very literate
> themselves…Learn about the constrictions and limitations that could affect the
> information – ie the categories of employment that an Enumerator could use on
> census records. When Census records were done. Remember that people didn't
> always give accurate information, either deliberately, or because they didn't know it.
> For example, people didn't always know where or when they were born; many people
> distrusted the taking of census information and were suspicious of how the
> information would be used; some didn't want to be found; also people didn't always
> know details about the lives of relatives, as in when they gave details re the
> registration of birth, marriage or death. That's not everything by any means but it is
> what comes to mind at present. (survey respondent #476)

With more FHR experience, many study participants appear to employ the knowledge they had

gained of recordkeeping practices and evidence found in other documents to make such

determinations.

In contrast, novices who participated in the think-aloud sessions appeared to have little

knowledge of how to interpret more complex documents. Most did not examine actual digitized

documents, but they did examine transcribed data from documents, such as data from census

records. The one novice participant who examined an actual census record appeared to lack an

understanding of the structure and the information contained within the document: "I am not

sure what some of these columns mean…is that an occupation?"(think-aloud participant #13).

Another novice (think-aloud participant #3) perused transcribed data from two census records

and was perplexed as to why census records for Pauline Mai did not include all of her children

from year to year. The participant was poised to dismiss the second census, although it clearly

looked to be the same Pauline Mai, until I explained to her that the census only documents who

is currently living in the household, and that given the children's ages, some may have moved

out. Novices appear to be unaware of the importance of examining images of the original source

material and may not have the knowledge to interpret these originals sources if they do

examine them.

Transcription Literacy

As noted earlier, many study participants indicated that as they became more experienced, they

stopped relying on transcriptions and began examining the original documents when possible.

They did so because, "ANY record can be wrongly transcribed" (survey participant #219), and

therefore, transcriptions should not be trusted. In response to the survey question, "What does

it take to be accurate at family history research?" one survey participant commented:

> Don't accept transcripts of records as being accurate. Get information from as many
> sources as possible. View original documents." (survey respondent #207)

Another participant similarly asserted, "You need to see the Primary Source, not just someone's

interpretation or transcription of that source" (survey respondent #266). During a think-aloud

session, one of the participants (think-aloud participant #11) encountered a transcription that

contained a gross misspelling: instead of "Pauline" it was "Poraline." After viewing the original

and realizing the transcriber had misread the handwriting and, therefore, misspelled Pauline's

name, he exclaimed: "That is an example of a really crappy transcription. That really bothers

me!

Experienced FHRs described how they managed transcription errors. One way was to be familiar

with the common mistakes in transcriptions:

> Transcriptions often have mistakes. And especially when it comes to spellings, some of
> the handwriting can really throw you off. When you are searching for a particular
> person or family, it's usually e's, i's and l's that get thrown together and confused by
> transcribers. (initial interview participant #4)

Another tactic was to use different spellings of surnames, such as replacing the "e" in the

surname with an "a" in their online searchers. Such a tactic was used to overcome the common

error of transcribers mistaking a cursive "e" for cursive "a."

Not only can transcriptions be inaccurate, they often lack data or are incomplete transcriptions

of original documents. Many of the crowdsourcing projects on websites such as FamilySearch

and Ancestry entail transcribing only what is deemed key data, such as names and life dates,

from original documents. Other information contained in the documents is not transcribed. As

one participant explained, "You get lots of information from the original documents that you

never get from the transcriptions. …It's like browsing the shelves in a library rather than just

checking the catalog" (initial interview participant #9). Another participant discussed the

importance of examining original documents for additional data not contained in the

transcription:

> You look at the document to verify that the keyed-in information is right. From my
> own research, like with the census, it will have more than they typed in, like what their
> occupation was, and because they're handwritten you can get some tidbits of
> additional information. (think-aloud participant #10)

Archer et al. (2009) found that the Web blurs distinctions between primary and secondary

sources by combining them seamlessly into a variety of different search tools. The same appears

to be true for transcriptions and primary/original source materials. On many family history

websites, as well as in the search results generated by a variety of all-purpose search tools (e.g.

Google), no clear distinction is made between original sources and transcribed data extracted

from those sources. Given transcriptions often contained errors — the transcriptions do not

accurately reflect the information found in the original — the lack of a distinction between the

transcribed data and the original source can lead to inaccurate information being copied from

the transcriptions and added to online family history content. According to data collected from

all three instruments, novices, unlike experienced FHRs, are reliant on transcriptions, possibly

because they are unaware of the great potential for errors in the transcriptions and their limited

knowledge of documentary sources.

User-Contributed Content Literacy

Family history websites are awash with user-contributed content. Data obtained through all

three collection methods suggest that knowing how to approach and assess user-contributed

content is a critical skill in the online world of family history research. One experienced FHR

after another spoke or wrote of the need to vet online user-contributed family history content.

One survey respondent described how she modified her research process as she became more

experienced:

> Now that I know more, I do not use information from any compiled source (i.e.,
> Ancestry, Family Search, Find A Grave) without thoroughly researching the provided
> information myself. Unfortunately, too much of what is currently being posted on Find
> A Grave is not documented or is just plain wrong. I've come across my own relative's
> graves where information provided by a cemetery was "updated" by someone, or that

a "biography" has been written that has no source reference whatsoever and is absolutely inaccurate. (survey respondent #75).

Another FHR described how he changed his view of the reliability of user-contributed content as he became more experienced:

At first, I believed all information I found in other published trees. I later discovered that much of the information was false and could not be backed up with sources/documentation. I now, accept all info as leads, but do not add to my tree unless I can fully provide documentation to validate the information. (survey respondent #332).

Yet another participant echoed these sentiments:

At first, I didn't realize how interesting, that's a good way to put it, these family trees were. And if I found something that looked good, I thought it was good. And then as I did more research I would discover that ok this isn't good. So now I approach everything with a more doubt. I try to find a record that is independent of a previous one that will have the same information. Then I can say at the point ok, the census says this person was born in you know 1812 in Ireland, and then I have a birth certificate or a marriage certificate that has the same information, then I assume it's basically true.

FHRs mentioned the various ways they manage the inaccuracy problem with user-contributed content, including: 1) ignore user-contributed content altogether; 2) only use user-contributed content that is well-sourced (has attached original documents or citations that can be followed to the original source); 3) use such content as a lead, but only accept as fact once more proof is found.

Novices who participated in the think-alouds appeared to lack awareness that search engines on sites such as Ancestry included user-contributed content in the search results. Further, when asked about the user-contributed content they had encountered, some of the novices appeared

to be completely unaware that they had perused such content; others were aware that some of

their search results were links to user-contributed content, but they voiced little concern about

the potential inaccuracy of such content or what they might look for in user-contributed content

to determine its accuracy.

Production and Sharing Literacies

The interviews and think-aloud sessions provided insight into the types of literacies associated

with online production and public sharing of family history content on the web. In this section,

three aspects of production and public sharing of family history content will be discussed:

production outcomes of the think-aloud sessions, knowledge related to when and what to

publicly share, and knowledge of public/private features of online production tools.

Production Outcomes of Think-Aloud Sessions

The content created by the think-aloud participants followed predictable patterns. Without any

help from the study's researcher, all of the experienced FHRs were able to create accurate

memorials for Pauline Mai that included her life dates (birth year and death date), location of

her grave, her immediate family members (husband and children), and some of her residences.

Six of the experienced FHRs included additional information in the narrative portions of their

memorials, such as where she resided at death, the names of some of her children's offspring,

when and from where she immigrated to the U.S., and her maiden name.

The novices, on the other hand, likely would not have produced accurate memorials without the

researcher's interjections. All seven novices were confident they had found information that

was relevant to Pauline Mai/May; however, in most cases, the found information/records

pertained to another Pauline Mai or to one of the many Pauline Mays. Even when novices

happened upon relevant results, they were unsure of their relevance and, accordingly, they did

not pursue those search results further. Only after I helped them to more wisely use the

provided information in their searches and relevance judgments, did they successfully identify

records for Pauline Mai. Due to my help, all the novices were able to create memorials for

Pauline that included an approximate birth year, some of her children's names, and some of her

residences. Four of the novices also included Pauline's year of death, and another identified all

four of Pauline's children and her grave location. In the end, experienced FHRs were able to

overcome the many obstacles that were built into the research scenario, which allowed them to

produce accurate and, in some cases, rich memorials. The novices, however, needed my help to

do so. This suggests that providing even a small bit of help, interjected into the FHRs actual

research process, can go a long way in enabling novices to produce more accurate family history

content.

When and What to Share Publicly

Many experienced FHRs who participated in this survey and the initial interviews said that as

novices they had shared inaccurate family trees and other family history content. The think-

aloud production outcomes provided further evidence that novices have a propensity for

creating inaccurate family content, as they do not yet possess the knowledge or the skills to

successfully navigate the many pitfalls that exist in the online research environment. The

production of inaccurate family history content would not be so problematic if it was not shared

on the web, but novices' online sharing of such content may be commonplace. For example,

44% of the survey respondents said they had shared family history content they created within the first eight months of researching their families. Slightly over 60% of the experienced FHRs who participated in the think-aloud sessions or initial interviews indicated that they, too, shared content online as novices. Novices appear to be unaware of the impact that sharing inaccurate genealogical data may have within the online FHR community, or its role in the proliferation of inaccurate content.

Experienced FHRs were aware of the need to improve the accuracy of the content they first produced. Many mentioned the "do-over" movement that has gained some traction in the family history community -- a movement that encourages FHRs to redo their family trees based on good research and sourcing practices. Experienced FHRs who were redoing their trees mentioned that they are now hesitant to share poorly researched content until they are more certain of its accuracy. Their earlier family history content was already out on the web, and they were reconciled with that reality. As it is not easy to share smaller portions of their family trees -- the portions they were confident about -- rather than large portions or whole trees, some participants were leery about removing entire family trees from the public view. Given this, they have decided to leave their trees in public view for now. Others told me they had removed trees that needed "work" from public view, as they did not want to "spread lies" (think-aloud participant #4). When asked about what content is worth sharing online, many experienced FHRs indicated that a tree should have at least three sources per ancestor, and some indicated at least five, before sharing that content with the public.

Privacy Settings Literacy

Many FHRs who participated in this study voiced concerns about sharing poorly sourced

information, as they feared it may be copied and reused by other FHRs. Nine of the experienced

FHRs who participated in the think-aloud sessions or initial interviews mentioned this concern,

but were uncertain whether their family trees on Ancestry were private or public. The opaque

nature of the privacy controls within the Ancestry family tree production system, and possibly

other online family history production systems, are facilitating the sharing of inaccurate content

or content that has not yet been fully researched. The fact that many online family history

production systems (Ancestry, Geni.com, Wikitree.com) make user-contributed publicly

viewable by default — the user must select "private" for their content to remain private — may

also be encouraging the sharing of content before it's ready to be consumed by other FHRs.

Understanding the privacy controls that are part of many online family history production

systems is an important literacy, but one that even experienced FHRs may not possess.

Three of the novices had started building trees on Ancestry prior to participating in the study.

When asked about whether their trees were private or public, they were not certain. It seems

likely that many novices, like many of their experienced counterparts, are not knowledgeable

about privacy settings in family history production systems. Family history content is likely being

shared or "published" to the web for other's consumption, without many FHRs, experienced or

novices, being cognizant of their actions.

Analyzing Found Data: Data Triangulation/Alignment Literacy

Experienced FHRs employ many literacies — search tool, document, transcription, and user-

content literacies — to help them make relevance and accuracy judgments about found data,

including data in search results, other transcribed data, and data they glean themselves from

original documents. These many literacies make it possible for FHRs to find relevant data and

documents in a sea of digital information about deceased people, many of whom shared the

same names. Experienced think-aloud, interview, and survey participants frequently referred to

a relevance judgment process that can be best described as "data triangulation or alignment," in

which FHRs compare known data with found data and make relevance and accuracy judgments

based on those comparisons. They use their knowledge of documents and historical context,

document transcriptions, and user-contributed content to aid in determining whether found

data and documents are accurate and relevant to their ancestors. One study participant's

description of his research practices provided a glimpse into the triangulation process:

> I try to find a record that is independent of a previous one that will have the same
> information. Then I can say at that point, ok, the census says this person was born in
> you know 1812 in Ireland, and then I have a birth certificate or a marriage certificate
> that has the same information, then I assume it's basically true. (initial interview
> participant #9)

Another simply described this process as "Double checking, cross checking EVERYTHING...don't

assume you have the right person just because it's the right name" (survey respondent #494).

Experienced FHRs triangulate already-known and newly found data points (e.g. life dates, places

of residence, occupations, immigration dates) about an ancestor, while also comparing known

and found data about family members related to that individual. They decide whether the newly

found data aligns with what is already known about their ancestors. One participant described

this process: "Usually, I try to triangulate, for a lack of a better word, the people. Is there

someone else in the records I can use to verify the information?" (initial interview participant,

#4). A think-aloud participant described this same process of comparing known information

about several ancestors to determine the relevance of newly found information. After

discovering what appeared to be a relevant obituary for a relative (Ella Pfenninger) of Pauline

Mai on the web, he explained the triangulation process:

> Ella was the wife of the late Herman, and she was born in Idaho, which is a link to
> Pauline. She has three deceased siblings, Anna, Oscar and Arthur, and her parents are
> Otto and Pauline Mai. This probably is the right person. New Jersey also is linked to
> Pauline. Now we have triangulation of a few pieces. One is Herman and Ella, in New
> Jersey. I would get this kind of stuff (referring to the information sheet about Pauline
> Mai provided to him) when I was doing this from my relatives -- who was so and so,
> when was he born. I'd interview three different people about the same cousin and get
> three different sources of where that person was born, when they died, so I have to
> kinda piece together. So when I am doing this research, I am looking for multiple
> pieces of information that line up. (think-aloud participant #14).

Another participant described triangulation in a similar fashion:

> Usually, does all the information match up with what you already know about the
> individual? So is the location correct, is the name correct, you know the dates in the
> same realm of possibility or are they close enough or are they exact? In some cases, I
> may have the birth dates, but I may not know the day. Does the rest of the
> information match up? You know, are the parents or the siblings all the same? If there
> are differences, that would always raise a red flag for me, where you might have, say
> there was an individual named Charles – would the wife have the same name, are the
> ages similar? I always try to match up as much information as I possibly can. And if it's
> a good match, then generally it probably is something that will work for you. But I
> would always want to track down the original documents or other documents that
> help support the document that you are able to get to online. (initial interview
> participant #2)

In addition to triangulating known and new data about an individual (and data about

related ancestors), the participant is also alluding to the importance of comparing

transcribed data found online with data found in original sources. The assessment of data

not only involves decisions about whether the newly found data aligns with known data,

but also whether the data is accurate. Transcribed data can incorrectly represent data in

original documents, so experienced FHRs prize original documents. Asone participant

explained:

> With a better appreciation of the records that were available, and how they might usefully be used to cross-validate one another, I began to plan a more structured and coherent method of researching my ancestors. I became less willing to make "leap in the dark" assumptions, and more intent on considering the possibilities, and saying "If that is right, then it ought to have produced this record -- I will go and look to see if such a record exists" (survey respondent #483).

As many original documents are not online and accessing them may require traveling to physical

locations, many FHRs indicated that data/document collection and the data triangulation

process can happen over days, months or even years. Many participants indicated that they had

been collecting data and documents, off and on, about the same group of family members over

the course of several years. Some mentioned having to wait years to visit brick-and-mortar

archives that possessed documents containing elusive information about their ancestors. Others

referred to the temporal nature of their research — they did FHR when they had time, which

meant that for some, there were large gaps in time between research sessions.

Possibly due to the temporal nature of the hobby for many and the fact that most historical

original sources are not online, experienced FHRs who participated in the interviews and think-

alouds seemed to accept that at any given time, they may not possess sufficient data about an

ancestor to make the best relevance inferences. This is often true early in the process of

researching an ancestor, particularly for ancestors about whom little is known. Experienced

FHRs spoke of keeping data that had some plausibility of being correct because it was unclear if

it might be relevant to their ancestors.

This finding suggests that the data triangulation process is not always black and white. Data gathered from the interviews and think-alouds suggest that experienced FHRs may classify and treat found data in different ways based on the quantity and quality of the information which is already known about an ancestor. FHRs appear to categorize data and act upon that categorization in three ways:

- They deem the data implausible or too far out of line with what is known, which, in turn, leads them to discard data.

- They deem the data somewhat plausible -- it doesn't align perfectly with what is known but there are some significant commonalities between the new and previously known data -- which leads them to keep the data in hopes that more data will be found to either "prove" or disprove" it's relevance to the ancestor/family.

- They deem the data plausible, as it lines up neatly with what is known, which leads them to keep the data.

If the data neatly aligns with known data and documents, FHRs have some confidence they have landed upon accurate data and documents, and they tend to treat such data as "ground zero" truth and employ that data to make future relevance inferences. However, since genealogical data doesn't always neatly align and because known data about individuals in one's family can be sparse (making triangulation difficult and less definitive), FHRs appear to collect data that is not perfectly aligned with what is known. They may eventually deem such data irrelevant and discard it, but until they can make relevance decisions based on several data points, they are likely to keep it. In some cases, enough new evidence is found that data that was once thought to be infallible is deemed incorrect and discarded.

These findings suggest that the FHR data and document collecting process, coupled with triangulation, resemble Bate's berry-picking information-seeking model. In this model, the researcher picks different bits of information from here and there using multiple strategies to locate the information and ultimately acquire a sufficient answer to her question. This model suggests that information-seeking is an iterative process in which relevance is an evolving, rather than static, construct. Similarly, FHRs berry-pick data and documents from a myriad of online and offline sources. FHRs employ triangulation, a highly iterative process, in which several sets of data are compared and then assessed for consistency each time new data is found. Other literacies, such as knowledge of documents, transcriptions, and user-contributed content, are leveraged by FHRs to help determine the relevancy and accuracy of the found data, as visualized in Figure 9. Newly found data may change what the FHR deems relevant or accurate in future searches and may lead her to discard previously collected data. Thus, relevance is an evolving construct in the FHR process.
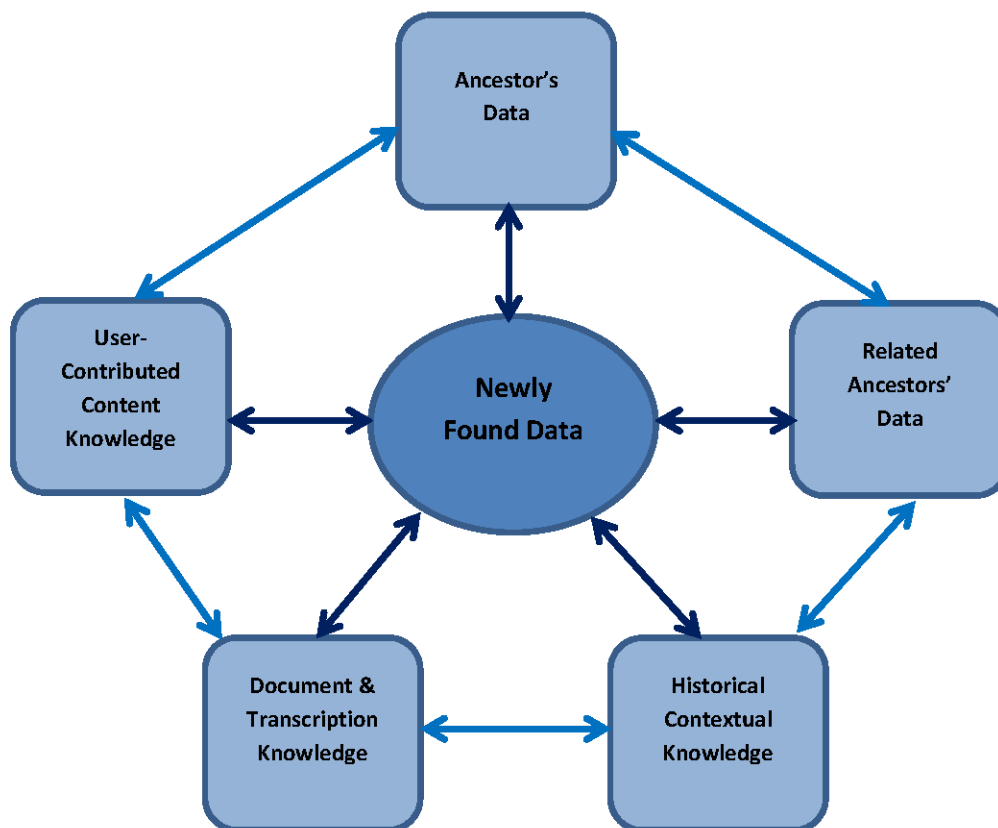
Figure 9: Data and knowledge Utilized in FHR Data Triangulation Process

6.5 Model of FHRs' Information Behaviors

Drawing from the findings from all three collection instruments, a model of **experienced** FHRs'

information behaviors was developed (Figure 10). The model includes both online and offline

searching and information collection, as these activities are closely integrated activities. For

example, FHRs may find documents offline at brick-and-mortar archives, photograph the

documents, and then store the digital copies online. The model also demonstrates the

interconnected nature of FHRs' information-seeking and information-production behaviors in online environments.

Model of FHRs' Information Behaviors

Figure 10: Model of FHRs' Information Behaviors

Searching

As the model visualizes, FHRs search for data and documents that are relevant to their

ancestors. They may search on the web or at physical locations such as brick-and-mortar

archival repositories or graveyards. FHRs may collect data (e.g. life dates) and/or citations for
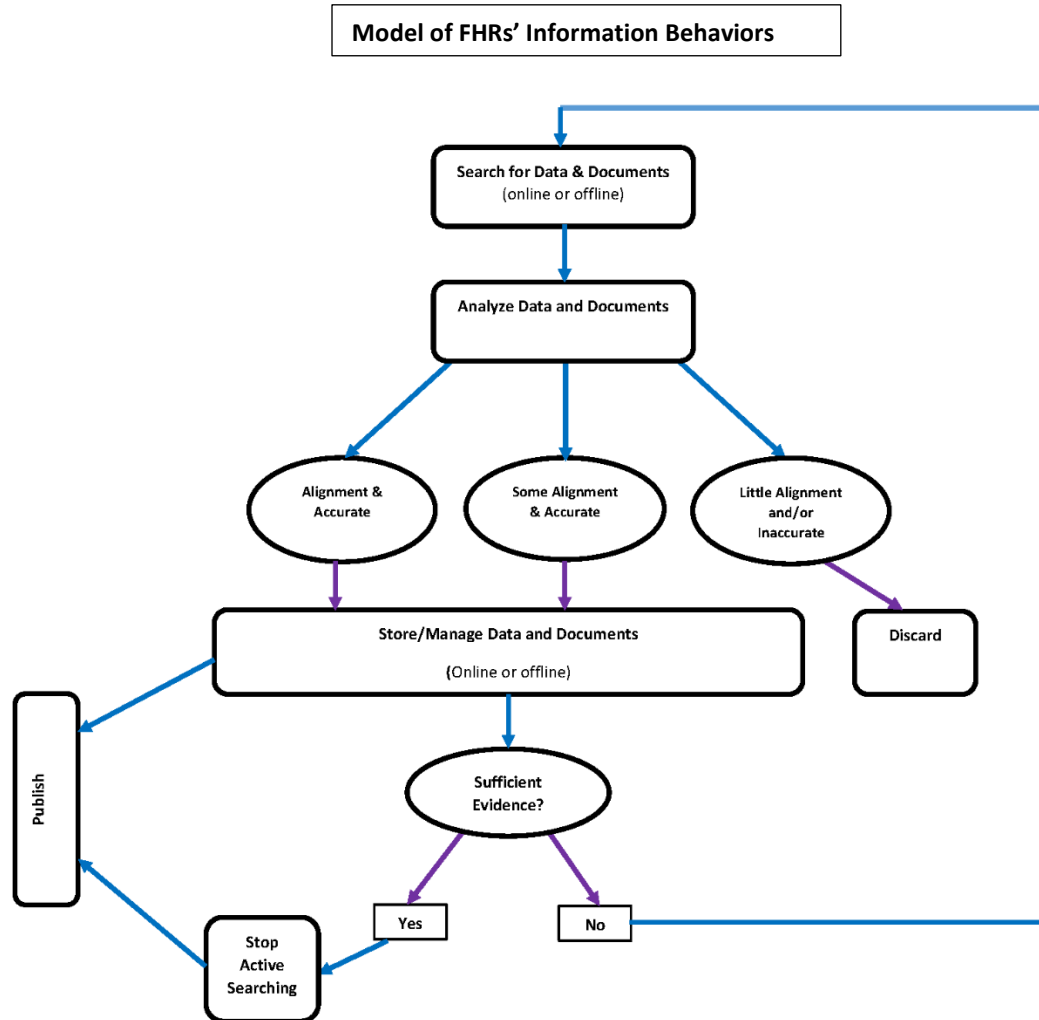
sources, and/or the documents themselves. The documents may be digital or analog.

Analyzing

FHRs analyze indexed data (such as data found in search results), transcribed data, and data

found in original documents. They also analyze the document — who/what created the

document and for what purpose, the document's structure or layout, and the document's

credibility — to help arrive at decisions about whether the data contained in the document is

relevant or aligns with what is known and is **accurate**. If found data and/or documents align

poorly with what is already known about an ancestor (e.g. an immigration document predates

the known arrival date of an ancestor by 30 years) and the FHR suspects that the information is

inaccurate (not trustworthy), FHRs are likely to discard the information. If the found data and/or

documents neatly align with what is already known about the ancestor or some critical data

points align with what is already known and appears to be accurate (trustworthy), FHRs tend to

retain the data and/or documents. As more data is found and more data points are available to

employ in the triangulation process, FHRs become increasing discriminate about accepting new

data and documents that do not align neatly with the "evidence" they have collected.

Storing and Managing Data and Documents

Once data/documents have been found to be potentially relevant, FHRs store and manage the data and documents in a number of ways: online or offline, in family tree software, in spreadsheets, in textual documents, in power point presentations, and/or databases of their own making. As was discovered in this study, such storage methods are not well-suited for online environments. While one might assume that online family tree and other family history production tools would better support FHRs' information-seeking and collection behaviors in web environments, these tools appear to be focused on encouraging the "publishing" of family histories, while neglecting to provide adequate support for the FHR research process.

Resuming Active Searching or Ending Active Searching

Experienced FHRs tend to continue their active searching for more data and documents until a threshold is reached. What that threshold is may vary from FHR to FHR, but for many it is at least three original sources of import that are deemed to contain more accurate data, along with less trustworthy content, such as user-contributed data and transcribed data from original sources. If the threshold is reached, FHRs tend to stop their active search for data and documents pertaining to an ancestor or ancestors and move on to researching other ancestors. An active search for an ancestor can occur over many years given that many sources that may provide important confirmatory evidence are only in analog form and may be scattered in archives across large geographical areas.

The end of an active search for information about a particular ancestor does not mean that additional information may not be found for the ancestor. Experienced FHRs mentioned finding

data about ancestors they had stopped actively pursuing when searching for information about

other ancestors. This suggests that FHRs are ever vigilant in looking for new data for all of their

ancestors, even if the search is not specifically focused on those ancestors.

Publishing

Many FHRs "publish" and share their findings using online tools, such as Ancestry's or Wikitree's

family tree production tools or Find A Grave's memorial production tools. This study found that

such tools also were being used by many FHRs as data collection tools, and as a result, data was

being added to trees before ample evidence had been collected to verify the data's accuracy.

Such data was then being shared online with others. This was partly due to the design of the

production tools. Many online family history production tools default to "public" in terms of

privacy settings, and the option to change the default to "private" for contributed content is

often buried and not prominently placed in the sites' production areas. This means that FHRs,

even those with extensive research experience, are often unwittingly "publishing" their content

to the web for others to view. Most family history production tools also do not allow FHRs to

publish data at a more granular level (i.e. publishing the parts of the family tree that are "well-

sourced") which compels many FHRs who want to share their research with others to

indiscriminately publish their content, well-researched and not yet well-researched, for public

consumption. Additionally, Ancestry's family tree production tool incorporates a searching tool,

which further blurs the line between the often messy research process and the "publishing" of

one's findings to the web for others to see and use.

This publishing behavior, facilitated by the design of FHR websites, is not captured in other

information-seeking models such as Ellis's (1989) or the enhanced version of his model

developed by Meho and Tibbo (2003). Meho and Tibbo include "writing" in their model, but

publishing to the web is not included. Yet in the online world of FHR, publishing family history

content online is commonplace and often closely intertwined with the information-seeking

process. Thus, "publishing," a production activity, is included in the FHR information-seeking

model.

Novices and the Model

Novice FHRs do engage in the majority of activities that are represented in the FHR Information

Behavior Model, with some exceptions. They tend to overlook or only collect a few documents,

and opt instead for collecting basic life data about their ancestors. Novices also tend to be less

discriminate in their analysis of data, which leads them to collect data that is not well-aligned

with known information or may contain inaccuracies (such as transcriptions and user-

contributed content). They also appear to not be aware of the importance of amassing evidence

for their claims by collecting data from several reliable original documents or sources. This may

result in novices unwittingly "publishing" and sharing inaccurate content with others on the

web.

6.6 Exceptions

A few American participants described researching their families abroad in countries, such as

Ireland and Scotland. One participant, however, was a Swiss citizen who had conducted most of

his research in Switzerland. He had mined one of Ancestry's websites and was heavily involved

as a volunteer contributor on Wikitree.com and Find A Grave. His experiences with family history research differed significantly from the experiences of Americans in this study. The starkest difference was that his family had primarily resided in two Swiss cantons for hundreds of years. This meant that there were extensive historic records on his family that could easily be found in the cantons. The participant's family also had family bibles and other histories that went back close to 500 years.

In contrast, most of the Americans who participated in the study belonged to families whose ancestors had emigrated to the U.S. from different countries. Many of those ancestors had resided in many different locales across the U.S. The Americans also seemed to possess far less information about their families' histories. This may mean that for American FHRs the research process is more difficult and mistakes are easier to make. However, additional comparative research is needed to determine whether other Europeans' experiences with FHR are similar to that of the Swiss national who participated in this study or whether they diverge.

CHAPTER 7: DESIGN IMPLICATIONS FOR FAMILY HISTORY WEBSITES

This chapter delineates several design suggestions for a constellation of tools that support FHRs as they conduct research and produce content online. Those tools, which in some cases may need to be stand-alone and in other cases can be integrated, include research and production tools, search tools, and learning support tools.

Some of the suggested tools and enhancements directly address the need to better support novices as they learn the research process. White, Dumais, and Teevan suggest that such tools are needed on other websites as well: It is "worthwhile for search tools to consider how they can better help domain non experts become domain experts over the course of time" (White et al., 2009, p. Section 6.2). It is not only search tools that may need enhancements to better support FHR novices. This chapter suggests that other tools and enhancements to family history websites are needed to support the adoption of research practices that are more likely to produce accurate family history content.

7.1 Situational Learning Support

Ancestry and FamilySearch, as well as many other family history websites, offer didactic learning resources on their sites. A small minority of survey respondents indicated that they perused these resources when they first started engaging in FHR; most indicated they jumped right in and started researching their family without consulting how-to resources. These learning resources may be under-utilized for number of reasons, some of which were revealed in the survey and interviews. Some participants indicated that when they first started out they were

only dabbling in the hobby, trying it out to see if it was an enjoyable activity for them. They wanted to try their hand at it, not commit time to reading or viewing how-to resources, many of which are long-winded and complex. Other participants suggested that they didn't have a good understanding of how difficult FHR can be — that it was a learned craft requiring the acquisition of certain skills and knowledge — and so they initially did not see the need to explore how-to resources. Ancestry commercials are certainly not disabusing novices of such ideas; if anything, the ads suggest that no skills or knowledge are needed to conduct FHR.

In addition to didactic learning resources, message boards abound on family history websites, but as was found in an earlier study, they do not appear to be effective venues for teaching the FHR craft (Willever-Farr & Forte, 2014). Existing learning resources and message boards do not appear to be adequately curbing the production of inaccurate content on family history websites. This suggests that some other form of support for learning the FHR craft — one that may be more amenable or attractive to novices — may be needed on family history websites.

The think-aloud sessions revealed that even a few targeted and situationally based suggestions from a knowledgeable researcher (in this case, myself) could compel novices to change their search terms and trajectories and to more carefully compare found data with known data, leading them to relevant records (and the production of more accurate Find A Grave memorials for Pauline Mai). The key reason why such learning support may be effective is the situational nature of my "suggestions." My suggestions were targeted and specific to the problems the novices were facing. Novices could easily act on those comments without having to find and then read how-to sources that may not provide the needed advice for the specific problems at hand. All three collection tools provided data that suggest that novices just want to get on with doing their research, but may need help initially to be accurate. So why not integrate the

learning support into the research process? This could be accomplished by experienced

volunteers:

1. Using a user-friendly online research tool (which will be discussed below), the novice

   researcher uploads and shares known data about their family with an experienced

   FHR volunteer.

2. Through synchronous technology platforms, an experienced FHR then shadows the

   novice as he searches for relevant data, giving the experienced FHR opportunities, via

   audio or textual chat, to offer helpful comments, particularly if the novice appears to

   be going astray in his search by not adequately triangulating, ignoring known data,

   misinterpreting found documents, etc.

Such a synchronous system would enable experienced FHRs, many of whom are committed to

helping others (W. M. Duff & Johnson, 2003; Veale, 2004; Willever-Farr & Forte, 2014; Yakel,

2004b; Yakel & Torres, 2007), to share their search, document, contextual, and triangulation

knowledge with novices in a targeted and tailored manner so that novices may overcome

specific challenges inherent in FHR. Synchronous systems or tools can be expensive to support

and may require larger bandwidths. To start, it may make sense for a company with deeper

pockets, such as Ancestry, to support such situational learning support on a first-come, first-

served basis — novices or others who want help sign up for it in hour-increments, and the first

ones to sign up receive such support. If this approach leads to the production of more accurate

user-contributed content, Ancestry may be willing to invest more resources in supporting

synchronous learning venues on its site. Another route would be to enlist the Open Source

community to build lower-cost tools to support synchronous learning on non-profit and/or

community-operated sites, such as Wikitree.com. If several family history websites offer

synchronous, situational learning assistance, it may be possible to support a wider online

learning community for FHR, one that is not as reliant on the big, for-profit websites, such as

Ancestry.

7.3 Other Ways to Support Collaboration

Many of the interview and think-aloud participants indicated that they had shared information

on their trees that they were uncertain about and were worried it had been copied by other

FHRs. If family tree production tools were designed to allow users to mark "public" those

ancestors for whom they had adequate evidence and "private" those they were uncertain

about, this problem of sharing dubious information might be avoided. Most of the examined

online family tree tools (with the exception of Geni.com's) do not allow for this level of

granularity in marking specific portions of a tree as private or public. Certainly, tree production

tools could be designed to allow for such gradations in viewing. But what if a FHR wants help

collecting more proof that they have the right person(s) on their tree? Marking these

inadequately "sourced" persons as "private" will not allow others to help. I propose another

option: For contributors struggling to confirm the accuracy of found data, a tagging feature

would allow them to ask others for help in finding confirmatory evidence for specific ancestors.

Such a tagging feature may encourage cooperation among content contributors and increase

awareness of the importance of finding evidence to support the relevancy of found data.

7.4 Family History Production Tools Are Not Enough

Findings from this study suggest that the copying and use of inaccurate content may be

exacerbated by the fact that websites like Ancestry have muddled the research and content

"publishing" processes. The focus of many web-based tools for family history is on publishing

content, whether family trees, memorials, or other types of family history content. There

appears to be far less focus on tools that provide support to the FHR in the research process.

Given the complexities of FHR, it may be best to imagine family history tools as research tools

first, with publishing as a secondary and voluntary function, rather than imagining these tools as

being primarily for content production. Existing family history or genealogical software provide

little guidance on how to produce accurate history and appear to be doing a poor job at

supporting the FHR process. These tools are focused on providing the means to visually display

simple data (and often images) about a person and their relationships to others in the family,

but they provide little support to the researcher as she collects and evaluates data and

documents.

Some attempts have been made to incorporate research functions in production software. For

example, Ancestry has integrated search into its production tool, but the search tool is little

more than a simple search thrown into the tree — not much help to the FHR attempting to find

and evaluate found information. Ancestry has also tried to provide some automated research

help to users in the form of "shaking leaves," or an automated suggestion system, but this

appears to have increased the amount of inaccurate content in the trees (Willever-Farr & Forte,

2014). Shaking leaves appear on a user's tree when Ancestry's search algorithms uncover what

may be a related person. These suggested connections appear to be frequently wrong, and

many novices seem to fall prey to them, populating trees with suggested data without assessing

the relevance to their ancestors (Willever-Farr & Forte, 2014). The incorporation of this

automated research help into the production tool, which by default is creating publicly facing

content, is adding to the inaccuracy problem with user-contributed content.

Ancestry has introduced a separate tool, the "Shoebox," which can be used to collect digital

documents from its website before they are committed to a family tree, but the tool does little

more than warehouse documents — the saved documents can only be browsed in the order

they were saved. No search engine is integrated into Shoebox and there is no way to connect

the documents to specific ancestors. Many of the experienced think-aloud and initial interview

participants cited Shoebox's limitations as a research support tool. As one participant

complained:

> If you delete one thing, it brings you to the first document in your shoebox. Fifteen
> pages you got to click through each at a time to get back to where you were before
> you deleted a document. It's a pain in the neck. And we are saving a lot in Shoebox.

As a research tool, Shoebox appears to poorly support the document-gathering process that

many FHRs engage in. In addition to Ancestry tools, I examined and tried out several web-based

genealogical software tools that were mentioned in the survey, interviews, and think-alouds,

including from Webtrees, GeneoTree, Geneweb, Wikitree.com, and Geni.com. These web-based

genealogical tools also were production-oriented, with little support for the research process.

An Online Research Tool Is Needed

Existing family history production tools and Ancestry's Shoebox tool are not adequately

supporting the FHR process. Better support of the research process may facilitate the

production and sharing of more accurate user-contributed content. Such a research tool should

support data and document collection and then assist the researcher in relating and

triangulating found data. The family tree appears to be an important means of organizing data

and documents specific to individuals and relating individuals to others in the family.

Accordingly, the following design suggestions for a research tool are premised on the idea that

the family tree should be the underlying organizational structure for data, documents, and other

collected information.

Supporting Data and Document Collection

Many study participants indicated that they collect data and digital documents from multiple

websites (and other sources). As one participant remarked: "I jump all over the Internet,

because there are so many websites out there now to get information from" (initial interview

participant #9). Where and how to store collected data and images, as well as documenting

where the information came from, can be challenging given the current online family history

digital tools available for such collecting. An online research tool that enables FHRs to collect

URLs, research notes pertaining to the websites, genealogical data, and other content

(webpages and images of original documents) from a multitude of websites would support them

in the collecting process. Possibly, a research tool with some of the same functions that are

present in the browser-based application Zotero — the ability to easily collect URLs and

webpages and other types of online content (e.g. PDFs or image files) from multiple websites —

combined with features that would uniquely support FHR, such as the ability to connect URLs

and online content to specific locations in a family tree, would better support FHRs' collecting

behaviors and organizational needs. Such an online tool would allow users to easily capture and

then organize website URLs and online content and relate this information/content to specific

individuals in a family tree. In addition, the ability to capture research notes about individual

websites would be helpful. With such a feature, FHRs could include notes about websites that

did not contain information on specific ancestors, so as to avoid searching for such information

again on those sites.

It is also the case that many FHRs visit brick-and-mortar repositories and collect analog

materials, as well as offline digital images of analog materials. The research tool should also

support collecting information about these locations/repositories and content collected from

those physical sites and the ability to upload and relate offline digital content. An integrated

annotation tool for the collected documents would also allow researchers to add notes or

additional information directly to specific areas of the document. This tool would come in

handy, for example, if researchers wanted to include notes regarding abbreviations used in

certain types of documents, or notes about how to interpret information in different columns or

rows in a historic record.

In addition to collecting documents and information about where a source was found, FHRs also

collect data, but this can be challenging when the FHR is confronted with gathering and

grouping data from various sources for more than one individual with the same name. This

gathering of data from various sources for people with the same name is often needed, as the

FHR is initially uncertain which of the named individuals is the ancestor. Only after finding

additional information or evidence can the FHR make such a determination. Until that evidence

is uncovered, the FHR may want to keep data on several people bearing the same name. Many

of the think-aloud participants experienced this challenge. From their comments, it was evident

that they were finding it challenging to keep all of key relevant data points in mind as they

searched for new information. On a provided note pad, some wrote down found data, but it

became burdensome for them to write down all the key data for multiple Mais, many of whom

were not Pauline's actual ancestors. An online research tool that better supports this process of

grouping and then capturing data and documents pertaining to all of the potential ancestors

(who share the same name) in an organized fashion is needed. Family tree production tools do

not support this activity well, as only one set of data can be added for one place in the family

tree. In the same spot on a family tree, a FHR cannot collect, for example, a set of data for John

Smith 1, a set of data for John Smith 2, and a set of data for John Smith 3 (and associated

documents) and then retain this information in that spot in the family tree until sufficient

evidence has been found to determine which of the three John Smiths is the correct one. With

a production tool that is poised to publish content online, such a function may not be desirable;

however, with a tool specifically for research, such a function could be included with less

concerns about "incorrect" information being published.

Automated Features of Online Research Tools

In addition to better supporting the capture and organization of URLs/location information,

data, and documents, an online research tool could feature automated functions that support

the creation of accurate content. Many participants in this study and the second pilot study

(Willever-Farr & Forte, 2014) mentioned user-contributed trees that had glaring errors, such as

adult men married to very young children or female spouses who had died before their

supposed spouse was born. It is quite likely that automated tools could be built that would

search user-contributed content, before it is published, for incongruent data pairs, or data pairs that defy typical social practices in the U.S. (such as an adult man marrying a 3-year-old or a dead woman). Parameters could be created that would help an algorithm automatically identify such incongruences, and then alert the researcher. The researcher could choose to heed or ignore the alert, and could turn off this feature as she becomes more experienced.

Another automated feature that may help facilitate the production of accurate content is to include helpful pop-ups or other methods of signaling (such as automatic coloring coding), which provide guidance to novices. For example, some online family tree construction tools allow the user to upload and relate original sources (either digital images of the source material or original source citation information) to specific individuals in a tree. To encourage the researcher to provide more sources or evidence that the person added to the tree should be there (because they are truly related to others in the tree), pop-ups could be used. If the user hovers over an individual (included in his tree) who is not well-sourced or has fewer than three related original sources (such a birth certificate, a military record, an obituary, etc.), a pop-up could be automatically generated indicating that more evidence is needed for this individual. Users could opt to turn off such help features as they became more experienced.

7.5 Design of Related Online Tools

Search Tools

The outcomes of the think-aloud sessions demonstrated that adding multiple data points in a search is critical in surfacing relevant results. Given that many of our ancestors shared the same names, providing more data in a search can help push relevant results to higher placements in

search results. Simple search forms, such as the simple search on Ancestry (a single search box), do not prompt the researcher to add additional data points. Novices who participated in the think-aloud sessions used simple searches and did a poor job at identifying what data points to include along with a first name and surname in these searches. By providing search boxes for specific types of data, as well as other types of filtering functions, advanced search forms on websites such as Ancestry and FamilySearch give the novice a sense of what data should be included in a search, whereas the simple search does not. Experienced FHRs who participated in the think-aloud session adeptly used the advanced search to surface relevant results and find additional sources related to Pauline Mai and some of her family members. This suggests that simple search forms may not be a useful feature for family history websites. Given that advanced search forms can be complex, novices may benefit from the integration of pop-ups or other types of help into the advanced search forms.

Other helpful additions could be made to the search results display on sites that have both original documents and user-contributed family history content. This study highlighted the problems with researchers extracting data from poorly sourced user-contributed family trees and then reusing that incorrect data in their own family trees. Search results displays should do a better job at differentiating user-contributed content from original source material. For example, Ancestry's search result display does this poorly, as many of the novices who participated in the think-aloud sessions appeared to be unaware that some of their search results included user-contributed content. Suspecting this from comments they made during the think-alouds, I decided to specifically ask them whether they knew that some of the search results they had pursued were user-contributed trees. Five of the novices admitted that they had no idea.

Tools for Supporting Document Literacy

This study has found that "sourcing" is at the heart of creating accurate family histories.

Sourcing requires knowledge of the kinds of original sources that contain biographical data and

how to interpret those sources. As a starting place, a simple but important improvement to

family history websites that contain digitized original sources or document, is to make the links

to documents prominent in search results and to include language in those results that

encourages the researcher to examine the documents, not only the transcriptions.

Helping the researcher make sense of the data in the original document is the next step. From

the survey results pertaining to how respondents learned to conduct FHR, it is likely that didactic

learning resources to increase artifactual literacy would not be utilized by many novice FHRs.

Another approach is to directly annotate groups of documents (for example census documents

from a certain year) with helpful information to assist the researcher in making sense of the

data contained in the document. Ancestry has added such annotations for a few types of

original documents on its website. This kind of document literacy support should be extended to

other documents and other websites. This work may not have been done because of the

resources involved, but if money and manpower are to be spent, this is a critical area of need.

While crowdsourcing can lead to inaccurate transcriptions, crowdsourcing document

annotations may be a starting place. A more rigorous but efficient review process for

crowdsourced document annotations could possibly be implemented, such as the peer review

process suggested by Hansen et al. (Hansen et al., 2013).

Assessment Tools for User-Contributed Content

Family history websites offer little help to users as they attempt to assess the accuracy of user-

contributed content. These sites offer no easy way for users to know how green the FHR is who

contributed the content. Additionally, they offer no means for users to rate user-contributed content. The addition of a robust reputation system on family history websites would help users of family history websites to understand the expertise level of those whose contributions they wish to use or link to.

Online Space for Stories

This study focused on factual family history content. However, it's important to note that not all family stories are "provable" through original sources, and such stories add color and interest to family histories. A narrative space to share such stories and connect them to more factually oriented displays, such as family trees or grave memorials, is needed. One study participant shared a particularly compelling story that would appear to be important to capture in a family history. The participant had been conducting FHR since grammar school and considered herself a skilled hobbyist. For generations, her husband's family considered themselves among the American Irish, living in predominately Irish communities for most of the 20th century. Through careful research and sourcing, the participant discovered that her husband's family hailed from France not Ireland. They were Catholics, but not Irish Catholics. However, for this family, their supposed Irish ancestry was a part of the families' identity for years and influenced what communities they lived in and who they married. This suggests that even less-than-factual stories shared within a family are important to document and share. Ancestry has added such a "story" space in its family tree tool, but many other online family history production tools have not.

CHAPTER 8: CONTRIBUTIONS AND DISCUSSION

The findings from my pilot studies and dissertation study contribute new knowledge about the characteristics of online family history research and a popular online production environment and how that environment influences the information behaviors of FHRs. In addition, the study characterized the literacies needed to research and produce family history content in an increasingly digital research and production environment, in which the production of public resources becomes a feature of the novice FHR experience. A summary of these contributions and how they extend existing literature will be discussed in the following sections.

8.1 A Better Understanding of the Family History Digital Landscape

 Understanding FHRs' information behaviors in an increasingly digital research environment requires close study of that environment — an environment dominated by non-archival repository, data-rich sites, such as Ancestry, FamilySearch, and Findagrave. Such commercial and non-profit enterprises are playing progressively important roles in providing direct access to historic materials and data without FHRs needing to interact with archivists and archival information systems. One published study (Friday, 2014) explored FHRs' online research needs in the context of local records repositories, but no research to date has studied the influence of non-repository family history websites on the online information behaviors of FHRs. My dissertation study, along with the pilot studies, addresses this gap and identifies the salient features of the online family history environment that are shaping FHRs' information behaviors. In the preceding chapters, I described important features of that environment including:

1) Non-repository, family history websites have become influential intermediaries between FHRs and a sea of archival data and documents, providing sophisticated online search tools for mining online historic data and user-contributed family histories.

2) Non-repository websites have lowered the bar for entry into the hobby, creating an environment where novices, in great numbers, are researching their families and producing and sharing family history content online.

3) Easy access to a myriad of online production tools, crowd-sourcing platforms, and the norm of reuse within the FHR community has facilitated the proliferation of user-contributed family history content on the web.

4) The design of popular FHR websites has blurred the lines between the research process and the creation of public-facing family history content, often leading to the online publishing of inaccurate family histories.

5) Lastly, the design of online family history production tools is limiting the ability of FHRs to collaborate on content production, and in turn, places the onus for accurate content production on the individual, rather than supporting collaboration in content production between more experienced and less experienced FHRs.

In addition, this dissertation describes how these realities of the online world of family history research and production are influencing the information behaviors of FHRs and, in turn, are leading FHRs to develop new literacies.

Non-repository, family history websites provide access to hordes of online historic data and documents, sophisticated online search tools, easy-to-use family history production tools, and the means to "publish" and share family histories on the web. Millions of American FHRs are using these websites to conduct research and publish their family histories. Previous studies

have not closely examined FHRs' interactions with the information and publishing systems on these family history websites, and how these systems are impacting FHRs' research and production behaviors. No studies compared novices vs. experienced FHRs' interactions with those web-based systems.

8.2 Online Information Behaviors of Novice vs. Experienced FHRs

Many seasoned FHRs who participated in this study asserted that it was much more difficult, and required serious commitment and learning, to conduct family history research before the emergence of websites such as Ancestry and FamilySearch. Before the emergence of online forums and message boards, many FHRs learned about the craft by joining genealogical societies, which gave them the opportunity to learn from more seasoned practitioners in face-to-face settings. Such face-to-face gatherings provided a space for collaborative learning and problem-solving (Yakel, 2004b; Yakel & Torres, 2007). Family history websites have lowered the bar for entry into the family history hobby by providing access to an abundance of online historical data and documents, as well as easy-to-use family history production tools. However, these websites have not provided digital platforms that effectively support collaborative learning and problem-solving between more and less experienced FHRs, leaving much of the onus for learning and creating accurate family history content on the individual (Willever-Farr & Forte, 2014; Willever-Farr et al., 2012). This study found that novices tend to **not** consult didactic, educational resources on many family history websites, before researching, producing and sharing family history content online. This has translated into a surfeit of inexperienced FHRs researching and sharing their family history content, often of questionable accuracy, online. This finding -- lots of inexperienced FHRs participating in the hobby in a public-facing

manner online — is important for three reasons: 1) it has implications for how online family history systems should be designed; 2) the presence of novices and the content they produce on family history websites has made experienced FHRs more skeptical of user-contributed content and compelled them to develop ways to assess the accuracy of that content; and lastly, 3) user-contributed content information on these sites may influence our understandings of the past, and as such its accuracy matters.

Developing online systems that better support novices as they learn the family history craft requires close study of novice research behaviors. This study found that experienced FHRs are skilled at locating data and documents, triangulating data to determine the relevance of found data/documents, using their knowledge of documentary forms and recordkeeping processes to determine the accuracy of data found in original sources, and producing well-sourced family histories (and knowing when to "publish" them). In sharp contrast, this study found that novices lacked or had limited knowledge of:

- Online search tools for family history research

- Data triangulation

- Assessing the accuracy and relevance of found data

- Assessing the accuracy of user-contributed content (both family history content and transcribed data employed by search engines)

- Historic document types

- Sourcing/citing

- Production and publishing systems

In general, novices were less skeptical than experienced FHRs, accepting the accuracy and relevancy of found data even when not aligned with what was known about their ancestors. This finding complements the results of other studies on the search behaviors of non-domain experts (e.g. Bhavnani, 2002; Jenkins et al., 2003). FHR novices tended to embrace non-relevant data that appeared in the top results of their searches. This behavior was also found in previous studies of non-domain experts (Bar-Ilan et al., 2009; Huang & Wang, 2004; Xu & Wang, 2008). However, this study uncovered an information activity that had not been identified in previous studies. Once incorrect data was accepted by novice FHRs, it had a snowballing effect on their search and relevance judgements: Accepting incorrect data led them to examine other irrelevant records and then accept those records, and so on. Their persistence in following the wrong search paths made them more committed to those paths and to the accuracy of their search results. Based on findings presented in earlier chapters, I hypothesize that this negative feedback loop would be found in other online information search and production environments where increased autonomy of novices and decreased influence of gatekeepers is accompanied by a lack of opportunities for community-building and information-quality practices to develop.

Novices' inability to "triangulate," and a corresponding tendency to accept data that was not well aligned with what was known, was further exacerbated by a lack of knowledge of family history search tools. In their seminal study of FHRs, Duff and Johnson (2003) suggested that archival information systems should be designed to provide name, place, life event, and document-type access points. While many archival repositories appear to have not responded to this call, websites including Ancestry and FamilySearch have developed complex search engines that incorporate these access points and many more. This study found that such

complex search tools are needed to surface relevant results given the vast amount of online historic data. Such search tools are critical in enabling FHRs to locate records that are truly about their ancestors, rather than other deceased individuals who shared the same names. This study further found that specialized knowledge is needed to successfully employ these search tools, and that novices appear to lack the knowledge to exploit the search tools to their best advantage, leading them to often gather irrelevant data. Experienced FHRs were able to successfully utilize these search tools because they possessed knowledge of the mechanics of the search tools (i.e. how to conduct a fuzzy date search vs. a specific date search) as well as the data triangulation process. Given the importance of these search tools in aiding the FHR in unearthing relevant records, more help or search tips in the form of pop-ups or similar "push" methods of immediate help delivery could be incorporated into the search tool to assist those without adequate knowledge of the tool and triangulation process. Search tools cannot teach novices everything about the FHR triangulation process but they can provide hints on the how to use different types of data (i.e. ancestor's name, life date, and child's name) as a means of surfacing more relevant records. Help on the mechanics of the search tool could also be incorporated, as the search forms have many search options, some of which are not commonly found in web-based search tools.

In addition, this study found that other digital tools were critical for producing accurate family histories through online family history research. For example, I observed the impact of transcribed data on the online family history search process and the resulting need to master the use of image viewers. Historical data is extracted from original sources (much of which has been crowdsourced), then indexed so that search tools or engines can exploit such data in search. Transcribed data is the only way for search engines to provide access to digitized original

sources. This study found that errors abound in transcriptions and that experienced FHRs, unlike

novices, knew to peruse linked images of historic sources to gather data, rather than relying

exclusively on transcribed data found in the search results and image "records." To extract data

from digitized records, FHRs needed to learn how to use image viewers (computer programs

that display stored graphical images), which can vary greatly from website to website. As a

means to overcome errors in transcribed and indexed data, many experienced FHRs also

developed a browsing approach in which they browsed groupings of documents and data

clusters within documents to gather relevant data that may have not appeared in search results

due to transcription errors. Many image viewers do a poor job at supporting this browsing

behavior, providing little help to the FHR as she attempts to navigate through pages of a

digitized document and gather data from those pages.

8.3 Web-based Publishing and Assessment of User-Contributed Content Literacies

One unique influence of the web on family history research is the prevalence of web-based

"publishing" platforms that enable FHRs to easily build and share family history content online.

Millions of FHRs are publishing family history content online ("Ancestry.com LLC Reports Fourth

Quarter and Full Year 2015 Financial," 2016, "WikiTree: The Free Family Tree," 2017; Arrington,

2007). For increasing numbers of FHRs, access to online production tools has made web-based

publishing an integral part of the family history research process. The design of many family

history websites (and their production tools) has created an environment in which research,

production, and web-based publishing of user-created content have become closely related,

often integrated processes. This integration is partly facilitated by websites such as

Ancestry.com and Wikitree that provide access to easy-to-use, online family history production

tools. Such sites do not provide separate research tools that support FHRs as they collect data, documents, and other important research information, and then determine the relevance of found data. Rather these websites provide only production tools, many of which default to a "public" view of the user-created content. For many FHRs, this reality makes publishing an almost seamless part of the family history research process in the web environment -- the online production tools are the research tools and vice a versa. Online production and publishing are inextricably tied to the research process.

Thus, many FHRs are "publishing" family history content on websites, such as Ancestry, Wikitree, Genicom, and Find A Grave, that translates into vast quantities of user-contributed family history content being accessible to other FHRs. This means that FHRs encounter this content frequently when conducting research online. Many search engines, both those tailored for family history research and more generic ones, retrieve user-contributed family history content and place it on equal billing with online original source material. Such search tools do not discriminate between the two different types of content – if it appears relevant, it is included in the search results. Due to the sheer quantity of user-contributed family history content and the likelihood that it will be retrieved by search engines/tools, FHRs need to learn how to manage and assess such content. In some cases, such content provides much-needed data that is not accessible online, data that may have been unearthed offline at brick-and-mortar repositories. But often, the content is inaccurate. This study found that an understanding of "sourcing" (the inclusion of citations for original sources or linked images of original sources in family histories) was critical in FHRs' ability to determine the accuracy of user-contributed content, a skill more experienced FHRs had mastered. Sourcing has been alluded to in previous studies, but never fully explored. This study fills that gap.

These findings related to novice search and relevance judgments, along with the statements of the majority of FHRs who participated in this study, suggest that novices are likely to produce inaccurate content. Additionally, this study found that novices were quick to share or "publish" family history content online well before they had learned how to be accurate in their research. Novices appeared to be unaware of how easy it is to produce inaccurate family histories, so they shared what they thought was accurate content. Only later did many of these FHRs realize that only through a rigorous research process can accurate content be produced. As mentioned earlier, the presence of inaccurate, user-contributed family history content online has pushed more experienced FHRs to become adept at assessing the accuracy of user-contributed content, by checking for key information that will validate the accuracy of the content, such as original source images and/or citations to original source material. Assessing the accuracy of user-contributed content, given the amount of it on the web, has become an important component in online family history research. This is a novel finding that extends our understanding of the skills needed in online family history research.

8.4 Comprehensive Model of FHRs' Information Behaviors

No existing study has provided a comprehensive model of FHRs' information behaviors. This study filled that gap by presenting an information model that captured both online and offline information activities of FHRs. Further, the model introduces a new type of information behavior – online publishing—that should be considered in future studies of web-based information behaviors. This publishing behavior, facilitated by web-based participatory platforms, is not captured in other information-seeking models such as Ellis's (1989) or the enhanced version of his model developed by Meho and Tibbo (2003). Meho and Tibbo include

"writing" in their model, but publishing to the web is not included. Given that the web has

revolutionized "publishing"— making it possible for anyone with a computer and access to the

internet to publish their own content online — it is plausible that the information behaviors of

other groups are also being influenced by this feature of the web environment.

8.5 Literacies Framework

To create accurate family history content in this environment — an environment characterized

by access to copious amounts of online historic data and documents via sophisticated search

tools, as well as easy access to online tools for producing and publishing content — requires

several literacies or specialized skills and knowledge. Previous studies identified FHRs'

information activities primarily in the context of brick-and-mortar archives and their archival

information systems (e.g. Darby & Clough, 2013b; Duff & Johnson, 2003; Friday, 2014; Yakel,

2004b; Yakel & Torres, 2007). Some of these information activities were not fully fleshed out.

This study endeavored to extend our knowledge of those information activities through the

practical lens of web-based information systems and search tools, and through the conceptual

lens of literacy and accuracy. In particular, this study extended our understanding of already

identified knowledge and skills associated with family history research – artifactual knowledge,

broader historical knowledge, and the collecting of data and documents – and identified and

detailed additional literacies needed for family history research and production, including web-

based searching knowledge, data triangulation knowledge, sourcing knowledge, transcription

knowledge, user-content knowledge, and publishing knowledge -- that are needed to create

accurate content in a web-based research and production environment. It is the first

comprehensive study of the literacies involved in the family history hobby in an increasingly web-based research and production environment.

The constellation of literacies identified in this study suggest that experienced FHRs possess knowledge that has been traditionally associated with two professional groups: archivists and historians. Much of the research knowledge employed by historians — collecting evidence from original sources, analyzing/interpreting those sources, placing their findings in historical context, and creating a narrative from those interpretations (Furay & Salevouris, 1988; Hockett, 1955; Wineburg, 1991, 1998; Wineburg & Fournier, 1994) — are activities that experienced FHRs engage in. Additionally, Duff, Yakel and Tibbo (2013) proposed a model of archivist reference knowledge that shares commonalities with the knowledge needed to be accurate in family history research. Their model is divided into three main knowledge areas — interaction, research, and collection. The model was a good starting place for building a model of family history research knowledge, but was modified to capture production/publishing knowledge and to emphasize the importance of digital skills and knowledge. The resulting model (figure 11) has four main quadrants—research knowledge, production knowledge, digital knowledge, and interaction knowledge.
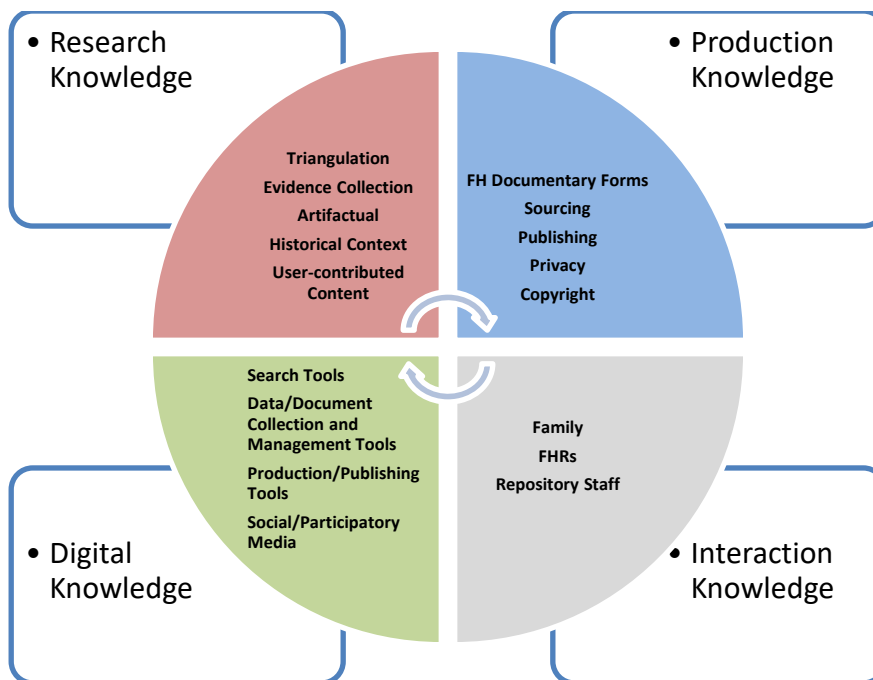
Figure 11: Family History Research and Production Literacies

The Duff et al. model included engagement with information systems in "interaction" knowledge, along with other types of interaction knowledge. Given the ubiquitous use of digital tools for almost every aspect of family history research and production in the digital age, the proposed family history and production model surfaces digital knowledge as a literacy in its own right to highlight its importance.

Research Knowledge

Many skills and knowledge are needed to build accurate family histories. First, FHRs need to

know how to collect evidence about their ancestors and from what sources to collect that

information. This requires an understanding of primary or original sources — where to find the

sources and how to interpret them. Many FHRs who participated in this study used their

knowledge of broader history, geography, and recordkeeping and legal processes to make sense

of original sources and to determine their accuracy. They knew that transcribed data and user-

contributed family histories were often faulty and, as result, relied on original sources whenever

possible. Experienced FHRs employed a data-triangulation process to determine whether found

data and documents were relevant to their ancestors. This study is the first to extensively

document how that process is employed in search.

Production Knowledge

Constructing family histories, such as family trees and narrative histories, has long been a part of

the family history hobby. However, the web has made it easier to produce family history

content and to share that content with others. This study found that "publishing" histories has

become so integrated into the research process on sites such as Ancestry — production and

publishing or sharing family trees are part and parcel of the same the process — that production

literacies are a necessity in the online world of family history research. Thus, the literacies model

includes a group of production-oriented literacies, including knowledge of family history

documentary forms (e.g. family trees, narratives, memorials); knowledge of the importance of

providing sources, either by attaching images of original sources or providing original source

citations, in family histories; knowledge of when to publicly share content and an corresponding

knowledge of privacy settings on production sites; and lastly, knowledge of copyright. While the data was not presented, this study captured limited data on copyright and family history "publishing" on the web. The use of copyrighted material in online family histories appears to be a growing issue and source of conflict on some family history websites. A basic understanding of the risks associated with including images of copyrighted material in user-contributed family histories is an important area of knowledge for FHRs engaging in online publishing. However, more study is needed to fully understand what this literacy entails.

Digital Knowledge

In the increasingly digitally based, even web-based, landscape of family history research and production, digital skills have become critical. Understanding how to effectively use a constellation of digital tools --from search tools to production tools -- is becoming an essential skill set. Search tools on websites such as Ancestry and FamilySearch are complex. Employing these search tools to surface truly relevant data requires the user to input several data points and modify search parameters based on their confidence about known data (e.g. when to conduct a fuzzy date search vs. an exact date match). This takes some knowledge of how search engines operate (even experienced FHRs could benefit from knowing more about this) and how to interpret the search results. Understanding that search engines use transcribed data to retrieve digitized sources, and knowing that transcribed data is served up in search results, is important given that such data can be erroneous. Experienced FHRs knew to be skeptical of that data – not take it as "truth." Lacking this knowledge, many novices included such data in their family histories without further vetting.

Having knowledge of how to use family history or genealogy software to collect data and documents and produce family history content (e.g. family trees) is also important, whether the software is online or offline. As this study found, most web-based systems are primarily production software, providing little support to FHRs as they collect, manage and assess data and documents. Regardless of the quality of this software, its use is ubiquitous, and FHRs need to possess specialized knowledge to employ such software successfully.

For those who are building family histories on the web, an understanding of privacy settings is essential. Without this knowledge, FHRs may inadvertently share their content well before it is ready for "prime time." Even those who choose family history software that is not web-based often upload their data to online systems. Thus, sharing or publishing family history content to the web is an increasingly important digital skill for FHRs.

Interaction Knowledge

Many FHRs gather information and/or seek assistance from family members, fellow FHRs, and information professionals. Before the web, experienced FHRs interacted with one another and archival repository staff to learn the craft and locate relevant material (W. M. Duff & Johnson, 2003; Yakel, 2004b). The web has made it easier for FHRs to interact with other FHRs on social media, message boards and other participatory platforms. Yet it appears that these participatory platforms may be doing a poor job of supporting the kind of knowledge transfer that happened in face-to-face interactions. It is unclear how the massive data stores on websites such as Ancestry are impacting FHRs' interactions with repository staff — whether it is decreasing such interactions and/or changing how archivists handle reference services for onsite

or offsite FHRs. Nonetheless, this study found that more experienced FHRs are still interacting with archivists, both online and offline. Communicating with information professionals continues to be an important part of family history research.

The web has provided avenues for individuals to ask fellow FHRs to help them find data and even conduct research work that they themselves are unable or unwilling to do. One venue for such requests is the website Random Acts of Genealogical Kindness (https://www.raogk.org/). In some cases, the web has provided a platform for FHRs to work collaboratively on building content (e.g. famous individuals' memorials on Findagrave). The web also makes it easy to reuse user-contributed content, and many FHRs contend that there should be community protocols surrounding reuse (Willever-Farr & Forte, 2014, p. 0). This suggests that there may be social protocols that FHRs may need to understand in order to interact successfully in online FHR communities and production spaces.

Implications of the Literacy Framework

This literacy framework will be useful in designing family history websites and research and production tools that fully support FHRs' many information activities. It also may prove helpful to archivists as they provide reference services to one of the largest archival user groups, family history researchers. Understanding that family history research is also about history production, and that "good" family histories include primary source materials and/or citations, may help shape the assistance that archivists provide to FHRs, either online or offline.

8.6 Design Implications for Family History Information and Publishing Systems

Existing studies that suggest system design changes to meet the needs of FHRs are almost

exclusively focused on archival systems (W. M. Duff & Johnson, 2003). Most of these archival

systems differ from those on sites like Ancestry and Family Search, which are designed with

FHRs in mind. It is not surprising, then, that websites such as Ancestry and FamilySearch

dominate the family history research landscape. This study explored the FHRs' use of these

websites in the research and production process. Given its unique focus, the study presented

novel design suggestions for web-based family history information and content production

systems.

Instructional Help Embedded in Systems and a True Research Support Tool

Given the number of inexperienced FHRs conducting research online and their low use of more

didactic instruction resources (at least initially), it makes sense to embed instructional help in

the search and production tools, rather than placing instructional help outside of these systems.

Additionally, a digital tool that better supports the family history research process is much

needed. Such an online tool would allow users to easily capture and then organize website URLs

and digital content (as well as offline analog materials in the FHR's possession) and relate this

information/content to specific individuals in a family tree. In addition, the ability to capture

research notes about individual websites and brick-and-mortar repositories would be helpful.

Annotation tools for collected images would also be a useful feature. The tool should also

provide helpful hints on how to be accurate in the research process.

Tools to Support Assessing Accuracy of Content

Novices, including dabblers who may not have the inclination to master all the literacies

associated with family history research, are likely to create less-accurate family history content.

Given the large numbers of novices who may be producing and publishing content on the web,

systems that help users identify the experience level of contributors and the quality of their

contributed content would provide FHRs with some help in assessing the accuracy of user-

contributed content. Systems that indicate the experience level of contributors (provided by the

contributor or determined by set criteria) and that automatically mark data lacking sufficient

supporting evidence in trees and other FH content would support FHRs as they assess the

accuracy of such content.

Collaborative Tools

Existing family history production websites provide few avenues for collaboration. This study

asserts that some form of online synchronous system is needed that would enable experienced

FHRs, many of whom are committed to helping others (W. M. Duff & Johnson, 2003; Veale,

2004; Willever-Farr & Forte, 2014; Yakel, 2004b; Yakel & Torres, 2007), to share their search,

document, contextual, and triangulation knowledge with novices in a targeted and tailored

manner so that novices may overcome specific challenges inherent in FHR.

Including additional tagging features in production systems may also support collaboration. For

instance, to help any contributor struggling with confirming the accuracy of found data, a

tagging feature would allow them to ask others for help in finding confirmatory evidence for

specific ancestors. Such a tagging feature may encourage cooperation among content

contributors and increase awareness about the importance of finding confirmatory evidence to

support the relevancy of found data.

CHAPTER 9: FUTURE RESEARCH GOALS AND CONCLUSION

9.1 Future Research Goals

This study provided an in-depth understanding of the research practices of novice and experienced FHRs and how these practices are impacting the creation and sharing of inaccurate family history content on the web. As an exploratory study, it also provided a framework for future research. While not explored fully in this study, it appears that the inaccuracy problem with user-contributed content may be coloring reuse norms and attitudes pertaining to online user-contributed content. Further data collection and analysis is needed to understand the impact that inaccurate content is having on reuse norms and practices.

While not related to the accuracy of family history content, some study participants mentioned the need to understand copyright laws when sharing images and other family history-related content online. Knowledge of copyright laws may be an additional literacy needed in the context of online family history production. Additionally, some participants needed assistive technologies to help them find and utilize the many information resources available online. Given that many FHRs are older and heavily using the web for their research pursuits, they are an apt group to study to determine how assistive technology can support adult researchers.

This study also highlights the need for more research on how non-academic users make relevance inferences in the web environment. More-targeted studies that further examine relevance inferences of FHRs could help fill this research gap. Such studies would help researchers determine the kinds of improvements that might be made to search tools and the presentation of search results to better support FHRs in their attempts to assess the relevance

of historic data and documents. For example, a research study could examine whether the inclusion of thumbnail images of original sources in the search results might better support users in making relevance judgments.

9.2 Conclusion

Web 2.0 technologies lower barriers to producing and sharing content and support large-scale collaboration. This has facilitated the production of a collaboratively built encyclopedia that rivals traditional encyclopedias (Chesney, 2006; Giles, 2005; Rajagopalan et al., 2010; Reavley et al., 2012). In the realm of family history, Web 2.0 has facilitated the creation of crowd-sourced information resources, such as Find A Grave and Ancestry's family trees, that are used by millions of FHRs. However, the ease with which family history content can be produced and shared on the web has also created problems. Online family history websites have so lowered the bar to entry into the hobby that anyone, no matter how experienced or inexperienced they are at FHR, can share content with others on the web. This has led to the production and sharing of inaccurate content. Given that sharing and reuse of genealogical information is a hallmark of the FHR community (W. M. Duff & Johnson, 2003; Veale, 2004; Willever-Farr & Forte, 2014; Willever-Farr et al., 2012; Yakel, 2004b), the reuse of such inaccurate content has led to the proliferation of erroneous content on many family history websites.

As experienced FHRs have made clear in this study, creating accurate family history content requires many competencies and literacies, including online search tool literacy, data and document collection and management literacy, interpretive document literacy, transcription literacy, user-contributed content literacy, production and sharing literacies, and data

triangulation literacy. Yet this study found that novices tend to jump right into conducting FHR

online without any preparation, such as reading how-to resources. Findings from this study

suggest that novices do not possesses the literacies to create accurate family history. That lack

of knowledge would not be as problematic if novices were not sharing their family history

content. However, this study found that many novices begin sharing the content they created

soon after they start researching their families, partly because many online production privacy

settings default to a public setting. As a result, user-contributed content is publicly available

unless the user changes their privacy settings to "private." Some novices also appear unaware of

the fact that user-contributed data is being offered up in search results on many family history

websites, or that reuse of user-contributed content is commonplace. Even if novices knew they

were sharing their content with the public (or, at least, Ancestry subscribers), they might not

initially grasp the potential impact that sharing poorly researched content could have on the

online FHR community.

Web 2.0 technologies are at the crux of the inaccuracy problem, but also offer potential

solutions. Web tools that go beyond helping FHRs collect and display basic data about ancestors,

to supporting researchers as they collect and analyze genealogical data and original sources

across many websites, may promote better research practices and lead to the production of

more accurate family history content. Embedding educational hints within online search,

research, and production tools may help novices learn some of the literacies needed to be

accurate in family history research, without being overly didactic or separated from the process

of searching for and collecting data and documents. Providing spaces that support situational

learning between expert and novice FHRs, along with spaces for cooperative research and

production activities, may be another effective way to teach the craft to newcomers. Lastly,

family history websites should make it clearer to users that they have an option to share or keep their content private, perhaps coupled with simple guidance about the importance of sharing content that is well-researched.

Millions of Americans are researching their families' pasts and sharing their findings with others. With improvements to family history websites, the craft of family history research likely could be mastered by many more participants, potentially leading to the production and sharing of good-quality, accurate family history content for generations to come.

List of References

Agosto, D. E. (2002). Bounded rationality and satisficing in young people's Web-based decision making. *Journal of the American Society for Information Science and Technology*, *53*(1), 16–27.

Ancestry.com LLC Reports Fourth Quarter and Full Year 2015 Financial. (2016). [commercial]. Retrieved July 2, 2016, from http://corporate.ancestry.com/press/press-releases/2016/02/ancestrycom-llc-reports-fourth-quarter-and-full-year-2015-financial-results/

Ancestry.com Traffic, Demographics and Competitors - Alexa. (2016). Retrieved September 9, 2016, from http://www.alexa.com/siteinfo/ancestry.com

Anderson, T. D. (2005). Relevance as Process: Judgements in the Context of Scholarly Research. *Information Research: An International Electronic Journal*, *10*(2), n2.

Anthony, D., Smith, S. W., & Williamson, T. (2009). Reputation and Reliability in Collective Goods The Case of the Online Encyclopedia Wikipedia. *Rationality and Society*, *21*(3), 283–306.

Arazy, O., Morgan, W., & Patterson, R. (2006). Wisdom of the crowds: Decentralized knowledge construction in Wikipedia. In *16th Annual Workshop on Information Technologies & Systems (WITS) Paper*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1025624

Archer, J., Hanlon, A. M., & Levine, J. A. (2009). Investigating primary source literacy. *The Journal of Academic Librarianship*, *35*(5), 410–420.

Arrington, M. (2007). MyHeritage Takes 180 Million People Profiles To War With Geni | TechCrunch. Retrieved May 29, 2017, from https://techcrunch.com/2007/08/28/myheritage-takes-180-million-people-profiles-to-war-with-geni/

Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced web users. In *Proceedings of the 14th international conference on World Wide Web* (pp. 583–592). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1060831

Aula, A., & Nordhausen, K. (2006). Modeling successful performance in Web searching. *Journal of the American Society for Information Science and Technology*, *57*(12), 1678–1693.

Bandura, A. (1977). Social learning theory. 1977. *Englewood Cliffos, New Jersey*.

Bar-Ilan, J., Keenoy, K., Levene, M., & Yaari, E. (2009). Presentation bias is significant in determining user preference for search results—A user study. *Journal of the Association for Information Science and Technology*, *60*(1), 135–149.

Barry, C. L. (1994). User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, *45*(3), 149.

Barry, C. L. (1998). Document representations and clues to document relevance. *Journal of the American Society for Information Science*, *49*(14), 1293–1303.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, *13*(5), 407–424. https://doi.org/10.1108/eb024320

Bates, M. J. (2009). Information Behavior. In *Encyclopedia of Library and Information Sciences, Third Edition* (Vols. 1–0, pp. 2381–2391). CRC Press. Retrieved from http://www.crcnetbase.com/doi/10.1081/E-ELIS3-120043263

Batson, C. D., Ahmad, N., & Tsang, J.-A. (2002). Four motives for community involvement. *Journal of Social Issues*, *58*(3), 429–445.

Baytiyeh, H., & Pfaffman, J. (2010). Volunteers in Wikipedia: Why the Community Matters. *Educational Technology & Society*, *13*(2), 128–140.

Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information-retrieval. *Canadian Journal of Information Science-Revue Canadienne Des Sciences De L Information*, *5*(MAY), 133–143.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, *38*(2), 61–71.

Bhavnani, S. K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 610–611). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=506508

Bransford, J. D., Brown, A. L., & Cocking, R. L. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work* (pp. 1–10). Retrieved from http://dl.acm.org/citation.cfm?id=1099205

Case, D. O., & Given, L. M. (2016). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Emerald Group Publishing.

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, *11*(11). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/viewArticle/1413

Cho, H., Chen, M., & Chung, S. (2010). Testing an integrative theoretical model of knowledge-sharing behavior in the context of Wikipedia. *Journal of the American Society for Information Science and Technology*, *61*(6), 1198–1212.

Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. (2010). Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 107–116). Retrieved from http://dl.acm.org/citation.cfm?id=1718940

Choi, Y., & Rasmussen, E. M. (2002). Users' relevance criteria in image retrieval in American history. *Information Processing & Management*, *38*(5), 695–726.

Choo, C. W., Detlor, B., & Turnbull, D. (2000a). Information seeking on the Web: An integrated model of browsing and searching. *First Monday*, *5*(2). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/729

Choo, C. W., Detlor, B., & Turnbull, D. (2000b). Information seeking on the Web: An integrated model of browsing and searching. *First Monday*, *5*(2). https://doi.org/10.5210/fm.v5i2.729

Cole, C. (2000). Name collection by Ph. D. history students: Inducing expertise. *Journal of the American Society for Information Science*, *51*(5), 444–455.

Connaway, L. S., Dickey, T. J., & Radford, M. L. (2011). "If it is too inconvenient I'm not going after it:" Convenience as a critical factor in information-seeking behaviors. *Library & Information Science Research*, *33*(3), 179–190. https://doi.org/10.1016/j.lisr.2010.12.002

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage. Retrieved from http://books.google.com/books?hl=en&lr=&id=0TI8Ugvy2Z4C&oi=fnd&pg=PT7&dq=Basics+of+qualitative+research:+Techniques+and+procedures+for+developing+grounded+theory.&ots=bkUdgnUWMx&sig=pRUwv86CqbXLF-21ZccAzZRV3Ng

Cothey, V. (2002). A longitudinal study of World Wide Web users' information-searching behavior. *Journal of the American Society for Information Science and Technology*, *53*(2), 67–78.

Darby, P. (2011). *The Information seeking behavior of genealogist and family historians* (masters thesis). University of Sheffield, Sheffield, UK.

Darby, P., & Clough, P. (2013b). Investigating the information-seeking behaviour of genealogists and family historians. *Journal of Information Science*, *39*(1), 73–84.

Dervin, B. (1983). Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. Retrieved from https://philpapers.org/rec/DERIAA

Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, *2*(2), 36–46.

Dervin, B., Foreman-Wernet, L., & Lauterbach, E. (2003). *Sense-making methodology reader: Selected writings of Brenda Dervin*. Hampton Pr.

Dong, P., Loh, M., & Mondry, A. (2005). Relevance similarity: an alternative means to monitor information retrieval systems. *Biomedical Digital Libraries*, *2*(1), 6.

Duff, W. M., Craig, B., & Cherry, J. (2004). Historians' use of archival sources: promises and pitfalls of the digital age. *The Public Historian*, *26*(2), 7–22.

Duff, W. M., & Johnson, C. A. (2002). Accidentally found on purpose: information-seeking behavior of historians in archives. *The Library Quarterly*, 472–496.

Duff, W. M., & Johnson, C. A. (2003). Where is the list with all the names? Information-seeking behavior of genealogists. *American Archivist*, *66*(1), 79–95.

Duff, W., Yakel, E., & Tibbo, H. (2013). Archival reference knowledge. *The American Archivist*, *76*(1), 68–94.

Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 39–48). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1357062

Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing*. Intellect books. Retrieved from https://books.google.com/books?hl=en&lr=&id=4lge5k_F9EwC&oi=fnd&pg=PR5&dq=Dumas,+++J+.,+++%26++Redish,++J.++A+++Practical+++Guide+++to++++Usability+Testing+&ots=vql69GfaxG&sig=-hfnr-6sgEUY0DUzOmPMT62n920

Dziadosz, S., & Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 365–366). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=564446

Eckert, S., & Steiner, L. (2013). Wikipedia's gender gap. *Media (Dis) Parity: A Gender Battleground*, 87–98.

Eisenberg, M., & Barry, C. (1988). Order effects: a study of the possible influence of presentation order on user judgements of document relevance. *Journal of the American Society for Information Science*, *39*(5), 293.

Ellis, D. (1989). A behavioural model for information retrieval system design. *Journal of Information Science*, *15*(4–5), 237–247.

Ellis, D. (2005). Ellis's model of information-seeking behavior. In K. Fisher, S. Erdelez, & L. E. F. McKechnie (Eds.), *Theories of information behavior: a researcher's guide* (pp. 138–142). Information Today.

Erdelez, S. (1997). Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context* (pp. 412–421). Taylor Graham Publishing. Retrieved from http://dl.acm.org/citation.cfm?id=267217

Ericsson, K. A., & Simon, H. A. (1985). *Protocol analysis*. MIT press. Retrieved from http://ammonwiemers.com/IdetPortfolio/articles/Assessment/Protocol%20Analysis%20--%20Verbal%20Reports%20as%20Data.pdf

Falconer, B. (2012a). Ancestry.com's Genealogical Juggernaut. Retrieved September 9, 2016, from http://www.bloomberg.com/news/articles/2012-09-20/ancestry-dot-coms-genealogical-juggernaut

Falconer, B. (2012b, September 20). Ancestry.com's Genealogical Juggernaut. *BusinessWeek: Lifestyle*. Retrieved from http://www.businessweek.com/articles/2012-09-20/ancestry-dot-coms-genealogical-juggernaut

Find A Grave - Millions of Cemetery Records. (n.d.). Retrieved January 13, 2017, from http://www.findagrave.com/

Fisher, K. E., Erdelez, S., & McKechnie, L. (2005). *Theories of Information Behavior*. Information Today, Inc.

Fitzgerald, M. A., & Galloway, C. (2001). Relevance judging, evaluation, and decision making in virtual libraries: A descriptive study. *Journal of the American Society for Information Science and Technology*, *52*(12), 989–1010.

Forte, A., & Bruckman, A. (2005). Why do people write for Wikipedia? Incentives to contribute to open–content publishing. *Proc. of GROUP*, *5*, 6–9.

Forte, A., Humphreys, M., & Park, T. H. (2012). Grassroots Professional Development: How Teachers Use Twitter. In *ICWSM*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/download/4585/4973

Forte, A., & Lampe, C. (2013). Defining, Understanding, and Supporting Open Collaboration Lessons From the Literature. *American Behavioral Scientist*, *57*(5), 535–547.

Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, *26*(1), 49–72.

Friday, K. (2014, December 15). Learning from e-family history: a model of online family historian research behaviour [text]. Retrieved July 2, 2016, from http://www.informationr.net/ir/19-4/paper641.html#author

Fulton, C. (2009). Quid pro quo: Information sharing in leisure activities. *Library Trends*, *57*(4), 753–768.

Furay, C., & Salevouris, M. J. (1988). *The methods and skills of history: A practical guide*. H. Davidson. Retrieved from http://www.raz7.com/columbia/research/methods.doc

Garrett, C. (2010). *Genealogical research, Ancestry. com, and archives*. Auburn University. Retrieved from http://etd.auburn.edu/etd/handle/10415/2014

Gasson, S., & Waters, J. (2011). Using a grounded theory approach to study online collaboration behaviors†. *European Journal of Information Systems*, *22*(1), 95–118.

Geiger, R. S., & Ribes, D. (2010). The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 117–126). New York, NY, USA: ACM. https://doi.org/10.1145/1718918.1718941

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901.

Glaser, B., & Straus, A. (1967). *The Discovery of Grounded Theory.* New York, NY, USA: Aldine Publishing Company.

Halfaker, A., Geiger, R. S., & Terveen, L. G. (2014). Snuggle: Designing for Efficient Socialization and Ideological Critique. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 311–320). New York, NY, USA: ACM. https://doi.org/10.1145/2556288.2557313

Hansen, D. L., Schone, P. J., Corey, D., Reid, M., & Gehring, J. (2013). Quality control mechanisms for crowdsourcing: peer review, arbitration, & expertise at familysearch indexing. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 649–660). Retrieved from http://dl.acm.org/citation.cfm?id=2441848

Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: characterizing survey response bias with propensity score estimation. *PloS One*, *8*(6), e65782.

Hockett, H. C. (1955). *The critical method in historical research and writing*. Macmillan. Retrieved from http://ahr.oxfordjournals.org/content/62/3/local/back-matter.pdf

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, *33*(1), 337–346.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, *44*(3), 161.

Huang, M., & Wang, H. (2004). The influence of document presentation order and number of documents judged on users' judgments of relevance. *Journal of the American Society for Information Science and Technology*, *55*(11), 970–979.

James, A. (1996). Incremental relevance feedback for information filtering. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 270–278). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=243274

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, *42*(1), 248–263.

Jenkins, C., Corritore, C. L., & Wiedenbeck, S. (2003). Patterns of information seeking on the Web: A qualitative study of domain expertise and Web expertise. *IT & Society*, *1*(3), 64–89.

Joyce, E., Butler, B., & Pike, J. (2011). Handling flammable materials: Wikipedia biographies of living persons as contentious objects. In *Proceedings of the 2011 iConference* (pp. 25–32). Retrieved from http://dl.acm.org/citation.cfm?id=1940765

Katzer, J., & Fletcher, P. T. (1992). The Information Environment of Managers. *Annual Review of Information Science and Technology (ARIST)*, *27*, 227–63.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. *Journal of the Association for Information Science and Technology*, *58*(7), 999–1018.

Kelly, D., Yuan, X., Belkin, N. J., Murdock, V., & Croft, W. B. (2002). Features of documents relevant to task-and fact-oriented questions. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 645–647). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=584907

Kinney, K. A., Huffman, S. B., & Zhai, J. (2008). How Evaluator Domain Expertise Affects Search Result Relevance Judgments. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 591–598). New York, NY, USA: ACM. https://doi.org/10.1145/1458082.1458160

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 37–46). Retrieved from http://dl.acm.org/citation.cfm?id=1460572

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 453–462). Retrieved from http://dl.acm.org/citation.cfm?id=1240698

Kuhlthau, C. C. (1988). Developing a Model of the Library Search Process: Cognitive and Affective Aspects. *RQ*, *28*(2), 232–242.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361.

Kuznetsov, S. (2006). Motivations of contributors to Wikipedia. *ACM SIGCAS Computers and Society*, *36*(2), 1.

Lave, J. (1991). Situating learning in communities of practice. *Perspectives on Socially Shared Cognition*, *2*, 63–82.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press. Retrieved from http://books.google.com/books?hl=en&lr=&id=CAVIOrW3vYAC&oi=fnd&pg=PA11&dq=Situated+learning:+Legitimate+peripheral+participation&ots=OArGuuZMCh&sig=nif1LekUTJm82ZitHgh9kaRtvw4

Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. *Nature*. Retrieved from http://www.ufrgs.br/limc/participativo/pdf/wikipedia.pdf

Lin, B., & Serebrenik, A. (2016). Recognizing gender of stack overflow users. In *Proceedings of the 13th International Conference on Mining Software Repositories* (pp. 425–429). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2901777

Liu, J., & Ram, S. (2009). Who does what: Collaboration patterns in the wikipedia and their impact on data quality. In *19th Workshop on Information Technologies and Systems* (pp. 175–180). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1565682

Lucas, S. A. (2008). *The Information Seeking Processes of Genealogists*. Emporia State University.

Maglaughlin, K. L., & Sonnenwald, D. H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and Technology*, *53*(5), 327–342.

Mansourian, Y. (2008). Coping strategies in web searching. *Program*, *42*(1), 28–39.

Marchionini, G., Lin, X., & Dwiggins, S. (1990). Effects of search and subject expertise on information seeking in a hypertext environment. In *Proceedings of the 53rd Annual Meeting of the ASIS, 1990*. Retrieved from http://ci.nii.ac.jp/naid/10013112060/

Meho, L. I., & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, *54*(6), 570–587.

Morgan, J. T., Bouterse, S., Walls, H., & Stierch, S. (2013). Tea and sympathy: crafting positive new user experiences on wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 839–848). Retrieved from http://dl.acm.org/citation.cfm?id=2441871

Morrison, J. B., Pirolli, P., & Card, S. K. (2001). A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. In *CHI'01 extended abstracts on Human factors in computing systems* (pp. 163–164). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=634167

News, A. B. C., & Farnham, A. (2012, October 24). Genealogy Enjoys Unlikely Popularity. Retrieved July 2, 2016, from http://abcnews.go.com/Business/genealogy-hot-hobby-worth-16b-mormons/story?id=17544242

Nicholas, D., Rowlands, I., Huntington, P., Jamali, H. R., & Hernández Salazar, P. (2010). Diversity in the e-journal use and information-seeking behaviour of UK researchers. *Journal of Documentation*, *66*(3), 409–433. https://doi.org/10.1108/00220411011038476

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010a). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2381–2390). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1753685

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010b). Think-aloud protocols: Analyzing three different think-aloud protocols with counts of verbalized frustrations in a usability study of an information-rich Web site. In *Professional Communication Conference (IPCC), 2010 IEEE International* (pp. 60–66). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/5529815/

Online Family History Trends Report. (2011). Retrieved December 3, 2013, from http://www.archives.com/blog/miscellaneous/online-family-history-trends-1.html

Patton, M. Q. (1990). *Qualitative evaluation and research methods* . SAGE Publications, inc. Retrieved from http://psycnet.apa.org/psycinfo/1990-97369-000

Pettigrew, K. E., Fidel, R., & Bruce, H. (2001). Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology (ARIST)*, *35*(43–78). Retrieved from http://faculty.washington.edu/fidelr/RayaPubs/ConceptualFrameworks.pdf

Piaget, J. (1976). *Piaget's theory*. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-46323-5_2

Prasarnphanich, P., & Wagner, C. (2011). Explaining the sustainability of digital ecosystems based on the wiki model through critical-mass theory. *Industrial Electronics, IEEE Transactions On*, *58*(6), 2065–2072.

Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, *1*(1), 13–32.

Purgailis Parker, L. M., & Johnson, R. E. (1990). Does order of presentation affect users' judgment of documents? *Journal of the Association for Information Science and Technology*, *41*(7), 493–494.

Rainie, L., Lenhart, A., Spooner, T., & Horrigan, J. (2000). *Tracking Online Life*. Pew Internet and American Life Project. Retrieved from http://www.pewinternet.org/Reports/2000/Tracking-Online-Life.aspx

Rajagopalan, M. S., Khanna, V., Stott, M., Leiter, Y., Showalter, T. N., Dicker, A., & Lawrence, Y. R. (2010). Accuracy of cancer information on the Internet: A comparison of a Wiki with a professionally maintained database. In *ASCO Annual Meeting Proceedings* (Vol. 28, p. 6058).

Reavley, N. J., Mackinnon, A. J., Morgan, A. J., Alvarez-Jimenez, M., Hetrick, S. E., Killackey, E., … Jorm, A. F. (2012). Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. *Psychological Medicine*, *42*(8), 1753–1762. https://doi.org/10.1017/S003329171100287X

Rieh, S. Y. (2003). Investigating Web searching behavior in home environments. *Proceedings of the American Society for Information Science and Technology*, *40*(1), 255–264.

Robyns, M. C. (2001). The archivist as educator: Integrating critical thinking skills into historical research methods instruction. *American Archivist*, *64*(2), 363–384.

Rodriguez, G. (2014, May 30). How Genealogy Became Almost as Popular as Porn. *Time*. Retrieved from http://time.com/133811/how-genealogy-became-almost-as-popular-as-porn/

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, *58*(13), 2126–2144.

Savolainen, R. (2007). Information Behavior and Information Practice: Reviewing the "Umbrella Concepts" of Information-Seeking Studies. *The Library Quarterly: Information, Community, Policy*, *77*(2), 109–132. https://doi.org/10.1086/517840

Schoberth, T., Preece, J., & Heinzl, A. (2003). Online communities: A longitudinal analysis of communication activities. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on* (p. 10–pp). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1174576

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002). How knowledge workers use the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 227–234). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=503418

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*(1), 161–176.
Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information Processing & Management*, *34*(5), 599–621.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. *Cambridge Handbook of the Learning Sciences*, *2006*. Retrieved from http://gerrystahl.net/cscl/CSCL_English.htm

Stebbins, R. A. (2009). Leisure and Its Relationship to Library and: Information Science: Bridging the Gap. *Library Trends*, *57*(4), 618–631. https://doi.org/10.1353/lib.0.0064

Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, *78*(6), 981–996.

Straus, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory. *2 Nd Ed*.

Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005). Information quality discussions in Wikipedia. In *Proceedings of the 2005 international conference on knowledge management* (pp. 101–113). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.3912&rep=rep1&type=p

df

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001.

Suh, B., Chi, E. H., Pendleton, B. A., & Kittur, A. (2007). Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (pp. 163–170). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4389010

Talja, S., & Hansen, P. (2006). Information sharing. In *New directions in human information behavior* (pp. 113–134). Springer. Retrieved from http://link.springer.com/chapter/10.1007/1-4020-3670-1_7

Tang, R., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: on the complementarity of experimental and naturalistic studies. *Journal of the American Society for Information Science and Technology*, *52*(8), 676–685.

Taylor, R. S. (1986). *Value-added processes in information systems*. Greenwood Publishing Group. Retrieved from https://books.google.com/books?hl=en&lr=&id=bYYoHV7bqvAC&oi=fnd&pg=PR5&dq=taylor+information+values+in+decision+contexts&ots=9owmwJUr8O&sig=FHV2IjJz7TDhtMz6QlMupcB_IIs

Tennant, M. (1997). *Psychology and adult learning*. London: Routledge. Retrieved from http://books.google.com/books?hl=en&lr=&id=j5muRQ9Yc0YC&oi=fnd&pg=PR1&dq=Ps ychology+and+Adult+Learning&ots=hhE8kW6kzq&sig=U6C3FoeDYEDPpYUJ5BG8GJIyTM c

The Ancestry Insider: RootsTech Attendee Demographics. (2016). Retrieved September 9, 2016, from http://www.ancestryinsider.org/2015/06/rootstech-attendee-demographics.html

Thom-Santelli, J., Cosley, D., & Gay, G. (2010). What do you know?: experts, novices and territoriality in collaborative systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1685–1694). Retrieved from http://dl.acm.org/citation.cfm?id=1753578

Thom-Santelli, J., Cosley, D. R., & Gay, G. (2009). What's mine is mine: territoriality in collaborative authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1481–1484). Retrieved from http://dl.acm.org/citation.cfm?id=1518925

Tibbo, H. R. (2003). Primarily history in America: How US historians search for primary materials at the dawn of the digital age. *American Archivist*, *66*(1), 9–50.

Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess web pages for information seeking. *Journal of the American Society for Information Science and Technology*, *56*(4), 327–344.

Toms, E. G., O'Brien, H. L., Kopak, R., & Freund, L. (2005). Searching for relevance in the relevance of search. In *Context: Nature, impact, and role* (pp. 59–78). Springer. Retrieved from http://link.springer.com/chapter/10.1007/11495222_7

Top 100 Genealogy Websites of 2014 (page 2). (2014). *GenealogyInTime Magazine*. Retrieved from http://www.genealogyintime.com/articles/top-100-genealogy-websites-of-2014-page02.html

Top 100 Genealogy Websites of 2015. (2015). *GenealogyInTime Magazine*. Retrieved from http://www.genealogyintime.com/articles/top-100-genealogy-websites-of-2015-page01.html

Top 100 Genealogy Websites of 2016. (2016). *GenealogyInTime Magazine*. Retrieved from http://www.genealogyintime.com/articles/top-100-genealogy-websites-of-2016-page01.html

Tucker, S. N. (2009). The most public of all history: family history and heritage albums in the transmission of records. Retrieved from http://dare.uva.nl/document/156792

Tuominen, K., Talja, S., & Savolainen, R. (2005). *The social constructionist viewpoint on information practices*. na.

U.S. Census Bureau. (2015). Population estimates, July 1, 2015, (V2015). Retrieved September 9, 2016, from //www.census.gov/quickfacts/

Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, *56*(5), 540–562.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management*, *39*(3), 445–463.

Vasilescu, B., Capiluppi, A., & Serebrenik, A. (2012). Gender, representation and online participation: A quantitative study of stackoverflow. In *Social Informatics (SocialInformatics), 2012 International Conference on* (pp. 332–338). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6542459/

Vasilescu, B., Posnett, D., Ray, B., van den Brand, M. G., Serebrenik, A., Devanbu, P., & Filkov, V. (2015). Gender and tenure diversity in GitHub teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3789–3798). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2702549

Veale, K. J. (2004). Discussing our Family Trees: Finding Community-Based Communication in Genealogical Newsgroups. *The Journal*. Retrieved from http://www.getcited.org/pub/103409174

Viegas, F. B., Wattenberg, M., Kriss, J., & Van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on* (pp. 78–78). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076527

Vygotsky, L. S. (1931). *Mind in society*. Cambridge, MA: Harvard University Press.

Wang, P., & White, M. D. (1995). Document Use during a Research Project: A Longitudinal Study. In *Proceedings of the ASIS Annual Meeting* (Vol. 32, pp. 181–88). ERIC. Retrieved from http://eric.ed.gov/?id=EJ513873

White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 132–141). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1498819

Whittaker, S. T., & Terveen, L. (1998). L., Hill, W., and Cherny, L.(1998). The dynamics of mass interaction. In *Proceedings of Conference on Computer Supported Cooperative Work* (pp. 257–264).

WikiTree: The Free Family Tree. (2017). Retrieved May 29, 2017, from https://www.wikitree.com/
Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, *55*(3), 246–258.

Wilkinson, D. M., & Huberman, B. A. (2007). Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis* (pp. 157–164). Retrieved from http://dl.acm.org/citation.cfm?id=1296968

Willever-Farr, H., & Forte, A. (2014). Family Matters: Control and Conflict in Online Family History Production. *CSCW 2014*.

Willever-Farr, H., Zach, L., & Forte, A. (2012). Tell me about my family: a study of cooperative research on ancestry.com. In *Proceedings of the 2012 iConference* (pp. 303–310). New York, NY, USA: ACM. https://doi.org/10.1145/2132176.2132215

Williamson, K. (1998). Discovered by chance: The role of incidental information acquisition in an ecological model of information use. *Library & Information Science Research*, *20*(1), 23–40.

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, *37*(1), 3–15.

Wilson, T. D. (1997). Information behaviour: an interdisciplinary perspective. *Information Processing & Management*, *33*(4), 551–572.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73.

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346.

Wineburg, S. S., & Fournier, J. (1994). Contextualized thinking in history. *Cognitive and Instructional Processes in History and the Social Sciences*, 285–308.

Xu, Y., & Wang, D. (2008). Order effect in relevance judgment. *Journal of the American Society for Information Science and Technology*, *59*(8), 1264–1275.

Yakel, E. (2004a). Information literacy for primary sources: creating a new paradigm for archival researcher education. *OCLC Systems & Services*, *20*(2), 61–64.

Yakel, E. (2004b). Seeking information, seeking connections, seeking meaning: Genealogists and family historians. *Information Research*, *10*(1), 10–1.

Yakel, E., & Torres, D. A. (2003). AI: archival intelligence and user expertise. *American Archivist*, *66*(1), 51–78.

Yakel, E., & Torres, D. A. (2007). Genealogists as a" Community of Records". *American Archivist*, *70*(1), 93–113.

Yeow, A., Johnson, S. L., & Faraj, S. (2006). Lurking: Legitimate or illicit peripheral participation. In *Proceedings of Twenty-seventh International Conference on Information Systems*.

Yin, R. K. (2015). *Qualitative research from start to finish*. New York, NY, USA: Guilford Publications. Retrieved from https://books.google.com/books?hl=en&lr=&id=DvpPCgAAQBAJ&oi=fnd&pg=PP1&dq=y in+qualitative+research+&ots=HTaoqWgp0O&sig=Ir2ghXiYNjxdd3OlXh5M7f653GU

Zhang, X., Anghelescu, H. G., & Yuan, X. (2005). Domain Knowledge, Search Behaviour, and Search Effectiveness of Engineering and Science Students: An Exploratory Study. *Information Research: An International Electronic Journal*, *10*(2), n2.

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., & Grover, V. (2010). Investigating online information disclosure: Effects of information relevance, trust and risk. *Information & Management*, *47*(2), 115–123.

Appendix A: Solicitation Message – Initial Interviews

Dear Family History Researchers,

I am a doctoral student at the College of Computing and Informatics, Drexel University, Philadelphia, Pa., and am conducting research to learn more about how to support people as they go about finding information on ancestors or deceased individuals and then use that information to create content for family history and memorialization websites. This is my dissertation research, and my research is not profit-oriented.

I am looking for Individuals to Participate in My Study

Specifically, I am looking for:

- **Newcomers to family history** (anyone with little experience conducting family history research and developing content for family history websites)

- **Experienced family history researchers and website contributors** (anyone who has advanced knowledge of family history research and has experience contributing to family history or memorialization websites).

What I Would Like You to Do

- First, take an online survey.
- Then if contacted, participate in a phone interview that will last from 30 to 60 minutes, depending on what you wish to share.

What You Get

Helping to advance our knowledge of the ways family history researchers are using web technologies for research and collaboration. This knowledge may help archives, historical societies, and other memory institutions to better support collaboration among researchers.

If you are interested in participating or want further information, please contact:

Heather Willever-Farr

hlw29@drexel.edu

You can read more about my research at http://www.heatherwilleverfarr.com

Thank you,

Heather Willever-Farr, MA

College of Computing and Informatics

Drexel University

Philadelphia, PA

Appendix B: Solicitation Message –Think-aloud Study

Dear Family History Researchers,

I am a doctoral student at the College of Computing and Informatics, Drexel University, Philadelphia, Pa., and am conducting research to learn more about how to support people as they go about finding information on ancestors or deceased individuals and then use that information to create content for family history and memorialization websites. This is my dissertation research, and my research is not profit-oriented.

I am looking for Individuals to Participate in My Study

Specifically, I am looking for:

- **Newcomers to family history** (anyone with little experience conducting family history research and developing content for family history websites)

- **Experienced family history researchers and website contributors** (anyone who has advanced knowledge of family history research and has experience contributing to family history or memorialization websites).

What I Would Like You To Do

- First, take an online survey.
- Then if contacted, participate in an in-person study that will last 60 to 90 minutes. Participants will be asked to use a computer to find information on deceased individuals (using the Internet) and then build family history content.

What You Get

Helping to advance our knowledge of the ways family history researchers are using web technologies for research and collaboration, and $20.00 for participating in think-aloud sessions.

If you are interested in participating or want further information, please contact:

Heather Willever-Farr

hlw29@drexel.edu

You can read more about my research at http://www.heatherwilleverfarr.com

Thank you,

Heather Willever-Farr, MA

College of Computing and Informatics

Drexel University

Philadelphia, PA

Appendix C: Initial Survey on Experience Level

(Only used for initial interviews)

WELCOME

Thank you for agreeing to take this survey about your experiences with family history research. For the survey you are about to take, please read the following information carefully and indicate your consent to proceed. The survey should take you roughly 10-15 minutes to complete.

By completing this survey and submitting it, you indicate that you understand the data will be used for research purposes and that you are allowing the researchers, Andrea Forte and Heather Willever-Farr, to use your data in their study.

Participation in the survey is voluntary, and you may stop participating at any time or skip any questions that you don't want to answer. No personal or identifying information about you or any participant will be included in reports or publications about this work. We are asking for information about your age, occupation, and race to get a sense of the demographics of contributors to family history and memorialization websites.

There are no foreseeable risks to you, and there is no compensation for taking the survey.

The survey is the first step in this research study. The second step is conducting interviews with individuals who have engaged in family history research. The interviews will be conducted remotely using Skype. If you are interested in participating, please indicate below.

If you have further questions about this survey or the research project, please contact Heather Willever-Farr (hlw29@drexel.edu) or the Institutional Review Board at Drexel University (ORCA@drexel.edu).

1) Please check one of the following before taking survey:

☐ I understand and consent to the above information.

☐ I do not understand or do not consent to the above information.

2) If you might be interested in participating in a phone interview (using Skype) for this study, please provide a name, contact number or email address I can reach you if selected for interview study:

_____

_____

3) Please state your age in years:_____

4) Please indicate your gender:

☐Female

☐Male

☐Other

5) Please state your occupation (if retired your previous occupation):_____

6) Please indicate your race (more than one category can be selected):

☐American Indian or Alaska Native

☐Asian

☐Caucasian or White

☐Black or African American

☐Hispanic or Latino

☐Native Hawaiian or Other Pacific Islander

☐Other

7) For how long have you been doing family history research or research on deceased individuals?

☐0-11 months

☐1-2 years

☐3-5 years

☐6-10 years

☐11 years or more

8) What do you consider to be your experience level with finding information on your ancestors or other deceased individuals?

☐ Beginner

☐ Intermediate

☐ Advanced

9) What do you consider to be your experience level with selecting information or records that accurately represent or relate to your ancestors or other deceased individuals?

☐ Beginner

☐ Intermediate

☐ Advanced

10) What do you consider to be your experience level with interpreting historic records, such as census records or military records, related to your ancestors or other deceased individuals?

☐ Beginner

☐ Intermediate

☐ Advanced

11) What do you consider to be your experience level with creating content on your ancestors or other deceased individuals, such as family trees or memorials?

☐ Beginner

☐ Intermediate

☐ Advanced

12) If you have you used websites to find information about your ancestors or other deceased individuals, how many different websites have you used?

☐ 0-2

☐ 3-5

☐6-10

☐11 or more

13) Have you used Findagrave.com to locate information on your ancestors or other deceased individuals?

☐Yes

☐No

14) Please list the websites you use to find information about your ancestors or deceased individuals.

_____

_____

_____

_____

15) How frequently do you use websites for finding information on your ancestors or other deceased

individuals (please pick the answer that best fits)?

☐Couple times a year

☐Once a month

☐Once a week

16) If you have visited brick and mortar institutions or organizations, such as court houses, archives,

historical societies, libraries, etc., to find information about your ancestors or other deceased

individuals, how many times?

☐0 times

☐1 time

☐2-5 times

☐6-10 times

☐11 or more times

17) Please list institutions or organizations you have visited to find information about your ancestors or other deceased individuals.

_____

_____

_____

_____

18) If you have visited graveyards to find information about your ancestors or other deceased individuals, how many times?

☐None

☐1 time

☐2-5 times

☐5-10 times

☐11 or more times

19) If you have used original sources (such as documents or records) to locate information on ancestors or deceased, please select types of records you used (please check all that apply):

☐census military service records   ☐ship manifests   ☐social security records   ☐ obituaries

☐graveyard records   ☐other, please list:_____

20) Have you created content (such as family trees, histories or memorials) about your ancestors or other deceased individuals for websites?

☐Yes

☐No

21) How many family trees have you built?

    ☐0

    ☐1

    ☐2

    ☐3

    ☐4 or more

22) How many Find A Grave memorials have you created?

    ☐0

    ☐1

    ☐2-5

    ☐6-10

    ☐11 or more

23) How many images have you uploaded to memorial or family history websites?

    ☐0

    ☐1

    ☐2-5

    ☐6-10

    ☐11 or more

24) If you have contributed other types of content to family history or memorialization websites, other than family trees or memorials, please list below.

_____

_____

_____

_____

25) Do you participate in any family history groups?

☐Yes

☐No

26) Are these groups face-to-face and/or virtual (please select all that apply)?

☐Virtual or online groups

☐Face-to-face groups

Appendix D: Final Survey

(New questions that were added to the final iteration of the survey are italicized)

WELCOME

Thank you for agreeing to take this survey about your experiences with family history research. For the survey you are about to take, please read the following information carefully and indicate your consent to proceed. The survey should take you roughly 20-25 minutes to complete.

By completing this survey and submitting it, you indicate that you understand the data will be used for research purposes and that you are allowing the researchers, Andrea Forte and Heather Willever-Farr, to use your data in their study.

Participation in the survey is voluntary, and you may stop participating at any time or skip any questions that you don't want to answer. No personal or identifying information about you or any participant will be included in reports or publications about this work. We are asking for information about your age, occupation, and race to get a sense of the demographics of contributors to family history and memorialization websites.

There are no foreseeable risks to you, and there is no compensation for taking the survey.

If you have further questions about this survey or the research project, please contact Heather Willever-Farr (hlw29@drexel.edu) or the Institutional Review Board at Drexel University (ORCA@drexel.edu).

1) Please check one of the following before taking survey:

☐ I understand and consent to the above information.

☐ I do not understand or do not consent to the above information.

2) Please state your age in years:_____

3) Please indicate your gender:

☐Female

☐Male

☐Other

4) Please state your occupation (if retired your previous occupation):_____

5) Please indicate your race (more than one category can be selected):

☐American Indian or Alaska Native

☐Asian

☐Caucasian or White

☐Black or African American

☐Hispanic or Latino

☐Native Hawaiian or Other Pacific Islander

☐Other

6) For how long have you been doing family history research or research on deceased individuals?

☐0-11 months

☐1-2 years

☐3-5 years

☐6-10 years

☐11 years or more

7) What do you consider to be your experience level with finding information on your ancestors or other

deceased individuals?

☐Beginner

☐Intermediate

☐Advanced

8) What do you consider to be your experience level with selecting information or records that accurately

represent or relate to your ancestors or other deceased individuals?

☐Beginner

☐Intermediate

☐Advanced

9) What do you consider to be your experience level with interpreting historic records, such as census records or military records, related to your ancestors or other deceased individuals?

☐Beginner

☐Intermediate

☐Advanced

10) What do you consider to be your experience level with creating content on your ancestors or other deceased individuals, such as family trees or memorials?

☐Beginner

☐Intermediate

☐Advanced

11) If you have you used websites to find information about your ancestors or other deceased individuals, how many different websites have you used?

☐0-2

☐3-5

☐6-10

☐11 or more

12) Have you used Findagrave.com to locate information on your ancestors or other deceased individuals?

☐Yes

☐No

13) Please list the websites you use to find information about your ancestors or deceased individuals.

_____

_____

_____

_____

14) How frequently do you use websites for finding information on your ancestors or other deceased

individuals (please pick the answer that best fits)?

☐ Couple times a year

☐ Once a month

☐ Once a week

15) If you have visited brick and mortar institutions or organizations, such as court houses, archives,

historical societies, libraries, etc., to find information about your ancestors or other deceased

individuals, how many times?

☐ 0 times

☐ 1 time

☐ 2-5 times

☐ 6-10 times

☐ 11 or more times

16) Please list institutions or organizations you have visited to find information about your ancestors or

other deceased individuals.

_____

_____

_____

_____

17) If you have visited graveyards to find information about your ancestors or other deceased individuals,

how many times?

☐None

☐1 time

☐2-5 times

☐5-10 times

☐11 or more times

18) If you have used original sources (such as documents or records) to locate information on ancestors or

deceased, please select types of records you used (please check all that apply):

☐census military service records   ☐ship manifests   ☐social security records   ☐

obituaries

☐graveyard records   ☐other, please

list:_____

19) Have you created content (such as family trees, histories or memorials) about your ancestors or other

deceased individuals for websites?

☐Yes

☐No

20) How many family trees have you built?

☐0

☐1

☐2

☐3

☐4 or more

21) How many Find A Grave memorials have you created?

☐0

☐1

☐2-5

☐6-10

☐11 or more

22) How many images have you uploaded to memorial or family history websites?

☐0

☐1

☐2-5

☐6-10

☐11 or more

23) If you have contributed other types of content to family history or memorialization websites, other

than family trees or memorials, please list below.

_____

_____

_____

_____

24) Do you participate in any family history groups?

☐Yes

☐No

25) Are these groups face-to-face and/or virtual (please select all that apply)?

☐Virtual or online groups

☐Face-to-face groups

26) *Please select the answer that best describes how you first started doing family history research.*

☐ *At first, I conducted research on my own.*

☐ *At first, someone helped me do the research.*

27) *What answer would best describe how you FIRST learned to conduct family history research?*

☐ *Jumped right in and started searching for family information using the Internet.*

☐ *Quickly looked over some how-to resources on websites.*

☐ *Closely studied how-to resources on websites.*

☐ *Read books or magazines about family history research.*

☐ *Learned from other sources.*

Please describe other sources below:

_____

_____

28) *Please describe how you learned to do family history research.*

_____

_____

_____

_____

29) *If you changed how you researched and produced family histories (family trees, memorials, etc.) as you became more experienced, what changed?*

_____

_____

_____

_____

30) *What do you need to know to be accurate at family history research?*

_____

_____

_____

_____

31) *Have you made your family histories (family trees, memorials, etc.) open for viewing by others on the Internet?*

    ☐*Yes*

    ☐*No*

32) *If you have made your family histories (family trees, memorials, etc.) open for viewing by others on the Internet, did you do so in:*

    ☐ *0-2 months after starting your family history research*

    ☐ *3-5 months after starting your family history research*

    ☐ *6-8 months after starting your family history research*

    ☐ *9-11 months after starting your family history research*

    ☐ *12 months or more after starting your family history research*

33) *The second part of this study involves in-person research sessions. The in-person research sessions will involve participants taking part in an exercise in which they build family history content and then participate in a brief interview about their family history research experiences. These sessions will last an hour to two hours depending upon how much the participant wants to share. Each participant who takes part in an in-person research session will receive $20.00. The in-person research session will take place at one of the following locations based on the participant's preference: Drexel University's main campus in Philadelphia, the participant's home, or a public library convenient to the participant. If you live in Philadelphia or the Philadelphia suburbs, are you interested in participating in an in-person research session?*

☐Yes

☐No

If you would like to participate in an in-person research session, please provide your first name and a

contact number or email address:

_____

___

Appendix E: Interview Guide – Initial Interviews

Thank you for participating in this interview. I am a doctoral student at Drexel University in Philadelphia. I am studying family history websites and how people contribute to them. Please know that your name or any other personally identifying information will not be used in my notes or any publications that may come out of this research study, and that this interview is only for research purposes. Do I have your permission to record this conversation?

Learning About Family History Research and Information-seeking on the Web

- Did you learn about how to research before you started searching for information about your ancestors or deceased individuals? What resources did you use to learn how to research?

- For experts: do you research any differently now that you have more experience with family history or research on deceased people?

- What sorts of websites do you use to find information about your ancestors or deceased individuals?

- Do you use websites that are for historical institutions or organizations? What were your experiences like with those websites?

- Do you use websites that managed by for-profit companies? What were your experiences like with those websites?

- Please tell me about your most recent experience with online searching for family history information or information about deceased individuals. What did you need to know to be successful at this? Please describe any obstacles you may have faced.

- When finding information on the internet or web, how do you determine if that information is correct for your ancestor or the deceased individual you are researching?

- What do you do with information that you are not sure is relevant to your family or the deceased person you are researching?

- Can you remember one of your first experiences searching for family history information on the Web? Please describe this experience. What sorts of obstacles, if any, did you experience?

Information-sharing

- If you have asked other researchers to help you find information, please describe your experiences.

- If you have asked helped other researchers find information, please describe your experiences.

Information Production on the Web

- Please tell me about the first time you contributed information, images, family trees, etc. to a family history websites, considering the following questions:

  o What kind of contribution did you make?

  o What inspired you to contribute?

  o How did you learn how to create the family tree, memorial, etc.?

  o What was it like contributing?

  o Do you remember any problems you ran into?

  o How do you make sure the content you contributed is right or correct for the family or deceased individual?

  o Did you to converse with other contributors in the process of contributing? Why did you do so?

- Please tell me about your most recent contribution, considering the following questions:

- o   What kinds of contributions did you make?

- o   What inspired you to contribute?

- o   How did you learn how to create the family tree, memorial, etc.?

- o   What was it like contributing?

- o   Do you remember any problems you ran into?

- o   How often do you contribute?

- o   How do you make sure the content you contributed is right or correct for the family or deceased individual?

- o   Did you to converse with other contributors in the process of contributing? Why did you do so?

Experiences at Memory Institutions

- Please tell me about your most recent use of a brick and mortar institution or organization that has historic materials, such as an archives, historical society, etc.?

- Please tell me about your first use of brick and mortar archives, historical society, etc.?

- What did you need to know to find records or other materials in these brick and mortar institutions/organizations?

- What sorts of difficulties, if any, did you encounter?

Experiences with Primary Documents

- Please tell me about your most recent experiences using historical records, such as census records, military records, photographs, etc. Please describe any difficulties you encountered while interpreting these records.

- Please tell me about the first time you used historical records, such as census records, military records, photographs, etc.? Please describe any difficulties you encountered while interpreting these records.

- How important is it to look at the original records?

Family History Websites and Learning the Craft of Family History Research and Production

- What do you think is the best way to learn how to conduct family history research?

- How do you think websites for family history can support learning the craft of family history research?

- How can family history websites better support you in the research process?

Appendix F: Think-Aloud Research and Production Scenario

Thank you for participating in this research session. I am a doctoral student at Drexel University in Philadelphia. I am studying family history websites and how people contribute to them. Please know that your name or any other personally identifying information will not be used in my notes or any publications that may come out of this research study, and that this think-aloud and interview session is only for research purposes. During this session, I will ask you to conduct research on a deceased individual using the Internet or Web; create a memorial for that person on the website Findagrave.com; and then I will interview you about this research experience as well as other family history research experiences you may have had. At the end of this session, you will receive $20.00 for participating. Do I have your permission to record this session, which will include audio and video taping you and capturing your screen activity on the provided computer? Please sign this form if you are in agreement.

Instructions

During this session you will work alone to:

- to find information on the Internet/Web pertaining to the lives of deceased individuals
- construct one memorial for a deceased person on the website Findagrave.com
- explain out loud what you are doing as you search for information and construct the memorial
- participate in short interview after you complete the think-aloud activity

The think-aloud sessions and interviews will be audio-video recorded, and you screen activity will be captured.

Deceased Individual to Research

Name: Pauline Mai or May (not sure of spelling)

Born: 1880s or 1890s

Lived in: Wisconsin, Idaho, and north east USA

Immigrated from: Germany

Children names: Arthur and Oscar (but there may be more children)

Pauline and the following man (Herman) are related to one another, but are not blood relatives.

Name: Herman Pfenninger

Born: early 1900s

Lived in: New Jersey and possibly New York City

Wife's name: Ella Pfenninger

Using this information, you are asked to do the following:

- Use the Internet to find more information on Pauline Mai/May

    o   Please explain out loud what you are doing as you search.

    o   You can use any site or resource you want.

    o   You can make notes about the information you find in whatever way makes sense to you.

- The website Findagrave.com is a place where people can create memorials for their deceased relatives. Using the information you just collected, go to findagrave.com and build memorials for Pauline.

    o   Please explain out loud what you are doing as you build the memorial.

    o   The memorial you create will be captured using a computer application; after the memorial you create are captured and stored on the researchers' computer, the memorial will be deleted from the Findagrave.com website.

After you finish constructing the memorial, the researcher will briefly interview you about your experiences during the think-aloud activity, as well as your family history research experiences.

Appendix G: Interview Guide for Think-Aloud Participants

Learning About Family History Research and Information-seeking/Sharing

- Did you learn about how to research before you started searching for information about your ancestors or deceased individuals? What resources did you use to learn how to research?

- For experts: do you research any differently now that you have more experience with family history or research on deceased people?

- Do you research your own family? Non-related deceased individuals? Or both? Please describe who you have researched.

- Besides the websites you used for the think-aloud study, what other websites, if any, do you use to find information about your ancestors or deceased individuals? What were your experiences like with those websites?

- When finding information on the internet or web, how do you determine if that information is correct for your ancestor or the deceased individual you are researching?

- Have you ever visited a historical archives such as a state archives, a historical society, a court house, to find information? What were your experiences like at these places?

- Besides this study, have you asked other researchers to help you find information? If yes, please describe your experiences.

- Besides this study, have you helped other researchers find information? If yes, please describe your experiences.

Think-Aloud Experience

- What are your impressions of the think-aloud study?

- What did you find fairly easy to do in terms of researching and constructing the memorial?

- What was difficult?

- If you found original documents related to the family you were researching in this study, please tell me about that experience. How important is it to look at the original records?

- Are there any additional resources you might have used to locate information for the memorial that were not accessible to you during the study? If yes, please describe.

- If I shared how-to information with you or other hints, was it helpful and in what way?

- Is there anything family history websites could do to help people learn how to conduct family history research? If yes, please describe.

- How can family history websites better support you in the research process?

Vita

Heather Lynn Willever-Farr

Education

Drexel University, College of Computing and Informatics Doctor of Philosophy, Information Science

The American University, Master of Arts, American History

McDaniel College, Bachelor of Arts, History, Magna Cum Laude, Phi Beta Kappa

Selected Publications

Duff, W. & Willever-Farr, H. (2015). User behavior. In L. Duranti & P. Franks (eds.), Encyclopedia of Archival Concepts, Principles and Practices. Lanham, Maryland: Rowman & Littlefield Publishing Group.

Willever-Farr, H. & Forte, A. (2014). Family matters: Control and conflict in online family history production. In Proceedings of Computer-Supported Cooperative Work and Social Computing (CSCW). New York, NY: ACM.

Willever-Farr, H., Zach, L., & Forte, A. (2012). Tell me about my family: A study of cooperative research on Ancestry.com. In Proceedings of iConference 2012. New York, NY: ACM.

Professional Experience

Head of Technical Services, Abington Township Public Library, Abington, PA (4/2017-present)

Manager of Digital Services, Historical Society of Pennsylvania, Philadelphia, PA (5/2015 – 4/2017)

Instructor, Drexel University, College of Computing and Informatics, Philadelphia, PA (9/2010 – present)

Manager of Archives and Record Management, American College of Physicians, Philadelphia, PA (8/1997-25/2001)

Director of Records Management and State Electronic Records Archivist, Indiana Commission on Public Records, Indianapolis, IN (7/1995 -7/1997)

Reserve Officer, Army Quartermaster Corps, 1st Lieutenant (1992-1998)