

Information Management in the Humanities: Scholarly Processes, Tools, and the Construction of Personal Collections

Ciaran B. Trace (direct correspondence to this author)
Assistant Professor
School of Information
The University of Texas at Austin
1616 Guadalupe Street, Suite 5.202
Austin, TX 78701-1213
(512) 232-3508; cbtrace@austin.utexas.edu

Unmil P. Karadkar
Assistant Professor
School of Information
The University of Texas at Austin
1616 Guadalupe Street, Suite 5.202
Austin, TX 78701-1213
(512) 471-9292; unmil@ischool.utexas.edu

“The breakthrough for mass usage of digital devices came with the introduction of smartphones and tablets and the new way the standardised applications were distributed. It is currently common for the general user to choose across thousands of apps for special purposes. Similar facilities are not, however, available to the research community, especially those working in the Humanities” (Moulin et al., 2011, p. 19).

Abstract

The promise and challenge of information management in the humanities has garnered a great deal of attention and interest (University of Minnesota Libraries, 2006; Freiman et al., 2010; Bulger et al., 2011; Wilson & Patrick, 2011; Trace & Karadkar, 2013). Research libraries and archives, as well as groups from within the humanities disciplines themselves, are being tasked with providing robust support for information management practices, including helping to engage humanities scholars with appropriate digital technologies in ways that are sensitive to disciplinary-based cultures and practices. However, significant barriers impede this work, primarily because the infrastructure (services, tools, and collaborative networks) to support scholarly information management is still under development. Under the aegis of the Scholars Tracking Archival Resources (STAR) project we are studying how humanities scholars gather and manage primary source materials with a goal of developing software to support their information management practices. This article reports the findings from our interviews with twenty-six humanities scholars, in conjunction with a set of initial requirements for a mobile application that will support scholars in capturing documents, recreating the archival context, and uploading these documents to cloud storage for access and sharing from other devices.

Introduction

With a history of studying human culture and experience dating back to antiquity, the humanities encompass an array of disciplinary areas including classics, history, languages, musicology, literature, archaeology, visual and performing arts, philosophy, and religion. While perceptions about the importance and centrality of humanistic inquiry (characterized by “critical analysis and logical thinking”) vacillate with changing political, cultural, and economic climates, today the humanities remain the “distinctive underpinning of the modern research university” (Davidson & Goldberg, 2004, para. 13). Re-energized by the emergence of a digital turn, the humanities are embracing the use of digital sources and the development of digital tools as a way of supporting and augmenting intellectual inquiry and the scholarly work process (Cohen et al., 2009). These developments

exist as part of a larger effort to help strengthen the overall research infrastructure (“technology, services, practices, and policy”) in the humanities (Borgman, 2009, para. 9).

Advances in digital technology have provided humanities scholars with new means and modalities for conducting research. Web sites like *Archives Portal Europe* (<http://www.archivesportaleurope.net/>) facilitate the finding of archival materials by aggregating descriptions about collections from multiple archives, thus facilitating scholars’ visits and aiding them in maximizing research output during the time spent at an archive. Digitization efforts, such as the Library of Congress’ *American Memory Project* (<http://memory.loc.gov>), have facilitated any-time, anywhere access to a large corpus of historically and culturally significant documents that often form the cornerstone of humanities research. Echard’s (2010) assessment that digital technology has freed medieval manuscripts from ‘house arrest’ applies more broadly to all archival documents, in principle. However, the reality is that vast document collections will remain undigitized for years, if not decades, to come. For the interim, scholars must peruse these documents in paper form, housed in their brick and mortar homes, and solutions must be sought that acknowledge this reality.

Indeed, scholars are increasingly taking digitization into their own hands, looking for solutions that allow them to create their own digital surrogates of documents in a cost effective and efficient manner. The digital camera, is one example of a technology that has revolutionized scholars’ document capture practices, much like the photocopier did in the early 1990s (Miller & Galbraith, 2010; Rutner & Schonfeld, 2012). Such is the interest in the use of these technologies that US National Archives’ Chief Innovation Officer, Pamela Wright, has issued a call for developing a mobile app, a ‘Pocket Archivist’, which will help researchers capture digital images of holdings and facilitate community engagement with these documents (Fretwell, 2013). However, while digital cameras and smart phones facilitate activities such as document capture and note-taking, they leave the scholar with the challenge of managing images and text stored on these various devices, and in several formats, for the entire duration of what are often multi-year research projects.

Under the aegis of the Scholars Tracking Archival Resources (STAR) project, we are studying the complexities involved in creating and managing distributed, multi-format, personal digital primary source collections and exploring mechanisms to support the information management activities associated with this

scholarly work. Borgman argues that the suitability of data for subsequent “curation, reuse, and sharing,” is dependent on “capturing data as cleanly as possible and as early as possible in its life cycle” (Borgman, 2009, para. 44). In this article, we argue that such data capture can be achieved by harnessing the emerging digital hardware and software infrastructure to create a ‘gathering tool’ for humanities scholars (a tool to “gather, arrange, organize, store, and share objects”) (University of Minnesota Libraries, 2006, p. 50). Although such a tool has been imagined as an online resource (University of Minnesota Libraries, 2006), we imagine this tool as a mobile application that gives scholars the ability to capture, organize, and track digital images while working in an archive and, in the process, form a uniquely personal, sharable, and reusable personal archive. In the pursuit of widespread adoption, we recognize that any such software tool needs to be “easy to adapt to, and should fit well with established practices and habits” of humanities scholars (University of Minnesota Libraries, 2006, p. 18).

As a first step in understanding the landscape of challenges faced by humanities scholars in creating and managing their personal digital collections of primary source documents, we surveyed and interviewed twenty-six humanities scholars to elicit their experiences both with capturing documents (physical and digital) and managing this collection of material over the duration of a significant research project. This article presents our findings and their implications for designing software to support scholarly practices in this regard. The rest of the article is organized as follows: the next section provides an overview of prior information science and digital humanities research that has examined the nature of the research process, technology adoption, and information management practices of humanities scholars. We then describe the design of our research study followed by a presentation of our findings. We address the implications of our findings for designing software to support scholars’ information management practices and then conclude the article with a discussion of future directions for this work.

Related Work

While the literature on the information behavior of humanities scholars is modest in comparison to comparable studies for the sciences and social sciences (Nghah & Sze, 1997; Wildemuth & Case, 2010), it is voluminous enough to have warranted the publication of a number of integrative literature reviews since the 1980s (Stone, 1982; Watson-Boone, 1994; Nghah & Sze, 1997). Over the past forty years, studies of humanities scholars have

focused on various aspects of their information behavior, including the nature of work practices, use of library and information resources, and the design and evaluation of tools for supporting scholarly practices. Care must be taken in making generalized claims about these information behaviors given the wide range of disciplines and methods represented in the humanities (Lönnqvist, 2007). Yet, the research has uncovered some commonalities in the humanities research process that are worthy of exploration here.

Humanities scholarship has long been portrayed as highly individualistic in nature (Stone, 1982; Watson-Boone, 1994), but is also increasingly recognized as having a communal and collaborative component (Bulger et al., 2011). Humanities research is characterized by a process in which researchers draw from “theory, source, and text,” in the course of “introducing new perspectives and reflections” (Ochsner, Hug, & Daniel, 2013, p. 80). A key characteristic of humanities scholarship is said to be the “personal interaction between the scholar and his material” (Stone, 1982, p. 295), with data being mobilized as “evidence in support of an argument” (Owens, 2011). Unlike science, where data is largely generated and sourced internally and contemporaneously, many humanities scholars work with retrospective data sources, sources that are created by others and subsequently collected and curated as part of a larger humanities research infrastructure. Humanities data encompass both primary sources (the focus of much humanities research) and the secondary sources which inform that work (Brockman et al., 2001).

The importance of the library to humanities work has been widely acknowledged. Research has shown that the library serves as an access point from which to browse and select published secondary sources, provides a physical space in which the scholar can work and, ultimately, operates as an outlet and repository for completed scholarly work (Stone, 1982; Wiberley & Jones, 1989). Although the subject is less studied (Duff & Johnson, 2002), the archive is often just as central as the library, providing, as it does, the “point of origin” for much of the raw data in the humanities and its subsequent “interpretation, comparison, and validation” (Dalbello, 2011, p. 486). While reference librarians are underutilized as a resource in the scholarly process, research shows that humanities scholars are heavily reliant upon archivists for finding and locating primary source material (Wiberley & Jones, 1989; Delgadillo & Lynch, 1999; Johnson & Duff, 2005; Rutner & Schonfeld, 2012). While humanities scholars have embraced the use of electronic information resources for searching, some skepticism remains about

the use of digital texts and online digital collections of archival material as source data (Brockman et al., 2001; Palmer & Neumann, 2002; Bulger et al., 2011; Chassanoff, 2013; Sinn & Soares, 2014).

Various models have been created to understand and delineate the overall research process in the humanities. Stages in the humanities research process include idea generation and topic selection; extensive background reading and contextualization of the research topic; identifying, studying, analyzing, and interpreting source material; and drafting, editing, and disseminating a finished scholarly work product (Stone, 1982; Wiberley & Jones, 1989; Chu, 1999; Brockman et al., 2001). Throughout this iterative process, humanities scholars exhibit a number of common research behaviors or activities (so called ‘scholarly primitives’), the identification of which is seen as a precursor to building appropriate tools to aid in the humanities enterprise. Perhaps a disproportionate amount of the information behavior literature has focused on how humanities scholars seek and search for information (whether in the library and archive or online), with ‘browsing’ long seen as a key behavior in this respect (Stone, 1982; Brockman et al., 2001). Unsworth’s concept of ‘scholarly primitives’ expands the scope of humanities scholars’ information behaviors to include other common functions that form the basis of scholarly work: discovering, annotating, comparing, referring, sampling, illustrating, and representing (Unsworth, 2000).

The related concept of ‘information work primitives’ (Palmer & Cragin, 2008; Palmer, Teffeu, & Pirmann, 2009) highlights the fact that much of the work of humanities scholarship is, at its heart, information work. Within this framework, the notion of the ‘scholarly primitive’ has been further refined such that information work primitives are now viewed as subordinate to (or the beginnings of) broader scholarly activities (Palmer, Teffeu, & Pirmann, 2009). These broad scholarly information activities are searching (with the associated primitives of direct searching, chaining, browsing, probing, accessing), collecting (with the associated primitives of gathering and organizing), reading (with the associated primitives of scanning, assessing, rereading), writing (with the associated primitives of assembling, co-authoring, disseminating), and collaborating (with the associated primitives of coordinating, networking, and consulting). While the information work primitives listed above are clearly core to a specific research process or activity, information work primitives have also been identified which cut across scholarly activities. These so-called cross-cutting primitives include monitoring, notetaking, translating, and data practices (Palmer, Teffeu, & Pirmann, 2009). The concept of scholarly primitives is also at the center of

a formal conceptual model of scholarly research activity and information requirements drawn up by Benardou et al. (2010a). Drawing from cultural-historical activity theory, and expressed in terms of the CIDOC Conceptual Reference Model, this Scholarly Research Activity Model (SRAM) provides a framework in which scholarly research activities are understood in terms of entities such as actors, goals, methods, procedures, activities, information objects, tools and services.

The association between the humanities, computers, and computing dates back to the mid-twentieth century (Loughridge, 1989; Toms & O'Brien, 2008). In particular, the beginnings of humanities computing have been traced to Robert Busa's concordance work of the 1940s, work that influenced the areas of philology, linguistics, and literary computing. Historians' use of computing technologies also dates to this period, with mathematical methods being applied to large datasets from the 1940s onward, and large-scale quantitative analysis of sources forming part of the new social and political history of the late 1960s. The late 1980s and early 1990s, ushered in a new age in the humanities with the advent of personal computers and desktop software, the large scale digitization of primary source material, and the creation of a new scholarly infrastructure facilitated by the internet and the world wide web (McCarty, 2003; Schreibman et al., 2004; Sabharwal, 2015).

Although a trajectory of technology adoption is evident in this timeline, the large-scale embrace of technology as a tool for the humanities was slow to develop. Writing in the early 1980s, Stone (1982) felt that it "may be part of the humanistic tradition to be anti-machine" (p. 300). A decade later, Wiberley and Jones (1994) agreed, arguing that humanities scholars were slow to expand their use of information technology, a fact attributed to the unique nature both of humanities research and the research process. Recent research has emphasized that humanities scholars, while still more reticent in technology adoption than their social science and scientific colleagues, are nevertheless capable of adapting well to "rapid technical change" (Brockman et al., 2001, p. vi). Today, as Sternfeld notes, "academic and intellectual boundaries separating humanists, information specialists, and computer scientists fade away as members of the various fields rush to make use of the new technologies" (2011, p. 548).

While the potential of computing to transform the humanities has long been understood (Wulf, 1995), it has arguably only come to widespread fruition with the rise of ubiquitous computing and the coalescing of expertise

around the field of digital humanities. At its core, the digital humanities are bound up with “the practice of using technology to create new objects for humanistic interrogation” (Schmidt, 2011). In particular, digital humanities involves researchers creating and accessing digital archives (collections of digitized and born-digital content), and creating and using computational tools in support of scholarly research and analysis on those collections. Such computational tools have a wide spectrum of use as part of the intellectual endeavor of the humanities: enabling the exploration, discovery, and interpretation of sources; assisting in the process of collaboration among scholars, including the production of joint works and analysis; and providing the functionality to visualize data across time and space (Summit on Digital Tools for the Humanities, 2006; Nguyen & Shilton, 2008). As a result of the proliferation of these tools, a number of humanities cyberinfrastructure projects--for example, Project Bamboo, Implementing New Knowledge Environments (INKE), Digital Research Infrastructure for the Arts and Humanities (DARIAH), Infoclio, and Data Service Infrastructure for the Social Sciences and Humanities (DASISH)--have created registries in order to help scholars find and compare these resources. These registries include Digital Research Tools (<http://dirtdirectory.org/>), DH-Tools (<https://de.dariah.eu/dh-tools>), Digital Toolbox (<http://www.infoclio.ch/de/node/130300>) and the Tools E-Registry for E-Social Science, Arts and Humanities (<https://github.com/DASISH/TERESAH>). Research suggests that humanities scholars can adapt well to these tools and technologies, as long as they are tied to “tried, tested, and somewhat traditional research functions” (Brockman et al., 2001, p. vi; Massey-Burzio, 1999). Such functions include keeping abreast of literature in the field; locating, accessing, and using primary sources; developing personal digital collections of primary and secondary source materials; and writing and editing of the scholarly work product (Brockman et al., 2001, p. v).

In the field of digital humanities, various strands of the discourse on the uptake of new technologies have avoided the air of technological determinism by paying attention to the larger social and political forces shaping the future of the humanities. Nowadays, the adoption and use of technologies in the humanities play out against a backdrop of evolving expectations, including the demand that academics embrace new and more efficient forms and models of scholarship (Ayers, 2013). These expectations are driven by concerns about the cost of graduate education, the average time-to-degree for doctorates, the nature of scholarly research output, and a flat or

diminishing research funding structure for the humanities. This attention to the external forces that are shaping the humanities is accompanied by an ongoing concern for the ways in which technology itself (in the form of digital mediation) is changing the methods, theory, and practice of research in the humanities (Berry, 2011). This more theoretical approach constitutes a framework in which the software and computational tools developed as part of the work of digital humanists are recognized and studied as theoretical objects, objects that exist within particular social, cultural, and political contexts. In this worldview, technologies and tools are not viewed as objective, but as socio-technical objects that function both to reify knowledge and to communicate and embody particular arguments (Berry, 2011; Thomas, 2011; Ramsay & Rockwell, 2012).

One practical area where the impact of technology has been felt in the humanities is within the immediate physical research environment of the archive. From the foundation of the modern archive service in the first decades of the twentieth century, researchers have utilized various technologies in the archive reading room. Pencils, note paper, and index cards remain ubiquitous for transcribing and taking notes. In the pursuit of an efficient means of copying primary source material, archivists and scholars have also looked to mechanical and electronic technology including the typewriter, microfilm, and photocopier (Cox, 2011). If in the early 1990s the photocopier was singled out for its impact upon humanities research (Case, 1991a), today that role has been usurped by digital cameras and other mobile devices capable of photographing and scanning materials (Cox et al., 2007; Dooley & Luce, 2010; Miller & Galbraith, 2010; Rutner & Schonfeld, 2012; Long & Schonfeld, 2014). Studies have found that archivists are increasingly welcoming the use of such technological devices in their reading rooms, with their advertised policies in this regard often more restrictive than their actual implementation (Dooley & Luce, 2010; Trace & Karadkar, 2013). Indeed, the use of such technologies is becoming so ubiquitous that as Keeling and Sandlos note, “these days, when you walk into many archives or library reading rooms, you are likely to see as many people with cameras in their hands as with pencils or even keyboards” (2011, p. 423). Thus, the work process is changing from that of reading and extensive note-taking to that of scanning, information gathering, and annotating. In this new model, the scholars’ aim is to accumulate as much information in as short a time as possible, with the filtering of the information and analytical work generally happening away from the archive itself (Rutner & Schonfeld, 2012; Trace & Karadkar, 2013). The challenge with this approach lies in the

resulting information management work that must be carried out in order to organize and integrate the research data so that it remains accessible and usable to the scholar over the long term (Rutner & Schonfeld, 2012).

The capture, organization, and storage of research data and research related materials (such as notes and citations) are part of the information behavior of humanities scholars (Case, 1991a; Case 1991b; Brockman et al., 2001; University of Minnesota Libraries, 2006; Antonijević & Cahoy, 2014; Long & Schonfeld, 2014). Research has established that both graduate students and faculty in the humanities are engaged in “structured and intentional” information behavior such as gathering (collecting, acquiring, and organizing) primary source material and general work products (email, bookmarks, and publications), as well as scanning and digitizing primary sources and, subsequently, building substantial collections of digitally accessible archival materials (University of Minnesota Libraries, 2006, p. 24; Bulger et al., 2011; Wilson & Patrick, 2011). Indeed, such is the extent of this information work that humanities scholars have been called “curators par excellence” (Benardou et al., 2010b, p. 28) and “practiced ‘archive and collection makers’ in their own right” (University of Minnesota Libraries, 2006, p. 24). However, much of the capture, naming, and organization of scholarly materials remains far from ideal, characterized as “haphazard, idiosyncratic, and often bordering on untenable” (University of Minnesota Libraries, 2006, p. 24). When technology is used to aid in the process, researchers often resort to cobbling together solutions to organize, annotate, store, and share research data using a combination of existing software and external services (Bulger et al., 2011). The net effect is that content is often organized and managed in ways that only make sense to the researcher, and which hamper subsequent use or reuse of the data (University of Minnesota Libraries, 2006).

Methodology

The specific research questions for this study (see below) are centered on understanding how scholars capture and manage the primary source materials that they locate in the archives, and how this process could be supported by harnessing emerging digital technology.

- How do scholars working in archive reading rooms gather primary source documents for their research projects?

- How do scholars create a structure for managing the collected primary sources, and associated notes, in order to use this material in their own work environment?
- What are the perceived unmet needs in primary source materials management, and what opportunities exist for harnessing emerging technology (devices and software) to simplify materials management?
- Which commonalities in strategies and tasks, if any, could serve as guidelines for designing software to help scholars manage their primary source materials?

Although the research questions were tailored to requirements gathering, a particular theoretical perspective shaped our overall understanding of the role of technology in the research process. In particular, this research is situated within the broader framework of the social shaping of technology (SST) approach (MacKenzie & Wajcman, 1999; Williams & Edge, 1996). Rejecting the standpoint of both technological and social determinism, the SST perspective holds that the relationship between technology and society is such that one cannot be divorced from the other. Technology is understood as both socially constituted and constructed, a “carrier” and an embodiment of particular political, economic, cultural, and organizational interests (Howcroft, Mitev, & Wilson, 2004, p, 334). Thus, in our research, the relationship between technology (as created and instantiated) and the research process is understood as one of mutual shaping (Williams & Edge 1996).

We designed the investigation to gather data in two stages. We began by administering a short questionnaire that collected demographic information as well as data about the scholars’ technology use practices. Then, we conducted semi-structured interviews to get an in-depth understanding of the scholars’ work practices, with an emphasis on the tools and strategies they use for collecting and organizing primary source documents for long-term use. Our aim was to obtain rich data to aid in eliciting software requirements. The qualitative approach enabled us to uncover new insights into the processes and technologies at play in the collection and management of primary source materials and provided insights into the scholars’ view of these information management practices and associated tools.

Process

Our professional background and connections allowed us to recruit participants who were engaged in archival research in three countries: USA, Germany, and Canada. We solicited individual participants through a number of mechanisms: university bulletin boards, email lists maintained by archives, word-of-mouth publicity (such as asking participants to invite other potential participants), and direct person-to-person contact. We screened respondents by requesting information about the nature of their research projects, the particular archives they had visited, and the type of primary source documents with which they had worked. As the term ‘archive’ is used in different disciplines to describe document collections with different qualities (for example, the physics publications preprint ‘electronic archive’, arxiv.org), this screening process ensured that our respondents had experience with conducting research in archive reading rooms. Respondents who had conducted research only online or those who had worked exclusively with widely published materials (such as books or peer-reviewed publications) were excluded from this study. Each participant met with us for a session lasting approximately ninety minutes. During the session, the respondent signed an informed consent document, completed a questionnaire, and then answered questions in a semi-structured interview format.

The questionnaire consisted of three sections. The demographic section recorded the scholar’s gender, age, discipline, specialization, current position or academic status, and research experience in terms both of the number of years he/she had conducted research in archives as well as the number of archives visited to date. The final question in this section gauged the scholar’s comfort level on a 5-point scale with the general use of digital tools such as digital cameras, smart phones, tablets, notebook and desktop computers, audio and video recorders, and document scanners. Questions in the other two sections paralleled each other in substance, while differing in context. These sections focused on the scholars’ technology use in the archive reading room and in their workspaces for capturing and managing documents in their personal research collections. The scholars identified the devices as well as software used, the providers of these tools, and the frequency of their use.

The semi-structured interview was designed to gain a deeper understanding of the scholars’ research process, the nature of their activities in archive reading rooms as well as in their workspaces, and the tools they employ to capture and manage their personal collections of materials discovered in the archive. The interview was divided into three sections: the first section inquired about a significant research project that the scholar had recently

completed, placing their experiences within the context of this project. The second section focused on the activities performed and their experiences with capturing documents in the archive reading room, while the final section asked about scholars' experiences with using this captured data in their workspaces, after the reading room visit and during the remainder of their project. Participants were encouraged to bring samples of their documents and associated notes if they were willing to share these with us. As the goal of our project is to design software for assisting scholars in the management of their personal collections, our final question asked participants to describe their 'dream app' for managing primary source documents.

We continued interviewing scholars until we met two criteria: (1) we had gathered input from scholars over a broad spectrum of experiences - from undergraduate students to senior professors and (2) we had reached a point of saturation with our data - when scholars' narratives of data management experiences started to overlap significantly, and little or no new information was gleaned from additional interviews. Quantitative analysis of the questionnaire data and qualitative analysis of the interview data (approximately 510 pages of transcripts, or about 200,000 words) allowed the research team to generate composite interpretations for the patterns of management activities employed by scholars to manage their personal primary source collections. Following qualitative data analytic processes (Saldaña, 2013) an initial coding of the interview data was carried out utilizing word processing software. The coding generated preliminary labels, followed by emergent categories centered on project types, work processes, activities, encounters, characterizations (about people, research material, nature of research, archival institutions etc.), and emotional states that surfaced during the research process. A second round of coding helped to refine, deepen, and analyze the initial categories; to compare categories; and to cluster categories in order to create major themes. For example, scholars' particular work processes and activities were further examined through the lens of sequence, behaviors, activities, strategies, and conditions or constraints. Our analysis was informed not just by the data but also by insights from the extant literature on the information behavior of humanities scholars and our own background knowledge and experience with the processes under study.

Demographic Data

As a group, the participants had the common experience of working on projects that required visits to one or more archival institutions in order to consult primary source materials. We surveyed and interviewed twenty-six humanities scholars (R1 – R26). While the study was conducted in three countries, the participants were affiliated with institutions in eight countries: USA, Canada, Great Britain, France, Germany, Switzerland, Austria, and Italy. Of the twenty-six participants, twelve are male and fourteen female, ranging in age from 18 to over 65. The group was nearly equally divided between those below (14 participants) and those above (12 participants) 35 years of age. This spread helped us capture the experiences of those who grew up in the age of personal information devices as well as those who adopted these devices in adulthood.

The diversity in participants' ages is also indicative of their professional status: full professor or equivalent (6 participants), associate professor or equivalent (5 participants), assistant professor, post-doctoral scholar or equivalent (5 participants), doctoral student (5 participants), and master's student (3 participants). Two participants were not affiliated with an academic unit – one respondent was a recent graduate from a baccalaureate program and the other held the title of scientific project leader. The participants' research areas spanned a diverse range of perspectives, methods and knowledge bases, with disciplinary affiliations at times crossing more than one area of expertise: classics, community and regional planning, German studies, literature, religious studies, rhetoric and writing, radio-television-film, Spanish, history of science and medicine, renaissance studies, American studies, art and art history, and history. The fact that half of our participants identified themselves as historians was an expected result given the prevalence of archival research in this community.

The participants also split almost evenly between those who had less (14 participants) and those who had more (12 participants) than twenty years of experience conducting research in archive reading rooms. That humanities scholars are quick to acquire this kind of experience is unsurprising given that they generally begin visiting archival institutions as undergraduate students and accumulate significant experience working in archives prior to graduation. Most of these scholars had also visited archives in multiple countries, and in the interviews they compared experiences across archival institutions as well as national boundaries.

Results

Introduction

Comparable to prior research (Stone, 1982), our data show that the typical work process of humanities scholars encompasses a number of interconnected phases or activities that includes exploratory research (using secondary sources), identifying primary (archival) sources, conducting research in the archive, data collection and management, analysis, and writing. The time that participating scholars spent on a research project was dependent on factors such as professional norms and expectations, the scope of the research, skills acquisition necessary for the process, other work commitments during the time period, and the amount of available research funding. Research projects varied in length from a few weeks in the case of a doctoral student writing a report for a non-profit organization, several months in the case of scholars working on a conference paper, several months to a year in the case of an undergraduate or master's student writing a paper or report, several years for a professor working on an article, up to eight years in the case of doctoral students working on their dissertations or post-doctoral scholars working on follow-up projects, and up to a decade or longer for professors working on book-length projects.

Although we know from prior research that humanities scholars are increasingly utilizing primary source materials that have been digitized and placed online (Chassanoff, 2013), visiting the archive, in person, in search of primary sources is still an integral part of the research process for these humanities scholars. Indeed, such is the importance of the archive to the humanities research process that a participant [R01] dubbed historians, 'archive brats,' due to the amount of time they spent doing research in these institutions. The number of archives that each of our scholars visited over the duration of a project varied, ranging from one to more than thirty institutions. Funding considerations meant that trips to archives out of state or out of the country tended to result in more intensive research visits (measured in days, weeks, or months), occasionally followed by repeat visits to the same archive.

Echoing prior research, we found that the particular research process in which each scholar engaged was individualized, emergent, and non-standard. A history post-doc [R01] referred to her process of doing research as akin to a "giant scavenger hunt" with "lots of different moving parts to it." Further complicating the matter was the fact that in some cases, as Borgman (2015) has noted, humanities scholars are never really 'done' with their

data. These scholars work on research projects and their offshoots over extended periods of time. The longevity of these projects meant that our scholars had to adapt their work processes continually to deal with increasing aggregations of material as well as to adapt to new generations of technologies (hardware and software) that aided them in the research process. It is understandable, therefore, that many of the scholars bemoaned the system (or lack of system) that formed their overall method, a system that was described as exciting and an intellectual puzzle, but also referred to as messy, chaotic, ad hoc, and one based on a lot of trial and error. Nevertheless, the interviews revealed the excitement that scholars felt in working with primary source material, both in terms of engaging with the intellectual content of the documents and the element of surprise that came from what these sources could reveal. A scholar of history of science and medicine [R23] commented: *“I always say that reading primary sources is like talking to children. You never know what they will tell you.”*

For our scholars, having knowledge of the policies and procedures in operation in each archive was seen as a crucial part of preparing for a research trip. While scholars were generally appreciative of the role that archivists played in supporting the research process, the interview data confirmed prior research (Duff et al., 2004) that has established that the scholars’ research process is heavily mediated by practices, policies, and rules in place in the archive reading room. Some scholars expressed frustration at what they saw as a general lack of transparency and consistency about the rules and procedures in and among archival institutions. The situation was only exacerbated when scholars had to take national differences into consideration.ⁱ One policy that met with resistance was the restrictions placed on how many volumes/books or boxes and folders of archival material the scholars could have access to or copy at any one time, a practice that slowed down the research process and hindered the activities of browsing, comparing, and cross referencing multiple sources. On the other hand, some archival practices allowed for a more efficient research process. One example was the increasing availability of internet access in archive reading rooms—a boon for scholars who wanted to check facts, order documents, or look up sources online. The ability to request materials in advance via online request systems was also generally appreciated, although scholars were hampered by the inability to port information across institutional systems.

Interview and questionnaire data shed particular light on the scholars’ information management practices with regard to primary source material. These information management practices took place under a number of

conditions: in parallel with visits to an archive, after visits to an archive or archives were concluded, or through a combined approach. Interview data on information management practices for primary source materials were clustered around the activities of collecting (gathering and organizing), note-taking, and tracking, with digital technology often playing a key role in these processes.

Technology Use – Hardware

Like Palmer and Neumann (2002), we found that these humanities scholars were “deliberate adopters of technology,” selectively taking on systems that “fit into their mutable and creative approaches to scholarship” (p. 110). Data from the questionnaire (see figure 1) established that the most popular general-purpose digital technology that scholars used when working with primary sources was the laptop. The exception was a doctoral student in history [R04] who, due to constraints on her time and finances, limited her activities in the archive to obtaining copies of materials. The laptop was followed in popularity by the photocopier and the scanner, technologies that scholars relied on the archive or their workplace to provide.

Where permitted, many scholars also used digital cameras and smartphones in the archive reading room, primarily in order to gain control over the copying process. In instances where smart phones were allowed in the archive reading room, it was up to the scholar to provide them. This was not always so in the case of digital cameras, with some repositories allowing scholars to use only digital cameras supplied by the archive. Interview data provided a more nuanced look at the perceived affordances of digital cameras and smart phones, and how those affordances affected their use in the scholarly work process. Some scholars preferred to use a digital camera for photographing materials. Reasons included the scholars’ familiarity with digital cameras and their perceived ease of use, a sense that the quality of images captured with digital cameras was superior to that of the smart phone, a concern that there was insufficient storage space on a smart phone (compared to a digital camera), and a preference for the file naming conventions and the import features of the digital camera. Digital cameras were also thought to be better for those scholars, such as art historians and literature scholars, who were interested in capturing material aspects of the documents. In particular, the digital camera’s zoom feature and ability to handle

variable (particularly low) light environments were thought to be better suited to taking very detailed, high quality, close up pictures of a primary source document.

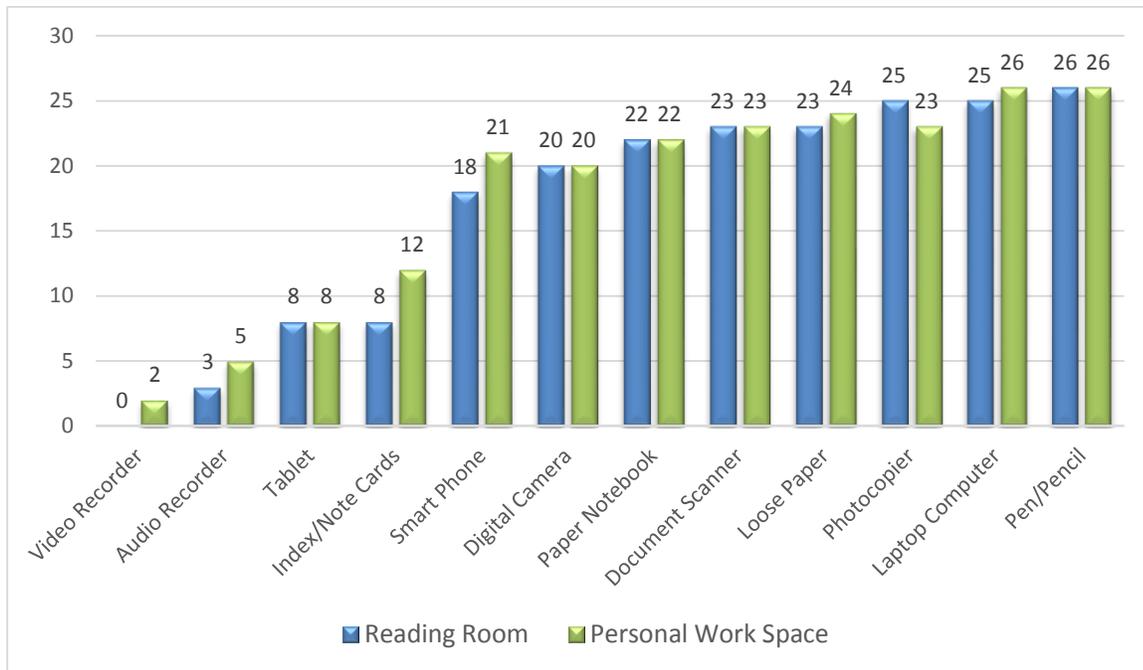


FIG. 1. Technology use by humanities scholars.

Five of the twenty-six scholars did not utilize smart phones at any stage of the research process, a factor that could be attributed to neither age nor gender. In other instances interview data highlighted scenarios where smart phone use was necessarily curtailed. A post-doctoral scholar in art history [R20] owned a smart phone but did not use it as part of the research process due to limitations imposed by the foreign archives (in Germany and Austria) in which she worked. Two other scholars [R13, R22] (a full and an associate professor in history) used smart phones as part of the research process but chose not to use them while working in the archive reading room. Even in cases where smart phones were used in the archive reading room a number of caveats came into play. In some instances the interview data show that archivists as well as scholars treated smart phones solely as communication devices, and nothing more. This finding is supported by the authors’ prior research that established that some archives allow smart phones to be brought into the archive reading room but do not allow them to be used, on site, as image capture devices (Trace & Karadkar, 2013). Similarly, a doctoral student in American studies [R07]

reported carrying a smart phone with him in the reading room but did not view it as a tool for data capture or note-taking. In other instances [R03, R04], smart phones were viewed primarily as backup devices for taking images (in case the battery on the camera died, for example) or as devices useful for taking and demarcating a few key images.

However, a majority of the scholars used smart phones as an integral part of the research process, and in doing so were drawn to specific perceived affordances of these devices. These perceived affordances included the ability to access the device quickly and easily (portability and ubiquity of the smartphone as compared to a digital camera), the ability to use apps on the phone to take a photo or to scan a document, and the utility of the embedded metadata (such as time stamps and location) that was captured by the software as part of the process of taking the image.

Technology Use – Software

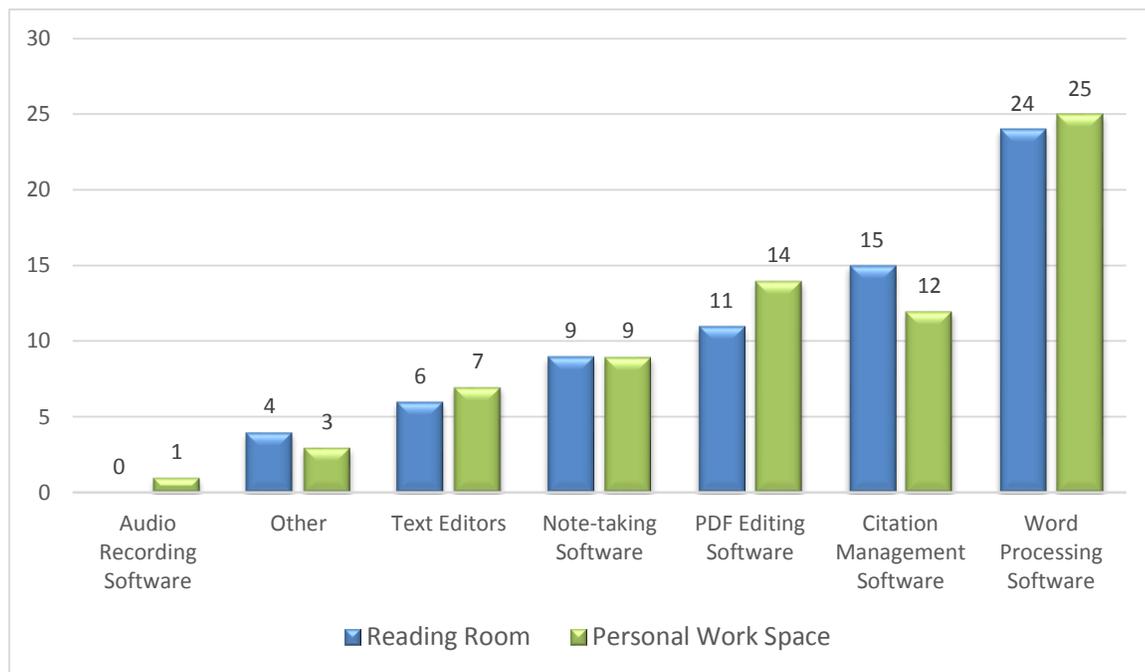


FIG. 2. Software used by humanities scholars for note-taking.

Data from the questionnaire also established that software was ubiquitous for note-taking, both in the archive and in the scholars’ personal workspace, and that a variety of software was utilized in this regard (see figure 2).

The most popular software for note-taking was word processing software (such as Microsoft Word, Word Perfect, Mellel, and Nota Bene), followed by citation management software (such as Zotero, Endnote, Citavi, and Bookends), PDF editing software, note-taking software (such as Evernote), and text editors. Databases, spreadsheet and task management software were also utilized as note-taking tools, both in the archive and in the scholar’s personal workspace. Audio recording software and voice recognition software were only used for note-taking outside of the archive reading room.

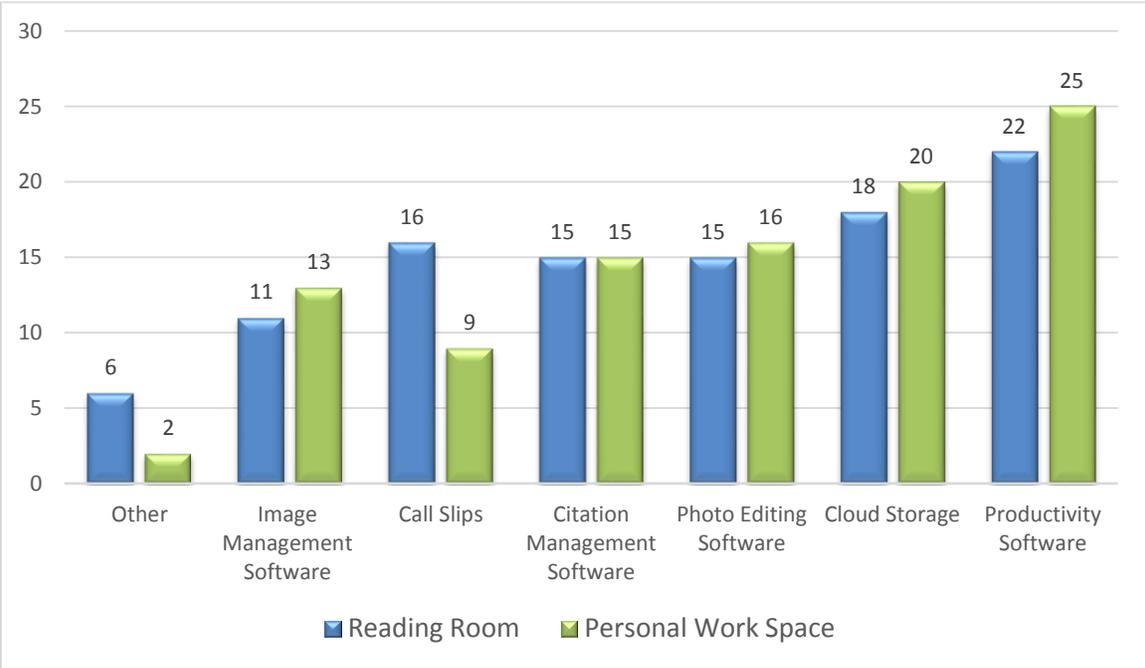


FIG. 3. Software and tools used by humanities scholars to track and manage primary source materials.

In addition (see figure 3), the scholars used software to track and manage their primary source materials and associated notes, with similar software being utilized by scholars whilst working in archive reading rooms and in their personal workspaces. Scholars reported that productivity software (such as word-processing software, spreadsheets, and databases) was used for the organization of sources and notes, cloud storage software for storage and backup, photo editing and image management software for managing and editing digital images, and citation management software for the organization and citation of notes and sources.

Collecting – Gathering

Prior research has established that an integral part of the research process in the humanities involves scholars building personal collections of primary and secondary source materials. Personal collections are a “necessity” for humanities scholars not only for reasons of convenience and accessibility, but because of their need for sustained engagement with their set of sources, with re-reading a “significant part of their interpretive work” (Palmer, 2005, p. 1144; Palmer, Teffeau, & Pirmann, 2009). Our research data shed additional light on how the scholarly activity of collecting (and its two information work primitives - gathering and organizing) plays out in the realm of primary source research.

We found that, in the archive, the scholarly process of gathering usually includes acquiring, when possible, copies of primary source material (texts, manuscripts, letters, photographs, etc.) such that this material can be available as part of the ongoing work of data interpretation and analysis. However, in some of the archival institutions visited by our scholars, the photocopying and/or photographing of documents was strictly prohibited. In such a scenario, scholars had the option either to take notes or to fully or partially transcribe documents, activities that usually lengthened the time scholars had to spend in the archive, and reduced the overall number of documents consulted. Copies of documents were generally preferred over a sole reliance on substitutes such as transcriptions. Where available, the traditional system in place for copying documents (photocopying), was unpopular with the scholars, with scanning generally considered a more acceptable alternative. However, both photocopying and scanning (whether done by the archivist or by the scholar, using machines available in the repository) were considered expensive and time consuming (although the digital delivery of images by the archive ameliorated this to some degree). The end result of photocopying was also viewed as problematic due to the low quality of the image and the difficulty in transporting the extra paper back from a research trip (particularly when traveling abroad). Given the number of scholars who managed their sources digitally, the use of photocopying technology added yet another step to the research process, requiring the scholar to subsequently scan or photograph the photocopies after he/she left the archive. One advantage of scanned documents was that scholars had the ability to subsequently amalgamate the images into one file and annotate the compiled document using PDF reader software.

The ability to gain control over the copying process was a perk that had already been granted to some of the scholars when dealing with microfilmed sources and audio files. For many of the scholars, more recent policy changes in some archive reading rooms allowing for the use of digital cameras and other image capture devices (such as smart phones) had transformed how they made use of the archive. The freedom to copy documents, and associated materials (such as finding aids and call slips), meant that scholars were far more likely to make copies of material even if they were not immediately sure of its relevance to their research. The freedom to create copies (rather than relying on notes or transcripts) also allowed scholars to pay more attention to things other than the content of the document, including visual aspects such as design, layout, handwriting and typography. It also gave scholars more control over how documents were copied, how many images were taken of the same thing (for quality control), and what contextual information was included along with the image (a ruler for measurement, or a color balance card, for example). The result was a process that was generally considered more efficient and cost effective.

However, the degree to which scholars embraced the copying of source material was directly related to how the scholar wished to use his/her time in the archive. The traditional model of work in the archive (what we are calling an ‘*in situ*’ approach) required copious reading and analysis of the source material, and the judicious use of note-taking and transcription in support, or in lieu, of copying. One senior historian [R13] hewed solely to this model - she did not use a digital camera and rarely requested copies of material, instead preferring to read sources on site and take any notes she needed digitally. It was her position as a senior scholar which gave her both the time and the resources to spend this extended time in the archive. However, our data confirm that a new model has emerged (an ‘*ex situ*’ approach) in which the work in the archive is largely centered on information gathering, with the filtering and analysis of the information generally happening away from the archive itself. As one scholar described it [R26]: “*When I did my PhD research – did my thesis research, there weren’t digital cameras. I had to sit in the archives and take notes – a long time. Nowadays, I take as few notes as possible. In fact, often I don’t take any notes at all.*”

For our scholars, the *in situ* and *ex situ* models generally coexisted side-by-side, with scholars generally adopting one or both as circumstances required. As one scholar [R22] noted, the advantage of the traditional

model is that the scholar takes home, not a whole pile of copies, but a draft of a scholarly work. On the other hand, scholars noted that having copies of primary source materials was invaluable when it came to the process of verifying information from source documents (quotations, page numbers etc.). Yet, some scholars expressed mixed feelings about an archival work process devoted mainly to information gathering. For some scholars it was important to have the option of processing the documents on site, in the archive reading room. This processing work generally involved both reading the primary sources (an enjoyable activity because these sources are often “full of surprises” [R23]) and analyzing them through the process of writing. Channeling advice from Umberto Eco,ⁱⁱ one Italian scholar [R23] stated her preference for engaging with the materials on site: “*but I don’t like photographs, not even photocopies. I mean, I do have them. Everybody has them. But I prefer to take notes because this is the way you can really look at things. If you just take a picture, then you never look at it anymore. Maybe they’re useful. I prefer to take notes.*”

Indeed, the point at which such so-called ‘processing’ of sources and associated materials took place was non-trivial. An American studies master’s student [R06] talked about his frustration at not being able to find certain documents when he was in the main write-up phase of his work. The root of this problem was his work process in the archive, where he focused on accumulating material rather than on the immediate processing of the documents. Given the chance to do it over, the scholar stated that he would have used a content-generation tool or word processing software to immediately flag documents that he believed would later figure heavily in analysis and writing.

Collecting - Organizing

For humanities scholars, the value of personal collections of source materials is such that this data has been referred to as their ‘capital’ (Brockman et al., 2001) or ‘dowry’ (Borgman, 2015). Over the course of a research project, these personal collections become “larger and more complex assemblages,” and as a result scholars need to devise “organizational systems and tools for storing and managing the content” (Palmer, Tefteau, & Pirmann, 2009, p. 18). Our interviews with scholars shed detailed light on the nature of these organizing practices. With our scholars generally capturing sources digitally, these practices often involved the creation of a file structure stored

either on a computer hard drive or a cloud storage service. However, before organizing could fully commence, scholars indicated that some initial clean-up or pre-processing of the accumulated paper documents and digital files (primary sources and related documents and notes) was sometimes necessary. This work ensured that scholars had complete intellectual and physical control over the data in advance of any extended analysis. An example was an art historian [R14] who, every evening after returning from the archive exported the images from her camera, relabeled the images with collection information, while at the same time color-balancing the photos and deleting any poor quality images. Similarly, a religious studies scholar [R17], returned from a trip to the archive, downloaded images of sources from his phone, saved them to his laptop, and then backed up all the files. Another method of pre-processing involved scholars aggregating material into a set sequence (chronologically, for example) through converting or exporting images to PDF. By combining numerous sources into one document or digital book, the scholars could then annotate and take notes on the documents as part of the analysis and writing process.

The interview data showed that organizing could happen on the fly in the archive reading room and in a more systematic way before or during the research trip. Organizing helped to create a structure that allowed materials to be subsequently found and also formed one part of the intellectual work of data analysis. The general method of organization involved the creation of hierarchical categories, manifested as a file structure on a laptop or a cloud storage service into which the scholar filed their notes and copies of materials. Categories created by our humanities scholars included project, geographic location, archival repository, call number, topic or theme, source or document type, and/or the semester and date of the research trip. As is clear from the following quotation from a senior historian [R26], these file structures were created in a way that facilitated the scholar's work practice and in a manner that aligned with the nature of the materials with which they were working. *"My data folder tends to be organized by archival repository. In other words, it's what I try and reproduce on my computer, nominally, because I go at things by understanding where they fit in the archival sources. I tend to reproduce the structure of the archival repositories that I'm looking at in terms of gathering data."* In instances where notes (paper and digital) were part of the scholar's personal collection, these documents were also internally structured and formatted in ways that squared with the unfolding research process.

The interview data also shed light on the work of organizing that happened in a scholar's workspace, after the research trip had been completed. In this instance, the organizing process served as a part of the ongoing work of managing the research material during an extended period of writing and analysis. As scholars moved on to a more intensive reading and writing stage, these organizational structures evolved to capture this new reality. One example is the master's student in film [R03] who, after completing his research trip, organized document images into folders on his computer according to topics he intended to use in his final paper. Another example is of the doctoral student in American studies [R07] who created a folder structure on his computer for the raw primary sources (unedited images and associated unfiltered notes, organized by publication or collection), alongside a separate folder structure for edited notes on various themes or subjects related to each dissertation chapter. Several of the scholars created sophisticated relational databases to manage their digital notes and sources at the document level, a system that allowed for the organization of the content, as well as the capture of relevant metadata about the documents themselves and associated information about the repository, call number, etc.

Note-taking (Cross-Cutting Primitive)

In the humanities, as in the sciences, note-taking is an integral part of the scholarly work process, forming a significant part of activities such as organizing, tracking, analysis, and writing (Case, 1991b; Brockman et al., 2001; Palmer, Teffeu, & Pirmann, 2009; Audenaert & Furuta, 2010). All of the scholars in our study took notes on paper, and all but one scholar [R16] took notes using a computer and various kinds of software. For one senior historian [R22], however, paper (a notebook) was the medium of choice for important notes and copies of sources. His skepticism of digital media was a result of a past loss of key research data due to software obsolescence, along with his preference for working with paper when doing close reading and annotating sources.

In the archive, some notes fulfilled an immediate and very practical need, including recording what material had been paged and used, what material a scholar had requested to be scanned or photocopied, or where the researcher paused while working through a collection. If copies were not wanted, warranted, or feasible, notes also took the form of transcribed and translated content as well as the recording of specific metadata related to the documents. In addition, notes took the form of interesting quotations or references to important documents, or

notes on how particular sources could be used in the final research product. For scholars following a more traditional research model, note-taking in the archive also took on more intellectual overtones, acting as a form of analysis through writing. One history post-doc [R01] emphasized the importance of note-taking as an intermediary, non-linear, self-organized space where primary sources, secondary sources, and scholar's notes could be combined in advance of a more sustained writing stage. In this and in other instances, note-taking was one of the types or levels of writing that would eventually form the final written product.

In the scholars' personal workspace, note-taking again supported activities such as tracking, organizing, analysis, and writing. Note-taking encompassed everything from transcription of documents to the creation of metadata about sources that were captured in excel spreadsheets, databases, citation software, or on index cards. With an emphasis on efficiency, at least one scholar [R26] generated notes automatically, bypassing the traditional transcription process through the use of technologies such as OCR (within Adobe Acrobat, or a standalone program) and voice recognition software. In all cases, annotating and making notes on documents was a key part of analysis, and used as a prelude to a more extended writing process. In the case of a master's student in American studies [R06], his note-taking system supported his chronological approach to analysis and writing, with a Japanese accordion notebook being put to use to create a timeline to link primary sources with key events and themes. During the analysis and write-up stages, an interplay or synergy existed between the sources and the notes as the scholars brought both together in the process of constructing new knowledge. While it is easy to compare paper based sources and notes side-by-side, our scholars indicated that the use of split screens and dual monitors make this part of the work of analysis equally feasible in the digital realm.

Tracking (New Cross-Cutting Primitive)

Another notable information management practice that scholars engaged in while working with primary sources is what one scholar [R23] called "keeping traces," and we are calling 'tracking.' At the most basic level, scholars needed an explicit reminder function, a way of denoting which archives were visited and which documents consulted. This was often done using hand written lists or lists created by word processing, spreadsheet, database, or citation management software. Color coding (or other methods such as bolding or

strikethrough, or a table structure) was then used as a way of tracking what had been ordered, what still needed to be ordered, what could not be found, what had been read and transcribed, etc.

With regard to sources, scholars kept track of them as they existed within two main contexts. One was their original archival context (their signature/call number and collection, box, and folder number), a context that scholars needed to capture in order to understand the document as evidence and to aid in its citation. This information could be captured separately from the source through note-taking, by keeping copies of calls slips and finding aids, or by taking a photograph of the relevant citation metadata (metadata which existed on the containers themselves or which scholars recorded on strips of paper and placed beside the documents). This information could also be captured by embedding the information as part of the metadata of the digital image. Scholars also had to keep track of documents within the context of their own personal collections, an activity mainly geared toward subsequently finding information so that it could form part of the analysis and writing process. Techniques here included renaming digital images using photo management software in a way that aided future retrieval, and adding tags or keywords within documents. Other scholars mentioned the importance of the search/find computer application as an auxiliary tool in helping to locate sources and notes.

In both scenarios, scholars also sometimes wished to flag particular sources or notes they thought to be of importance so that they could later retrieve these documents quickly and easily. There were various ways to do this. Objects could be placed in the frame when photographing a document, the background color of an image could be changed using photo-editing software, and notations or flags (symbolic markings such as an asterisk or marks in the form of color highlighting) could be placed besides references to a document within a text. Another scholar [R03] was able to keep track of important notes thanks to the distinctive color (yellow) of the notepaper provided by the archive.

Software Features

The questionnaire and the interview primed the humanities scholars to reflect on their activities working with primary source documents in archive reading rooms and in their workspaces. The final question in the interview provided scholars with an explicit opportunity to describe a ‘dream app’ that would assist them in their scholarly

work with primary source materials (see table 1). In requesting features for such an application, we asked the scholars to ignore any issues of implementation feasibility and focus solely on those features that they would value highly in accomplishing their scholarly objectives.

TABLE 1. Mobile app features requested by humanities scholars.

Category	Feature
Archive policies and workflow	Request documents online from an archival repository
	Access archive's copying policies and acceptable use permissions
	Get information about surroundings (food, coffee)
Scholarly processes and primitives (<i>Collecting – Gathering</i>)	Capture document images
	Compare documents
	Measure and record size of objects
Scholarly processes and primitives (<i>Data Pre-Processing</i>)	Convert among different file/image formats (for example, save as PDF)
	Save multi-page documents
	Relate documents to each other
	Perform OCR to recognize text in documents
	Analyze metadata and find patterns
Scholarly processes and primitives (<i>Collecting – Organizing</i>)	Organize documents
	Provide an overview of stored documents
	Share documents and metadata across projects
Scholarly cross- cutting primitives (<i>Note-taking</i>)	Take notes (text, audio, image)
	Annotate document images
	Automatically capture metadata
Scholarly cross- cutting primitives (<i>Tracking</i>)	Search and locate sources and notes
	Create citation metadata for documents
	Generate citations in various formats
	Keep track of viewed documents from archive
	Scan call numbers
	Demarcate significant documents
	Track cross-references between documents
	Save rights and permissions for documents
Create to-do lists	
User interface features	Customize user interface
	Autocomplete for search
	Provide visualizations of documents and data
Integration and platform support	Provide cloud support (Dropbox)
	Provide platform-specific support (iPad, Android tablet)
	Integrate with Web services

Table 1 outlines the features requested by our participants, which we grouped into three categories during our data analysis phase: (1) features that relate to archive policies and procedures, (2) features that can be coupled to common research behaviors or activities (scholarly processes and information work primitives), and (3) features that connect to the actual design and implementation of the mobile application, including user interface and integration and platform support issues. These findings solidified the results from the interview data, indicating that while scholars are adept at cobbling together existing software to support their practices they would highly value software features that support the information management practices of gathering, pre-processing, organizing, note-taking, and tracking. In addition, this data highlighted the specific features that scholars most wanted from a mobile application: the ability to capture a document image, to take notes (text, audio, image), to synchronize with cloud based services, to annotate documents images, and to be able to locate specific documents within their growing personal collection.

Discussion

Grounded in specific research questions, this qualitative study aimed at capturing the workflow and process of scholars as they worked in the archive in order to help design a piece of digital infrastructure for the humanities. In particular, our research focused on understanding the perceived unmet needs in primary source materials management, and the opportunities that exist for using technology (devices and software) to simplify this process. Yet, in doing so, we were wary of adopting an approach in which tools (as “objects of knowledge”) are understood as “self-identical, self-evident, ahistorical, and autonomous” (Drucker, 2012, p. 86). As a counterpoint, we paid attention to the external and internal forces that are helping to shape the work processes of humanities scholars, including attitudes towards technology and technology adoption. In the case of tool design for the humanities, it is evident that any such work needs to be understood against the backdrop of political, economic, and disciplinary pressures to adopt new models of scholarship that conform to notions of efficiency and progress.

In the process of designing tools, researchers are also encouraged to work within a framework that “embody specific theoretical principles drawn from the humanities” (Drucker, 2012, p. 86). For Unsworth (2000; 2003), tool building for the digital humanities should be based on an understanding of the recurring and interrelated research behaviors (scholarly primitives) at the heart of scholarly work. Adopting this model for our own research, we have uncovered the commonalities in strategies and tasks that can serve as guidelines for designing software to help scholars manage their primary source materials. Furthermore, the notion of mutual shaping (Williams & Edge 1996), means that any technology we design for humanities scholars is seen as both influencing and being influenced by the extant and emerging scholarly research processes.

As members of an academic tribe (Becher & Trowler, 2001), many humanists are inculcated into the idea of the archive as a laboratory of ideas, a place where scholars come to amass evidence of past human activities in the form of primary source documents. For many humanists, the archive reading room is the geographic space in which scholar and archivist engage, with the scholar utilizing the archival infrastructure in order to search for and locate relevant sources. While the reading room was once regarded as a space where scholars would spend their time reading, transcribing, and analyzing documents, the drive for efficiencies and economies in the research process is changing the way that scholarship is conducted in the archive, turning archival visits into what has been dubbed “more of a collection mission” (Rutner & Schonfeld, 2012 p. 12). The introduction of portable digital devices is an integral part of this *ex situ* trend. Digital cameras and smart phones, in particular, have given researchers the ability to copy documents, almost at will, as part of the process of creating and managing their personal collections of primary source documents.

While digital capture technology has matured quickly, our research supports the assertion that the infrastructure for managing the associated document images and metadata is not yet sufficiently developed. Scholars in the present study reported a lack of widely accepted systems for managing their primary source materials, in the absence of which, they have cobbled together mechanisms based on available hardware and software infrastructure. These mechanisms were remarkably similar, whether scholars were working in an archive reading room or in their personal workspace. Overall, scholars’ practices combine diverse, unrelated devices such as digital cameras and smartphones for capturing document images; paper-based or digital notes to contextualize

these images; spreadsheets, small databases, and folder structures to manage and integrate the images and notes; and cloud-based storage for accessing these documents from various devices, regardless of their location.

Although these ad-hoc systems are functional, scholars expressed support for, and interest in, an integrated solution that seamlessly supports their scholarly information activities of gathering, pre-processing, organizing, note-taking, and tracking.

With the US National Archives calling for the development of a mobile app, a ‘Pocket Archivist’ to facilitate the research process in the archive, our research shows that such an application has merit, under certain circumstances. In particular, a mobile application would seem to be a viable tool for those scholars who, from preference or necessity, adopt the *ex situ* model of working in the archive centered on information gathering, where information capture and annotating are the chief order of the day. Our research data show that the smart phone, in particular, is becoming a ubiquitous device in the archive reading room, with scholars already integrating this device into the scholarly research process, at least to some degree. In particular, scholars have been drawn to certain affordances of the smart phone (including portability and the ability to use software to copy or scan documents) as they use this device both in their personal and professional lives.

However, scholars also expressed concerns with the quality of images captured by smart phone cameras. These concerns were manifested along two dimensions: image size or resolution (for example, smart phone sensors typically ranged from 5 to 8 mega-pixels while stand-alone cameras sported 12 to 16 megapixels) and image quality in the low light conditions present in the archive reading room. While these concerns are justified, the specifications for smart phone cameras are improving rapidly. Typical smart phone camera sensor size has improved to between 8 and 15 megapixels recently and phone cameras are increasingly available with image stabilization and continuous focusing, which would enable improved photography in conditions without much ambient light. Stand-alone digital cameras now also come equipped with full-fledged mobile operating systems such as Android-enabled Samsung Galaxy cameras (<http://www.samsung.com/us/photography/galaxy-camera>), providing scholars with the choice of adopting a camera-phone or a full-featured camera with mobile app support. We anticipate that devices such as these will enable scholars to use mobile apps for document management as well as take high quality photographs.

Therefore as mobile phones continue to evolve from being communication devices alone to being computing and information devices that can run multiple applications, access Web-based content, take high-quality photographs, or record videos, their usefulness as a research tool will likely increase (Trace & Karadkar 2013). Furthermore, our data show that humanities scholars can already envision how such an app could fit into their scholarly work process and what such an app could entail. Whatever the particular mobile device, our data show that in order to facilitate the information management practices of humanities scholars, and the subsequent use and reuse of this data, the app should, at minimum, support the following features:

- Document capture - allow for direct capture of images of primary source documents, along with a robust infrastructure to relate those images to two key contexts: the archival context in which the scholar first encountered the document, and the scholarly context in which the document forms part of a personal digital archive.
- Multipage documents – scholars capture images of single page letters as well as brochures, diaries, magazine articles, and myriad other multi-page documents. Any app will need to be able to combine multiple image files into a single multi-page document.
- Metadata – support for the capture of metadata related to the documents is needed. This metadata could include the physical context of the document, such as the box and folder number, and its order within the folder, as well as the intellectual structure in which it resides, such as the series, and sub-series, and file within the collection.
- Multi-modal metadata – mobile devices are not ideal for textual metadata entry. Accepting image or audio metadata options in addition to text will support existing scholarly practice such as taking images of a box and folder prior to capturing the documents within.
- Flexible workflows – while it is tempting to ensure that all available metadata is captured, enabling scholars to enter metadata at the end of a reading room session, or when capturing an image, may provide alternatives that will help scholars find workflows that optimize, or even reduce the time they need to spend in the reading room.

- Metadata upload – generate metadata files in familiar, human-readable formats such as CSV, or Excel. This will enable scholars to visually browse the generated metadata rather than rely on specialized software.
- Organizing sources and notes – give scholars the ability to create a predetermined or emergent structure that associates the images with important scholarly, archival, and bibliographic contexts (project, archive, collection, citation, etc.).
- Note-taking – support the taking of notes at both at the document and the work process level.
- Flagging and annotation – as they take photographs, scholars demarcate documents that have special significance and that the researcher wants to return to easily and quickly. A successful app will enable scholars to tag or classify documents as these are stored.
- Flexible synchronization – storage space on mobile devices can be limited, especially when capturing high-resolution images. It is tempting to upload the device data to remote storage, freeing local space for additional capture. However, it may help scholars to retain some data locally - if only the metadata - to act as a starting point for a future research session. Another aspect to such flexibility would be the ability to upload data to cloud storage or to a connected computer, especially if an internet connection is not available in some reading rooms.
- Automatic citation generation – the metadata captured could be used to generate citations for captured materials, and then uploaded to a cloud service such as Zotero or Mendeley.

Conclusion

As technology has become enmeshed in today’s scholarly workflow, humanities scholars now find themselves “utterly dependent on many different digital tools and resources” (Prescott, 2012, p. 65). No existing application, whether desktop, Web-based, or mobile, currently supports integrated information management for humanities scholars with regard to capturing, curating, and accessing their primary source documents. In order to manage their primary source documents, individual scholars must currently cobble together ad-hoc solutions including the use of digital cameras (capture), DropBox (cloud storage for access from multiple devices), Zotero (bibliography management), plain text or Word documents (note-taking), and Excel Spreadsheets, Adobe FileManager, or

Microsoft Access (document and metadata management). By providing a model of the scholarly work process in the archive, and a set of required features for software development, the STAR project has taken the first step along the path to design integrated software for supporting reliable digital information management practices in the humanities.

The next step in this research agenda is prototype development. Utilizing the findings of this research study a prototype mobile application that streamlines the management, storage, and retrieval of primary sources and facilitates cross-device data integration can be created. The utility and usability of the software can then be tested through an ongoing process of experimental evaluation, as well as by embedding and studying the technology as it is used as part of the scholars' work process. In particular, the usability of the designed software can be evaluated in terms of whether the user interface meets the criteria for usable software (appropriateness of interface components, appropriateness of interaction techniques); whether the internal data storage and organization features of the application meet the quality guidelines for robust software (reliability, repeatability, predictability); and whether the application interoperates as required with external services. Evaluating the impact of the software on scholarly practices can be measured according to recognized quality components for usable software, including learnability, efficiency, memorability, error minimization, and satisfaction (Nielsen, 1993).

Endnotes

ⁱ Such was the impact of policies and procedures on the work of these scholars that it fed into their characterization of the culture of the archive. German and English archivists were singled out as being particularly helpful to patrons (even though it was noted that German and Austrian archives rarely permit researchers to photograph documents), Italian libraries and archives were described as having abundant primary sources but were described as poorly managed, while the French archivist was described as picky, mean, and bureaucratic. For one researcher [R11], the stereotype of the 'pushy American' was actually useful in that he found some European archives were willing to make exceptions to the rules for people, like him, visiting from abroad.

ⁱⁱ Eco has this to say about the use of photocopies as part of the research process: "*Beware the 'alibi of photocopies'!* Photocopies are indispensable instruments. They allow you to keep with you a text that you have

already read in the library, and to take home a text you have not read yet. But a set of photocopies can become an alibi. A student makes hundreds of pages of photocopies and takes them home, and the manual labor he exercises in doing so gives him the impression that he possesses the work. Owning the photocopies exempts the student from actually reading them. This sort of vertigo of accumulation, a neocapitalism of information, happens to many. Defend yourself from this trap: as soon as you have the photocopy, read it and annotate it immediately. If you are not in a great hurry, do not photocopy something new before you *own* (that is, before you have read and annotated) the previous set of photocopies. There are many things that I *do not know* because I photocopied a text and then relaxed as if I had read it” (Eco, 2015, p. 125).

Acknowledgements

This research has been supported by the School of Information at The University of Texas at Austin through the John P. Commons Teaching Fellowship and by the Herzog August Bibliothek, Germany. We thank the scholars who participated in this study, the anonymous reviewers for their constructive comments, and Catherine “The Great” Schneider for her gift of mobile phones for testing software prototypes.

References

- Antonijević, S., & Cahoy, E.S. (2014). Personal library curation: An ethnographic study of scholars’ information practices. *portal: Libraries and the Academy*, 14(2), 287-306.
- Audenaert, N., & Furuta, R. (2010). What humanists want: How scholars use source materials. In *Proceedings of the 10th Annual Joint Conference on Digital libraries (JCDL ’10)* (pp. 283–292). New York: ACM.
- Ayers, E. L. (2013). Does digital scholarship have a future? *Educause Review*, 48(4), 24.
- Becher, T., & Trowler, P.R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Milton Keynes [England]: Society for Research into Higher Education.
- Benardou, A., Constantopoulos, P., Dallas, C., and Gavrilis, D. (2010a). A conceptual model for scholarly research activity. In J. Unsworth, H. Rosenbaum, and K. E. Fisher (Eds.), *iConference 2010 Proceedings* (pp. 26-32). Urbana-Champaign, IL: University of Illinois
- Benardou, A., Constantopoulos, P., Dallas, C., & Gavrilis, D. (2010b). Understanding the information

- requirements of arts and humanities scholarship. *International Journal of Digital Curation*, 5(1), 18-33.
- Berry, D.M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1-22.
- Borgman, C. L. (2009). The digital future is now: a call to action for the humanities. *Digital Humanities Quarterly (DHQ)*, 3(4), 1-30.
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, Massachusetts: The MIT Press.
- Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. J. (2001). *Scholarly work in the humanities and the evolving information environment*. Washington, D.C. Digital Library Federation, Council on Library and Information Resources.
- Bulger, M. E., Meyer, E.T., De la Flor, G., Terras, M., Wyatt, S., Jirotko, M., & Madsen, C.M. (2011). *Reinventing research? Information practices in the humanities*. *Information Practices in the Humanities*. London: Research Information Network.
- Case, D. (1991a). The collection and use of information by some American historians: A study of motives and methods. *Library Quarterly*, 61(1), 61-82.
- Case, D. (1991b). Conceptual organization and retrieval of text by historians: The role of memory and metaphor. *Journal of the American Society for Information Science*, 42(9), 657-668.
- Chassanoff, A. (2013). Historians and the use of primary source materials in the digital age. *American Archivist*, 76(2), 458-480.
- Chu, C. (1999). Literary critics at work and their information needs: A research-phases model. *Library and Information Science Research*, 21(2): 247-273.
- Cohen, D., Fraistat, N. Kirschenbaum, M., & Scheinfeldt, T. (2009). *Tools for data-driven scholarship: past, present, future*. Ellicott City, MD: Maryland Institute for Technology and the Humanities. Retrieved August 30, 2015, from <http://mith.umd.edu/tools/final-report.html>
- Cox, R.J. (2011). Technology's promise, the copying of records, and the archivist's challenge: A case study in documentation rhetoric. In T. Cook (Ed.), *Controlling the past: Documenting society and institutions* (pp. 131-149). Chicago: Society of American Archivists.

- Cox, R.J., Alderfer, L., Arden, B., Barry, T., Bence, A., Berdahl, S., Bleich, B., et al. (2007). Machines in the archives: Technology and the coming transformation of archival reference. *First Monday*, 12(11).
- Dalbello, M. (2011). A genealogy of digital humanities. *Journal of Documentation*, 67(3), 480-506.
- Davidson, C.N., & Goldberg, D.T. (2004). A manifesto for the humanities in a technological age. *Chronicle of Higher Education*, 50(23), B7-B9.
- Delgadillo, R., & Lynch, B.P. (1999). Future historians: their quest for information. *College & Research Libraries*, 60(3), 245-259.
- Dooley, J.M., & Luce, K. (2010). Taking our pulse: The OCLC research survey of special collections and archives. Dublin: OCLC Research.
- Drucker, J. (2012). Humanistic theory and digital scholarship. In M.K. Gold (Ed.), *Debates in the digital humanities* (pp. 85-95). Minneapolis: University of Minnesota Press.
- Duff, W., 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*, 72(4), 472-496.
- Echard, S. (2000). House arrest: Modern archives, medieval manuscripts. *Journal of Medieval and Early Modern Studies*, 30(2), 185-210.
- Eco, U. (2015). *How to write a thesis*. Cambridge, Massachusetts: The MIT Press.
- Fretwell, L. (2013, February 1). Innovation is prologue for National Archives' future. *Fedscoop*. Retrieved August 30, 2015, from <http://fedscoop.com/innovation-is-prologue-for-national-archives-future>
- Freiman, L., Ward, C., Jones, S., Molloy, L., & Snow, K. (2010). Incremental scoping study and implementation plan. Retrieved August 30, 2015, from http://www.lib.cam.ac.uk/preservation/incremental/Incremental_Scoping_Report_062010.pdf
- Howcroft, D., Mitev, N. & Wilson, M. (2004). What we may learn from the social shaping of technology approach. In J. Mingers & L. Willcocks (Eds.), *Social theory and philosophy for information systems* (pp. 329-371). Chichester: Wiley.

- Johnson, C.A., & Duff, W.M. (2005). Chatting up the archivist: Social capital and the archival researcher. *American Archivist*, 68(1), 113-129.
- Keeling, A., & Sandlos, J. (2011). Shooting the archives: Document digitization for historical–geographical collaboration. *History Compass*, 9(5), 423-432.
- Long, M.P., & Schonfeld, R.C. (2014). Supporting the changing research practices of art historians. Ithaca S+ R. New York, N.Y.: Ithaca S + R.
- Lönnqvist, H. (2007). The research processes of humanities scholars. *Advances in Library Administration and Organization*, 25, 175-202.
- Loughridge, B. (1989). Information technology, the humanities and the library. *Journal of Information Science*, 15(4-5), 277-286.
- MacKenzie, D. A., & Wajcman, J. (1999). *The social shaping of technology*. Buckingham: Open University Press.
- Massey-Burzio, V. (1999). The rush to technology: A view from the humanists. *Library Trends*, 47(4), 620-639.
- McCarty, W. (2003). Humanities computing. In M. Drake (Ed.), *Encyclopedia of library and information science* (pp. 1224-1235). New York, Marcel Dekker.
- Miller, L., & Galbraith, S.K. (2010). “Capture and release”: Digital cameras in the reading room. Dublin: OCLC Research.
- Moulin, C., Nyhan, J., Ciula, A., Kelleher, M., Mittler, E., Tadić, M., Ågren, M., Bozzi, A., & Kuutma, K. (2011). *Research infrastructures in the digital humanities*. Technical Report. European Science Foundation.
- Ngah, Z.A., & Sze, G.S. (1997). Information needs and use of humanities researchers: a bibliometric analysis and review of literature. *Kekal Abadi*, 16(3), 1-15.
- Nielsen, J. (1993). *Usability Engineering*. Boston: Academic Press.
- Nguyen, L., & Shilton, K. (2008). Tools for Humanists. In D. Zorich (Ed.), *A survey of digital humanities centers in the United States*. Washington, D.C.: Council on Library and Information Resources.
- Ochsner, M., Hug, S.E., & Daniel, H.D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2), 79-92.

- Owens, T. (2011). Defining data for humanists: Text, artifact, information or evidence? *Journal of Digital Humanities*, 1(1).
- Palmer, C.L. (2005). Scholarly work and the shaping of digital access. *Journal of the American Society for Information Science and Technology*, 56(11), 1140-1153.
- Palmer, C.L., & Cragin, M.H. (2008). Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, 42(1), 163-212.
- Palmer, C.L., & Neumann, L.J. (2002). The information work of interdisciplinary humanities scholars: Exploration and translation. *Library Quarterly*, 72(1), 85-117.
- Palmer, C. L., Tefteau, L. C., & Pirmann, C. M. (2009). Scholarly information practices in the online environment: Themes from the literature and implications library service development. Report commissioned by OCLC Research. Retrieved August 30, 2015, from <http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf?urlm=162919>
- Prescott, A. (2012). Consumers, creators or commentators? Problems of audience and mission in the digital humanities. *Arts and Humanities in Higher Education*, 11(1-2), 61-75.
- Ramsay, S., & Rockwell, G. (2012). Developing things: Notes toward an epistemology of building in the digital humanities. In M.K. Gold (Ed.), *Debates in the digital humanities* (pp. 75-84). Minneapolis: University of Minnesota Press.
- Rutner, J., & Schonfeld, R. C. (2012). Supporting the changing research practices of historians. Final Report from ITHAKA S+ R. Retrieved August 30, 2015, from <http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf>
- Saldaña, J. (2012). *The coding manual for qualitative researchers*. London, England: Sage Publications Ltd.
- Sabharwal, A. (2015). *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Waltham, MA: Chandos Publishing.
- Schmidt, B.M. (2011). Theory first. *Journal of Digital Humanities*, 1(1). Retrieved August 30, 2015, from <http://journalofdigitalhumanities.org/1-1/theory-first-by-ben-schmidt/>

- Schreibman, S., Siemens, R. G., & Unsworth, J. (2004). *A companion to digital humanities*. Malden, Mass: Blackwell Pub.
- Sinn, D. & Soares, N. (2014). Historians' use of digital archival collections: The web, historical scholarship, and archival research. *Journal of the Association for Information Science and Technology*, 65(9), 1794-1809.
- Sternfeld, J. (2011). Archival theory and digital historiography: Selection, search, and metadata as archival processes for assessing historical contextualization. *American Archivist*, 74(2), 544-575.
- Stone, S. (1982). Humanities scholars: Information needs and uses. *Journal of Documentation*, 38(4), 292-313.
- Summit on Digital Tools for the Humanities. (2006). *Summit on Digital Tools for the Humanities: Report on summit accomplishments*. Charlottesville, V.A: The Organizing Committee.
- Thomas, W.G. (2011). What we think we will build and what we build in digital humanities. *Journal of Digital Humanities*, 1(1). Retrieved August 30, 2015, from <http://journalofdigitalhumanities.org/1-1/what-we-think-we-will-build-and-what-we-build-in-digital-humanities-by-will-thomas/>
- Toms, E.G, & O'Brien. H.L. (2008). Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation*, 64(1), 102-130.
- Trace, C.B., & Karadkar, U.P. (2013). Technological devices in the archives: A policy analysis. In *Digital libraries: Social media and community networks* (pp. 35-44). Springer International Publishing.
- University of Minnesota Libraries. (2006). A multidimensional framework for academic support. Retrieved August 30, 2015, from http://www.lib.umn.edu/about/mellon/UMN_Multidimensional_Framework_Final_Report.pdf
- Unsworth, J. (2000). Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? Paper presented at the Symposium on Humanities Computing: Formal methods, experimental practice, Kings College, London. Retrieved August 30, 2015, from <http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html>
- Unsworth, J. (2003). Tool-time, or 'haven't we been here already?' Ten years in humanities computing. *Transforming Disciplines: The Humanities and Computer Science*. Retrieved August 30, 2015, from <http://people.brandeis.edu/~unsworth/carnegie-ninch.03.html>

- Watson-Boone, R. (1994). The information needs and habits of humanities scholars. *RQ*, 34(2), 203-215.
- Wiberley, S.E., & Jones, W.G. (1989). Patterns of information seeking in the humanities. *College & Research Libraries*, 50(6), 638-645.
- Wiberley, S.E., & Jones, W.G. (1994). Humanists revisited: A longitudinal look at the adoption of information technology. *College & Research Libraries*, 55(6), 499-509.
- Wildemuth, B.M., & Case, D.O. (2010). Early information behavior research. *Bulletin of the American Society of Information Science and Technology*, 36(3), 35-38.
- Williams, R., & Edge, D. (1996). The social shaping of technology. *Research policy*, 25(6), 865-899.
- Wilson, J. & Patrick, M. (2011). *Sudamih researcher requirements report: Supporting data management practices in the humanities*. Oxford: Oxford University Computing Services. Retrieved August 30, 2015, from http://sudamih.oucs.ox.ac.uk/docs/sudamih_JISC_Data_Management_Oxford_dissemination.pdf
- Wulf, W.A. (1995). Warning: Information technology will transform the university. *Issues in Science and Technology*, 11(4), 46-52.