Data Practices in Digital History

Rongqian Ma
University of Pittsburgh

Fanghui Xiao
University of Pittsburgh

Abstract

This paper presents an exploratory research project that investigates data practices in digital history research. Emerging from the 1950s and '60s in the United States, digital history remains a charged topic among historians, requiring a new research paradigm that includes new concepts and methodologies, an intensive degree of interdisciplinary, inter-institutional, and international collaboration, and experimental forms of research sharing, publishing, and evaluation. Using mixed methods of interviews and questionnaire, we identified data challenges in digital history research practices from three perspectives: ontology (e.g., the notion of data in historical research); workflow (e.g., data collection, processing, preservation, presentation and sharing); and challenges. Extending from the results, we also provide a critical discussion of the state-of-art in digital history research, particularly in respect of metadata, data sharing, digital history training, collaboration, as well as the transformation of librarians' roles in digital history projects. We conclude with provisional recommendations of better data practices for participants in digital history, from the perspective of library and information science.

International Journal of Digital Curation
2020, Vol. 15, Iss. 1, 21 pp.

1

http://dx.doi.org/10.2218/ijdc.v15i1.597
DOI: 10.2218/ijdc.v15i1.597

Introduction

Traditional historical research has over the long term built its own system of argumentation, documentation, evaluation, and research processes. Recent developments in information and communication technologies, however, have transformed the established conventions of historical research and called for new research paradigms and practices. A "digital history," as coined in the 1990s (Ayers, 1999), becomes a focus of intensive discussions in historical research. This paper seeks to provide a critical review of data practice issues in digital history, so as to raise awareness of the current challenges in historical data practices, compose a possible set of best data practices, and provide recommendations for a more rewarding collaboration mechanism for digital history researchers and project partners. Specifically, this paper addresses the following research questions:

1. From an ontological perspective, what does the notion of "data" and the digital approach add to historical research?

2. What are the current data practices in digital history research? How do data practices in digital history differ from conventions in analog history and add to the modes of scholarship production in the history discipline?

3. What are the major challenges that current digital history researchers face in terms of data practices?

4. Considering the current data practice issues and challenges in digital history, what can be done to improve data practices in digital history?

To investigate the research questions, this paper starts with a review of the evolution of digital history as a field of historical research since the rise of computational tools and digital methods, demonstrating the major factors that have shaped various stages of digital history and justifying the current significance of studying data practices in digital history scholarship. Following the literature review, we proceed to illustrate the overall design of this research study, major methods utilized during the research process, and techniques applied in data collection and analysis. The results section presents the findings of the research from the perspectives of ontology (e.g., the notion of data and use of the digital approach in historical research), workflow (e.g., data collection, processing, preservation, and presentation and sharing), and challenges, before highlighting some essential observations from the findings in the discussion section. This paper concludes by proposing multiple recommendations for better data practices in digital history research, building upon the results of our research study.

Literature Review

During the 1960s and 1970s, with the rise of computing and computational methods, historians started to rethink the history profession and to question if there was a "new history," which relied more on quantitative methods, statistical analysis, and historical computing, rather than the craft and art of historical analysis (Thomas, 2004). Despite

the severe criticism from historians such as Robert Swierenga (1970) that fought against the trend of "quantification" in historical research, proponents for the computational history argued for its promising future and experimented with computational methods to implement digital projects, especially utilizing the new medium of the Internet.

The term "digital history" first appeared in the name of the "Virginia Center for Digital History" established in 1997 and was used by historians Ed Ayers (1999) and William Thomas in projects such as *The Valley of the Shadow: Two Communities in the American Civil War*.[1] The first attempt to define "digital history" took place in 2008 when pioneer scholars such as Daniel Cohen, Michael Frisch, Patrick Gallagher, Steven Mintz, Kirsten Sword, Amu Murrell Taylor, William G. Thomas II, and William J. Turkel presented an online discussion roundtable titled "Interchange: The Promise of Digital History." From this discussion came an early definition for digital history:

> "Digital history is an approach to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems. On one level, digital history is an open area of scholarly production and communication, encompassing the development of new course materials and scholarly data collections. On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human record of the past. To do digital history, then, is to create a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem." (Cohen et al., 2008).

This early definition demonstrated several aspects of imagination that historians had towards a distinct genre of historical scholarship. From a reader's perspective, digital history, in contrast to conventional non-digital historical research, encouraged readers to form their *own* understanding of the past by actively engaging with historical sources within a multimedia virtual environment and explore histories from a hypertextual environment, which "offer[ed] readers multiple ways to navigate digital historical [materials] by following preferred pathways through the Web of hyperlinked texts" (Sabharwal, 2015). For digital historians, doing digital histories also meant the continual adding, annotating, editing, and refining of resources and narrative, which was not particularly familiar to historians of the time.

While pioneer projects of the time, such as *Race and Place: An African American Community in the Jim Crow South* by the University of Virginia[2] and *Victoria's Victoria* at the University of Victoria in British Columbia[3], attempted to critically present the past and engage with the public through the use of information and communication technologies (e.g., the Web), scholars such as Roy Rosenzweig, Douglas Seefeldt, William G. Thomas, and Abby Smith argued for a future model of digital history that emphasizes interdisciplinary collaboration with programmers, information architects, designers, and publishers and "tooling up" the historians to deepen historical analyses (Smith, 2003; AHA, 2004). As Seefeldt and Thomas (2009) demonstrated, digital history as a field must "endeavour to shift the focus of digital historical scholarship away from the product-oriented exhibit or 'web site' and move it more toward the process-oriented work of employing new media tools in our research and analysis – 'doing' digital

---

[1] The Valley of the Shadow: http://valley.lib.virginia.edu/
[2] Race and Place: http://www2.vcdh.virginia.edu/afam/raceandplace/
[3] Victoria's Victoria: https://web.uvic.ca/vv/

history." The choice and implementation of digital tools, and management of historical information and data, as well as the building of research partnerships, therefore became the major concerns for digital historians. There were not only peer-reviewed journals and websites such as *The Programming Historian*[4] targeted at equipping history researchers with necessary programming languages and digital methodologies for critical analysis, historical research centers and cultural heritage institutions such as archives, libraries, and museums also collaborated to create "intentional archives" (Cohen et al., 2008) and various new forms of digital historical scholarship.

Ten years into the evolution of digital history, it is now time to ask what changes have (or have not) taken place and in which direction digital history scholarship might be heading. One aspect is the renewed emphasis on argumentation in digital history. As Blevins (2016) demonstrated, there exists a gap in digital history where many projects "may incorporate scholarly claims and interpretations… but argumentation is rarely their central purpose." The recent white paper (2018) published by the Roy Rosenzweig Center for History and New Media at George Mason University also reinforced that "[d]igital history in various forms often represents a commitment to expanding what history is and can do as a field" (p.1), and therefore, "[i]ncorporating digital history into the profession's historiographical conversations about the past requires historians be able to recognize, read, and engage with those various forms of argument, as well as to incorporate digital components into existing forms" (p.1). In addition to its argumentative aspect, digital history research has also shifted its focus to data, discussing how a data-driven approach influenced the landscape of historical research (ter Braake, Fokkens, Ockeloen, and van Son, 2016) and how the utilization of linked data principles "foster[s] cross-researchers and cross-project collaborations, …allowing for data integration and new types of integrated analysis" (de Boer, Merono Penuela, and Ockeloen, 2016).

Building upon the development of digital history and the increasing importance of data as a subject matter of study in digital history, our study acknowledges the significance of data practices as an emerging but crucial dimension to achieve a more fulfilling digital history scholarship and aims to examine ways to facilitate digital history in this respect. This paper also aims to broadly communicate our findings and discussions to digital history researchers, digital humanists, and librarians, who have been important stakeholders of the digital history enterprise.

Methodology

The research used an iterative exploratory-sequential mixed-methods design (Creswell and Plano Clark, 2018), beginning with and prioritizing the collection and analysis of qualitative data in the first phase, in order to design and develop a quantitative instrument based on our exploratory results, and then deploy the instrument with a larger sample to test and elaborate our initial findings. The reason why we started with a small-scale qualitative phase was that our review of prior research on data practices found little published work on data practices in digital history and thus insufficient guiding theory to frame a larger-scale quantitative survey. We used semi-structured interviews with historians, librarians, and digital humanists to build the foundation for designing a subsequent web-based questionnaire survey administered to a larger

---

4   The Programming Historian: https://programminghistorian.org/

population via the Qualtrics platform (our institution's preferred online survey tool), in order to gather additional case studies and opinions about digital history research.

Interview Design

Acknowledging the high level of idiosyncrasy in historical research, which can be classed as either a humanities or social science discipline, we developed an interview guide based on a thorough and broad investigation of research literature, covering six questions about specific data practices (e.g., data collection, processing, analysis, preservation, presentation, and sharing), the attitudes, thoughts, and concerns of participants towards the "digital history," and the collaboration practices among librarians, history researchers, digital humanists, and other partners. Each interview lasted for about 30 minutes, audio-recorded, and transcribed in its full length for analysis.

Questionnaire Design

Following the interviews, we designed a questionnaire that contained ten multiple choice and one open-ended questions slightly modified based on the results from the interviews. The questionnaire started with a demographic question that asked participants to indicate their research status. It then proceeded to ask participants about their types of digital history projects; the specific data practices in data collection, processing, analysis, preservation, presentation, and sharing; metadata practices; and challenges in digital history research. This questionnaire also contained two questions specifically for librarians, which investigated the roles of librarians in digital history projects and the types of support they can provide.

Data Collection

We used snowball sampling to identify eight interviewees suitable for this study, including five historians and three librarians at various institutions, and conducted both in-person and Skype interviews. Regarding the data collection for the questionnaire, we distributed the questionnaire via three email lists (ACRL history librarians interest group, ACRL digital humanities discussion group, and RUSA History section discussion list) and personal networks, and received a total number of 40 valid responses. Among the 40 responses we gathered, 28 participants identified themselves as librarians (with or without faculty status), five as PhD students in history, two as digital humanists (with one also as history researcher), one as a master's student in history and digital humanist, one as faculty member, two as other (library specialist and paraprofessional staff), and one unknown (i.e., the participant did not indicate their research status). Among the 28 librarians, 19 solely chose librarian, but five also identified as history researchers, two as digital humanists, and two as both.

Data Analysis

We randomly assigned an identification number (e.g., S1, S2, and S3) to each interview and coded the transcripts from three categories: 1) ontology, 2) workflow (i.e., data collection, processing, analysis, preservation, presentation, and sharing), and 3) challenges of digital history. We summarize the key points and identify key words and

sentences in each interview transcript, and classify them into the corresponding categories. For the questionnaire, we apply descriptive statistics to analyze the results.

Results

This section presents our results from interviews and survey responses, organized around the four research questions and covering aspects of ontology (e.g., the notion of data and use of the digital approach in historical research), workflow (e.g., data collection, processing, preservation, and sharing), and challenges of digital history research.

RQ1: From an ontological perspective, what does the notion of "data" and the digital approach contribute to the landscape of historical research?

When it comes to data practices in historical research, the first challenge that historians face is to define "data" in a digital history research context. According to our interview results, "data" in the digital history context is mainly utilized from three perspectives. First, "data" refers to numerical data or data that can be stored in tables; videos or images are not called data. One interviewee demonstrates that they do not use "data" to describe materials such as videos in a digital collection: "We call them video archives" (S6). The second perspective views data and sources as interchangeable, but acknowledges the differences in the use of the terms in various academic cultures. As S4 pointed out, "data and sources are interchangeable to me; my sources are my data. But I recognize that data make more sense in technological contexts…[So] I do often make the distinction for cultural reasons rather than principle reasons." The third perspective separates data from evidence and identifies data as "some kind of interpretation of evidence" (S3). As a digital historian said, "data is not an actual thing; it is something that humans create as a way to help us make comparisons, when we are looking at many things that we know are unique, but we also know share certain traits and qualities, and we want to be able to structure that" (S3). As we can see from the interview results, there has not been a commonly agreed definition and use of "data" in digital history research; people have different perspectives on data based on its relationships with terms such as sources, collections, and evidence. Nevertheless, based on the interviews, our research identified four major areas where the digital adds to the historical research, namely, the scale, analytical methods, collaboration, and the capacity of tracking historical uncertainties.

Scale

Compared with analog historical research, a digital approach makes it possible to conduct research using large-scale datasets or digitized archives. For example, as S3 indicated, "we are able to look at many more records and bring them into analysis in a way that is impossible when you are without using other digital methods." The increase in scale is also demonstrated by the ability of the digital approach to help raise new research questions (S1), identify new resources (S2), and share the results more broadly (S2).

Analytical methods

Digital history also differs from analog history in its use of computational methods for data analysis. S5 pointed out that "one way of looking at digital is that we are talking about computational methods within the field… such as coding or computational analysis of quantitative data." Additionally, due to methodological complexity in digital history research, the analysis has gradually evolved into a collaborative effort. One interviewee stated the necessity of working with research assistants (RAs) and other professionals (e.g., digital humanists and computer scientists) on data analysis (S7).

Collaboration

Digital history research raises the importance of collaboration. Due to its interdisciplinary nature, digital history increasingly calls for collaborations among researchers and practitioners to implement various types of skill sets and expertise (e.g., technical skills, data analytical skills, domain knowledge). Collaboration provides a way to optimize individual skills across the team. For example, to collect and analyze large-scale information about people living in the 1500-1700 Britain, S4, who is a scholar of European histories, collaborated with statisticians, computer programmers, and information professionals to deploy machine learning and crowdsourcing methods.

Tracking uncertainties

Historical narratives are used to be constructed based on fragmented evidence. However, by means of digital approaches and methods, researchers are able to track the large amount of data on the topic across a long period of time, so as to obtain a more comprehensive idea of the evidence landscape, analyze and compare various historical sources, and draw the most convincing and reliable historical conclusions. As illustrated by an interviewee (S3) who is an art historian engaged with data-driven analyses, using statistical modeling helped researchers stay confident about their conclusions. To some extent, the application of digital methods empowers a form of macroanalysis of historical questions, providing a way to track and evaluate the accountability of historical narratives.

Despite all these accounts suggesting the special traits of digital history, the line between the digital and the analog remains fuzzy as it has always been. One participant, who is also a digital historian, claimed during the interview that digital and analog histories are not distinct from each other. "I think the work of history is always trying to find materials that provide some kind of insight about the human past and how it changes over time," said the interviewee, and "digital work is just part of that same process" (S2). We believe this response represents a typical opinion towards digital history that should not be ignored, especially when we discuss "digital history" within both the historical research context and the digital scholarship sphere.

RQ2: What are the current data practices in digital history research? How do data practices in digital history differ from conventions in analog history and add to the modes of scholarship production in the history discipline?

Our interview and survey results also demonstrate the current data practices in digital history research, particularly from the following perspectives.

Data collection

Our research points to diverse data collection methods and techniques in digital history research. Figure 1 summarizes the categories of data collection methods, tools and techniques identified from the interviews. Figure 2 shows the percentage and frequency of using different methods by survey participants, with archival/bibliographic research being the most widely reported approach (30%, n=29), followed by digitization and extracting data from databases. The results suggest a mixed use of conventional and emerging digital methods in digital history research, including traditional archival research, digitization, database search, and crowdsourcing. The diversified landscape of skill sets in digital history demands more in-depth interdisciplinary collaborations, or the upscaling of individual skills.

Given that the majority of survey respondents are librarians, our results also suggest the emerging roles of librarians in digital history research. On one hand, they provide archival research support as conventionally required by historical research; while on the other hand, librarians also provide assistance in the use of digitization (28%) and extracting data from databases (23%), demonstrating their new roles in the transdisciplinary collaborations.
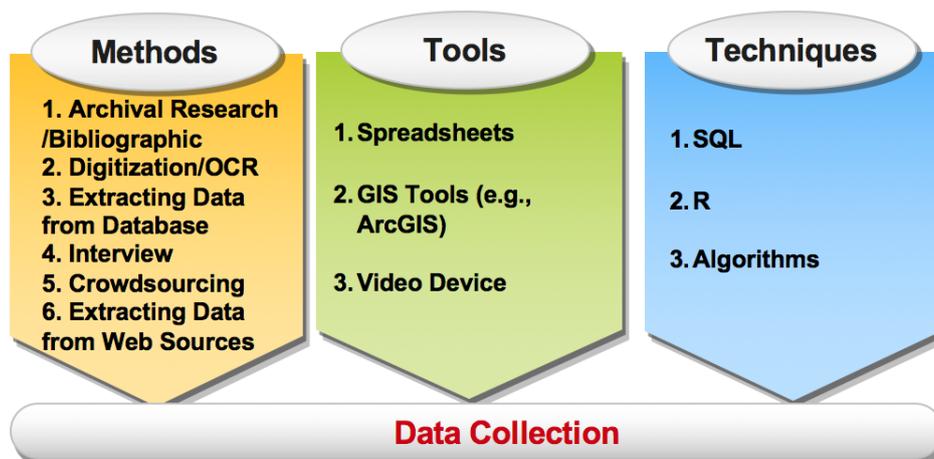
Figure 1. Data collection methods, tools, and techniques.
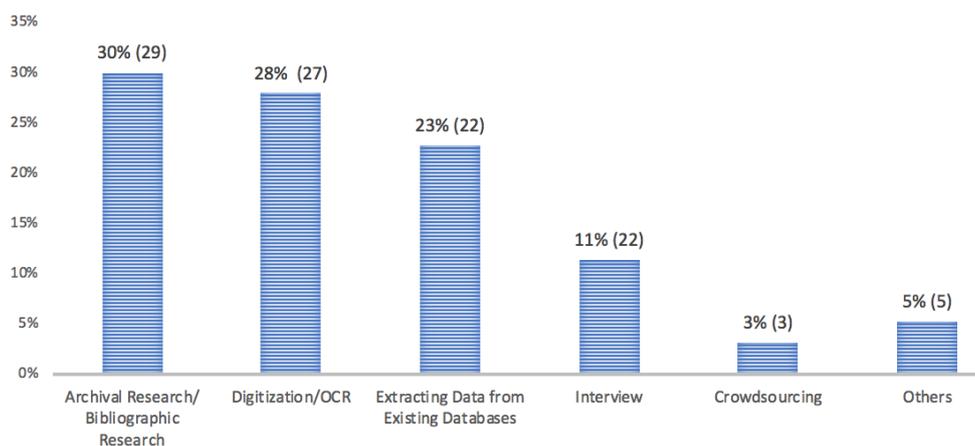
Figure 2. Distribution of data collection methods.

Data processing

Most participants acknowledged that it was an indispensable step of research in digital history to perform quality data processing, and they emphasized three important components of it, namely, the data cleaning, data transformation, and data anonymization (or de-identification). Data cleaning is a time-consuming and essential part of data processing in digital history. For instance, S1 claimed that "the vast majority of digital humanities projects, at least in history, is about cleaning data. You got the data, but you spent the majority of your time cleaning them, getting them into correct format to work with whatever software that you use for analysis or visualization." Data transformation refers to data format conversion. As S3 stated, they performed "a lot of data *enhancement*," to "get a regular export of the data in large .csv files in their original formats, run scripts to reshape the data, and get them into a format that is easier for analysis." Data anonymization and de-identification is defined as an attempt to protect the privacy of subjects. For example, in a digital oral history project, practitioners conducted different forms of de-identification and anonymization practices, such as voice distortion and image mosaic, as a means of protecting the identities of interviewees (S6). Such data processing practices mark the distinction of digital history research from conventional analog historical research, where historians tend to use majority of their time analyzing original archival materials.

Data analysis

With the development of computational methods, an increasing number of digital history projects choose to apply these methods to data analysis. Our interview and survey results show that visualization (22%), textual analysis (22%), and spatial analysis (15%) are the most prevalent methods utilized by participants (Figure 3). In addition to the methods, iteration is another special trait of digital history research, where analyses scatter around the whole data curation cycle. S4 claimed that digital history research involves "repeatedly recollecting and re-analyzing the data," suggesting the iterative nature of digital history research. The various forms of digital analyses shown in the results also emphasize the increasing needs and requirements for coding skills among digital history researchers and practitioners. The quality of data analyses begins to be more closely related to the choice and use of effective technologies and methods in digital history.

Despite the various forms of analysis, certain types of digital history projects initiated by libraries and other cultural heritage institutions do not aim to provide data analyses, but rather, intend to collect and present data for academic and public use (Figure 3). Distinct from research-oriented projects, library-initiated digital history projects focus on curating materials and facilitating access, rather than providing analyses with regard to data. On one hand, such practices intend to promote use by various parties and cater to different needs; on the other, it is also an attempt for libraries to maintain a neutral stance in digital history projects. (S6).
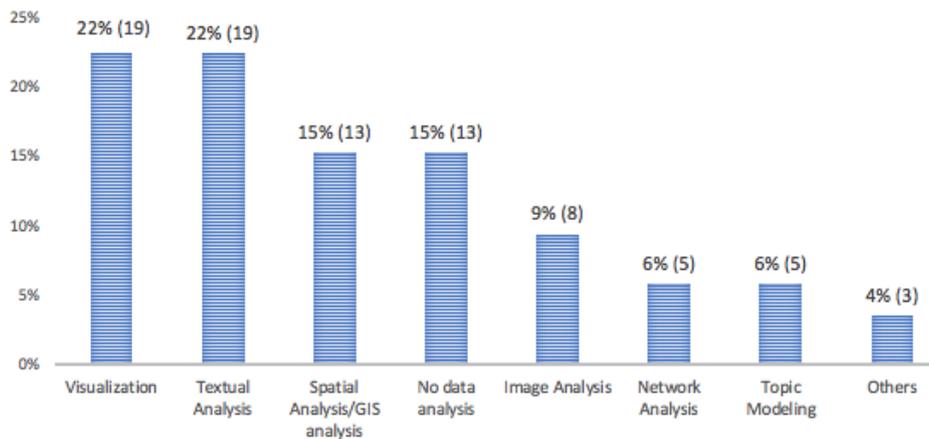
Figure 3. Distribution of data analysis methods.

Data preservation

Our results demonstrate the lack of established preservation plans and awareness in digital histories. Most researchers and practitioners preserve data for their personal use (e.g., 31% of respondents use personal computers and hard drives to preserve data), and there has been no demonstrated needs for complex preservation practices. Some use of metadata (32%) suggests that librarians are promoting more established, sophisticated, and effective preservation practices for digital history; however, the lack of infrastructural support, e.g., the low availability of data repositories, especially for social sciences and humanities, prevents the further development of data preservation in digital histories. Compared with traditional analog historical research, digital history research should not merely preserve the final results data, but also the data generated during the iterative digital processes. The reproducibility of historical research, or more broadly the humanities research, will as a result increase in the near future.
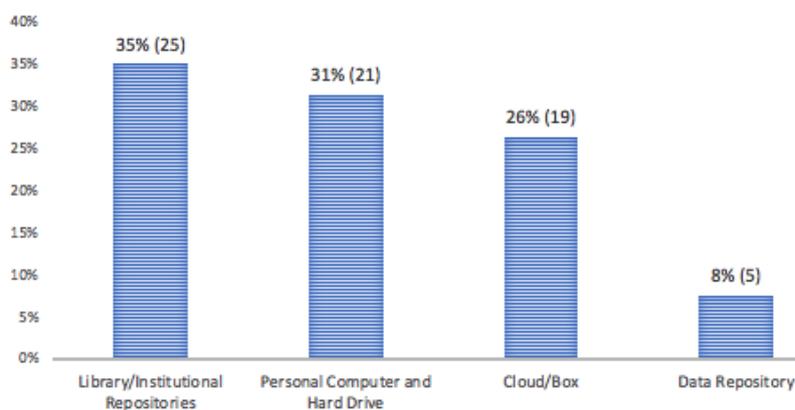
Figure 4. Distribution of preservation locations.
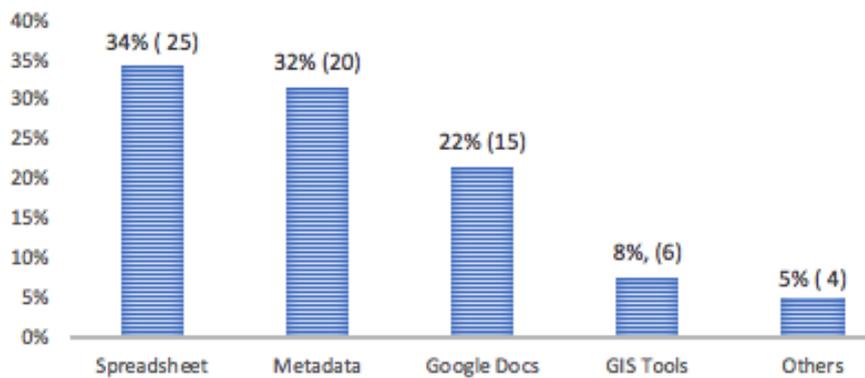
Figure 5. Distribution of preservation tools or techniques.

Data presentation

Monographs, peer-reviewed journal articles, conference papers and presentations have been dominating the publishing landscape of historical research for a long time. Our research findings suggest that websites, digital collections, and conference papers and presentations have become the most commonly used channels for data presentation in digital history research, followed closely by public presentations and journal articles (Table 1). New data presentation channels to communicate research outcomes, such as interactive visualization, maps, databases, and Github, are also gaining popularity.

Table 1. Channels of data presentation and sharing.

Channel	Number	Percentage
Website	23	18%
Digital collection	22	17%
Conference paper and presentation	19	15%
Public presentation and lecture	16	12%
Journal article	14	11%
Interactive visualization	9	7%
Dataset	8	6%
Database	8	6%
GitHub	3	2%
Monograph	2	2%
Dissertation	2	2%
Others	3	2%

RQ3: What are the major challenges digital history researchers face in terms of data practices?

The survey results (Table 2) demonstrate that "lack of technical skills" is the top challenge in digital history, which is consistent with our interviews results. Most interviewees defined their projects as collaborative; particularly for historians, they work with people equipped with computational skills. In the interview session, S5, an experienced digital scholarship librarian, used the example of Geographic Information System (GIS) to illustrate the complexity of using digital tools. "GIS is complicated

enough," as he said, "if you do not have formal training in GIS, there will be a learning curve for individuals to learn how to use it." In addition to the demand for better technical capacity, finding supports and data quality issues have also been top concerns for people involved in digital history projects. As we can see from the table, challenges exist in almost all stages of a data cycle and cover various aspects of data management.

Table 2. Common challenges in digital history projects.

Challenges	Number	Percentage
Lack of technical skills	21	23%
Support (e.g., institutional, departmental, financial)	17	19%
Data quality issues	14	16%
Limitations in existing preservation platforms	12	13%
Difficulties of finding needed data	10	11%
Communication problems among team members	6	7%
Limitations in existing presentation platforms (e.g., interactive visualization)	6	7%
Others	4	4%

Discussion

The collaborative nature of digital history research requires a more complex workflow that incorporates and coordinates the work of various research partners. Our research demonstrates that digital history projects tend to have more iterative and recursive research processes. Figure 6 presents a general workflow model of digital history we constructed based on discussions with interviewees and the survey results.

In this model we can see that most digital history projects go through stages of data collection, data processing, data analysis, data preservation, data presentation, and data sharing. Some stages, however, such as data analysis and sharing, are absent for certain projects. Our results suggest that particularly for digital history projects initiated and maintained by libraries and other cultural heritage institutions, such as museums and archival centers, the main purpose of creating digital history projects is to enrich the resources for research and public use, rather than to provide analyses. Individual projects led by historians and researchers, in contrast, tend to put more emphasis on data analysis. In addition, as the progress of data preservation, presentation, and sharing tend to be more linear and phasic, the earlier stages of the research data cycle from the data collection to data preservation are highly iterative. As illustrated in the general workflow graph, data collection, processing, and analysis keep renewing each other during the process, enabling multiple rounds of thinking, refining, and working with data. Although the general workflow model proposed in this paper does not claim to cover all the complexities within the process of digital history research, we think this model contains the necessary procedures of data practices for fulfilling a digital history project and can also serve as a guiding tool for people involved in digital history to locate issues for future improvement in data practices. In this section, the researchers use this model as a guide to further highlight and discuss additional issues in digital history data practices.

Figure 6. General workflow of digital history research (based on interviews and survey results).

Data Sharing

One problem that has been raised in our study is the lack of data sharing practice in digital history. By "data sharing," we specifically refer to the sharing of original research data, which in the digital history context, may contain archival materials, objects, documents, numeric data, or other forms of resources. Compared with the existence of diverse channels for data presentation, our research found that digital historians usually cannot publicly share their data due to copyright concerns and restrictions on institutional policies. When asked about their data sharing plans, most digital historians responded that they would be willing to share the data if they were allowed. For example, S7 and S1 based their research on data from copyrighted texts and archival collections, and they both mentioned legal concerns and institutional policy issues in relation to sharing the original data. Cultural heritage centers (i.e., libraries, museums, and archives), in contrast, expressed fewer concerns in this respect when engaging with digital history. S6 told the researchers during the interview that one of their major objectives for the oral history project is to create new historical materials for open access; and therefore, the practice of data sharing is a crucial component of the project. From an infrastructural and technical perspective, the lack of data repositories and portals (S6) as well as the inconsistency in data sharing policies also contributes to the low rate of data sharing in practice. As S2 indicated during the interview, one of the challenges they faced was "to ensure a stable home" for the project. With the emerging notion of Linked Data and its increasing application in digital libraries and digital humanities studies, we may expect more promising solutions for data sharing problems; but meanwhile, efforts and commitments to facilitate change in copyright principles, policies, and infrastructure are also important to increase the accessibility and influence, as well as the sustainability of digital history.

Awareness of Metadata

Metadata is significant in helping systems maintain content and facilitating users to discover, organize, access, and share information with others (Riley, 2017). Although presumably useful for complex digital history projects involving multiple participants, our research found that metadata are not commonly practiced in digital history. Digital historians have limited awareness of using metadata to organize, maintain, and share their data. Although some created metadata during their research processes, the purpose

was mainly to help themselves organize and locate information. Rarely was the case where metadata was created based on established schema and mechanisms, or shared along with research data for public reuse. We contend that it is imperative to raise awareness of the importance of metadata among digital historians, which will help digital history participants manage data effectively, preserve data for longer terms, and increase data accessibility among wider audiences.

Digital History Training

The problems associated with current data practices in digital history demonstrate that the training of historians may need to be updated and to incorporate teaching of new knowledge and skill sets for historical research, such as coding and programming skills, collaborative project management and communication, and data literacy. As suggested by a digital scholarship librarian, despite several online programs (e.g., *Programming Historians*) and a few digital humanities initiatives at universities, necessary skill sets for historians to "do" digital history has not been integrated into the formal training for historians (i.e., the PhD program in history). One of our interviewees, who is also a doctoral student in history, demonstrated that she needed to learn from digital humanists at the institution and teach herself to master the necessary skills so as to complete her digital history project (S1). A faculty member in history also stated the need to collaborate with technicians and students with various backgrounds for the project (S7). With digital methods and data practices becoming crucial elements in humanities research, providing necessary training for historians will open up new research opportunities and therefore enrich the landscape of historical research.

Transformation of the Roles of Libraries and Librarians in Digital History

The research findings also illustrate the transformation of libraries' and librarians' roles in digital history research. In addition to providing supporting services, libraries have gradually become an active partner and stakeholder in digital history and librarians start to take up leadership roles in digital history. This transformation has been embodied in digital history projects initiated and maintained by librarians (S6) and has been a rising phenomenon for digital scholarship as well.

Deeper Collaboration

The continued development of digital history also encourages a new mode of scholarship production that heavily relies upon "deep" collaborations. By "deep" we mean the type of collaboration that can potentially achieve a shared understanding of research subjects among participants and one that has transformative influences on disciplinary research conventions. Researchers and practitioners shaped through disciplinary knowledge and skillsets will be able to increase their capacity in addressing transdisciplinary concerns and research topics through deeper collaborations. Data practices in digital history, as we believe the results have shown to some extent, demonstrate the potential to engage in such a level of collaboration and further (re)shape historical scholarship.

Conclusion

As a conclusion, we ask: what can be done to improve data practices in digital history? And more specifically, from the perspective of library and information science, how can libraries contribute? Based on the study, we identify the following ways in which librarians may contribute to data practices in digital history.

1. **Understanding:** The first step towards a better librarian-researcher collaboration is to understand the specific data needs and requirements in the digital history field, and to speak the language of historians. The increased mutual understandings among librarians and digital history researchers on a disciplinary ground facilitate more effective collaborations with potential partners.

2. **Outreach:** It is necessary to train liaison librarians and humanities data specialists to take initiatives and reach out to researchers, so as to work closely with them over the course of the research and assist them with their data problems and challenges.

3. **Specifically-tailored data practice guidelines for historical research:** By means of participating in the collaborative process, librarians will develop a better capability to compose a research data management and curation guideline that is specifically tailored for historical research and which promotes the best data practices for history studies.

4. **Consulting services:** Librarians may provide project management consultancy and technical support for digital history projects, addressing issues such as where to store and preserve historical data, what are the best methods for data sharing, and how to get their research delivered most effectively. Additionally, consultancy over the data ownership and copyright issues will also be rewarding to the development of digital history research.

5. **Data infrastructures:** Responding to the lack of data repositories and sharing platforms, libraries may endeavor to construct and promote quality data infrastructures, so as to further support and strengthen data practices in digital history research.

6. **Training:** Librarians can offer workshops tailored to digital historians that introduce digital methods, teach technical skills such as coding and programming, clarify resources, policies, or principles, and showcase project management, equipping historians with necessary skill sets to perform digital history research.

7. **Leadership in digital history:** Librarians can also deepen engagement with digital history by developing leadership roles in such projects, not merely providing consulting services and supporting services, but rather, initiating and guiding the paradigm shift of digital history research.

For the next step of the research, we aim to create a more comprehensive picture of data curation practices in the digital history field. Specifically, we aim to collect a larger number of data through interviews and the survey instrument to formulate a representative demographic pool of digital history researchers and practitioners.

Afterward, the study aims to portray the data curation landscape of the field and create a Data Curation Profile (DCP) for digital history research (Brandt and Kim, 2014; Witt, Brandt and Cragin, 2009; Wright et al., 2013).

References

American Historical Association (AHA). (2004). AHA workshop: Entering the second stage of online history scholarship. Retrieved from https://www.historians.org/annual-meeting/past-meetings/118th-annual-meeting

Ayers, E.L. (1999). *The pasts and futures of digital history.* Retrieved from http://www.vcdh.virginia.edu/PastsFutures.html

Blevins, C. (2016). Digital history's perpetual future tense. In Lauren F. Klein & Matthew K. Gold (Eds.), *Digital humanities: The expanded field.* Minnesota: University of Minnesota Press.

Brandt, D.S., & Kim, E. (2014). Data curation profiles as a means to explore managing, sharing, disseminating or preserving digital outcomes. *International Journal of Performance Arts and Digital Media, 10*(1), 21-34. doi:10.1080/14794713.2014.912498

Cohen, D. J., & Rosenzweig, R. (2006). *Digital history: A guide to gathering, preserving, and presenting the past on the web*. Philadelphia: University of Pennsylvania Press.

Cohen et al. (2008). Interchange: The promise of digital history. *Journal of American History, 95*(1/2). Retrieved from https://academic.oup.com/jah/article/95/2/452/707613

Creswell, J.W., & Plano Clark, V.L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J.W. & Poth, C.N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage.

de Boer, V., Merono Penuela, A., & Ockeloen, C.J. (2016). Linked data for digital history: Lessons learned from three case studies. *Anejos de la Revista de Historiografía*, (4), 139-162. Retrieved from http://www.victordeboer.com/wp-content/uploads/2016/04/LinkedDataforDigitalHistory_submitted.pdf

Riley, J. (2017). Understanding metadata: What is metadata, and what is it for?: A Primer. *National Information Standards Organization*. Retrieved from https://www.niso.org/publications/understanding-metadata-2017

Robertson, S., & Mullen, L. (2018). *Digital history and argument*. Retrieved from https://rrchnm.org/argument-white-paper/

Roy Rosenzweig Center for History and New Media. (2018). Digital history and argument (white paper). Retrieved from https://rrchnm.org/wordpress/wp-content/uploads/2017/11/digital-history-and-argument.RRCHNM.pdf

Sabharwal, A. (2015). *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Waltham, MA: Chandos Publishing. doi:10.1016/C2014-0-02808-0

Seefeldt, D., & Thomas, W.G. (May, 2009). What is digital history? *Perspectives on History*. Retrieved from https://www.historians.org/publications-and-directories/perspectives-on-history/may-2009/what-is-digital-history

Smith, A. (2003). New-model scholarship: Destined for the dustbin of history. *Perspectives, 47*(7). Retrieved from www.historians.org/Perspectives/issues/2003/0310/0310vie1.cfm

Swierenga, R.P. (ed.) (1970). *Quantification in American history*. New York, NY: Atheneum.

ter Braake, S., Fokkens A., Ockeloen N., van Son C. (2016). Digital history: Towards new methodologies. In Bozic B., Mendel-Gleason G., Debruyne C., & O'Sullivan D. (Eds.), *Computational history and data-driven humanities* (pp. 23-33). Springer, Cham. doi:10.1007/978-3-319-46224-0_3

Thomas, W.G. (2004). Computing and the historical imagination. In Susan Schreibman, Ray Siemens, & John Unsworth (Eds.), *A companion to digital humanities* (pp. 56-68). Malden, MA: Blackwell Publishing.

Witt, M., Carlson, J., Brandt, D.S., & Cragin, M.H. (2009). Constructing data curation profiles. *International Journal of Digital Curation, 4*(3), 93-103. Retrieved from doi:10.2218/ijdc.v4i3.117

Wright, S.J., Kozlowski, W.A., Dietrich, D., Khan, H.J., Steinhart, G.S., & Mclntosh, L. (2013). Using data curation profiles to design the datastar dataset registry. *D-Lib Magazine, 19*(7/8), 37-49. doi:10.1045/july2013-wright

Appendix

Interview Questions

1. Please indicate your academic affiliation and choose your research status.

- Senior history major

- Master's student in history

- Ph.D. student in history

- Faculty member. (Please also indicate the title here _____)

- History researcher
- History liaison librarian
- Librarian (in general)

2. Have you ever participated in digital history research?

- Please introduce the project if the answer is yes, including information such as the duration of the research, research topic, and participants.
- What was your role(s) in the project?

3. Compared with conventional analog historical research, what do you think the "digital" adds to (or not) to historical research?

4. Discuss data practices in digital history research.

- What are your thoughts on the notion of "data," in the context of digital history?
- Could you discuss your data collection approaches, methods, and techniques?
- Could you discuss your data processing practices?
- Could you discuss your data analysis practices (if any)?
- Could you discuss your data preservation approaches?
- In terms of the research outcomes, could you discuss your data presentation, sharing, and publishing approaches?
- Have you encountered any data curation or management challenges in digital history research? Please elaborate.

5. What was the workflow of the digital history project that you participated in? Did you collaborate with other partners (e.g., other researchers, students, librarians, data scientists) on the project? What was your role(s)?

6. Could you discuss if you encounter any issues or challenges in your digital history projects?

6. For librarians:

1. What was your role(s) and work in the digital history project(s)?

2. How would you describe your relationships with partners in digital history research (e.g., historians, information and computer scientists)?

3. Do you have recommendations, from a librarian's perspective, for improving data practices in digital history research?

Questionnaire

This study aims to examine data practices in digital history research. In this questionnaire, you will be asked a few questions with regards to your experiences in digital history projects, covering data management approaches and techniques in various stages of a data cycle. The completion time of the questionnaire is about 5-10 minutes, and your answers will be completely anonymous. If you have participated in digital history research, please continue with the questionnaire. If not, you may discontinue it.

1. Please choose your research status from the following options. You can choose more than one option.

- Senior history major

- Master's student in history

- Ph.D. student in history

- Faculty member. (Please also indicate the title here _____)

- History researcher

- Digital humanist

- Librarian

- Others, please specify:_____

2. Please choose your research type. (Please answer with your current or most recent project.)

- Dissertation or thesis

- Research project (Individual scholar as PI)

- Library/Museum/Archival/Institutional project

- Others, please specify:_____

3. What are your data collection approaches, methods, and techniques? (multiple choices)

- Archival Research/Bibliographic Research

- Digitization/Optical Character Recognition (OCR)

- Crowdsourcing

- Extracting data from existing databases or other types of resources

- interviews

- others, please specify: _____

4. How are you approaching data analysis? (multiple choices)

- Spatial Analysis/GIS analysis

- Network Analysis

- Image Analysis

- Textual Analysis

- Topic Modeling

- Visualization

- No data analysis

- Others. Please specify: _____

5. Where did you (or will you) preserve your data? (multiple choices)

- Library/institutional repositories

- Cloud/Box

- Personal Computer and hard drive

- Data repository

- Others, please specify:_____

6. What techniques and technologies did you use?

- Spreadsheet

- Metadata

- Google Docs

- GIS tools

- Others, please specify:_____

7. If you created metadata for your project, what were your purposes? (multiple choices)

- For data discovery purposes

- For data organization purposes

- For data analysis purposes

- For data preservation purposes

- For data sharing purposes

- Other purposes, please specify: _____

8. n which form(s) are your data shared and presented? (multiple choices)

- Monograph

- Journal articles and conference papers

- Public presentations

- Dissertation

- Website

- Digital Collection

- Database

- Dataset

- Interactive Visualization

- GitHub

- Others, please specify: _____

9. What challenges did you encounter in digital history projects? (multiple choices)

- Data quality issues

- Difficulties of finding needed data

- Data quality issues

- Lack of technical skills

- Communication problems among team members

- Support (e.g., institutional, departmental, financial)

- Limitations in existing preservation platforms

- Limitations in existing presentation platforms (e.g., interactive visualization)

- Others, please specify: _____

If you are a librarian, please click "Continue"; if not, please click "Skip".

- Continue

- Skip

10. For librarians: What kinds of help do you provide for digital history projects? (multiple choices)

- Create new data (digitization, digital sources/collections)

- Assist finding data or locating data sources

- Technical consulting

- Provide advice on the project plan

- Provide data management services

- Copyright consulting

- Provide workshops to improve data literacy

- Others. Please specify: _____

11. For all the participants: What do you think libraries and librarians can do to better support digital history research?

12. Please let us know if you have other comments. Thank you very much!