



In The Library With The Lead Pipe

An open access, open peer reviewed journal

[HOME](#) [ABOUT](#) [AUTHORS](#) [ARCHIVES](#) [CONDUCT](#)

[SUBMISSION GUIDELINES](#) [CONTACT](#) [SEARCH](#)

2017
MAY
17

Malina Thiede

1 Comment

PRESERVATION IN PRACTICE: A SURVEY OF NEW

YORK CITY DIGITAL HUMANITIES RESEARCHERS

In Brief

Digital Humanities (DH) describes the emerging practice of interpreting humanities content through computing methods to enhance data gathering, analysis, and visualization. Due to factors including scale, complexity, and uniqueness, the products of DH research present unique challenges in the area of preservation. This study collected data with a survey and targeted interviews given to New York City metro area DH researchers intended to sketch a picture of the methods and philosophies that govern the preservation efforts of these researchers and their institutions. Due to their familiarity with evolving preservation principles and practices, librarians are poised to offer expertise in supporting the preservation efforts of digital humanists. The data and interviews described in this report help explore some of the current practices in this area of preservation, and suggest inroads for librarians as preservation experts.

By Malina Thiede (with significant contributions from Allison Piazza, Hannah Silverman, and Nik

Introduction

If you want a definition of Digital Humanities (DH), there are hundreds to choose from. In fact, Jason Heppler's whatisdigitalhumanities.com alone offers 817 rotating definitions of the digital humanities, pulled from participants from the Day of DH between 2009-2014. A few of these definitions are listed below:

Digital Humanities is the application of computer technology to make intellectual inquiries in the humanities that either could not be made using traditional methods or are made significantly faster and easier with computer technology. It can include both using digital tools to make these inquiries or developing these tools for others to use. –Matthew Zimmerman

DH is the study, exploration, and preservation of, as well as education about human cultures, events, languages, people, and material production in the past and present in a digital environment through the creation and use of dynamic tools to visualize and analyze data, share and annotate primary sources, discuss and publish findings, collaborate on research and teaching, for scholars, students, and the general public. –Ashley Sanders

For the purposes of this article, digital humanities will be defined as an emerging, cross-disciplinary field in academic research that combines traditional humanities content with technology focused methods of display and interpretation. Most DH projects are collaborative in nature with researchers from a variety of disciplines working together to bring these complex works to fruition. DH projects can range from fairly traditional research papers enhanced with computing techniques, such as text mining, to large scale digital archives of content that include specialized software and functionality.

Due to the range of complexity in this field and the challenges of maintaining certain types of digital content, long-term preservation of DH projects has become a major concern of scholars, institutions, and libraries in recent years. While in the sciences, large scale collaborative projects are the norm and can expect to be well funded, DH projects are comparatively lacking in established channels for financial and institutional support over the long term, which can add another layer of difficulty for researchers. As librarians at academic institutions take on responsibility for preserving digital materials, they certainly have a role in ensuring that these DH projects are maintained and not lost.

For the purposes of this paper, a digital humanities project will be broadly defined as cross-disciplinary collaboration that manifests itself online (i.e. via a website) as both scholarly research and pedagogical resource using digital method(s). Methods can include, but are not limited to, digital mapping, data mining, text analysis, visualization, network analysis, and modeling.

Literature Review

The Library of Congress's (n.d.) catchall definition of digital preservation is "the active management of digital content over time to ensure ongoing access." Hedstrom (1998) offers a more specific definition of digital preservation as "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable." Digital preservation is a complex undertaking under the most favorable conditions, requiring administrative support, funding, personnel, and often specialized software and technology expertise.

Kretzschmar and Potter (2010) note that digital preservation, and, in particular, digital humanities preservation, faces a "stand-still-and-die problem" because it is necessary to "continually...change media and operating environments just to keep our information alive and accessible." This is true of preserving most digital objects, but the complex, multi-faceted nature of many DH projects adds additional layers of complexity to the already challenging digital preservation process. Zorich (2008) lists other components of the "digital ecosystem" that must be preserved in addition to the actual content itself: "software functionality, data structures, access guidelines, metadata, and other...components to the resource."

Kretzschmar and Potter (2010) lay out three seemingly simple questions about preserving digital projects: "How will we deal with changing media and operating environments? Who will pay for it? And who will do the work?" whose

answers are often difficult to pin down. When working with DH projects, 'what exactly are we preserving?' may also be an important question because as Smith (2004) notes that "there are... nagging issues about persistence that scholars and researchers need to resolve, such as... deciding which iteration of a dynamic and changing resource should be captured and curated for preservation." In 2009, Digital Humanities Quarterly published a cluster of articles dedicated to the question of "doneness" in DH projects. Kirschenbaum (2009) notes in the introduction to the cluster that "digital humanities...[is] used to deriving considerable rhetorical mileage and the occasional moral high-ground by contrasting [its] radical flexibility and mutability with the glacial nature of scholarly communication in the fixed and frozen world of print-based publication." Unlike some digital assets that undergo preservation, DH projects and the components thereof are often in a state of flux and, indeed, may never truly be finished. This feature of DH projects makes their preservation a moving target. Kretzschmar (2009) detailed the preservation process for the Linguistic Atlas Project, a large scale DH project that spanned decades, explaining "we need to make new editions all the time, since our idea of how to make the best edition changes as trends in scholarship change, especially now in the digital age when new technical possibilities keep emerging." Another example of a DH project that has undergone and continues to undergo significant revisions is described in Profile #5 below.

In addition to the particular technological challenges of preserving often iterative and ever-evolving DH projects, there are structural and administrative difficulties in supporting their

preservation as well. Maron and Pickle (2014) identified preservation as a particular risk factor for DH projects with faculty naming a wide range of entities on campus as being responsible for supporting their projects' preservation needs, which suggested "that what preservation entails may not be clear." Bryson, Posner, St. Pierre, and Varner (2011) also note that "The general lack of policies, protocols, and procedures has resulted in a slow and, at times, frustrating experience for both library staff and scholars." Established workflows and procedures are still not easily found in the field of DH preservation, leading scholars, librarians, and other support staff to often attempt to reinvent the wheel with each new project. Other difficult to avoid problems noted across the literature are those of staff attrition and siloing.

Although rife with challenges, the preservation of DH projects is far from a lost cause, and libraries have a crucial role to play in ensuring that, to some degree, projects are successfully maintained. The data and interviews summarized in this paper reveal how some of these projects are being preserved as well as their particular difficulties. There are certainly opportunities for librarians to step in and offer their preservation expertise to help scholars formulate and achieve their preservation goals.

Methodology

The methodology for this project was influenced by time frame and logistics. Initially the project was slated to be completed within five months, but the deadline was later extended to nine months. Because it would have been difficult to interview multiple individuals across New York

City within the original time frame, we decided on a two phase approach to conducting the survey, similar to Zorich's methodology, where an information gathering phase was followed by interviews (Zorich, 2008). The survey involved (1) conducting an online survey of NYC faculty members engaged in digital humanities, and (2) performing in-person or phone interviews with those who agreed to additional questioning. The survey provided a broad, big picture overview of the practices of our target group, and the interviews supplemented that data with anecdotes about specific projects and their preservation challenges. The interviews also provided more detailed insight into the thoughts of some DH scholars about the preservation of their projects and digital preservation in general.

The subjects of our survey and interviews were self-selected faculty members and PhD candidates engaged in digital humanities research and affiliated with an academic institution within the New York City area. This population of academics was specifically targeted to reach members of the DH community that had access to an institutional library and its resources. We limited our scope to the New York City for geographic convenience.

We targeted survey respondents using the NYC Digital Humanities website as a starting point. As of October 2015, when the selection process for this project was underway there were 383 members listed in the NYC Digital Humanities online directory. An initial message was sent to the NYCDH listserv on June 3, 2015, and individual emails were sent to a subset of members in June 15, 2015. We approached additional potential survey respondents that we knew fit our criteria via email and Twitter.



Figure 1: NYC Digital Humanities Logo

Survey

The survey tool was a 34-item online Qualtrics questionnaire asking multiple choice and short answer questions about the researchers' work and their preservation strategies and efforts to date. The survey questions were developed around 5 specific areas: background information about the projects and their settings, tools used, staff/management of preservation efforts, future goals, and a query about their availability for follow up interviews. As all DH projects are unique, respondents were asked to answer the questions as they pertain to one particular project for which they were the Principal Investigator (PI).

Interviews

Interviewees were located for the second phase of the research by asking survey respondents to indicate if they were willing to participate in a more in-depth interview about their work. Interested parties were contacted to set up in-person or conference call interviews. The interviews were less formal and standardized than the survey, allowing for interviewees to elaborate on the particular issues related to the preservation of their projects. Each interview was recorded but not fully transcribed. Team

members reviewed the recordings and took detailed notes for the purpose of comparing and analyzing the results.

Limitations

Although the scope of this project was limited to a particular geographic area with a large population base, the sample size of the survey respondents was fairly small. The institutions of all but three respondents are classified as moderate to high research activity institutions according to the Carnegie Classifications. These types of institutions are by no means the only ones involved in DH work, but the high concentration of respondents from research institutions may indicate that there is greater support for DH projects at these types of institutions. As a result, this paper does not provide much discussion of DH preservation practices at smaller baccalaureate or masters institutions with a stronger emphasis on undergraduate education.

A Note about Confidentiality

Individuals who participated in the online survey were asked to provide their names and contact information so we could follow-up with them if they chose to participate in the interview. Individuals who took part in the interviews were guaranteed confidentiality to encourage open discussion. All findings are reported here anonymously.

Survey Results

The survey was live from June 3, 2015 to July 10, 2015. In total, 18 respondents completed the survey.

Demographics of the Faculty Engaged in Digital Humanities

Our survey respondents represented 10 New York City academic institutions, with the most responses coming from Columbia University. Department affiliations and professional titles are listed below (figure 2).

Institutional Affiliation	# of respondents
Columbia University	5
CUNY Graduate Center	3
New York University	2
Bard Graduate Center	1
Hofstra University	1
Jozef Pilsudski Institute of America	1
New York City College of Technology	1
Queensborough Community College	1
St. John's University	1
The New School	1

Department Affiliation	# of respondents
Library/Digital Scholarship Lab	7
English	4
History	3
Art History	2
Linguistics	1
Unreported	1

Academic Titles	# of respondents
Professor	4
Assistant Professor	3
Associate Professor	2
Adjunct/Lecturer	2
Digital Scholarship Coordinator or Specialist	2
PhD Candidate	2
Director	2

Chief Librarian	1
-----------------	---

Figure 2: Survey respondent demographics (n=18)

We asked respondents where they received funding for their projects (figure 3). Responses were split, with some respondents utilizing two funding sources.

Funding Source	# of respondents
Institutional funding	28%
Grant funding	22%
Personal funds	17%
Institutional and grant funding	17%
No funding	11%
Institutional and personal funds	6%

Figure 3: Funding Source

DH Project Characteristics

As previously mentioned, respondents were asked to choose one digital humanities project in which to answer the survey questions. Questions were asked to determine the number of people collaborating on the project and the techniques and software used. The majority of respondents (88%) were working collaboratively with one or more colleagues (figure 4).

# of collaborators	# of respondents
2-3 collaborators	33%
6+ collaborators	33%
0 collaborators	22%
4-5 collaborators	11%

Figure 4: Collaborators involved in DH project (n=18)

The techniques utilized are listed in figure 5, with 61% of projects utilizing more than one of these techniques.

Technique	# of projects
Data Visualizations	39%
Other*	32%
Data Mining and Text Analysis	28%
Geospatial Information Systems (GIS)	22%
Network Analysis	17%
Text Encoding	11%
3-D Modeling	6%

*maps, interactive digital museum exhibition,

audio (2), software code analysis, data analysis tools, OHMS (Oral History Metadata Synchronizer)

Figure 5: Techniques used in DH project (n=18)

The techniques mentioned above are created with software or code, which can be proprietary, open-source, or custom. Respondents utilized a mix of these software types, with 33% of respondents saying that they used proprietary software in their projects, 89% report using open-source software, and 33% used custom software. A list of software examples can be found in figure 6.

Proprietary Software	Open-Source Software
Adobe Photoshop (2)	WordPress (6)
Adobe Dreamweaver	Omeka (3)
Adobe Lightroom	Python (2)
Google Maps	MySQL (2)
TextLab	Timeline.js (2)
SketchUp	QGIS (2)
	Weebly
	DSpace

Figure 6: Software utilized by respondents

Knowledge of Preservation

33% of respondents reported that they had formal training in digital preservation, which the authors intended to mean academic coursework or continuing education credit. Informally, respondents have consulted numerous resources to inform preservation of their project (figure 7).

Source	Percent
Published scholarly research	72%
Colleagues or informal community resources	66%
Digital Humanities Center, library/librarian, archivist	50%
Grey literature	44%
Professional or scholarly association sponsored events	22%
Conferences	33%
Campus workshops or events	11%
None	6%

Figure 7: Sources consulted to inform Preservation

Project Preservation Considerations

Preservation of their DH project was considered by the majority (72%) of respondents. When asked who first mentioned preservation of their

project, 93% of those who had considered preservation said either they or one of their collaborators brought up the issue. In only one instance did a librarian first suggest preservation, and there were no first mentions by either funder or host department.

The majority of initial preservation discussions (53%) took place during the project, with 39% taking place before the project began, and 8% after project completion.

When asked to consider how many years into the future they see their project being usable and accessible, the majority (56%) said 5+ years, followed by 3-4 years (22%), and 17% were unsure. One respondent noted they were not interested in preservation of the project.

Preservation Strategy

Version control, migration, metadata creation, emulation, durable persistent media, and bit stream preservation are just a few strategies for preserving digital materials. We asked respondents to rate each strategy by importance (figure 8).

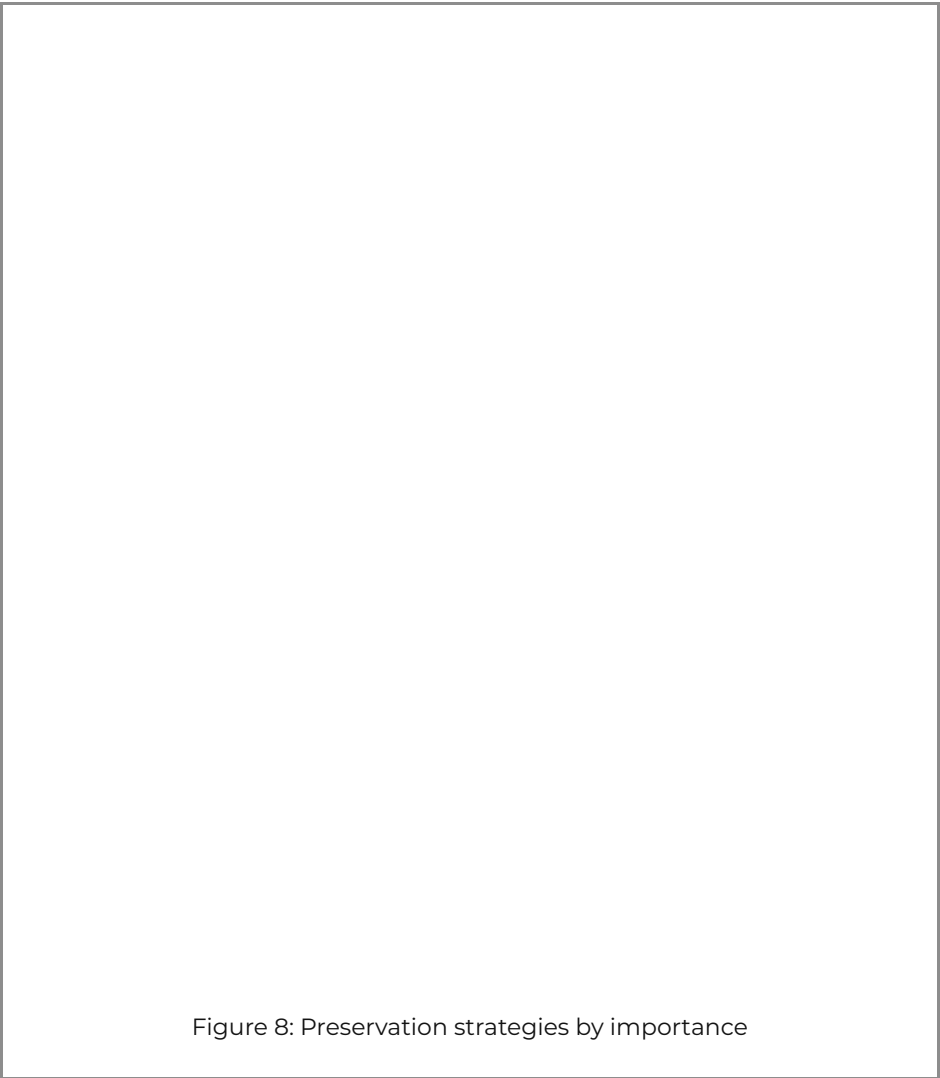


Figure 8: Preservation strategies by importance

All respondents reported that they backup their work in some capacity. The most respondents (78%) are using cloud services. Half report the use of institutional servers, and 44% use home computers. GitHub was mentioned by two respondents as a safe storage solution for their projects. The majority of respondents (66%) are utilizing more than one way of backing up their work.

Interview Findings

Through follow-up interviews with five respondents, we delved into several of these projects in greater detail. Interviewees gave us more information about their projects and their partnerships, processes, and policies for the

preserving the work.

Profile #1: DH Coordinator

Interview conducted and summarized by Nik Dragovic

Respondent 1 was a coordinator in a Digital Humanities Center at their institution and had undertaken the work in collaboration with librarian colleagues because the library works closely with researchers on DH projects at this particular institution.

This initiative was unique in that no preservation measures were being undertaken, a strategy that resulted from discussion during the conception of the project. The resulting life expectancy for the project, comprising a geography-focused, map-intensive historical resource incorporating additional digital content, was three to four years. The reason for the de-emphasis of preservation stemmed from a shared impression that the complexity of preservation planning acts as a barrier to initiating a project. Given their intention to produce a library-produced exemplar work rather than a traditional faculty portfolio piece, the initiative was well-suited to this approach. The technical infrastructure of the project included a PHP stack used to dynamically render the contents of a MySQL database. The general strategy incorporated elements of custom software and open source technologies including Neatline and Omeka.

The unique perspective of the respondent as an institutional DH liaison as well as a practitioner made the interview more amenable to a general discussion of the issues facing a broad set of digital humanists and their interaction with

library services. The overriding sentiment of the respondent echoed, to a large extent, existing literature's assertion that DH preservation is nascent and widely variable.

Specifically, the interviewee opined that no one framework, process, or solution exists for those seeking to preserve DH outputs, and that every project must have its own unique elements taken into account. This requires an individual consultation with any project stakeholder concerned with the persistence of their work. A primary element of such conversations is expectation management. In the respondent's experience, many practitioners have the intention of preserving a fully functional interface in perpetuity. In most cases, the time, cost, and effort required to undertake such preservation measures is untenable.

The variegated and transformative code stack environments currently underpinning DH projects is a leading issue in permanent maintenance of the original environment of a DH project. As a result, the respondent advocated for a "minimal computing" approach to preservation, in which more stable formats such as HTML are used to render project elements in a static format, predicated on a data store instead of a database, with languages like Javascript as a method for coordinating the front-end presentation. This technique allows not only for a simpler and more stable preservation format, but also enables storage on GitHub or Apache servers, which are generally within institutional resources.

Another preservation solution the respondent explained was the dismantling of a DH project into media components. Instead of migrating the

system into a static representation, one leverages an institutional repository to store elements such as text, images, sound, video, and data tables separately. The resulting elements would then require a manifest to be created, perhaps in that format of a TAR file, to explain the technology stack and how the elements can be reassembled. An Internet Archive snapshot is also a wise element to help depict the user interface and further contextualize the assets.

In the experience of the respondent, helping digital humanists understand strategic and scaled approaches to preservation is one of the greatest challenges of acting as a library services liaison. Students and faculty have an astute understanding of the techniques underpinning the basic functionality their work, but not the landscape of current preservation methodologies. Not only is the learning curve steep for these more library-oriented topics, but the ambitions of the library and the practitioner often diverge. Whereas the scholar's ambition is often to generate and maintain a body of their own work, the library focuses more on standardization and interoperability. This creates a potential point of contention between library staff and those they attempt to counsel. Often the liaison must exercise sensitivity in their approach to users, who themselves are experts in their field of inquiry.

The broader picture also includes emerging funding consideration for national grants. When asked about the intentions of the National Endowment for the Humanities to incorporate preservation and reusability into funding requirements, the respondent expressed skepticism of the agency's conceptualization of preservation, stating that a reconsideration and reworking of the term's definition was in order.

To apply too exhaustive a standard would encourage a reductive focus on the resource-intensive preservation methods that the respondent generally avoids. Like most facets of the DH preservation question, this warrants further inquiry from practical and administrative standpoints. In a general sense, realistic expectations and practical measures ruled the overall logic of the respondent, as opposed to adherence to any given emerging standard presently available.

Profile #2: Library Director

The impetus behind respondent 2's project was not to advance scholarship in a particular subject, so the preservation strategy and goals differed from projects that had a more explicitly scholarly purpose. The idea was hatched by a team of librarians as a means to help librarians learn and develop new skills in working with digital research with the ultimate goal of enhancing their ability to collaborate and consult with researchers on their projects. The learning and training focus of this project informed the team's preservation strategy.

A number of tools were used to plan, document, and build out this project, and some levels of the production were designed to be preserved where others were intended to be built out, but then left alone, instead of migrated as updates become available. The process was documented on a WordPress blog, and the ultimate product was built on Omeka. The team did preservation and versioning of code on GitHub, but they do not intend to update the code even if that means the website will ultimately become unusable.

What was very important to this team was to

preserve the “intellectual work” and the research that went into the project. To accomplish that, they decided to use software, such as Microsoft Word and Excel, that creates easy to preserve files, and they are looking into ways to bundle the research files together and upload them to the institution’s repository. Respondent 2 expressed that an early problem they had with the technology team was that they “wanted everything to be as well thought out as our bigger digital library projects, and we said that DH is a space for learning, and sometimes I could imagine faculty projects where we don’t keep them going. We don’t keep them alive. We don’t have to preserve them because what was important was what happened in the process of working out things.”

This team encountered some challenges working with Omeka. At one point they had not updated their version of Omeka and ended up losing quite a bit of work which was frustrating. “We need to be thinking about preservation all along the way” to guard against these kinds of losses of data. Working with the IT department also posed challenges because “technology teams are about security and about control” and are not always flexible enough to support the evolving technology needs of a DH project. The project had to be developed on an outside server and moved to the institutional server where the code could not be changed.

Profile #3: Art Professor

Respondent 3’s institution has set up a DH center with an institutional commitment to preserving the materials for the projects in perpetuity. The center relies on an institutional server and has a

broad policy to download and maintain files in order to maintain them indefinitely on the back end. Front end production of the project was outsourced to another institution, and the preservation of that element of the project had not been considered at the time of the interview.

This researcher's main challenge was that although many of the artworks that are examples in the project are quite old and not subject to copyright, certain materials (namely photographs of 3D objects) are copyrighted and can only be licensed for a period of 10 years. The front-end developer expressed that 10 years was a long time in the lifetime of a website (which would make that limitation of little concern), but being able to only license items for a decade at a time clashes with the institutional policy of maintaining materials indefinitely on the server and raises questions about who will be responsible for this content over the long term if the original PI were to move on or retire.

Profile #4: Archivist

Interview conducted and summarized by Hannah Silverman

Respondent 4, who has developed a comprehensive set of open source tools for the purpose of archiving documents and resources related to a specific historical era, sees their work within the sphere of Digital Humanities. The sense that their archival work was essentially related to the Digital Humanities came about over a period of time as their technical needs required them to connect with a larger set of people, first with the librarians and archives community through the Metropolitan New York Library Council (METRO), then as a DH activity

introduced at a METRO event. “I myself am writing a [DH] blog which originally was a blog by archivists and librarians...So, the way I met people who are doing similar things is at METRO. We are essentially doing DH because we are on the cross of digital technologies and archives. It is just a label, we never knew we were doing DH, but it is exactly that.”

The respondent goes on to describe the value of developing tools that can read across the archive, allowing researchers to experience a more contextual feel for a person described within the material – adding dimensionality and a vividness to the memory of that person:

What I am struggling with is essentially one major way of presenting the data and that is the library way. The libraries see everything as an object, a book is an object, and everything else is as an object. So they see objects. And if you look at the NY Public Library...you can search and you can find the objects which can be a page of an archive but it is very difficult to see the whole archive, the whole collection; it's not working this way. If you search for an object you will find something that is much in the object but it is not conducive to see the context and the archives are the context, so what I am trying to see if we can expand this context space presentation. We spent very little money on this project product which we use to display the data. There is a software designer... who built it for us, but if we could get more funding I would work on [creating] a better view for visualizing the data.

Several projects [like this] are waiting in line for funding here...We collect records, records are not people. Records are just names. We would like to put the records in such a way that all the people are listed and then give the information about this person who was in this list because he was doing something, and in this list because he was doing something else, and in this document because he traveled from here to here and so on. That would be another way of sort of putting all the soldiers and all the people involved in these three (volunteer) uprisings for which we have complete records of in part of the archive. We have complete records of all the people in such a way that you could follow a story of a person and also maybe his comrades in arms. It may be the unit in which he worked, and so on.

The respondent has addressed preservation with multiple arrays of hard drives that are configured with redundancy schemes and daily scrubbing programs for replacing any corrupted digital bits. Also copies stored on tape are routinely managed in multiple offsite locations, as well as quality assurance checks occurring via in both analog and digital processes.

Profile #5: English and Digital Humanities Professor

Interview conducted by Hannah Silverman and summarized by Malina Thiede.

The project discussed in this interview began as a

printed text for which an interactive, online platform was later created. The online platform includes data visualizations from user feedback (such as highlights) and a crowdsourced index, as no index was included in the original print text. The code for the project is preserved and shared on GitHub which the interviewee sees as a good thing. The visualizations of the data are not being preserved, but the data itself is. There is an intent to create and preserve new visualizations, but the preservation plan was not set at the time of the interview.

The initial project was conceived and executed in a partnership between an academic institution and a university press on a very short timeline (one year from call for submissions to a printed volume) with very rigid deadlines. Due to the rapid and inflexible timeline, preservation was not considered from the outset of the project, but a data curation specialist was brought in between the launch of the site and the first round of revisions to review the site and give advice on issues of preservation and sustainability. The institution supporting the project has strong support for digital initiatives; however, an informal report from the data curation specialist tasked with reviewing the project indicated that “precarity in the institutional support for the project could result in its sudden disappearance.”

The interviewee stated that “we are less focused on preservation than we should be” because “we’re looking towards the next iteration. Our focus has been less on preserving and curating and sustaining what we have” than on expanding the project in new directions. At the time of the interview, this project was entering a new phase in which the online platform was going to be adapted into a digital publishing platform that

would support regular publications. The interviewee indicated several times that more of a focus on preservation would be ideal but that the digital elements of this project are experimental and iterative. The priority for this project is moving ahead with the next iteration rather than using resources to preserve current iterations.

Analysis & Conclusion

Through this survey of NYC librarians, scholars, and faculty, our aim was to capture a sample of the work being done in the digital humanities, paying close attention to this population's preservation concerns, beliefs, and practices. Through this research, we offer the following observations regarding DH content creators and preservation:

1. Preservation is important to the researchers working on these projects, but it is often not their main focus.
2. Scholars working on DH projects are looking for advice and support for their projects (including their project's preservation).
3. Librarians and archivists are already embedded in teams working on DH projects.

Preservation Challenges

We noticed through textual responses and follow-up interviews that preservation rarely came up in the earliest stages of the project – sometimes due to tight deadlines, and other times simply because preservation is not generally in the conversation during the onset of a project. Researchers are typically not accustomed to thinking about how their work will be preserved. The workflows for traditional published research

leave preservation in the hands of the consumer of the research, which is often the library. However, DH and other digital projects often have less clearly defined workflows and audiences, making it less obvious who should be responsible for preservation and when the preservation process should begin. Our data indicates that most planning about preservation occurs sometime during the course of the project or after its completion, rather than at the beginning. Best practices for digital projects state that preservation should be a consideration as close to the beginning of the project as possible, but researchers may not be aware of that until they have done significant work on a project.

It is also noteworthy that just over half of our survey respondents set a goal of preserving their work for five or more years, and significant percentages (22 and 17, respectively) set goals of three to four years or were unsure of how long they wanted their work to be preserved. This indicates that not all projects are intended to be preserved for the long term, but that does not mean that preservation planning and methods should be disregarded for such projects.

As these projects go forward, respondents who do want their projects to be available long term grapple with the difficulties that surround preservation of digital content and the added time commitment it demands.

The following survey respondent illustrates this potential for complexity:

Unlike many digital humanities projects this project exists/existed in textual book format, online, and in an exhibition

space simultaneously. All utilize different aspects of digital technologies and are ideally experienced together. This poses much more complicated preservation problems since preserving a book is different from preserving an exhibition which is different from preserving an online portion of a project. What is most difficult to preserve is the unified experience (something I am well aware of being a theatre scholar who has studied similar issues of ephemerality and vestigial artifacts) and is something that we have not considered seriously up to this point. However, because books have an established preservation history, the exhibition was designed to tour and last longer than its initial five-month run, and the online component will remain available to accompany the tour and hopefully even beyond, the duration of the project as a whole has yet to be truly determined and I am sure that considerations of preservation and version migration will come up in the near future for both the physical materials and the digital instantiations of the project. It promises to provide some interesting conundrums as well as fascinating revelations.

And another survey respondent:

I feel like I should unpack the perpetuity question. Our project is text (and) images (and) data visualizations on a website. The text (and) images I'd hope would be accessible for a long time, the

data (visualization) relies on specific WordPress plugins/map applications and may not be accessible for a long time. Since we're self-administering everything we will take things forward with updates as long as we can, but...

Roles for Librarians and Archivists

As one librarian interviewee explained, preservation is a process that needs to be considered as a project is developed and built out, not a final step to be taken after a project is completed. Hedstrom noted as far back as 1998 that preservation is often only considered at a project's conclusion or after a "sensational loss," and this remains a common problem nearly 20 years later. Therefore, librarians and archivists should try to provide preservation support starting at the inception of a project. Considering preservation at an early stage can inform the process of selecting tools and platforms; prevent data loss as the project progresses; and help to clarify the ultimate goals and products of a project.

Nowwiskie (2015) posed the question: "is [digital humanities] about preservation, conservation, and recovery—or about understanding ephemerality and embracing change?" Humanists have to grapple with this question as it regards their own work, but librarians and archivists can provide support and pragmatic advice to practitioners as they navigate these decisions. Sometimes this may mean that information professionals have to resist their natural urge to advocate for maximal preservation and instead to focus on a level of preservation that will be sustainable using the

resources at hand. Librarians and archivists would do well to consider this advice from Nowviskie (2015):

We need to acknowledge the imperatives of graceful degradation, so we run fewer geriatric teen-aged projects that have blithely denied their own mortality and failed to plan for altered or diminished futures. But alongside that, and particularly in libraries, we require more a robust discourse around ephemerality—in part, to license the experimental works we absolutely want and need, which never mean to live long, get serious, or grow up.

Profiles #1 and #2 exemplified the ‘graceful degradation’ approach to DH preservation by building a website that was intended to be ephemeral with the idea that the content created for the site could be packaged in stable formats and deposited in an institutional repository for permanent preservation. The project discussed in profile #5, while not explicitly designed as an ephemeral project, has a fast moving, future focused orientation, such that any one particular iteration of the project may not exist indefinitely, or even for very long. Of course, an ephemeral final product may not be an acceptable outcome in some cases, but advice from librarians can inform the decision making process about what exactly will be preserved from any project and how to achieve the level of preservation desired.

Due to variations in the scale and aims of individual DH projects and the resources available

in different libraries, it would be virtually impossible to dictate a single procedure that librarians should follow in order to provide preservation support for DH projects, but based on our data and interviews, librarians who want to support preservation of DH research can take the following steps:

1. Keep up with existing, new, or potential DH research projects on campus. Depending on the type of institution, those projects may be anything from large scale projects like the Linguistic Atlas mentioned above to undergraduate student work.
2. Offer to meet with people doing DH on campus to talk about their projects. Begin a discussion of preservation at an early stage even if long term preservation is not a goal of the researchers. Establishing good preservation practices early can help to prevent painful data losses like the one mentioned in profile #2 as the project progresses.
3. Work with the researchers to develop preservation plans for their projects that will help them meet their goals and that will be attainable given the resources available at your institution/library.
 - In developing a plan, some of the questions from our survey (see Appendix I) may be helpful, particularly questions about the nature of the project and the intended timeline for preservation.
 - Also keep in mind what resources are available at your library or institution. Kretzschmar and Potter (2010) took advantage of a large, extant media archive at their library to support

preservation of the Linguistic Atlas. The interviewees in profiles #1 and #2 also mentioned the institutional repository (IR) as a possible asset in preserving some of the components of their work. (While useful for providing access, IRs are not a comprehensive preservation solution, especially at institutions that use a hosting service.)

– Coordinate with other librarians/staff that may have expertise to help with preservation such as technology or intellectual property experts. As discussed in profile #3, copyright can pose some challenges for DH projects, especially those that include images. Many libraries have staff members that are knowledgeable about copyright who could help find solutions to copyright related problems.

– For doing preservation work with limited resources, The Library of Congress Digital Preservation site has a lot of information about file formats and digitization. Another good, frequently updated source from the Library of Congress is the digital preservation blog The Signal. Although created in 2013 and not updated, the POWRR Tool Grid could be a useful resource for learning about digital preservation software and tools.

Conclusion

DH projects are well on their way to becoming commonplace at all types of institutions and among scholars at all levels from undergraduates to full professors. The data and interviews presented here provide a snapshot of how some digital humanists are preserving their work and about their attitudes toward preservation of DH projects in general. They show that there are

opportunities for librarians to help define the preservation goals of DH projects and work with researchers on developing preservation plans to ensure that those goals are met, whether the goal is long term preservation or allowing a project to fade over time.

Acknowledgements

Although this article is published under a single author's name, the survey and interviews were created and conducted by a team of four that also included Allison Piazza, Nik Dragovic, and Hannah Silverman. Allison, Nik, Hannah, and I all worked together to write and conduct the survey, analyze the results, and present our findings in an ALA poster session and to the Metropolitan New York Library Council (METRO). Writing and conducting the interviews was likewise a group effort, and all of them contributed to writing our initial report although it was never fully completed. The contributions of these team members was so substantial that they should really be listed as authors of this paper alongside me, but they declined when I offered.

This project was initially sponsored by the Metropolitan New York Library Council (METRO). Tom Nielsen was instrumental in shepherding this project through its early phases.

Special thanks also to the Pratt Institute School of Information for funding the poster of our initial results that was displayed at the 2015 ALA Annual Conference.

Additional thanks to Chris Alen Sula, Jennifer Vinopal, and Monica McCormick for their advice and guidance during the early stages of this

research.

Finally, thanks to publishing editor Ian Beilin, and to reviewers Ryan Randall and Miriam Neptune. Their suggestions were immensely helpful in bringing this paper into its final form.

References

Bryson, T., Posner, M., St. Pierre, A., & Varner, S.

(2011, November). SPEC Kit 326:

Digital Humanities. Retrieved from

<http://www.arl.org/storage/documents/publications/spec-326-web.pdf>

Carnegie Classifications | Basic Classification.

(n.d.). Retrieved from

http://carnegieclassifications.iu.edu/classification_descriptions/basic.php

Hedstrom, M. (1997). Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, 31(3), 189–202.

Kirschenbaum, M. G. (2009). Done: Finishing

Projects in the Digital Humanities, *Digital*

Humanities Quarterly, 3(2). Retrieved from

<http://www.digitalhumanities.org/dhq/vol/3/2/000037/000037.html>

Kretzschmar, W. A. (2009). Large-Scale

Humanities Computing Projects: Snakes Eating

Tails, or Every End is a New Beginning? *Digital*

Humanities Quarterly, 3(2). Retrieved from

<http://www.digitalhumanities.org/dhq/vol/3/2/000038/000038.html>

Kretzschmar, W. A., & Potter, W. G. (2010). Library

collaboration with large digital

humanities projects. *Literary & Linguistic*

Computing, 25(4), 439–445.

Library of Congress. (n.d.). About – Digital

Preservation. Retrieved from
<http://www.digitalpreservation.gov/about/>

Maron, N. L., & Pickle, S. (2014, June 18). Sustaining the Digital Humanities: host institution support beyond the start-up phase. Retrieved from
<http://www.sr.ithaka.org/publications/sustaining-the-digital-humanities/>

Nowwiskie, B. (2015). Digital Humanities in the Anthropocene. *Digital Scholarship in the Humanities*, 30(suppl_1), i4–i15.
<https://doi.org/10.1093/llc/fqv015>

Smith, A. (2004). Preservation. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.). *A Companion to Digital Humanities*. Oxford: Blackwell. Retrieved from
<http://www.digitalhumanities.org/companion/view?docId=blackwell/978140510313/9781405103213.xml&chunk.id=ss1-5-7&toc.depth=1&toc.id=ss1-5-7&branddefault>

Walters, T., & Skinner, K. (2011, March). New roles for new times: digital curation for preservation. Retrieved from
http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar

What is digital humanities? (2015, January). Retrieved from
<http://whatisdigitalhumanities.com/>

Zorich, D. M. (2008, November). A survey of digital humanities centers in the US. Retrieved from
http://f-origin.hypotheses.org/wp-content/blogs.dir/1834/files/2013/08/zorich_2008_asurveyofdigitalhumanitiesce

Appendix: Survey

*Preservation in Practice: A Survey of NYC
Academics Engaged in Digital Humanities*

Thanks for clicking on our survey link! We are a group of four information professionals affiliated with the Metropolitan New York Library Council (METRO) researching the digital preservation of DH projects. Contextual information is available at the myMETRO Researchers page. Our target group is New York City digital humanists working in academia (such as professors or PhD candidates) who have completed or done a significant amount of work on a DH project. If you meet this criteria, we'd appreciate your input. The survey will take less than 15 minutes. The information we gather from this survey will be presented at a METRO meeting, displayed on a poster at the annual conference of the American Library Association, and possibly included as part of a research paper. Published data and results will be de-identified unless prior approval is granted. Please note that your participation is completely voluntary. You are free to skip any question or stop at any time.

You can reach the survey administrators with any questions or comments:

Nik Dragovic, New York University,
nikdragovic@gmail.com

Allison Piazza, Weill Cornell Medical College,
allisonpiazza.nyc@gmail.com

Hannah Silverman, JDC Archives,
hannahwillbe@gmail.com

Malina Thiede, Teachers College, Columbia
University, malina.thiede@gmail.com

Is your project affiliated with a New York City-area institution or being conducted in the New York City area?

Yes

No

Title or working title of your DH project:

Does your project have an online component?

Yes (Please provide link, if available):

To be determined

No

What techniques or content types have you used or will you use in your project? Select all that apply.

Data visualizations

Data mining and text analysis

Text encoding

Network analysis

GIS (Geospatial Information Systems)

3-D modeling

Timelines

What date did you begin work on this project (MM/YY)

Approximately how many people are working on this project?

2-3

4-5

6+

I am working on this project alone

Has preservation been discussed in relation to this project?

Yes

No

Who first mentioned the preservation of your project?

Self

Librarian

DH center staff

Project member
Funder
Host department
Other:

At what stage in the project was preservation first discussed?

Before the project began
During the project
After project completion

Who is/will be responsible for preserving this project? Select up to two that best apply.

Self (PI)
Library
Host department
Another team member
Institution
Person or host to be determined
Campus IT
Another institution

How important are each of these processes to your overall preservation strategy for this project?

Bit-stream preservation or replication (making backup copies of your work)
Durable persistent media (storing data on tapes, discs, or another physical medium)
Emulation (using software and hardware to replicate an environment in which a program from a previous generation of hardware or software can run)
Metadata creation
Migration (to copy or convert data from one form to another)
Version control

Are there any other preservation strategies essential to your work that are not listed in the above question? If so, please list them here.

Do you have defined member roles/responsibilities for your project?

Yes

No

Not applicable, I am working on this project alone.

What is your main contribution to this project team? Select all that apply.

Technical ability

Subject expertise

Project management skills

Is there a specific member of your team that is responsible for preservation of the technical infrastructure and/or display of results?

Yes

No

Is there a DH center at your institution?

Yes

No

How often have you consulted with the DH center for your project?

Never

Once

A few times

Many times

DH center staff member is a collaborator on this project

My institution does not have a DH center

How is this project funded? Select all that apply

Institutional funding

Grant funding

Personal funds

Were you required to create a preservation plan for a funding application?

Yes

No

What kinds of resources have you consulted to inform the preservation of your project? Select all that apply.

Published scholarly research (such as books or journal articles)

Guides, reports, white papers and other grey literature

Professional or scholarly association sponsored events or resources (such as webinars)

Conferences

Campus workshops or events

Colleagues or informal community resources

None

DH Center, Library/librarian, archivist

Have you had any training in digital preservation?

Yes

No

How many years into the future do you see your project being usable/accessible?

1-2 years

3-4 years

5+ years

Not sure

Is your resource hosted at your own institution?

Yes

No

If no, where is it hosted?

How are you backing up your work? Select all that apply.

Cloud service

Institutional server

Home computer

DAM tools

Not currently backing up work

Other

Which of the following types of software have you used to create your project? Select all that apply.

Proprietary software (Please list examples)

Open-source software (Please list examples)

Custom software

If you would like to add any perspectives not captured by the previous questions, or clarify your answers, please use the comment box below:

Your full name

Email address

Institutional affiliation

Primary department affiliation

Academic title

If applicable, when did/will you complete your PhD?

Would you be willing to be the subject of an approximately 45-minute interview with a member of our team to talk more in-depth about your project and preservation concerns?

Share Post:

< [The Innovation in Libraries Awesome](#)



This work is licensed under a [CC Attribution 4.0 License](#).

ISSN 1944-6195

[About Us](#) | [Archives](#) | [Submissions](#) | [Conduct](#)