Developing Collaborative Best Practices for Digital Humanities Data Collection: A Case Study

Rachel Di Cresce

Information Technology Services, University of Toronto Libraries

rachel.dicresce@utoronto.ca

Julia King

Department of English and Drama, University of Toronto

jlm.king@utoronto.ca

Abstract

This case study explores the data management practices of medieval manuscript scholars working on the Digital Tools for Manuscript Study project at the University of Toronto. We chose this user group, despite their incredibly domain specific praxis, since the data challenges they face while doing digital humanities work are representative of the wider community. Our goal is to rethink how librarians can best assist researchers within a digital humanities centered environment. This paper first explores how data is conceived in the DH context and what insights can be drawn for data management. Next, focus shifts to the key characteristics of data collection and post-processing activities carried out by manuscript scholars during repository visits. Parallels are drawn between manuscript scholar practices and those of other humanities disciplines. Finally, the implications for information professionals are explored and best practices for assisting digital humanists defined. In particular, community engagement in the process is stressed throughout as it is the authors' belief it is necessary for success. The best practices are in no way exhaustive, and they are intended to be broadly applicable to a range of disciplines within the digital humanities and to librarians. Future work will involve validating a new data management approach informed by this study by testing in the field.

Keywords: *Data Management, Digital Humanities, Manuscripts, Scholarly Needs, Best Practices, Knowledge Organization*

Data and the Humanist

Following the scientific method, a researcher should pose a hypothesis, collect data, and test it against that hypothesis to declare it true or false. The data collected is often measurable in some way; it has gone through a rigorous experimentation process, been approved by ethics boards, repeated hundreds of times to ensure little variation, and published as evidence to shore up a hypothesis. The data in the scientific model is meant to be uniform; by doing the same experiment twice, if the research is sound and the experiment has been set up properly, the data should come out to be similar, if not exactly the same. Data that lends itself to measurability, like numbers, computerized data, or facts, is valued by the sciences, and this conception of visible and tangible data is what has shaped our modern understanding of numbers, charts, sets, and tables as more related to laboratory experimentation than humanistic study.

But what of humanities data? Unlike scientific studies, which seek to repeat answers to confirm their truth, humanistic inquiry takes an assumption and answers it in several different ways. A simple question can have multiple answers, and the value of a good research question is that it can produce a variety of responses. Compare this to the value of repeatable scientific data. How do you manage data that comes out of humanistic inquiry when it is not as mathematically measurable and regular as scientific data? How do humanists view and manage their own research output and do they conceive of it as manageable data? To truly tend to humanist data management needs it is important to understand these questions and look for answers within the community. One method of understanding the variety of data available to humanists is to recognize the different kinds of data humanities research can produce. For example, Research

Data Canada refers to the "Knowledge Map of Information Studies" study, which, among other things, collected 130 definitions of data formulated by forty-five scholars (Zins 2007). Within it, all data, regardless of format or medium, are recognized. Research Data Canada's broad definition of research data reads:

> Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records. (Research Data Canada 2017)

Humanities researchers produce most, if not all, of these types of data. The multimedia aspect of humanities research is only part of the complex puzzle of how to organize data management. One must understand the theoretical underpinnings of humanities research and the data it produces in order to appreciate the often much smaller and more nuanced data sets of humanists scholars and the unique nature of humanist inquiry. Taken from a professor of Digital Medieval Studies, the following excerpt explores this phenomenon:

> Humanities' data has depth in small universes. Our material has the capacity to unfold inwards, as it were, to disclose layer upon layer of insights and connections, within a comparatively tiny amount of data--almost an inverse matryoshka, as it were, where each inner doll is bigger and more complex than the one encasing it. (Bolintineanu 2016)

Humanities data requires a level of interference and analysis that is divergent from scientific inquiry. It is changeable, shaped by everything from the tools used to analyse or present it to the scholars who attempt to interpret it. This is why traditional understandings of data seem foreign or unfit for use in a humanities context. Perhaps Posner put it best in stating, "When you call something data, you imply that it exists in discrete, fungible units; that it is computationally

tractable; that its meaningful qualities can be enumerated in a finite list; that someone else

performing the same operations on the same data will come up with the same results. This is not

how humanists think of the material they work with" (Posner 2015).

In our case, whether digital or traditional humanities research is concerned, the data produced

often poses challenges to the information professional. Simply applying scientific understanding

and practices to the field of humanities data management ignores the theoretical underpinnings

of humanities research. Even when tools or analytical techniques from the sciences can be fit into

a humanities-esque mold, disagreement exists about their appropriateness:

> [DH visualization tools borrowed from the sciences] carry with them assumptions of
> knowledge as observer-independent and certain, rather than observer co-dependent and
> interpretative. […] To begin, the concept of data as a given has to be rethought through a
> humanistic lens and characterized as capta, taken and constructed. (Drucker 2011)

This does imply, however, a unified understanding of what constitutes data within the realm of

scientific research and beyond (Funari 2014 or 2015?). Definitions abound, with their own

inclusions and focus, even among scholars of the same university department (Whitmire, Boock,

and Shutton 2015). It has been shown that academic institutions, federal funding agencies, and

regulatory bodies all define 'data' uniquely (Joshi and Krag 2010). For example, the Tri-Council

Agencies of Canada, made up of the Canadian Institutes of Health Research (CIHR), the Natural

Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and

Humanities Research Council (SSHRC), provide a definition for data in its policies for all grant-

funded projects. The agencies note that research data, "include observations about the world that

are used as primary sources to support scientific and technical inquiry, scholarship and research-

creation, and as evidence in the research process" (Tri-Agency Statement of Digital Principles on

Digital Data Management 2016). A more agnostic definition, from ISO/IEC 2382-2015 (2015),

defines data as "a re interpretable representation of information in a formalized manner, suitable

for communication, interpretation, or processing." But even these definitions of data, rooted in

scientific modes of understanding research, cloud how humanities scholars interpret their own

research. It does little to bring the humanities or social sciences, which tend to not think of their

findings as tractable, finite, or identically reproducible, into the realm of research data.

In an effort to be more succinct, and to align ourselves more with humanistic data theory,

we wish to present one more definition of data: Data is "units of information observed, collected,

or created in the course of research" (Erway et al. 2013). Importantly, Erway's definition

presumes no scientific inquiry, quantitative analysis, or identically reproducible results. From

here, we are better placed to understand the data management needs of digital humanist scholars.

Research data management

As with all projects, it is imperative to invest in a data management strategy in the digital

humanities. As early as 1968, researchers were concerned that "librarians are less than ever

before keepers of books; they are coming to be managers of data" (Hays, 1968, 5). More

recently, literary scholars have become concerned with the 'computational turn', or the

increasing reliance on computer science techniques to perform humanities research. This is

necessarily different from the concept of the digital humanities, but it is responsible for what

Manovich has termed the 'cultural analytics paradigm', whereby one assumes that the "big data"

created by twenty-first century cultural production is vast, and, therefore unknowable (Hall

2013). Research data management, however, is all aspects of creating, housing, maintaining, and

retiring data (O'Reilly et al.2012) and therefore makes these vast amounts of data knowable, sortable, and manageable.

The data lifecycle, although originally conceived for science data, is also applicable to humanities data management and can provide helpful guidelines for structuring a data management plan. The California Digital Library defines the data life cycle as having eight steps: plan, collect, assure, describe, preserve, discover, integrate, and analyze (Strasser et al. 2012). By managing these steps, standardized and usable data is created; housed in a way that it is stable, searchable, and findable; maintained through various switches of file formats, permutations, and manipulations; and retired to an archive in a sustainable fashion. Although all of the different permutations of Manovich's big data are unknowable, research data management makes them possible and searchable.

Managing data created during the course of (digital) humanities research requires that the data manager pay attention to the special landscape which they navigate to create, conceptualize, and analyze their data. Humanities research data management is, as Awre et al. (2015) point out, an example of Rittel and Webber's (1973) 'wicked problem', that is, a problem that is seen differently to different stakeholders. As opposed to a 'tame problem', where there exists one answer to each problem, for example, "How do I execute a search strategy on the library catalogue", a wicked problem has multiple solutions that are neither true nor false, just a good solution or a bad solution. As Awre et al. point out, the first step in reckoning with managing any amount of research data is to recognize the complexity of the problem. Keeping this necessary complexity in mind, it becomes obvious that individual projects require an individualized plan, and, to that end, we have used the experience of one particular humanities research data problem as a lens through which to view the subject.

Method

Rimmer et al. point out that when designing digital resources for humanities scholars, "we need to better understand their research experiences and practices" (2008, 1378). This same principle extends to designing digital humanities data management strategies. The research experiences and practices of scholars heavily informed the work of this project. The case study arose out of collaborative work on the Digital Tools for Manuscript Study Project, based jointly out of the University of Toronto Libraries and Centre for Medieval Studies, to create modular, interoperable tools for scholars using digital medieval manuscripts. The project pairs a set of development outcomes with a scholarly counterpart to demonstrate the capabilities of the tools. One tool we wish to extend and improve upon, in particular, is called VisColl (Porter 2013). VisColl is designed to generate digital visualizations of the binding structure and physical makeup of a medieval manuscript. These digital visualizations are known to scholars as 'collation diagrams', and are of immense importance to scholars interested in the method, context, and afterlife of the creation of medieval codices. Traditionally, collation diagrams are produced by hand, as the scholar carefully analyzes the binding of each section of pages in a manuscript (known as a 'quire'), producing diagrams of the quire's structure and developing what is known as a collation statement. VisColl is intended to make this process easier and more robust.

Scholars want to use VisColl to produce multiple visualizations and statements of extant *Canterbury Tales* manuscripts. Data collected by researchers will need to interact with the VisColl tool, which, in turn, will need to interpret and represent the data. As such, from the

outset, we recognized the need for a research data management strategy to streamline collection

processes. We not only felt that this was essential to the success of the overall project, but we

also saw an opportunity for progress in the world of digital humanities data management.

Two researchers (referred to as Researcher A and Researcher B) were sent overseas to visit

multiple archives and libraries to examine several manuscripts. Instruction came from the lead

scholar only; no prior input was given to the researchers by an information professional. From

speaking with medieval scholars across several institutions prior to this research trip, it became

very apparent that, even among specialties, there is no standard data collection practice shared by

scholars. As the digital humanities continue to grow and develop in current and new fields,

practices most likely will not be standardized across or among disciplines.

Upon their return, researchers were interviewed separately about their experiences. At the same

time, we examined the data files, both analog and digital, and developed basic organizational

spreadsheets in which the researchers were to insert their data. The spreadsheets were created in

order to get a good understanding of what raw data we were dealing with while creating a

preliminary organizational scheme and preparing for data transfer to our collation tool.

Throughout the post-collection process we kept in close contact with researchers to ensure that

our assumptions and ideas were valid and representative of their experiences. Our findings from

this experience will be discussed in the following section.

Discussion

How can we as library professionals best aid humanities scholars in the area of data

management? We operated under the assumption that the data collected by researchers would be

input into a collation tool and used to develop a scholarly argument. By analyzing the data

produced by the researchers and speaking with them about their process we recognized four key

findings that characterize a researcher's approach to manuscript study and provide a roadmap for

information professionals: the influence of time, universality of pre-data collection practices,

reliance on mixed media data collection, and personalized information management.

i. Time: Scholars have very limited time to work with physical manuscripts. Any

implemented data management processes must be cognizant of this.

Time was by far the most influential factor to researchers during the data collection

process. One researcher's ideal data collection process was described simply as "More time."

During the research visit, most of the items had not been digitized, meaning if information was

missed or questions remained, and the researcher could not easily refer to the manuscript once

back home. In addition, researchers must operate within the fixed hours of the library or archive

they visit, resulting in their having an average of between six and eight hours per manuscript per

day. Given the size and complexity of many manuscripts, certain texts required more time to

analyze than others. This in turn affected research processes, data collection and data

management. Researcher B, for example, stated these timeframes were "not really enough time

to study a manuscript. It's just enough for collation and notes on interesting things". The more

time given to researchers, the more information and detail can be collected.

Both researchers stated that they spent twice as much time post-processing their data as

compared to time spent with a manuscript. This is significant because it frames the way in which

the researchers think of their work in repositories. Researcher B had even less time than normal

when looking at certain select manuscripts, which affected the type and quality of data they were able to collect. Both researchers described their time as being dominated by taking notes about what they felt were the most important aspects of a manuscript as quickly as possible. Researcher B stated, "If I know I'm running out of time, I take as many pictures as I can and hope they are sufficient later on". It seems, in this instance, that work done in a repository often entails collecting information that is interesting or has the potential to be interesting in the future, and relying on later information processing to make sense of the data that was gathered. Development of scholarly connections and arguments often happen far away from the material in question.

Ideally, any data management approach we develop for these scholars must not require excessive time. For this reason, any alteration to their research process must be minimal or we risk non-adoption or misuse. It should be noted that, through speaking with other manuscript scholars, there are instances where time may be less of a challenge (e.g., when interested in one specific manuscript or a few which are all housed at the same repository), but, for the most part, time is of the essence. Researchers want to spend their time examining a manuscript and opt for whichever collection method they feel is the fastest. In a broader context, all scholars operate under similar constraints and preferences. Digital tools and their associated workflows need to feel natural and easily work into the current research process, because if they do not it is a waste of valuable time (Antonijevic 2015).

ii. Pre-data collection Preparation: Researchers conduct basic to very in-depth research about their objects of interest prior to a repository visit. This should be the stage of

intervention for information professionals in which clarity of research purpose has been reached and time is not a stressor.

Both researchers engaged in pre-visit preparation for this and other projects. Other researchers with whom we have spoken over the last few months indicate that they follow the same practice. Actions range from checking bibliographic cataloguing records to reading previous scholarship about the manuscripts. The researchers seek out an understanding of the research that has already been completed on the object, note items of interest, and identify areas where research may be lacking. These preparatory practices are closely related to time limitations. As one researcher pointed out, "I prep in advance, try to figure out how much time each manuscript will take me, especially with a limited amount of time in an archive".

If there is time, or the research goal is very well articulated, researchers tend to think about organization, even in an abstract way, prior to their visit. For example, researcher B cobbled together checklists they came across over years of study. Researcher A found information to compare findings to the scholarly canon. Every trip teaches them something new about their data collection process, and they recognize holes in their preparation that affect results. What is interesting, however, is as they reflected back on their collection processes they consistently identified tactics well known to information professionals. For example, without using the information terms precisely, researchers recognized controlled vocabularies, pre-defined categories, improved workflows, tracked tags, and systematic file-naming as beneficial to their research. One researcher stated, "I wish I had thought about my categories prior to visits so my notes would have been organized and efficient".

Ultimately, this discussion was not prompted by the potential to reduce post-collection work on the part of the information professional, but the potential for the researcher to save time in the archive. In the researcher's mind, better quality data does not mean less post-processing. The goal is to decrease the inconsistency of data collection. Researchers lamented notes that became less clear depending on situational factors. Information deemed nonessential is often left out only to be missed later. It is their belief that with a more structured process, the frequency of these occurrences will decrease.

The often serendipitous nature of manuscript work is a concern for information managers and researchers alike. Researchers truly never know exactly what they will see when looking at a manuscript - their intention to study one aspect may be completely pushed aside upon the discovery of something unexpected. As with most research, what is fascinating to one researcher may not be worth a second glance from another. One simply cannot control for all of the possible variabilities in manuscripts and the whims of human nature. Any data management plans constructed prior to archival visits must reflect the potentially unstructured path of inquiry. Any attempts at imposing an immovably rigid system will risk serving a few of users and will ensure non-adoption from many others who do not trust the system or are not able to adapt their research practices around it.

iii. Mixed Media: Manuscript researchers tend to produce a multitude of both digital and analog files during their visits.

Not a trait solely of manuscript scholars, humanists of all disciplines subscribe to a "fusion of digital and 'pen and paper' practices" (Antonijevic 2015). Manuscript scholars rely

heavily on do-it-yourself images regardless of available digital surrogates. Based on responses from our two researchers, photos are often taken of details which were not caught in the digitization process, when something is too difficult to describe quickly, is an example of a particular phenomenon, when a feature looks interesting, or, as a last resort, to gather as much information as possible before running out of time. Researcher A even took a video of a part of a binding structure which was so different from the standard so that they could consult with colleagues about it later. Due to their volume, and difficulty to track, organize, and store, photos are a particular problem. Researchers often spend a lot of time naming image files and linking them to their notes in some way. These are often kept in greater disarray than other files, with non-descriptive file names and non-standardized tags.

Alongside images and photo data, researchers create textual notes about the manuscript they are examining. One researcher took analog notes completely while the other started with analog but switched to digital when they felt it was not efficient. Other researchers we talked to also report a mix of analog or digital notes depending on the individual scholar's preference, subject matter, and experiences. Often, certain items are interesting, but are not easily expressed digitally. For example, a collation statement, such as the one for Cambridge, Corpus Christi College MS 144, which is notated I^8-VII^8 $VIII^8$ (+1), is easier to write down manually than enter into a text document because of the superscript notation. The preferred method for collecting digital notes is in Microsoft Excel or Word whereas analog notes tended to have a loose structure of organization such as charts, columns and sub-headings that were unique to the researcher. Finally, researchers often create drawings of manuscript structures, either manually or digitally. These collation diagrams are essential to the researcher, and are most easily produced by hand. Often times, the structure of a binding will reveal oddities of book production or call into

question the textual content of a manuscript. These diagrams are often referred to countless times throughout research and used in publications. They are made most commonly with pencil and paper, but digital collation tools are becoming more usable. One researcher was able to visualize a binding pattern by creating a digital collation in Excel while keeping the data neat and organized.

iv. Personalized Information Management: All manuscript researchers create their own personalized approach to study which is reflected in every aspect of their personal information management practices.

Both interviews and analysis of raw data collected by researchers made very apparent that each researcher develops their own idiosyncratic data management system. There was a lack of standardized vocabulary, researchers disagreed on what labels to put on their data, and their organization grew organically as their data was produced. This presents a series of problems. The creation of standardized vocabulary is quite difficult within the field. "Things like how to record a manuscript's quire formulas are pretty standard, but the words we use are all over the place," said researcher A. For example, describing the cover of a manuscript can take many forms; one scholar might refer to the "boards", whereas another might call it a "cover", and another might lump it in with the general description "binding". As researcher B points out, "This is why pictures and diagrams are very useful as they can transcend the vagaries of language."

More difficult is the phenomenon of the organic development of a data management style. Researcher A commented, "Because I was collecting a whole pile of data and I wasn't sure what I would find I put everything into tiny categories; I started to refine a better system as I

went through. By that point I had missed earlier data." Because of the restrictions of different

repositories, it is difficult to return and retrieve the missing data. However, when asked if there

was a particular feature of their data management system that they did not like, the researcher

responded, "No, because if there was, I would change it. I wouldn't know [I didn't like a feature]

until I found the magic work around difference." This individualization of research processes

makes it extremely difficult for the information professional to create a pre-defined research

procedure. Since each researcher has created a method created through testing different strategies

to find what works for them and what does not, they will often be resistant to strategies that have

been deemed appropriate for the group which they have personally found ineffective.

Problems for the information professional

For the information professional, then, creating a data management strategy can be

difficult. For those who want data that is sortable and easily malleable, creating Microsoft Excel

tables or asking for checklists to be completed might clash with a researcher's desire to take

more photographs that cannot be sorted or to take notes with a more organic information

structure by hand. Time is always a factor in these decisions as it puts further constraints on a

data management plan. At some point in the research process, data collected on these trips will

need to take on a digital form. Whether for analysis, preservation, sharing, or publication, all data

will go through transformations to facilitate use. Given this inevitable outcome, information

professionals need to work with scholars to identify a suitable point of intervention while

communicating the benefit of such actions.

Our desire for order, through standardization, structure, and schemas often runs opposite to the more nuanced, organic, and personalized work of individual humanists. Terminology, itself sometimes a subject for scholarly argument, changes depending on the era of study or background of the researcher. Since humanities research is often a discipline given to individual study it leads to individual practices and vocabularies. As such, dreams of standardized workflows or even a taxonomy of vocabulary terms are fairly unrealistic in this climate. A compounding factor is the uniqueness of the material of study itself. No two manuscripts are exactly the same nor are the scholars who look at them. Attempting to predict every scenario, oddity, or change of interest is impossible.

Best Practices

A result of this study has been the development of general best practices that will work to serve the manuscript scholarly community and the greater digital humanities community simultaneously. In the near future, we plan to test our ideas in the field with the same subjects to determine whether the approach holds value. As Antonijevic states, "although generic tools have better potential to meet research needs of a broader set of humanists, there is also space for a smaller-scale and more experimental tool building" (97). Our hope is that by creating best practices that work within the context of our manuscript based-research project, these smaller-scale tools will have broader application to the wider digital humanities environment.

The first practice is to work with scholars during the planning phase of the data life cycle. Information professionals should promote early planning as both beneficial to the overall research process and compliant with university and funding agencies. Our researchers preferred

preparation methods, with one noting, "I think the main thing is the more prep-work beforehand to be honest." Scholars can lay out expectations, create resources that are mutually agreeable to both the scholar and the information professional, and address any concerns before reaching the repository. Information managers can and should create basic tables or checklists at this time to ensure that data is standardized, sortable, and searchable.

The second practice, and perhaps most important, is to follow a community approach to data management solutions. Information professionals should incorporate scholars during planning and use their insights to develop solutions. Providing them with a taxonomy or rigid, generalized rules does little to encourage scholars to make use of them, regardless of benefit. But working in a more interdisciplinary way, information managers can borrow from different research communities of practice that fit researchers' needs. For example, a field like archaeology - with its marriage of both scientific and artistic practices - could be used as a reference point for humanities data management practices. "In archaeology," writes Antonijevic, "there is no real distinction between digital and non-digital tools" (49).

Finally, the third practice is to develop an approach that aligns with scholarly practice as closely as possible. In her ethnographic study, Antonijevic recognizes, "humanities scholars envision tools that would enable seamless and multidimensional flow of research activities from one phase to another and back, across multi sided and multimedia corpora" (95). Indeed, our study participants imagined a futuristic world in which the collection of data in a library could be immediately organized, tagged, and connected to related information with little intervention. The first step in this direction would be careful consideration of the data and processes that surround it. The easier it is to incorporate protocols into research, the more likely scholars will make use of them and the greater the potential for data sharing, long-term preservation, and reuse.

Conclusion

Based on our findings, we are beginning to develop an approach for the next stage of our research. Still in the preliminary planning stage, our hope is for the beginnings of an ontology, which allows flexible changes to its collection and structure, a formalized checklist outlining the essential data that need to be collected, and a template, both in analog and digital form, which will add structure to their research notes and facilitate the use of tools later on in the research cycle. All of this will be developed and vetted with the close consultation of researchers to ensure their cooperation and our mutual success. This data will then be usable throughout our wider digital humanities project, and the structures and workflows that we develop for data collection and curation can be used for future digital humanities projects. It will serve to validate the tools we create for digital manuscript scholars and also test our framework against the wider field of digital humanities.

As the digital humanities grow and adapt to new environments and applications research data practices will come under necessary review. Although humanities scholars have always 'managed' their data, in that they track their research and use their own organizational systems, incorporating digital tools changes the way this process unfolds. In short, digital humanities research necessitates an approach perhaps more in line with the standardized scientific approach than the traditionally individualized nature of humanist inquiry. As information professionals, we need to understand these differences and reconcile them with current research data management practices. We must challenge our traditional notions of research data management by placing ourselves within the context of different fields and theories. Information professionals are well suited for this role since we understand both the potential and limitations afforded by different

data sets and practices. In short, we must understand and accommodate both the digital *and* the

humanities in our own work. Future efforts in the realm of DH data management will only be

successful if we stake out a path in which both sides of the digital humanities coin are recognized

and considered.

References

Abbas, June. 2010. "Structures for organizing knowledge: exploring taxonomies, ontologies, and other schemas". New York, NY: Neal-Schuman Publishers.

Antonijevic, Smiljana. 2015. "Amongst Digital Humanists. An ethnographic Study of Digital Knowledge Production". New York, NY: Palgrave MacMillan.

Awre, Chris, et al. 2015. "Research Data Management as a 'Wicked Problem'." *Library Review.* 356-371.

Baofu, Peter. 2008. *"*The future of information architecture: conceiving a better way to understand taxonomy, network and intelligence". Michigan: Chandos.

Bolintineanu, Alexandra. 2017. "DH History and Data". Lecture at Woodsworth College, CCR199H1S, Introduction to Spatial Digital Humanities, January.

Briney, Kristin. 2015. "Data management for researchers: organize, maintain and share your data for research success." Exeter, UK: Pelagic Publishing Ltd.

Crompton, C., Lane, R. J., Siemens, R. G. 2016." Doing digital humanities: Practice, training, research". New York, NY: Routledge.

Drucker. Johanna. 2011. "Humanities Approaches to Graphical Display". *Digital Humanities Quarterly.* 5(1)*.* Retrieved from http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Erway R. et al. 2013. "Starting the Conversation: University-wide Research Data Management Policy".   Retrieved from: http://www.oclc.org.myaccess.library.utoronto.ca/content/dam/research/publications/library/2013/2013-08.pdf

Funari, Maura. 2015. "Research data and humanities: a European context" *Italian Journal of Library and Information Science* 5(1): 209-236.

Goven, Abigail and Raszewski, Rebecca. 2016. "The data life cycle applied to our own data". *Journal of the Medical Library Association.* 103(1): 40-44.

Hall, Gary. Dec 2013. "Toward a Postdigital Humanities: Cultural Analytics and the Computational Turn to Data-Driven Scholarship." *American Literature* 85(4): 781-809.

Hays, David G. 1968. "Data management in the humanities". *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed October 4, 2016). http://www.dtic.mil/dtic/tr/fulltext/u2/668752.pdf

Heuser, Ryan and Le-Khac Long. 2011. "Learning to Read Data: Bringing out the Humanistic in the Digital Humanities" *Victorian Studies,* 54(1):79-86.

ISO/IEC 2382-2015. 2015. "Information Technology: Vocabulary". Retrieved from: https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en

Joshi, Margi and Krag, Sharon S. 2010. "Issues in Data Management" *Science and Engineering Ethics* 16:743-748.

Kanare, Howard M. 1985. "Writing the laboratory notebook". Washington, D.C.: American Chemical Society.

Krier, Laura and Strasser, Carly A. 2014. "Data Management for libraries". Library and Information Technology Associatio*n*, Chicago: Neal Schuman Publisher.

O'Reilly, Kelley et al. 2012. "Improving University Research Value: A Case Study" *SAGE Open* 2:3 (https://doi.org/10.1177/2158244012452576)

Porter, Dorothy. 2013. "Viscoll: Visualizing physical manuscript collation". Retrieved from: https://github.com/leoba/VisColl.

Posner, Miriam. 2015. June 25. "Humanities Data: A Necessary Contradiction". Retrieved from: http://miriamposner.com/blog/humanities-data-a-necessary-contradiction

Posner, Miriam. 2016, April 19. "Data Trouble: Why Humanists Have Problems with Datavis, and Why Anyone Should Care". Retrieved from: https://www.youtube.com/watch?v=sW0u1pNQNxc&t=209s

Research Data Canada. 2017. "Original RDC Glossary*"*. Retrieved from: https://www.rdc-drc.ca/glossary/original-rdc-glossary/

Richardson, Julie and Hoffman-Kim, Diane. 2010. "The Importance of Defining "Data" in Data Management Policies" *Science and Engineering Ethics* 16: 749-751.

Rittel, Horst W. J. and Melvin M. Webber. 1973. "Dilemmas in a General Theory of Planning" *Policy Sciences* 4: 155-169.

Rimmer, J. and C. Warwick, A. Blandford, J. Gow and G. Buchanan. 2008. "An examination of the physical and digital qualities of humanities research." *Information Processing and Management* 44: 1374–1392

Strasser, Carly; Cook, Robert; Michener, William; & Budden, Amber. 2012. Primer on Data Management: What you always wanted to know. UC Office of the President: California Digital Library. Retrieved from: https://escholarship.org/uc/item/7tf5q7n3

Tri-Agency Statement of Digital Principles on Digital Data Management. 2016. Retrieved from:http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

Whitmire, A. L., M. Boock., and S. C. Sutton. 2015. Variability in academic research data management practices. *Program,* 49(4): 382-407.

Zins, C. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the Association for Information Science and Technology,* 58(4): 479–493. doi:10.1002/asi.20508