“Yeah, I Guess That’s Data”: Data Practices and Conceptions among Humanities Faculty

Jennifer L. Thoegersen

abstract: The ability to interact with and properly manage data in a research setting has become increasingly important in all disciplines. Libraries are attempting to identify their role in providing data management services. However, humanities faculty’s conceptions of data and their data management practices are not well-known. This qualitative study explores the data management practices of humanities faculty at a four-year university and examines their perceptions of the term data.

Introduction

The discussion of how researchers interact with and manage data has become increasingly pertinent in the digital world. In addition to the ways technology has transformed how researchers collect, interpret, present, and store their data, there are increasing external pressures to adequately manage and promptly share research data.

A major catalyst for the sudden thrust of data management into the center stage of research was the 2011 implementation of a new guideline by the National Science Foundation (NSF) requiring the inclusion of a data management plan with all grant proposals. On a broader scale, several forces—including high-profile journals, research communities, public advocates, and funding agencies—pushed toward greater availability of research data both to ensure the validity of the research and to allow for data reuse. This effort cast a spotlight on researchers’ competencies in managing data and raised concerns in academic libraries about how to aid researchers in acquiring the knowledge and skills necessary to manage data.
Yet, even the term *data* may carry connotations that preclude researchers in certain disciplines, especially in the humanities, from recognizing the applicability of these concepts. Little is known of how humanities researchers perceive data related to their research or how they manage the products of their research.

This exploratory study had two goals. First, it aimed to gain greater insights into the data management practices of humanities faculty at a public four-year research university to better understand their data management needs and to identify areas in which library data management services might assist them. The secondary goal of this study was to explore humanities faculty’s conceptions of the term *data*. The author used in-depth interviews with faculty to collect qualitative data based on participants’ experiences.

**Data Management and Libraries**

Research data management, “the process of controlling the information generated during a research project,” has gained increased attention in the research community over the past several years. This heightened interest has largely resulted from federal funding agency requirements for data management plans.

Though the National Institutes of Health (NIH) began requiring data-sharing plans for grants exceeding $500,000 as early as 2003, the major event that led to greater interest in data management was the NSF’s requirement that all grant proposals submitted after January 17, 2011, include a data management plan. Other federal funding agencies required similar plans following this regulation, and, in 2013, the White House Office of Science and Technology Policy issued a memo highlighting its expectation that federally funded researchers make their findings publicly available.

Though varying in detail, data management plans submitted with grant proposals are expected to outline how the data collected and generated during a funded research project will be managed throughout the project’s duration as well as after its conclusion. These plans generally include information on types of data, how the researchers will store and back up data, how they will maintain security, how they will handle sensitive information, and how and when they will share data.

Driven by these events, data management has become an area of interest and growth for academic library research and services. Library data management services range from consulting on data management plans for grant proposals to managing repositories for data preservation and sharing. Projects intended to inform and guide libraries offering these services include the creation of data management curricula and the identification of data literacies.
Literature Review

Studies related to scholars’ data management practices have largely focused on researchers in the sciences and, to a lesser extent, social sciences. This focus is evident in Elizabeth Berman’s 2017 comparison of data management needs assessment studies.12 In her review, Berman compares 14 studies conducted over the last 10 years at a variety of U.S. public academic institutions. Most of these investigations narrowed their study populations to researchers in science or science, technology, and mathematics (STEM) fields, or some subset within these fields (for example, agricultural scientists or principal investigators of NSF/NIH grants). Two of the studies reviewed by Berman included all disciplines in their study population and had at least one research participant from the humanities.13

Some studies have included more about humanities researchers’ data management practices. In a 2015 study, Hélène Prost, Cécile Mallaret, and Joachim Schpfel focused on social sciences and humanities researchers. Aiming to better assist PhD students in research data management and to study data sharing and reuse, these researchers examined social science and humanities dissertations (N = 283) at Charles de Gaulle University—Lille III in Villeneuve d’Ascq, France. The investigators determined that 188 of the 283 dissertations included data in at least one of their appendices.14 The three authors noted a heterogeneity of data types between disciplines, as well as within some disciplines.

Katherine Akers and Jennifer Doty also noted a heterogeneity of data among humanities researchers. In 2012, they surveyed the data management practices and perspectives of faculty at Emory University in Atlanta, Georgia. The survey was limited to those faculty who self-identified as generating “some type of data” when conducting research (n = 330). Twenty-seven percent (n = 126) of respondents answered “No” to generating data and were not included in the survey.15 Akers and Doty noted that humanities faculty were more likely than those in other disciplines to be unsure of how much data they stored and speculated that this uncertainty was due to respondents being doubtful about what constituted data.16 They also noted that “arts and humanities researchers are more likely to rely on computer hard drives and internet-based storage services, such as Dropbox and Google Drive.”17 They concluded that arts and humanities scholars should engage in conversations about how research data in their disciplines may differ from data in other disciplines.18

Methodology

This study used in-depth interviews to explore humanities faculty’s conceptions of data and data management practices. The population for this study consisted of humanities faculty at a large, public, four-year research university. The author constructed a sampling frame (N = 105) from the faculty directories of these disciplines and used it to select a random sample of participants. The sample included faculty in history (n = 23;
23 percent), classics (n = 14; 13 percent), English (n = 41; 39 percent), philosophy (n = 12; 11 percent), and languages (n = 15; 15 percent). The author invited faculty members in these disciplines to be interviewed for this study via email.

A total of nine faculty agreed to participate, and the author conducted all in-depth interviews during the summer of 2016. Participants included faculty from history (two), English (four), languages (one), and philosophy (two). The academic ranks of participants were assistant professor (two), associate professor (four), and professor (three). Four males and five females participated.

The author conducted 30- to 60-minute in-person interviews with the faculty members. Interviews included open-ended questions and began with a discussion of the faculty members’ current research projects. The author then asked them to describe the research materials they collect, generate, and use for their research, and how they manage these materials. The interview concluded with a discussion of the term data and how they saw this term apply (or not) to their research. (See the Appendix for interview questions.)

Audio recordings were made of the interviews, and the recordings were transcribed. The author analyzed the transcriptions following an iterative approach like that outlined by Greg Guest, Kathleen MacQueen, and Emily Namey. She read the text of the interviews several times and identified potential themes, then developed the thematic areas into codes. The researcher coded a sample of the interviews twice, approximately one month apart; reviewed any discrepancies between the coding; and clarified the codes. She then recoded the text of all the interviews with the revised code definitions. The presence or absence of a code in each individual interview was noted.

**Limitations**

This study is exploratory in nature. Given the small sample size (nine) at a single institution, the results of this study are not generalizable. It is instead intended as exploratory, highlighting areas of interest for future research.
Findings

Research

At the beginning of the interview, faculty were asked to describe their current or most recent research projects. Five faculty (56 percent) reported that they largely focused on archival studies, while only one participant mentioned research with human subjects. The two philosophy faculty described their research as contributing and responding to arguments.

All the faculty were currently working on a book project. Eight were writing monographs, and one participant was editing a volume. Three (33 percent) also mentioned working on scholarly articles.

Though five faculty (56 percent) mentioned collaborating with other researchers on at least one project at some point, only one scholar frequently collaborated and was currently working on articles with others. One researcher commented, “Really, in the humanities, most of us work on our own.”

Storage and Backup

The author asked the faculty about their backup and storage practices. All but one stored their digital materials on either their work or personal computer. The remaining researcher stored all documents on a personal drive on a server managed by departmental information technology (IT) personnel.

All nine participants had some method of backup; however, the types, complexity, and extent of backup procedures varied considerably. Table 2 presents the backup methods used by each participant. Methods ranged from copying to a flash drive to employing different backup procedures for various digital content and different purposes. For example, one researcher used an automated backup service to back up to cloud storage. However, because the configuration for the backup software was complicated, that individual also manually copied locally stored e-mail content to external hard drives on occasion. Finally, that person also stored some documents in Dropbox for backup and sharing.

When asked about their backup procedures, five (56 percent) of the participants expressed a fear of losing materials and a desire to ensure that their work was secure, though only one mentioned having actually lost materials. A computer failure had caused the data loss, resulting in a significant destruction of research materials and adoption of CrashPlan backup software.

Those participants who fear loss but did not mention personally losing digital materials instead discussed anecdotal cautionary tales, for example, “All of those authors whose house burned down and they lost a novel,” and a general concern about losing work and materials. One stated, “I like having two backups, because I’m paranoid, so I just want to make sure that . . . if my computer ever completely crashes, I won’t lose that much.”
Table 2.
Backup methods used by researchers

<table>
<thead>
<tr>
<th>Backup method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flash drive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>2</td>
</tr>
<tr>
<td>Departmental server</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>3</td>
</tr>
<tr>
<td>External hard drive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>5</td>
</tr>
<tr>
<td>Cloud, Dropbox</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>3</td>
</tr>
<tr>
<td>Cloud, Box</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Cloud, other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>2</td>
</tr>
</tbody>
</table>

**Organization**

As part of discussions of material descriptions and storage, researchers were asked about file-naming conventions and organization. File management and organization appeared to worry most respondents. When asked how they organize materials, six researchers made disparaging remarks regarding their digital file organization. Though the author did not ask participants about the quality of their organization, their comments included such phrases as “I’m still not very good at organizing this,” “I’m not always consistent,” and “Not well.” One researcher’s initial response to the question was “Oh God,” followed by a discussion of how “unweedy” certain folders had become.

The main methods of organization were by project (especially, for book projects, by chapter) and theme. For three researchers, the main method of locating materials was through searching. One stored all materials in a citation database; the other two used Apple’s macOS Spotlight search software to locate files. One said, “I’ll label the files with the author’s last name and then something in the title and then kind of just know to search that way on my computer. If I were organized, I’d have different files, but really, they just get flung all over the place . . . So, no, I don’t do it particularly well.”

**Protection**

The interview asked faculty members whether they had any confidential information or any concerns involving privacy, intellectual property, or security related to their research materials. Most participants expressed little or no anxiety related to these protections but described a variety of concerns. Three faculty reported unease related to confidentiality. Only one of the three expressed extreme concern for confidentiality related to the sensi-
tivity of the topic. Two additional researchers identified instances where confidentiality was important in their work: One mentioned having written letters of recommendation, and the other described a single instance in which living relatives requested that specific information on a family member be kept private.

Two faculty expressed worry about protecting unpublished research ideas, with one participant noting, “I’ve been working on this a heck of a long time and I haven’t published it yet and I don’t want to get scooped, basically.” One philosophy professor felt frustrated with the stricter rights and open access fees some disciplinary journals impose on researchers and stated, “I just don’t write for those journals anymore.”

Two participants discussed terms of use of the archival materials they accessed at other institutions. One historian, in particular, found keeping track of the terms of use for each archive “extremely challenging” and added, “I’m getting better at that, always better at this. So, what I do as part of my archive with the documents organized for this project, I also scan the release form basically that I signed with the conditions that I have to fulfill for this archive . . . But it’s not always easy to keep track.”

Sharing

The interview asked faculty under what circumstances they might share research materials. All would share under some conditions. However, public sharing of research materials was limited. Two English faculty mentioned sharing information and resources on public websites. All other public sharing of research by participants took place via publications.

Most of the faculty (n = 7) expressed some willingness to share research materials with others on request. A general desire to help other researchers predominated, as did the notion that sharing was a positive action. However, several participants reported a desire to maintain control over what materials they share and with whom. Conditions under which the faculty would share on request included sharing with people who are working in an area I’m interested in,” lending published materials that are difficult to find, sharing with researchers who request materials or information, and sharing in a mutually beneficial way, especially if materials took a great deal of effort to gather or generate. One history faculty explained:

Sometimes, I’ve worked with people before, and it’s give and take. I help them, they help me in some way. Sometimes, I am not willing to share it just because it was asked, because it took me a lot of work and effort. These things don’t just grow on trees and you can pick them easily. It takes [a lot] of expertise.

Of the two who would not share on request, one viewed publication as the sole method of sharing research, and the other would share research materials “just [for] collaborative work.”
Keeping Materials

When asked how they decided what research materials to keep or discard at the end of a project, all participants indicated that they kept all or nearly all materials. One history faculty member expressed the necessity to keep everything: “I cannot discard anything. That would be very, I think, problematic in my field.”

Four of the faculty members mentioned discarding physical materials. All cited space issues as the main reasons for disposing of materials, and one also mentioned throwing out materials as required by the Institutional Review Board: “Usually, if I’m required to destroy something, I’ll destroy it. If I don’t have to destroy something, I’ll keep it. With the exception of space issues.”

Participants did express concerns about keeping everything. Concerns included maintaining organization as the number of physical and digital materials continued to grow, running out of space on a computer hard drive, and managing the eventual obsolescence of digital storage mediums. Other participants expressed no anxiety about digital storage because they generated and collected a small volume of digital materials in relation to the available storage.

Concerns included maintaining organization as the number of physical and digital materials continued to grow, running out of space on a computer hard drive, and managing the eventual obsolescence of digital storage mediums.

While discussing retention of research materials, one history researcher reflected on the challenges future scholars will face when attempting to use current researchers’ digital materials, stating, “We often work with materials that other researchers . . . produced over their lifetime . . . but how’s it going to work with the digital, with my material that I have?”

Conceptions of Data

The interview asked participants which, if any, of their research materials they would consider data as well as how they would define data. Table 3 summarizes the responses. The answers varied, but several themes emerged. Most participants stated, with varying levels of confidence, that they would consider most of the research materials discussed to be data.

Only the two participants from the Philosophy Department stated that, though they may occasionally draw on data to support arguments, they would not consider any of their materials to be data. Both described data as empirical evidence to support a hypothesis. Most other participants (n = 6) veered toward data broadly as information, especially information related to a research question or project. As such, they considered most or all the materials discussed in the interview to be data. The remaining participant viewed data as discrete facts—“taking a piece of information” from a source and giving it context.

In discussing the meaning of data, six of the nine participants referenced either their own discipline or other disciplines. Whether comparing themselves to the sciences, to
### Table 3.
Participant conceptions of data

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Which, if any, of your materials would you consider data?</th>
<th>Definition of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>History</strong></td>
<td>All materials related to my research project</td>
<td>All materials related to my research project, and it doesn’t matter what form it is in . . . Everything that relates to—in the context of research for a research project.\nAny kind of information that I’m going to use to answer the questions that I’m looking at as a historian.</td>
</tr>
<tr>
<td>History</td>
<td>I don’t know what that means exactly. I mean, they’re information. They’re not quantitative. They’re information that gives texture about how people lived and what they thought. So, yeah, I guess that’s data. I would consider that I abstract data from them [the research materials that I collect and generate]. Well, I guess I would. In a broad definition, I guess data is information that you use. Research information.</td>
<td>It becomes data . . . when I’m taking a piece of information and putting it into a narrative. Information. What’s been written down . . . Data is broadly any facts that help with the work. Shed light on what we’re trying to find out. In some ways, all of the information is data, right?</td>
</tr>
<tr>
<td><strong>English</strong></td>
<td>My sense is the popular meaning of the word would tend to be numbers, anything that could be converted into numbers. In which case, I have very little data. In terms of it being information, then I have a lot. So, I think definitely interviews, articles, personal experience, drafts, I don’t know if IRB [Institutional Review Board] materials are data, unless I was writing an article about the IRB, which I might someday . . . Summary sheets are data. Brainstorming materials, you’re thinking about data and the coded data is data.</td>
<td>Information that’s useful for generating knowledge. So, when you’re trying to answer a particular research questions, there’s certain information that you collect that would allow you to answer that question or to synthesize with other data to analyze that data.</td>
</tr>
</tbody>
</table>
Table 3. Continued.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Which, if any, of your materials would you consider data?</th>
<th>Definition of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>I think all of them I consider data.</td>
<td>For me, data is stuff that could possibly provide proof for something that's interesting.</td>
</tr>
<tr>
<td></td>
<td>Not really.</td>
<td>Data is anything that would be used in support of confirming some empirical hypothesis.</td>
</tr>
<tr>
<td>Philosophy</td>
<td>Really what I think of as data is nothing I produce.</td>
<td>Something that would get used as empirical evidence to try and establish or disprove some hypothesis.</td>
</tr>
</tbody>
</table>
social sciences, to other humanities disciplines, or even to other researchers in their own field, these participants expressed a feeling of “otherness” in how they approach or conceive data. Though they ultimately stated that they would consider their research materials data, three participants mentioned that their materials do not fit the “typical” numerical or quantitative description of data: “I always think of numbers when I think of data.” This conception also demonstrates a general uncertainty among the participants as to what falls into the scope of data. Every person interviewed expressed some level of doubt. For example, one asked, “Does that sound right?” after providing a definition of data.

Discussion

The heterogeneity of some of the data management practices of participants is consistent with the findings of previous studies, most notably the participants’ backup procedures.20 Given the minimal concern expressed about digital storage space and the ability to store most content on a personal computer, these individuals store only modest amounts of data, which provides more flexibility as to where to store and back up materials. The backup procedures for most of the participants as well as the concern expressed about the potential for data loss suggest this is an area where libraries can emphasize simple solutions and advertise to researchers more strongly. Most scholars handle backups individually, and many use manual backup procedures. The university’s IT department provides free, unlimited cloud storage and backup to all university faculty, staff, and students. Only one participant mentioned using Box for some research materials, though some faculty do use servers managed by their IT departments, which have robust and automated backup procedures. Whether this avoidance of university-supplied backup options results from lack of knowledge, lack of trust, or another reason should be investigated further.

Most of the participants put little emphasis on collaboration or sharing of research materials, and no participant mentioned sharing materials via public repositories, only through personal requests and websites (that are hosted external to the university and paid for by the researchers themselves). This attitude leaves the long-term preservation of materials solely in the hands of the individual researchers. Given participants’ concerns about data loss and reluctance to discard materials, outreach related to long-term preservation and preparing materials for reuse may be worth pursuing.

The uncertainty that participants expressed when discussing the definition and their conceptions of data supports Akers and Doty’s suggestion that humanities researchers are less likely than researchers in other disciplines to have a clear concept of the term.21
This uncertainty could become problematic in the promotion of library data management services among humanities faculty. However, whether humanities faculty’s conceptions of data align with a particular definition is of less importance than their understanding of the relevance of services provided by the library to their work. Understanding how participants view their own research materials in relation to data can help guide library outreach in this area. As one participant noted: “I suspect my story is different than my colleagues, so I will be interested to see if that’s the case. The first time I got your e-mail, I thought, ‘Oh, I feel like my work is pretty different, it might not fit,’ but then I thought, ‘Well, maybe that’s what’s needed, right?’”

**Conclusion**

The findings in this study demonstrate both the areas where library data management services could especially benefit humanities faculty and the need to stay cognizant of how these services are branded and presented. Participants in this study had concerns related to their ability to organize and back up data. They also expressed little interest in making research materials publicly available through repositories or in planning for long-term preservation and access to their materials, though they put great value on the materials themselves. Librarians could place additional focus on these areas when framing data management services. Though most of the participants did consider some or all their research materials data, they felt much uncertainty about the term *data*. This study did not identify a better term to use when referring to data management services. Instead, the findings suggest that libraries might engage humanities researchers in discussions about their research and about the term *data* to raise their awareness of both data management concepts and the library’s data management services.

Librarians interested in increasing data management outreach and services for humanities faculty may find the information gathered from these interviews useful. Further studies examining the practices and conceptions of humanities faculty at additional institutions would allow the findings to be applied more broadly. Libraries should highlight the data management topics of most relevance to humanities faculty and ensure that terminology does not inhibit promotion of their data management services.

Jennifer L. Thoegersen is an assistant professor and data curation librarian at the University of Nebraska-Lincoln; she may be reached by e-mail at: jthoegersen2@unl.edu.
Appendix

Interview Questions

1. To get started, can you describe your current research for me?
   Guiding: Could you tell me more about that? What do you mean by that? Could you give me an example? Could you go into a little more detail? Could you tell me how that looks in the class? Can you explain that in a different way? Could you explain that further?

2. I would like you to create a complete list as possible of the materials you used in your research. Please write down any materials you have generated or collected in the course of this research.
   Guiding: What other materials do you generate? Are there any other materials that you collect?

3. Can you describe each of these materials for me?
   Guiding: Where do you store the material? Do you keep any documentation on the material? Are there file formats that you use frequently? Are there any standards followed in your discipline for the material? Do you keep backups of these materials?

4. Can you describe who has access to these materials? Under what circumstances might you share these materials with other researchers? With the public?
   Guiding: Which materials would you share? Can you provide an example of sharing research materials?

5. Of the materials you have listed, are there any that contain confidential information? Are there any that have privacy concerns? Intellectual property concerns? Security concerns?
   Guiding: How do you handle this material? Do you have provisions for sharing this material?

6. How do you decide what research materials to keep or discard at the end of a research project? For those you keep, where do you keep them?
   Guiding: Could you tell me more about that? What do you mean by that? Could you give me an example? Could you go into a little more detail?

7. Think of the materials you generate and collect in the course of your research. Which, if any, of these would you consider “data”?
   Guiding: What makes you consider this material to be data? What do you mean by that? Could you give me an example? Could you go into a little more detail?

8. Could you describe or tell me what you mean by “data”?
   Guiding: Could you give me an example? Could you go into a little more detail? Why do you approach data in this way?

Notes


16. Ibid., 16.

17. Ibid., 9.

18. Ibid., 17.


21. Ibid., 16.