CHAPTER 4*

Practices Do Not Make Perfect

Disciplinary Data Sharing and Reuse Practices and Their Implications for Repository Data Curation

Ixchel M. Faniel and Elizabeth Yakel

Introduction

An unprecedented amount of data sharing and reuse is now possible, but disciplinary practices and traditions can create challenges for researchers wanting to meaningfully reuse data other researchers created for different purposes. Until recently in many disciplines, data sharing among peers occurred informally, in response to colleagues' requests. Now given the ability to generate digital data, federal mandates for data management and sharing, and the motivation to pose interdisciplinary questions that address critical social and environmental problems, some data producers are expected to formally share data via deposit into a repository with limited guidance about what to share and how to share it. However, the financial, technological, and human resources required to prepare data

Originally published in: Curating Research Data, Volume 1: Practical Strategies for your Digital Repository. Lisa R. Johnson, Ed. Chicago: Association of College and Research Libraries.

for sharing are limited or nonexistent. Yet even with these challenges, data sharing and reuse are growing. We contend additional growth can occur with repository staff's increased understanding of their designated communities of data reusers.

In this chapter, we draw from the results of the Dissemination Information Packages for Information Reuse (DIPIR) project. A multiyear investigation jointly funded by the Institute of Museum and Library Services, OCLC, and University of Michigan, DIPIR has investigated data sharing and reuse practices within three academic communities: quantitative social science (i.e., social science), archaeology, and zoology. Over the years we identified a number of interesting similarities and contrasts across the disciplines with regard to data sharing and reuse.[1] In this chapter we focus on three areas: (1) disciplinary practices and traditions surrounding data sharing and reuse within the three communities, (2) researchers' development of trust in the data they seek to reuse, and (3) sources of contextual information researchers rely on in addition to the repository. In the sections that follow, we describe our research methodology, discuss our findings, and conclude by describing the implications of this study for repository practice.

Overview and Methodology for the DIPIR Project

The DIPIR project aimed to identify significant factors affecting data reuse and to consider the implications they have for repository practice. We focused on the social science, archaeological, and zoological research communities because of the differences in their data sharing and reuse practices and the different repository infrastructures each had in place for archiving and disseminating data for reuse.

To conduct this research project we collaborated with three key individuals at three disciplinary repositories: (1) Nancy McGovern representing the social sciences discipline at the Inter-university Consortium for Political and Social Research (ICPSR), (2) Eric Kansa representing the archaeology discipline at Open Context, and (3) William Fink representing the zoology discipline at the University of Michigan Museum of Zoology (UMMZ). Although we used the repositories to gain access to data reusers, our research collaborators helped facilitate access to a broader disciplinary network of users beyond the repositories.

Our data collection plan employed a mixed-methods approach in that we used multiple data collection techniques, both qualitative and quantitative, to address our research questions. By using mixed methods, we were able to triangulate data from the different methods to more fully answer our research questions and address the limitations of each individual method. We conducted semi-structured interviews with data reusers in each discipline. We then employed a secondary data collection technique especially suited for each discipline (table 4.1). Specifi-

cally, we implemented a survey of social scientists because of the large population of data reusers in that area. We observed zoologists working with specimens in a museum because their research practice involves interactions with both physical specimens and digital repositories. Finally, we analyzed server logs from Open Context because we were interested in understanding how archaeologists who are new to data reuse navigated and worked with digital data. Mixed-methods studies are good for addressing complex environments and issues and can lead to increased validity and reliability.[2] Data were collected in the following ways:

- We recruited staff and data reusers for interviews from our three collaborating organizations. Interviewees also were recruited through disciplinary conferences and snowball sampling techniques, so in all cases our sample consisted of data reusers beyond our partner institutions.
- We surveyed social scientists using the *ICPSR Bibliography of Data-Related Literature* as our sample.[3] Social scientists were surveyed using Qualtics an online survey application.
- We collected server logs from the Open Context repository between August 2011 and December 2013.
- Finally, we observed zoologists interacting with physical specimens at UMMZ.

This staged approach to data collection aligns with Creswell's "sequential explanatory strategy" in which each different data collection method builds on the previous method to address broader research questions.[4]

TABLE 4.1
DIPIR Data Collection Methods and Final Participant Numbers by Discipline

	Archaeology	Zoology	Social Science
Phase 1: Project Start-Up (2011)			
Staff Interviews	4	10	10
Phase 2: Collecting and Analyzing Reuser Data (2011–2013)			
Interviews	22	27	43
Observations		13	
Survey			237
Server log entries	572,134		

In this chapter, we focus on findings from the interviews and observations. To analyze these data, we began by coding transcripts from the interviews and observations using NVivo, a qualitative data analysis software package. We created an initial codeset that was based on the data reuse literature and the interview pro-

tocol. During analysis, team members discussed the coding and added codes that emerged from the data. Coding began with two DIPIR team members working on the same transcript so we could test for interrater reliability (IRR). We used Scott's pi to calculate IRR. Our IRR scores were 0.73 for the archaeologists, 0.74 for the interviews with zoologists, 0.88 for the observations of zoologists, 0.77 for the expert social scientists, and 0.88 for the novice social scientists. When IRR was achieved, each person coded transcripts independently. Once the data were fully coded we went through several phases of analysis to delve more deeply into the findings related to each code as well as to identify relationships between codes.

Disciplinary Traditions for Data Sharing and Reuse

Increased interest in sharing and reusing data has several common drivers regardless of discipline. Computing power and communication bandwidth have enabled data to be generated, shared, and analyzed more easily and cheaply.[5] In addition, federal regulations and mandates have effectively mobilized attention and support for public access to the data and other research outputs. Since the OMB circular A-110 in 1999, federal funding agencies have issued data sharing mandates, required data management plans, and begun to allow budget items related to data management, preparation, and sharing.[6]

In response to a 2013 White House Office of Science and Technology Policy memorandum, many federal agencies have developed policies to increase public access to federally funded research outputs.[7] Initiatives within the higher education and research communities, such as SHARE (SHared Access Research Ecosystem), have been established to facilitate university compliance and to better meet stakeholders' research needs.[8] In addition, academic and research libraries have begun to develop services to support researchers' data management, sharing, and curation needs.

Accompanying these drivers are large-scale, interdisciplinary research studies in the sciences and humanities, where data reuse is vital. For example, in our DIPIR work we saw archaeologists—who once focused on a single site—reusing data from multiple sites in quantities larger than any one person could collect in a lifetime in order to examine regional social, economic, and cultural transitions between ancient civilizations.[9] Zoologists conducting biodiversity research were reusing data from repositories such as GenBank and the Global Biodiversity Information Facility (GBIF) to address questions about extinction or migration events, and social scientists were integrating government and academic research data to study household economic trends over time. In the following paragraphs, we discuss specific disciplinary practices and traditions as they relate to data sharing and reuse.

Social Scientists

Social scientists have the benefit of over fifty years of data sharing and reuse through repositories at institutions such as the ICPSR, the Howard W. Odum Institute for Research in Social Science, and the Roper Center for Public Opinion Research. The repositories curate data that tends to be well-structured and homogeneous; their data includes survey data, public opinion polls, administrative data, and international political, economic, and social indicators. The US federal government is one of the largest producers of social science data, followed by academic researchers, private survey and marketing firms, and research organizations.[10]

Given the longevity of social science data repositories, best practices in digital preservation and archiving have emerged. The Reference Model for an Open Archival Information System (OAIS) has been used as a guide, and the Data Seal of Approval has been awarded to several institutions as a signal that their repositories are trustworthy preservation archives, including ICPSR, the Odum Institute, and the Roper Center. The repository infrastructure has created a sound base for social scientists to build a disciplinary tradition around data sharing and reuse, but not without challenges.[11]

The DIPIR study revealed that in the social science community, data collection can be complex and dynamic, particularly for large-scale, longitudinal studies, which may involve a variety of sampling procedures, the attrition of survey respondents, and changes to survey questions over time. Privacy concerns also arise when collecting some types of personally identifiable data. However, our interviews suggested the repositories were well staffed and developed practices to address these issues. For instance, in some cases, ICPSR staff recruited data from major studies before the team's data collection had begun, which allowed the articulation of curation goals and a negotiation of needs to occur at the beginning of the data life cycle. In other cases, repository staff had long-standing relationships with data producers at various survey organizations, state and local governments, and federal agencies to archive data, which enabled a common understanding of needs to develop over time. Moreover, the social scientists studied were dealing with a select few data formats, so repositories could easily convert data into mainstream software packages (i.e., SPSS, Stata, SAS, Excel) as well as preservation friendly formats (i.e., CSV).

Building on the knowledge and experience within the community over time, data deposit and documentation requirements were explicit and detailed, such as in the case of the ICPSR. Codebooks evolved as a standard way of describing data within the community, and the DDI standard developed in turn as a way to compile, present, and exchange data documentation.[12] Moreover, research shows social scientists' satisfaction with data reuse is positively related to high-quality data documentation.[13] Given a long-standing culture of data sharing, a mature

repository infrastructure, and well-established relationships with data producers, data sharing and reuse have become well-established within the social science community.

Archaeologists

Archaeologists face internal and external pressures to change their data sharing and reuse practices and traditions. In addition to government mandates, data collection and dissemination practices and publication norms are changing for other reasons that are driven by cultural and political factors, in reaction to previous large-scale removal of cultural property from the country of origin. Data sharing and reuse practices in archaeology also are being adopted at different rates depending on the different sub-areas of the discipline.[14] Furthermore, the repository infrastructure to support data sharing and reuse is only recently emerging.

Legal and ethical mandates affect archaeology more than social science. For example, international and national legislation against the removal of cultural property means that archaeologists must document artifacts on-site and that they no longer have the luxury of shipping items home for further analysis and study.[15] In addition, professional organizations, such as the Society of American Archaeology, mandate that authors provide a "Data Availability Statement" detailing the "disposition and accessibility of the physical and digital data on which the research is based."[16] A final push toward new models of data sharing and reuse comes from the publishers. Traditionally, archaeologists have published books with large appendices listing artifacts, measurements, drawings of sites, and so on. Our interviewees noted that many publishers are no longer willing to print these, so archaeologists must identify other means for distributing these data tables, site information, and analyses. All of these factors have converged to move archaeology researchers into the early stages of practicing data sharing and reuse.

The DIPIR study identified two aspects of archaeological practice that make data sharing and reuse difficult. First, the variety of data types used to document an archaeological site presents a challenge. Archaeologists essentially destroy the context of field sites during excavation; therefore data collection best practice requires documentation of the physical surroundings in exhaustive detail.[17] As a result, archaeologists create and rely on different types of data (e.g., photographs, field notes, measurements) in a variety of formats; these types can range from hand-drawn maps and figures to proprietary files that require special software (e.g., CAD drawings and GIS shape files).[18] In turn, data reusers may need to contextualize the physical artifacts to a variety of analog and digital data as part of the reuse process.

Second, the DIPIR study revealed that archaeology lacks common data recording practices. The archaeological community has not yet developed a shared

understanding about the documentation and contextual information needed for data reuse, and there are few agreed-upon standards for data collection within the community.[19] Moreover, the lack of standards hampers any interoperability of archaeological data across sites and sometimes over the course of a single excavation when it takes place over a long period of time.[20] This makes data reuse difficult. In one study, several archaeologists who analyzed the same dataset reached different conclusions; despite inadequate documentation, each archaeologist concluded that the data were trustworthy enough to conduct the types of analyses they wanted to accomplish.[21]

The absence of standard data repository infrastructure also hampers data reuse in archaeology.[22] This applies to both the existence of trusted and sustainable repositories as well as agreed-upon standards for curation. There are no metadata standards to encode or encapsulate the different types of data collected in the discipline of archaeology, and there are no agreed-upon vocabularies or ontologies to link related materials.[23] This is especially problematic since archaeological data are dispersed worldwide. For example, artifacts may be in museums or remain at the discovery site, while field notes, images, and other documentation remain with the archaeologist or in a different type of repository. While museums traditionally house the physical artifacts, the availability of repositories to deposit the digital data collected in the field is relatively new in archaeology. For example, Open Context and The Digital Archaeological Record (tDAR) are both less than ten years old. In our interviews, archaeologists described performing separate searchers across many different sites to look for reusable data.

Zoologists

Zoologists have built a strong data sharing and reuse infrastructure over centuries,[24] but computerization and advanced analytical techniques recently have transformed the nature of research. Zoology has gone from an observational science where taxonomic identifications were made on the basis of visual inspection to a field where DNA is used to categorize and verify species. In parallel with this transformation, new standards for sharing data, such as Darwin Core (an expansion of Dublin Core for biological taxa), were developed and repositories began to emerge for data sharing and reuse.

The DIPIR study found that the repository infrastructure for data sharing is strong in zoology. In the past, both amateur naturalists and professional zoologists deposited physical specimens in museums.[25] In recent decades this practice has been formalized by requiring those individuals who are collecting specimens to get legal permits or licenses, which in turn mandate the deposit of physical specimens to a museum.[26] At the same time, zoologists (the primary data collectors) are affiliated with museums that have collection managers on staff to assist

with the preparation, management, and curation of the physical specimens and their accompanying data and documentation. Interviewees noted that having dedicated curatorial staff to create documentation as a part of the research workflow has made it easier for zoologists to share and reuse data.

Accepted standards for zoological data are entrenched in the repository infrastructures that span the different formats of specimen data—from labels on a physical specimen, to a Darwin Core metadata representation, to a DNA sequence. Data from individual museums are aggregated to national and international repositories, such as VertNet and GBIF. Standardization of metadata, particularly Darwin Core, has enabled a rich array of interconnected repositories with different metadata representations of the same specimen at various levels of granularity. Aggregating zoological collections makes data discovery and access more efficient, but the levels of metadata also vary, so provenance information that traces the different representations back to the museum that holds the original physical specimen is important. While basic specimen data are standardized often using Darwin Core (e.g., what was collected, who collected it, where was it collected, when was it collected), curating the deeper context is more difficult and requires access to other sources, such as field books or specimen images (including x-rays). For data reuse studies that require information beyond the basics, the lack of context can complicate the reuse process.

Data Reuse and Trust

Trust in both the data and the repository plays a major role in whether or not data are reused. In the digital curation community, trust in repositories is often conceptualized in terms of the Trustworthy Repositories Audit & Certification (TRAC) process. TRAC is based on evaluating the internal processes of repositories and trust is synonymous "with 'reliable', 'responsible', 'trustworthy', and 'authentic,' in relation to archival functions such as creating, managing, and using digital objects."[27] However, our DIPIR research was more interested in trust from data reusers' points of view. Therefore, we adopted a more classic definition of trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another."[28] We viewed trust as a multidimensional concept with both cognitive and emotional aspects that come into play as reusers search for, identify, and work with data that they did not originally collect.

In our DIPIR research, we examined data reusers' trust in repositories and found that data reusers assess trust through repository functions—particularly data processing, metadata application, and data selection—and to a lesser extent repository actions, such as transparency.[29] These findings coincide with Adolfo

Prieto's work that identified clear repository policies, customer service, and systematic processes as increasing users' confidence in a repository's authenticity, integrity, and accessibility.[30] We also found that a repository's guarantee to preserve digital data and its overall reputation increased trust in the repository for certain disciplinary communities.[31]

In this chapter, we focus on trust in the data, because research suggests trust is "a key mediating variable between information quality and information usage."[32] For example, Kelton, Fleischmann, and Wallace presented a general model of trust in information and contended that accuracy, objectivity, validity, and stability are important attributes leading to trust.[33] Donaldson and Conway confirmed these attributes but found that people considered authenticity, believability, coverage, currency, first-hand or primary nature, form, inaccurate information, and legibility important when assessing trust in archival documents.[34] Given these previous studies, we were interested in identifying which attributes data reusers might rely on to assess trust in data.

Across the three disciplines in the DIPIR study, we found that individuals used five trust markers when determining whether to reuse a dataset: the identity of the data producer, documentation, original peer-reviewed publications about the data, indications of prior reuse, and repository reputation (figure 4.1). Data reusers in each discipline mentioned data producers and documentation frequently, while only zoologists mentioned original peer-reviewed publications. Indications of prior reuse were primarily valued by social scientists, and repository reputation was important for both zoologists and social scientists.

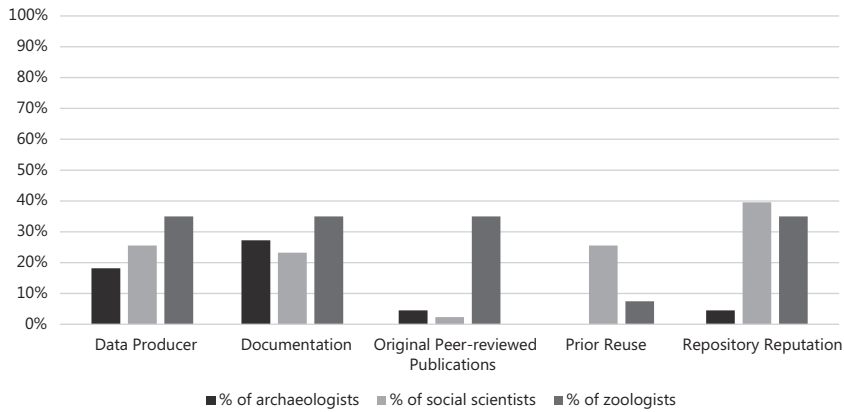

FIGURE 4.1
Top five trust markers DIPIR study participants considered when assessing trust in data based on interviews with archaeologists (*n*=22), social scientists (*n*=43), and zoologists (*n*=27).

Trust Marker: Data Producer

Information about the data producer ranked highly as a trust marker across the three disciplines. Trust in the data producer was often mentioned in tandem with some other characteristic, such as the university where the research took place, the repository housing the data, or the university where the data producer trained. Archaeologist 13 provided this example:

> Whose data do you trust? And it's primarily, it's sort of like, who do you know who does good work. So I go to the people from programs that are well known in the field, which for me is the Germans. People working out of handful of universities in Germany whether it's Munich or Tübingen, they are really well-taught. They know what they're doing, and they do it in a very standardized way.

Trust Marker: Documentation

The level or quality of documentation for the data scored as another important indicator of trust across the disciplines. Reusers tended to focus on *how* the data were documented, rather than *what* was documented about the data. Characteristics of the documentation that were important included completeness or thoroughness of the record, evidence of standardized or professional practice, and the reuser's perception of its correctness. Zoologist 11 discussed how researchers' notebooks could reveal whether they were being systematic during data collection:

> I used notebooks from multiple people… and some of them, through reading through them, I essentially did not fully trust the data they were collecting… I could tell they weren't doing it quite systematically enough.

Trust Marker: Publications and Prior Reuse Indicators

Original peer-reviewed publications about the data were seen as an important indicator of trust for the zoologists, but much less so for social scientists and archaeologists. Zoologist 3 discussed using peer-reviewed literature to double-check information from museums:

> For example, when I see records that look funny… like you know that's a mountain species, what would it be doing down there down by the sea, for example. I would then go to other published research about that group of species and see what people are saying if they… To try and cross-validate with the specimen data I'm using.

Indications of prior data reuse were most important for social scientists in assessing trust in the data. Social scientist 27 discussed prior reuse in reference to the Panel Study of Income Dynamics (PSID):

> It has been around since 1968. It's heavily used. The dataset has been examined for problems like people dropping out of the study.… And when you look at the PSID demographics there, since this was started in '68, how does that compare to a country now that has more immigrants, people where the demographics of the country has changed? Well, we've had a lot of work investigating all of these questions on PSID so these characteristics are pretty well known. It's an extremely trustworthy dataset.

Prior reuse was less important for zoologists and not a factor for archaeologists. Prior reuse has become easier to track given the emerging data citation practices and the availability of alt.metrics showing downloads on some social science data repository websites.

Trust Marker: Repository Reputation

Both social scientists and zoologists ranked repository reputation highly. Members of these disciplines had experience with a wide variety of repositories and were able to differentiate reputations among them. The archaeologists in our study did not discuss repository reputation as a factor in assessing trust in the data. Social scientist 14 summed up the interplay of repository reputation and trust in the data as follows:

> In general sort of I'm a lot more trustworthy of academic repositories and academically associated sources of data. Not that academics don't have political biases but they are subject to a lot of scrutiny in the academic community. So if you publish something with data that's clearly leaning towards one side you are going to get hammered or you are not going to get published really.

Sources of Additional Support for Data Reuse

Prior research shows reusers often need different types of contextual information about data during the data reuse process and that they use a variety of sources to get it.[35] Our findings showed that the data reusers relied on seven key sources—peer-reviewed publications, repository or museum records, data producer-generated records, people, documentation, codebooks, and specimens or artifacts (figure 4.2). Some of these sources were employed across all three disciplines to different degrees and others were discipline-specific. The way in which researchers accessed and brought the sources together also varied given disciplinary and repository practices.

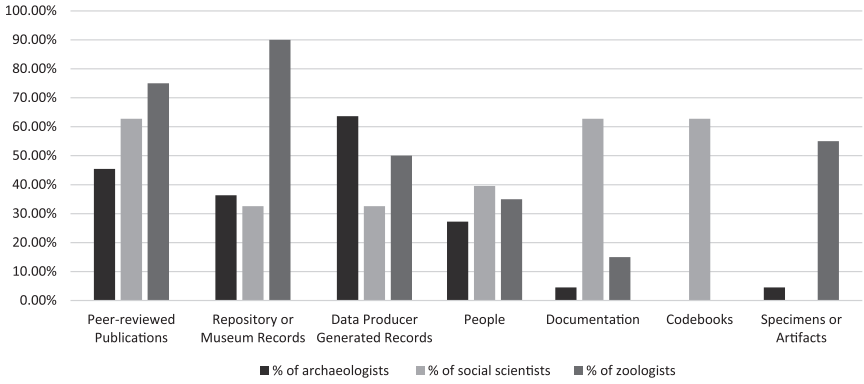

FIGURE 4.2
Seven key sources of contextual information that DIPIR study participants employed during the data reuse process based on interviews with archaeologists (*n*=22), social scientists (*n*=43), and zoologists (*n*=27).

Social Scientists

Social scientists primarily used codebooks, documentation, and peer-reviewed publications to facilitate data reuse. Our analysis showed the terms *codebook* and *documentation* were used interchangeably for this group, and they were the only group to specifically mention the term *codebook*. In order to create data records and documentation, ICPSR staff routinely used data collection and analysis information captured in the data producer–generated records. In addition, they added additional context such as missing data reports, descriptive statistics, data producer publications, and staff processing notes. By bringing all of this infor-

mation together, repository staff created one central point of access to much of the context information social scientists needed for data reuse. Social scientist 29 described some of the common types of contextual information social scientists sought in codebooks:

> The codebook for me is more like; first, to get the label of the variable. Second, to get the meaning of the different categories that they have inside the dataset, and sometimes, yes they give you some basic descriptive statistics there, so you could see, "Okay, this [is] the range, this is the mean."

ICPSR provides references to works citing the data held in its repository. Peer-reviewed publications were of particular interest to data reusers. Although the list was not exhaustive and the full text was not provided, social scientists were able to find publications that reference data of interest directly on the repository website. We found they used the publications in several ways during the data reuse process. For instance data producers' publications were used to clarify data collection and analysis information, while data reusers' publications were used to gauge the community interest and acceptance of the data. Social scientist 25 explained how peer-reviewed publications were used to discover data:

> Yeah well, the story goes, when you research the literature you find out what datasets are to be used. Through the literature you find out what authors and investigators used to answer these questions. So you find out through into the literature.…

Interestingly, even with long-existing repositories in place like the ICPSR, other people were a source of information for approximately 40 percent of the social scientists. For instance, social scientist colleagues provided opinions about reusing data and helped with data discovery. For novice data reusers like social scientist 09, professors provided reuse advice:

> Because I'm so novice in these areas, I would heavily value the opinions of professors… even if I didn't understand the reasons… I'm willing to accept that they know more about these areas than I do.

Archaeologists

Like the social scientists, archaeologists relied on contextual information from data producer–generated records. Unlike the practices in social science, however,

the records were not assembled and repackaged into one source of information, like a codebook. Instead, archaeologists searched, reviewed, and assembled a number of data producer–generated records to support their reuse of data, such as geographic maps, stratigraphy drawings, tables of numerical data, images, artifact sketches or photos, field notes, and field reports. These records contained contextual information that was recorded in the field during an excavation or survey. For older studies, much of this information remained in paper-based form and could be accessed only in museums or through the data producer.

To a lesser extent than the social scientists and zoologists, archaeologists also relied on the peer-reviewed publications to facilitate data reuse. Archaeologists typically used data producer publications to discover and access data and additional contextual information since sharing and reusing archaeological data were relatively new phenomena. Archaeologists consulted people during the data reuse process, relying on museum staff and data producers primarily. These people resources were used in the same way as data producer publications: to discover and access data and contextual information. Sometimes, archeologists sought data producers' help through collaboration on data reuse studies as well. Archaeologist 09 described a visit to a museum to gather more data and to meet with the original excavator in order to clarify the contextual information found in several sources:

> And so I started with the publications but I began to realize that I needed two things. One, I needed more data… I didn't have measurements on some of the artifacts and I needed that. And the other is I really needed to talk to the original excavator to find out some things that were confusing to me when I just looked at the photographs, or the maps, or the descriptions.… The materials were in Tulsa, Oklahoma. So I arranged to go to Tulsa and spend several days there getting the information I needed.… I sat down with the original excavator with maps and with everything else.… And he clarified a bunch of things for me.…

Zoologists

Most zoologists in our study mentioned using additional repository and museum records, followed by peer-reviewed literature, specimens, and data producer–generated records. All of these sources were used to access data collection and specimen information. The repository and museum records provided basic specimen information. Additionally, zoologists mentioned using the handwritten labels repository staff created when preparing specimens for preservation and photographs or x-rays of specimens if available.

The zoologists did not mention using peer-reviewed literature to gather information about prior reuse as frequently as social scientists, even though they commonly shared and reused zoological data. Instead zoologists used the literature, mainly journal articles, to discover and access data. Zoologist 07 talked about accessing data from a journal article that was "locked away" in a pdf format:

> The literature I'm using here is there are tables of fossil localities or presence in the taxa used for various analyses and those can prove to be useful. So I'm scraping, you know, I'm actually taking content that is locked away in pdf and converting that into a format so I can reuse for analysis of the data.

Zoologists also mentioned using physical specimens much more than archaeologists used the physical artifacts. Several zoologists discussed visiting museums or requesting that physical specimens be sent to them to gather additional sequence or morphometric (e.g., size, shape, color, etc.) data. Others requested physical specimens to verify identification of species. Zoologist 19 discussed the desire to examine the voucher (i.e., representative) specimen from which DNA was drawn:

> If somebody misidentifies the fish, or the tree or whatever it is you're looking at and uploads it to GenBank with an incorrect taxon identifier, that causes downstream problems, particularly if they didn't save a voucher so that someone can verify the ID. So when I do reuse specimens, I try to get a photograph of the voucher, or actually look at the true voucher itself, to verify that the original person that deposited the DNA sequence had correctly identified the species from which it was taken.

Zoologists also reported using data producer–generated records, but unlike archaeologists, they used fewer types, field notebooks primarily, to access the contextual information captured during data collection. Zoologists did not generate as many records during data collection. Zoologists also relied on people, collection managers at museums and data producers, to get additional contextual information. Collection managers were called on prior to a museum visit to get more information about a collection and to make arrangements to ensure a worthwhile visit as well.

The sources of support researchers in our study mentioned using depended primarily on how the data were documented in and out of the field. In both social science and zoology, there were dedicated repository and museum staff with expertise in particular types of collections that could help data producers manage and curate the data and the associated documentation. The same was not true

for archaeology. Fewer dedicated staff were available to help. These differences influenced whether and how documentation about the data were represented and disseminated via repositories or other information sources.

Implications for Repository Practice

Data sharing and reuse are increasing and will continue to do so for the foreseeable future. Our chapter aims to provide insight into the needs of data reusers, knowing that disciplinary practice is not always aligned with the changes afoot. It takes time. Not only do disciplinary practices need to change, but repository infrastructures need to mature. Both are well underway, but we believe that a broader understanding of the designated community of users, particularly data reusers, is needed. In the paragraphs that follow we discuss the implications our findings have on repository practice.

We found that the repository and museum staff within the social science and zoology disciplines play a key role in readying data for reuse. They manage, prepare, and curate data for deposit—steps that allow them to create central access points to data and other sources of contextual information that data reusers may need. The staff's work also lightens the data producers' load, especially when the staff intervene at the beginning of the data life cycle and can negotiate curation goals and needs of the data producers, repository staff, and data reusers concurrently. This level of support is beginning to happen within the archaeology discipline as well. Open Context staff is working on a project to intervene during the data collection process at the archaeological site. Staff plans to examine data producers' practices during excavations in order to provide guidance on recording and managing data in ways that make repository staff's downstream activities easier and better align data creation practices with meaningful reuse.[36] Similar to each of these three communities, we suggest that all repository staff find ways to center themselves within their designated community of users to better understand upstream and downstream needs in order to align them with repository staff's data deposit and curation activities.

We also found that reusers' trust in data is not informed only through their encounters with the data. Reusers rely on a variety of factors at play during the data life cycle, such as how the data are documented, where the data producers were trained, what university they represented when the research took place, and the trustworthiness of the repository housing the data. These characteristics are not always captured and disseminated through a repository. Furthermore, they are signifiers for reputational perceptions and opinions that get formed over time as one gains experience within the discipline, with the data, and with the repository.

We suggest that repository staff make themselves aware of these trust indicators and consider ways to more readily shape reusers' opinions about the data being offered from the repository. Repository reputation and documentation quality, in particular, can be shaped by repository staff to meet reusers' expectations.

Our findings indicate that all repositories do not have to house all of the contextual information associated with the data to be effective. However, they do have to provide data reusers access to provenance information and pointers to the additional contextual information about the data if housed elsewhere. Take Genbank, GBIF, and VertNet as examples. As repositories that aggregate zoological data across museums to facilitate easier search and discovery of species, they upload some, but not all of the contextual information associated with specimens. The museums where the data were originally deposited and the rich metadata originally created remain responsible for managing, curating, and preserving the data and contextual information. In this case, repository staff at multiple institutions made an informed decision about data stewardship and data services, given the needs of their respective data producers and reusers and the repository infrastructures in place at each other's institutions. We suggest other repository staff do the same. Consider the types of partnerships that can be formed with other repositories to complement and extend each other's capabilities and to add value to the designated community of users.

Regardless of the growth in repositories or how well established they have become, data producers remain an important source of contextual information for data reusers. By reaching out to and developing relationships with the data producers, repository staff can provide reusers with another way to learn about the data. By monitoring these engagements, repository staff also can benefit by understanding the unmet needs of their designated community of users and adapt accordingly, hopefully reducing its reliance on data producers' memories over time. Knowing that novice data reusers sought advice from expert data reusers, repository staff might want to talk to both groups to determine whether there are user interface design changes, instructional modules, or other scaffolding that the repository can provide to improve the novice data reusers' experience.

Our findings show that data reusers across the three disciplines supplement their data reuse with peer-reviewed publications. Archaeologists and zoologists rely on data producer publications primarily, whereas social scientists rely on data reuser publications. The differences are likely due to the maturity of data sharing and reuse within the disciplines. Archaeologists and zoologists are using data producers' publications to access data and contextual information, whereas social scientists are using data reusers' publications to gauge the social science community's interest and acceptance of data for reuse. We liken the latter to a dataset peer review, of sorts. In disciplines where data sharing and reuse are still in the early stages and peer-reviewed publications are limited, repository staff might consider assembling a team of experts to provide a peer review of the data. We

also suggest that repository staff capture data reuse metrics, provide DOIs and suggestions for data citation, and maintain bibliographies of data reuse studies for researchers to incorporate into their decisions to trust and reuse the data.

Conclusion

Disciplinary practices and traditions are the guiding forces in the development of a data sharing and reuse culture. Yet external forces, such as technological advances, federal mandates and policies, and repository infrastructure have shaped them further. We examined data reuse, a less-studied phenomenon, because we believe that the knowledge held by the designated communities of users, particularly reusers, is needed. We see it as a way to further develop repository infrastructure in ways that will align with the cultural changes needed in many disciplines to make data reuse a valued and viable alternative or supplement to original data collection.

Acknowledgments

The DIPIR Project was made possible by a National Leadership Grant from the Institute of Museum and Library Services, LG-06-10-0140-10, "Dissemination Information Packages for Information Reuse," and support from OCLC, and University of Michigan. We thank members of the DIPIR team, including University of Michigan students, research Fellows, institutional partners, and individual collaborators. We also thank manuscript reviewers and editors for their insightful comments and suggestions.

Notes

1. Morgan Daniels, Ixchel Faniel, Kathleen Fear, and Elizabeth Yakel, "Managing Fixity and Fluidity in Data Repositories," in *Proceedings of the 2012 iConference* (Toronto: ACM, 2012), 279–86, doi:10.1145/2132176.2132212; Rebecca D. Frank, Adam Kriesberg, Elizabeth Yakel, and Ixchel M. Faniel, "Looting Hoards of Gold and Poaching Spotted Owls: Data Confidentiality Among Archaeologists and Zoologists," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (Silver Springs, MD: American Society for Information Science, 2015), 37:1–37:10, http://dl.acm.org/citation.cfm?id=2857070.2857107; Rebecca D. Frank, Elizabeth Yakel, and Ixchel M. Faniel, "Destruction/Reconstruction: Preservation of Archaeological and Zoological Research Data," *Archival Science* 15, no. 2 (January 11, 2015): 141–67, doi:10.1007/s10502-014-9238-9; Adam Kriesberg, Rebecca D. Frank, Ixchel M. Faniel, and Elizabeth Yakel, "The Role of Data Reuse in the Apprenticeship Process," *Proceedings of the American Society for Information Science and Technology* 50, no. 1 (2013): 1–10, doi:10.1002/meet.14505001051.

2. Janice M. Morse and Linda Niehaus, *Mixed Method Design* (Walnut Creek, CA: Left Coast Press, 2009).
3. Inter-university Consortium for Political and Social Research, "ICPSR Bibliography of Data-Related Literature," May 2014, http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/.
4. John W. Creswell, *Research Design* (Los Angeles: Sage, 2009).
5. Tony Hey and Anne Trefethen, "The Data Deluge: An E-Science Perspective," in *Grid Computing*, ed. Fran Berman, Geoffrey Fox, and Tony Hey (Hoboken, NJ: John Wiley & Sons, 2003), 809–24.
6. Office of Management and Budget, "Circular A-110 Revised 11/19/93 As Further Amended 9/30/99," White House website, last updated September 30, 1999, http://www.whitehouse.gov/omb/circulars_a110.
7. John P. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research," Memorandum for the Heads of Executive Departments and Agencies, Office of Science and Technology Policy, Executive Office of the President, February 22, 2013, https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
8. Tyler Walters and Judy Ruttenberg, "SHared Access Research Ecosystem," *Educause Review* 49, no. 2 (2014): 56–57.
9. Benjamin S. Arbuckle, Sarah Whitcher Kansa, Eric Kansa, David Orton, Canan Çakırlar, Lionel Gourichon, Levent Atici, et al., "Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey," ed. Stephen Shennan, *PLoS ONE* 9, no. 6 (June 13, 2014): e99845, doi:10.1371/journal.pone.0099845.
10. Myron P. Gutmann, Mark Abrahamson, Margaret O. Adams, Micah Altman, Caroline Arms, Kenneth Bollen, Michael Carlson, et al., "From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data," *Library Trends* 57, no. 3 (2009): 315–37, doi:10.1353/lib.0.0039.
11. Kalpana Shankar, Kristin Eschenfelder, and Greg Downey, "Studying the History of Social Science Data Archives as Knowledge Infrastructure," *Science and Technology Studies* (in press).
12. Mary Vardigan, Pascal Heus, and Wendy Thomas, "Data Documentation Initiative: Toward a Standard for the Social Sciences," *International Journal of Digital Curation* 3, no. 1 (August 6, 2008): 107–13, doi:10.2218/ijdc.v3i1.45.
13. Ixchel M. Faniel, Adam Kriesberg, and Elizabeth Yakel, "Social Scientists' Satisfaction with Data Reuse," *Journal of the Association for Information Science and Technology* 67, no. 6 (June 2015): 1404–16, doi:10.1002/asi.23480.
14. Sarah Whitcher Kansa and Eric Kansa, "Beyond Bone Commons: Recent Developments in Zooarchaeological Data Sharing," *SAA Archaeological Record* 11, no. 1 (January 2011): 26–29.
15. Eric C. Kansa, Jason Schultz, and Bissell N. Ahrash, "Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a 'Some Rights Reserved' Model," *International Journal of Cultural Property*, no. 12 (2005): 285–314, doi:10.1017/S0940739105050204.
16. Society of American Archaeology, *Editorial Policy, Information for Authors, and Style Guide for* American Antiquity, Latin American Antiquity *and* Advances in Archaeological Practice (Washington, DC: Society of American Archaeology, last revised October 14, 2014), 5, http://www.saa.org/portals/0/saa/publications/styleguide/styleguide_final_813.pdf.

17. Frank, Yakel, and Faniel, "Destruction/Reconstruction."
18. Eric C. Kansa and Sarah Whitcher Kansa, "We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication," *Journal of Eastern Mediterranean Archaeology and Heritage Studies*, Project Muse, 1, no. 1 (March 16, 2013): 88–97.
19. Ixchel Faniel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel, "The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse," in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, (New York: ACM, 2013), 295–304, doi:10.1145/2467696.2467712.
20. Arbuckle et al., "Data Sharing Reveals Complexity."
21. Levent Atici, Sarah Kansa, Justin Lev-Tov, and Eric Kansa, "Other People's Data: A Demonstration of the Imperative of Publishing Primary Data," *Journal of Archaeological Method and Theory*, April 11, 2012, 1–19, doi:10.1007/s10816-012-9132-9.
22. Diane Harley, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and C. Judson King, *Assessing the Future Landscape of Scholarly Communication* (Berkeley: University of California Press, 2010).
23. Faniel et al., "The Challenges of Digging Data"; Kansa and Kansa, "We All Know That a 14 Is a Sheep."
24. Stephen T. Asma, *Stuffed Animals and Pickled Heads* (Oxford; New York: Oxford University Press, 2003); Susan Leigh Star and James R. Griesemer, "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39," *Social Studies of Science* 19, no. 3 (August 1, 1989): 387–420, doi:10.1177/030631289019003001.
25. Star and Griesemer, "Institutional Ecology, 'Translations' and Boundary Objects.
26. University of California Berkeley, Museum of Vertebrate Zoology, "MVZ Guidelines for Scientific Collecting Permits," accessed March 18, 2016, http://mvz.berkeley.edu/Sci_Coll_Permits.html. This website provides a good overview of the regulatory environment zoologists face when collecting specimens in the field.
27. Ayoung Yoon, "End Users' Trust in Data Repositories: Definition and Influences on Trust Development," *Archival Science* 14, no. 1 (March 2014): 19, doi:10.1007/s10502-013-9207-8.
28. Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer, "Not So Different after All: A Cross-Discipline View of Trust," *Academy of Management Review* 23, no. 3 (July 1, 1998): 395.
29. Elizabeth Yakel, Ixchel Faniel, Adam Kriesberg, and Ayoung Yoon, "Trust in Digital Repositories," *International Journal of Digital Curation* 8, no. 1 (June 14, 2013): 143–56, doi:10.2218/ijdc.v8i1.251.
30. Adolfo G. Prieto, "From Conceptual to Perceptual Reality: Trust in Digital Repositories," *Library Review* 58, no. 8 (2009): 593–606, doi:10.1108/00242530910987082.
31. Yakel et al., "Trust in Digital Repositories."
32. Kari Kelton, Kenneth R Fleischmann, and William A Wallace, "Trust in Digital Information," *Journal of the American Society for Information Science and Technology* 59, no. 3 (February 1, 2008): 363, doi:10.1002/asi.20722.
33. Ibid.
34. Devan Ray Donaldson and Paul Conway, "User Conceptions of Trustworthiness for Digital Archival Documents: User Conceptions of Trustworthiness for Digital Archival Documents," *Journal of the Association for Information Science and Technology* 66, no. 12 (2015): 2427–44, doi:10.1002/asi.23330.

35. I. M. Faniel and T. E. Jacobsen, "Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data," *Computer Supported Cooperative Work* 19, no. 3–4 (August 2010): 355–75, doi:10.1007/s10606-010-9117-8; Betsy Rolland and Charlotte P. Lee, "Beyond Trust and Reliability: Reusing Data in Collaborative Cancer Epidemiology Research," in *CSCW '13: Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, TX: ACM, 2013), 435–44; Ann Zimmerman, "Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse," *International Journal on Digital Libraries* 7, no. 1–2 (2007): 5–16, doi:10.1007/s00799-007-0015-8; Faniel et al., "The Challenges of Digging Data"; Kriesberg et al., "The Role of Data Reuse in the Apprenticeship Process"; Ixchel M. Faniel, Adam Kriesberg, and Elizabeth Yakel, "Data Reuse and Sensemaking among Novice Social Scientists," *Proceedings of the American Society for Information Science and Technology* 49, no. 1 (2012): 1–10, doi:10.1002/meet.14504901068.
36. Alexandria Archive Institute, "Bridging Data Creation and Reuse," accessed March 8, 2016, http://alexandriaarchive.org/projects/bridging-creation-and-reuse/.

Bibliography

Alexandria Archive Institute. "Bridging Data Creation and Reuse." Accessed March 8, 2016. http://alexandriaarchive.org/projects/bridging-creation-and-reuse/.

Arbuckle, Benjamin S., Sarah Whitcher Kansa, Eric Kansa, David Orton, Canan Çakırlar, Lionel Gourichon, Levent Atici, et al. "Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey." Edited by Stephen Shennan. *PLoS ONE* 9, no. 6 (June 13, 2014): e99845. doi:10.1371/journal.pone.0099845.

Asma, Stephen T. *Stuffed Animals and Pickled Heads: The Culture and Evolution of Natural History Museums*. Oxford; New York: Oxford University Press, 2003.

Atici, Levent, Sarah Kansa, Justin Lev-Tov, and Eric Kansa. "Other People's Data: A Demonstration of the Imperative of Publishing Primary Data." *Journal of Archaeological Method and Theory*, April 11, 2012, 1–19. doi:10.1007/s10816-012-9132-9.

Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles: Sage, 2009.

Daniels, Morgan, Ixchel Faniel, Kathleen Fear, and Elizabeth Yakel. "Managing Fixity and Fluidity in Data Repositories." In *Proceedings of the 2012 iConference*, 279–86. Toronto: ACM, 2012. doi:10.1145/2132176.2132212.

Donaldson, Devan Ray, and Paul Conway. "User Conceptions of Trustworthiness for Digital Archival Documents." *Journal of the Association for Information Science and Technology* 66, no. 12 (2015): 2427–44. doi:10.1002/asi.23330.

Faniel, I. M., and T. E. Jacobsen. "Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data." *Computer Supported Cooperative Work* 19, no. 3–4 (August 2010): 355–75. doi:10.1007/s10606-010-9117-8.

Faniel, Ixchel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel. "The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse." In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 295–304. New York: ACM, 2013. doi:10.1145/2467696.2467712.

Faniel, Ixchel M., Adam Kriesberg, and Elizabeth Yakel. "Data Reuse and Sensemaking among Novice Social Scientists." *Proceedings of the American Society for Information Science and Technology* 49, no. 1 (2012): 1–10. doi:10.1002/meet.14504901068.

———. "Social Scientists' Satisfaction with Data Reuse." *Journal of the Association for Information Science and Technology* 67, no. 6 (June 2015): 1404–16. doi:10.1002/asi.23480.

Frank, Rebecca D., Adam Kriesberg, Elizabeth Yakel, and Ixchel M. Faniel. "Looting Hoards of Gold and Poaching Spotted Owls: Data Confidentiality among Archaeologists and Zoologists." In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, 37:1–37:10. Silver Springs, MD: American Society for Information Science, 2015. http://dl.acm.org/citation.cfm?id=2857070.2857107http://dl.acm.org/citation.cfm?id=2857070.2857107.

Frank, Rebecca D., Elizabeth Yakel, and Ixchel M. Faniel. "Destruction/Reconstruction: Preservation of Archaeological and Zoological Research Data." *Archival Science* 15, no. 2 (January 11, 2015): 141–67. doi:10.1007/s10502-014-9238-9.

Gutmann, Myron P., Mark Abrahamson, Margaret O. Adams, Micah Altman, Caroline Arms, Kenneth Bollen, Michael Carlson, et al. "From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data." *Library Trends* 57, no. 3 (2009): 315–37. doi:10.1353/lib.0.0039.

Harley, Diane, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and C. Judson King. *Assessing the Future Landscape of Scholarly Communication : An Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley: University of California Press, 2010.

Hey, Tony, and Anne Trefethen. "The Data Deluge: An E-Science Perspective." In *Grid Computing*, edited by Fran Berman, Geoffrey Fox, and Tony Hey, 809–24. Hoboken, NJ: John Wiley & Sons, 2003.

Holdren, John P. "Increasing Access to the Results of Federally Funded Scientific Research." Memorandum for the Heads of Executive Departments and Agencies, Office of Science and Technology Policy, Executive Office of the President, February 22, 2013. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Inter-university Consortium for Political and Social Research. "ICPSR Bibliography of Data-Related Literature." May 2014. http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/.

Kansa, Eric C., and Sarah Whitcher Kansa. "We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication." *Journal of Eastern Mediterranean Archaeology and Heritage Studies*, Project Muse, 1, no. 1 (March 16, 2013): 88–97.

Kansa, Eric C., Jason Schultz, and Bissell N. Ahrash. "Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a 'Some Rights Reserved' Model." *International Journal of Cultural Property*, no. 12 (2005): 285–314. doi:10.1017/S0940739105050204.

Kansa, Sarah Whitcher, and Eric Kansa. "Beyond BoneCommons: Recent Developments in Zooarchaeological Data Sharing." *SAA Archaeological Record* 11, no. 1 (January 2011): 26–29.

Kelton, Kari, Kenneth R. Fleischmann, and William A. Wallace. "Trust in Digital Information." *Journal of the American Society for Information Science and Technology* 59, no. 3 (February 1, 2008): 363–74. doi:10.1002/asi.20722.

Kriesberg, Adam, Rebecca D. Frank, Ixchel M. Faniel, and Elizabeth Yakel. "The Role of Data Reuse in the Apprenticeship Process." *Proceedings of the American Society for Information Science and Technology* 50, no. 1 (2013): 1–10. doi:10.1002/meet.14505001051.

Morse, Janice M., and Linda Niehaus. *Mixed Method Design: Principles and Procedures*. Walnut Creek, CA: Left Coast Press, 2009.

Office of Management and Budget. "Circular A-110 Revised 11/19/93 As Further Amended 9/30/99." White House website. Last updated September 30, 1999. http://www.whitehouse.gov/omb/circulars_a110.

Prieto, Adolfo G. "From Conceptual to Perceptual Reality: Trust in Digital Repositories." *Library Review* 58, no. 8 (2009): 593–606. doi:10.1108/00242530910987082.

Rolland, Betsy, and Charlotte P. Lee. "Beyond Trust and Reliability: Reusing Data in Collaborative Cancer Epidemiology Research." In *CSCW '13: Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 435–44. San Antonio, TX: ACM, 2013.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. "Not So Different after All: A Cross-Discipline View of Trust." *Academy of Management Review* 23, no. 3 (July 1, 1998): 393–404.

Shankar, Kalpana, Kristin Eschenfelder, and Greg Downey. "Studying the History of Social Science Data Archives as Knowledge Infrastructure." *Science and Technology Studies* (in press).

Society of American Archaeology. *Editorial Policy, Information for Authors, and Style Guide for* American Antiquity*,* Latin American Antiquity *and* Advances in Archaeological Practice. Washington, DC: Society of American Archaeology, last revised October 14, 2014. http://www.saa.org/portals/0/saa/publications/styleguide/styleguide_final_813.pdf.

Star, Susan Leigh, and James R. Griesemer. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39." *Social Studies of Science* 19, no. 3 (August 1, 1989): 387–420. doi:10.1177/030631289019003001.

University of California Berkeley, Museum of Vertebrate Zoology. "MVZ Guidelines for Scientific Collecting Permits." Accessed March 18, 2016. http://mvz.berkeley.edu/Sci_Coll_Permits.html.

Vardigan, Mary, Pascal Heus, and Wendy Thomas. "Data Documentation Initiative: Toward a Standard for the Social Sciences." *International Journal of Digital Curation* 3, no. 1 (August 6, 2008): 107–13. doi:10.2218/ijdc.v3i1.45.

Walters, Tyler, and Judy Ruttenberg. "SHared Access Research Ecosystem." *Educause Review* 49, no. 2 (2014): 56–57.

Yakel, Elizabeth, Ixchel Faniel, Adam Kriesberg, and Ayoung Yoon. "Trust in Digital Repositories." *International Journal of Digital Curation* 8, no. 1 (June 14, 2013): 143–56. doi:10.2218/ijdc.v8i1.251.

Yoon, Ayoung. "End Users' Trust in Data Repositories: Definition and Influences on Trust Development." *Archival Science* 14, no. 1 (March 2014): 17–34. doi:10.1007/s10502-013-9207-8.

Zimmerman, Ann. "Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse." *International Journal on Digital Libraries* 7, no. 1–2 (2007): 5–16. doi:10.1007/s00799-007-0015-8.