Arch Sci (2015) 15:141–167
DOI 10.1007/s10502-014-9238-9

ORIGINAL PAPER

Destruction/reconstruction: preservation of archaeological and zoological research data

Rebecca D. Frank · Elizabeth Yakel · Ixchel M. Faniel

Published online: 11 January 2015

Abstract Archaeology and zoology are fields in which data collection and analysis involve destruction. In this study, we examine the results of 49 interviews with archaeologists and zoologists, focusing on researchers' discussions of internal or disciplinary norms and external factors affecting their attitudes and actions concerning preservation. We identified two categories of disciplinary practices: data collection and data management/recordkeeping as key to shaping attitudes and activities about preservation. Likewise, we found three external factors: funding, legal requirements, and the status of museums and repositories, influencing attitudes toward preservation. We found that while archaeologists and zoologists are uniquely positioned to appreciate the value of data preservation, because data collection in both disciplines involves destruction, they are sceptical about whether preservation is possible and that these attitudes are influenced by both internal and external factors.

Keywords Data preservation · Digital preservation · Data reuse · Archaeology · Zoology · Data collection

R. D. Frank (✉)
School of Information, University of Michigan, 3429 North Quad, 105 S. State St, Ann Arbor, MI 48109-1285, USA
e-mail: frankrd@umich.edu

E. Yakel
School of Information, University of Michigan, 4323 North Quad, 105 S. State St, Ann Arbor, MI 48109-1285, USA
e-mail: yakel@umich.edu

I. M. Faniel
OCLC Research, 6565 Kilgour Place, Dublin, OH 43017-3395, USA
e-mail: fanieli@oclc.org

Introduction

The nature of data collection and analysis as destructive acts has played a significant role in shaping the practices of each field. Archaeologists excavate a site, remove artefacts, and later may take pieces from artefacts to analyse their composition, forever changing the landscape or object. Zoologists trap specimens or take a small sample (such as a fin clipping) to be used in analysis. In both cases, the original object and the context of study are necessarily altered by the researcher. Given the inevitable destruction in the act of research, we were interested in exploring attitudes and actions concerning data preservation in each field.

We frame our discussion around theories of infrastructure development which hold that infrastructures are socially constructed, yet also shape the communities in which they operate (Hughes 2012; Van der Vleuten 2004). In this paper, we consider disciplinary norms for data collection and management as well as external factors, such as funding, legal requirements, and repositories, as the sociotechnical infrastructure supporting the work (research, knowledge creation) of the archaeological and zoological communities (Hughes 2012; Ribes and Finholt 2009; Van der Vleuten 2004). We find that the ways in which archaeologists and zoologists contribute to, interact with, and experience this sociotechnical system affect their attitudes about data preservation and the actions they take regarding the preservation of data. We argue that the similarities between the two disciplines—including their consciousness of data collection as a destructive act—are reflected in attitudes about the value of data preservation and actions taken to preserve data, but that differences in both internal and external factors result in divergent attitudes about the feasibility of long-term data preservation.

Our study is motivated by the following research questions:

1. How do disciplinary practices and norms affect how archaeologists and zoologists view/understand preservation as it relates to their own research data?
2. What external factors influence the attitudes of archaeologists and zoologists toward the feasibility of long-term preservation of research data?

In this paper, we examine the results of 49 interviews with archaeologists and zoologists. We focus on researchers' discussions of internal or disciplinary norms and external factors affecting their attitudes and actions concerning preservation. We identified two categories of disciplinary practices: data collection and data management/recordkeeping as key to shaping attitudes and activities about preservation. Likewise, we found three external factors: funding, legal requirements, and the status of museums and repositories, influencing attitudes toward preservation.

Literature review

Our theoretical framework is based on studies of infrastructure development, and in particular on the development of research and knowledge infrastructures. We draw

upon the definition of infrastructure developed by Bowker and Star (1999), which describes infrastructure as embedded into other structures, transparent, having reach or scope beyond a single event, being learned as part of inculcation into a community, having links with conventions of practice, embodying standards, built upon an installed base, becoming visible only upon breakdown, and changing in modular increments rather than all at once (Bowker and Star 1999, p. 35). Knowledge infrastructures are socially constructed, and in particular, the infrastructures supporting the work of archaeologists and zoologists discussed in this article are products of variation and selection, having been built incrementally over time to support the work of the scientists who rely upon them (Bijker et al. 2012). These infrastructures are shaped by the communities of researchers who use and build them, and in turn, the infrastructures shape the research practices and data preservation practices of those communities (Bowker and Star 1999).

Knowledge infrastructures are built around a series of disciplinary norms and external factors. We define a norm as "a voluntary behaviour that is prevalent within a given reference group" (Interis 2011, p. 425). Data collection and data management/recordkeeping can be considered practices governed by norms. In both archaeology and zoology, researchers engage in voluntary behaviours that are prevalent throughout their disciplinary communities. We have selected these two practices (data collection and data management/recordkeeping) because they encompass both physical and intellectual work and thus lead to questions and actions of destruction and preservation of data. The external factors that we examine are as follows: funding agency mandates, legal requirements, and the role of museums and digital repositories in the disciplines. We have selected these factors because they represent elements of the infrastructure that support academic research and are significant factors influencing the data practices of researchers.

Disciplinary norms

Data are fundamental to scientific endeavours; however, what constitutes "data" differs widely among different communities of practice (Borgman 2008). Borgman, Wallis, and Mayernik acknowledge that data can come from a wide variety of sources including: observations, computations, experiments, surveys, or archives, and in diverse analogue and digital formats including text, sound, images, models, genomic sequencing, observational data, and other forms that are produced either by humans or machines (Borgman et al. 2012). This conceptualization of data needs to be extended for archaeologists and zoologists because it limits the definition of data to information *about* phenomena. For archaeologists and zoologists, data include the objects of study: the artefacts and specimen themselves (Ilerbaig 2010). Hodder defined archaeological data as, "a set of dynamic, dialectical, unstable relations between objects, contexts, and interpretations" that involves both the physical world and our theories about it (Hodder and Hutson 2003; Hodder 1999, p. 84). In his definition, Hodder considered the artefact itself to be data, as well as the contextual information about an artefact and the interpretations or models made by researchers. Star and Griesemer's (1989) discussion of data in a museum of zoology provided a similar picture of data as encompassing specimens, contextual information, and

interpretations, and Kanfer et al. (2000) argued that data were deeply embedded in social systems and contexts and that separating data from context is difficult at best.

Thus, our working definition of data in this paper aligns with the communities of practice under study. Data encompasses both analogue and digital objects excavated, captured, or produced by archaeological and zoological researchers, with an emphasis on the importance of maintaining a link between the object and contextual information about that object. This approach reinforces the findings from Borgman et al. (2012) who noted that the distinction between data and context depends on the perspective of the data user.

Since we are studying perception and actions impacting preservation, we are interested in the disciplinary practices that most influence data: collection and management and recordkeeping. Despite the fact that the use of data is a fundamental practice across scientific disciplines, the particular practices surrounding data collection and management and recordkeeping are discipline-specific and based on historical and logistical factors, as well as on an understanding of what constitutes evidence in a particular field.

The literature provides some background on these activities. The process of collecting data in archaeology includes excavation, destroying the sites that researchers seek to examine and understand (Aitchison 2009), and survey, where sherds or bone materials are identified on the top of the ground (sometimes removing them and sometimes not). In these processes, archaeologists capture contextual information, such as the relationships between finds, geolocation, stratigraphy, and the conditions under which artefacts were collected, as this information is critical for later analysis and interpretation; however, this very information is often destroyed through the data collection process (Hodder and Hutson 2003). The fact that "excavation destroys evidence" means that "it is important for the archaeologist to assume that the contextual relationships are fixed" (Hodder 1999, p. 93). The destruction of context that occurs when artefacts are collected makes preservation of data—both the artefact and the contextual information that accompanies the artefact—of the utmost importance for archaeologists. "Artefacts are not best used when considered independent of the contexts from which they were recovered" (Orser 1996, p. 285).

For zoologists, the organism itself is the primary unit of study (Bartholomew 1986; Gilbert and Sarkar 2000). Greene (2005) argues that "an understanding of organisms in nature is… integral to studies at both lower and higher levels in the hierarchy of biological complexity" (p. 24). The field of zoology includes individuals who identify themselves with a range of disciplinary specialties, including evolutionary biology, ecology, systematics, and biodiversity. Each of these disciplines relies upon the specimen as a unit of data, either directly through examination, or indirectly by using international databases aggregating metadata, images, or genomic data about specimens. While some may choose to focus on a level of analysis smaller than the specimen, such as the genetic sequence, others choose to focus on a level of analysis larger than the specimen, such as the ecosystem. In both cases, the field of zoology broadly conceived has roots in the tradition of natural history as a method of data collection and inquiry (Arnold 2003;

Bartholomew 1986; Gilbert and Sarkar 2000; Greene 2005; Noss 1996; Wilson 1989).

Data collection in zoology involves a range of activities, from traditional specimen collection to the production of geolocation information, counts of species, or sensor data (Bezanson et al. 2013; Borgman et al. 2012; Brussard 1982). Often data collection involves the production of multiple types of data, for example capturing geolocation data from where a specimen was collected, writing field notes about the collection event, and/or taking photographs of the specimen in the field. All these activities are done to "fix" or preserve knowledge of the context of the collection event at a point in time, as this collection event cannot be replicated at a later date. These data are also important to maintain if the actual specimens are used for DNA analysis. As with archaeology, the changeability of the environment and destruction of specimens for data collection and analysis are primary concerns for this field (Bezanson et al. 2013; Taberlet et al. 1999; Waits and Paetkau 2005).

This destruction is significant both because of the potential ecological impact that researchers can have on the environments and populations that they wish to study, and also because the destructive nature of specimen collection means that data collection cannot be duplicated or replicated as one can repeat a laboratory experiment. Brower et al. (1998) noted, in their manual for field methods, that the dimensions of a habitat include temporality, physicality, and geographic, chemical, and biotic dimensions. These dimensions reinforce the idea that data collection is temporally and spatially bound and cannot be replicated. As with archaeology, the inherent destructiveness of data collection in zoological practice makes preservation of data, including contextual information, important for both the collectors of data and researchers who would seek to reuse these data at a later date (Borgman et al. 2012; Kanfer et al. 2000).

For both archaeology and zoology, the management of research data is an issue of ongoing importance (Aitchison 2009; Beagrie et al. 2009; Borgman et al. 2012; Gray et al. 2005; Plale et al. 2005). Data management/recordkeeping practices in archaeology are diverse and not standardized due to the widely varying research questions, methodologies, and perspectives among researchers (Atici et al. 2012). This high degree of variance increases the need for adequate contextual information to help the researchers themselves, as well as potential reusers, make sense of research data. These factors also contribute to the high volume of data that can be generated in archaeological research, which makes good data management and recordkeeping practices particularly important. Zoologists also collect high volumes of data and contextual information, although data collection practices in this field are more standardized, with guides for field methods providing researchers with sample forms for data collection (Brower et al. 1998).

Contextual information helps researchers make sense of the data they collect; it also helps others who reuse that same data later (Borgman et al. 2012). Gray et al. (2005) note that the management of data in a broad sense, specimen, as well as contextual information is important, "because data collection is now separated from data analysis, extensive metadata describing the data in standard terms is needed so people and programs can understand the data" (Gray et al. 2005, p. 40). Field notes

in archaeology are similar to those in anthropology, idiosyncratic and highly personal (Jackson 1990).

Data format is a factor that affects how researchers in the fields of archaeology and zoology manage their data and how others later curate and preserve that data. Researchers use a variety of formats for data, and archaeologists in particular use a wide range of software programs to create data, both proprietary and non-proprietary (Condron et al. 1999). The format of the data will have a strong effect on what types of data management and recordkeeping activities can take place and what level of support a repository can provide for any given dataset (Garrett and Waters 1996). Both archaeology and zoology are disciplines that work with a mix of analogue and digital data. While some objects can be digitized, and a great deal of contextual information is either captured digitally or digitized after the fact, researchers in both disciplines rely upon artefacts and specimens for much of their research (Greene 2005; Ilerbaig 2010; Kristiansen 1996). In addition to these physical objects, both archaeologists and zoologists rely upon field notes (both analogue and digital), hand drawings (analogue), and photographs (digital) as a way to capture information that can be considered either as original data, contextual information, or both (Belton 2009; Blair 2004). Other genres of data that researchers in these fields collect include such formats as geospatial information (GIS) and genetic sequence data, which are born digital. In some ways, the format of the data itself influences the ways in which preservation must be approached. Different repositories manage different types of data, and so researchers need to consider where their data will go when they decide how to capture it, as repositories seek to encourage researchers to create data in formats that will be easier to manage and preserve (Rombouts and Princic 2010).

External factors

In addition to the internal factors discussed above, disciplinary practices around data preservation in archaeology and zoology are influenced by external factors, such as funding agency mandates, legal requirements, and the role of museums and digital repositories in the disciplines. These external factors can be thought of as elements of the research infrastructure that support the work of archaeologists and zoologists. We are also focusing on these requirements because they most closely influence data collection and management and recordkeeping.

Researchers in both archaeology and zoology rely on grant funding to support data collection efforts. In the USA, funding organizations such as the National Science Foundation (NSF), the National Endowment for the Humanities (NEH), or the Andrew W. Mellon Foundation (Mellon) fund archaeological research, while the NSF is the primary funder of zoological research. Requirements for data sharing and data management are becoming increasingly common with grant funding for research, particularly from public or government institutions such as the NSF or NEH (National Endowment for the Humanities Office of Digital Humanities 2013; The National Science Foundation 2013). Researchers are required to submit data management plans as part of grant applications and are expected to share their data

with other researchers, "at no more than incremental cost and within a reasonable time" (The National Science Foundation 2013).

The data management requirements established by funding agencies, such as the NSF and NEH, are attempting to shift norms within the research communities regarding data, but we cannot yet say whether their efforts are successful. Researchers' plans for data management implicitly emphasize the value of data and thus how data are viewed as candidates for preservation (Akmon 2014). This is particularly important given the fact that funding for research is decreasing in the USA, making it increasingly important that researchers are able to get the fullest possible value from data. Additionally, new disciplines such as sustainability and climate science require data sharing across disciplines and over time (Edwards 2010).

Requiring, rather than suggesting, that researchers demonstrate a plan to manage their data helps to communicate shifting norms within the scientific community and the public, as well as specific disciplinary communities, both about the value and importance of data and also about the types of activities that researchers are expected to engage in with regard to their research data (Bicarregui et al. 2013; Cox and Pinfield 2013; Griffiths 2009). One particular challenge with regard to these data management requirements for researchers is that "it is difficult to know in advance which data are valuable to curate and preserve" (Faniel and Zimmerman 2011). It is also important to note that a significant challenge to the efforts of funding agencies to create new norms around data management is that even though the grant agencies require data management plans, they specifically do not allocate funds for the implementation of these plans: "It is NSF's strong expectation that investigators will share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grant" (National Science Foundation 2010).

Legal requirements influence data preservation in archaeology and zoology in a number of ways. Both the excavation of archaeological sites and the collection of zoological specimens often require permits or licenses that must be issued from the government in the area where the data collection is to take place. Researchers must comply with legal requirements that govern data collection and the subsequent deposit of the objects or specimens (Messenger and Fagan 1999). In archaeology, the UNESCO Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property changed practice in that previously artefacts could more easily be removed from the original country (United Nations Educational, Scientific, and Cultural Organizations 1970). Now treaties and national laws focusing on the protection of cultural property ban the removal of physical objects from some countries.

In zoology, researchers are required to obtain permits to collect specimens. When researchers deposit their specimens into museums, they are usually required to submit a copy of the permit in order to prove that the specimens were obtained legally. This requirement is considered to be both a legal requirement and a signal of professional and ethical conduct (Michener et al. 1997; Winker et al. 2010). In addition to granting permission for collection, the permits also comprise important

contextual information that must be preserved along with data (Michener et al. 1997).

Museums and repositories play a significant role with regard to data preservation in both archaeology and zoology. As we discussed above, researchers are often legally required to deposit artefacts and specimens in museums. Additionally, funding agencies often require that researchers deposit their data into a repository, such as GenBank for genetic sequence data in zoology, in order to promote data sharing and preservation (The National Science Foundation 2013). Finally, journal publishers also are beginning to require that data are available for scientists to replicate and better understand the evidential basis of articles (Whitlock 2011).

We argue that the internal and external factors discussed above are significant for understanding how archaeologists and zoologists understand preservation as it relates to their own research data, and how they understand the feasibility of long-term preservation of research data. Examining these factors, which are a part of the infrastructure that supports research and knowledge creation, will help us to understand the attitudes of researchers toward preservation.

Methods

This current study is part of a larger research project, Dissemination Information Packages for Information Reuse (DIPIR). The Institute for Museum and Library Services sponsored the 3-year project to examine data reuse among three disciplinary communities: zoology, archaeology, and quantitative social science (Faniel and Yakel 2011). This current study focuses on 49 interviews conducted with the zoologists (27) and archaeologists (22) in 2011 and 2012, with a primary emphasis on those researchers who engage in data collection and fieldwork.

Participants

Our sample consists of researchers from two disciplines: archaeology and zoology. These two disciplines were selected because of the similarities in their data collection practices and the differences in their data sharing norms. Both archaeology and zoology involve data collection activities that can be seen as destructive. In some cases, this destruction is literal, as when an archaeological site is excavated or when a zoological specimen is killed. And in others, this destruction can mean a destruction of the original context within which an object was found. This destruction—of either the object of study or of the context within which that object was found—means that the link between the object of study and contextual information about that object is of particular importance for both of these disciplines. This contextual information, along with the object itself, comprises data that researchers in archaeology and zoology collect and/or create and deposit in museums or repositories. This focus on the museum or repository as a place for data deposit, and the place where data are preserved, is an important similarity between the two disciplines for the purposes of this study. The differences between the two disciplines in terms of maturity of data sharing infrastructure emphasize the

differences in norms around data sharing (Faniel et al. 2013; Zimmerman 2008), which allows us to ask how these norms affect attitudes toward preservation and how external factors, such as museums and repositories, affect attitudes about the feasibility of long-term data preservation.

The similarities and differences between the two disciplines with regard to norms around data collection and management, the type of research questions asked, and the infrastructures that have developed in response to these norms and practices provide an important focus for this study. Namely, that while the similarities allow us to compare the two, the differences provide us with a locus for our examination.

Participant recruitment

We recruited subjects from both disciplinary communities using convenience and snowball sampling techniques. We used three recruitment methods. First, we worked with our research partners in zoology and archaeology to recruit a diverse group of interviewees (different sub-specialties, various levels of expertise, geographic focus of research). Second, we recruited at disciplinary conferences. Third, we asked interviewees to suggest other potential interviewees.

In the end, we conducted interviews with 22 archaeologists and 27 zoologists. The majority of these subjects were based at institutions in the USA. These semi-structured interviews lasted approximately 1 hour and respondents were compensated in the amount of $25 for their participation in the study. In the interview protocol, we asked participants about their data reuse practices, their data collection practices, and their attitudes toward the role of repositories in their respective disciplines. The interviews were audio recorded and transcribed for analysis. This study was reviewed and approved by the Institutional Review Board at the primary author's university.

Analysis

We analysed the interview transcripts using NVivo, a qualitative data analysis software. We developed an initial codeset based on themes from the literature and supplemented these with emergent codes arising from the coding and initial analysis process (Miles and Huberman 1994). This coding, which combined deductive and inductive approaches, enabled us to compare our data with existing themes in the literature as well as to recognize new findings from our study. The final codeset addressed topics such as data collection experience, data reuse practices, interaction with other researchers, and ethical considerations regarding data reuse. For each group, two coders worked independently coding the same transcript. We followed this process through several iterations until we reached an acceptable level of interrater reliability for independent coding. Using Scott's Pi, a statistic measuring interrater reliability for coding textual data, we achieved a score of .73 for the archaeologists, and .74 for the zoologists.

After this initial round of analysis, we delved further into the data. We isolated several codes (data collection, ethics, interaction with the designated community, legal issues, and preservation) for a second cycle of analysis focused on addressing

the topics of original data collection and examining attitudes toward preservation of research data in greater depth. First, we went through each transcript and applied one new code in order to identify all sections where respondents discussed original data collection. These passages were extracted from the transcripts as a report, and these two reports, one for archaeologists and another for zoologists, were used for further coding and analysis. Second cycle coding focused on respondents' experiences with original data, including properties of the data, data collection experience, and data management and preservation experience. As this level of coding was done by a single team member, no reliability rating is available.

Findings

In this section, we present the findings from our study. Specifically, we find that the disciplinary practices of data collection and data management and recordkeeping, and the external factors of funding, legal requirements, and the status of museums and repositories influence attitudes toward the long-term preservation of research data. Researchers recognize the importance of data preservation, but remain sceptical about its feasibility.

Data collection

Archaeologists and zoologists do not explicitly have preservation issues on their minds during data collection. Rather, it is destruction that poses the biggest challenge. Archaeologist18 described data collected through fieldwork as "a subset of the potential area of information that existed at one point for which we again, as I've said, gone in and destroyed". Archaeologist15 explained that, "archaeological excavation in fact destroys the context in which the evidence is discovered. It's this ironic thing in our discipline in which we actually destroy the laboratories in which we recover our data. So, there's an added burden of data recording placed on us in the field and sometimes under very difficult conditions". This particular archaeologist went on to discuss the debate within archaeology about preservation versus destruction, as data practices are evolving to include highly sophisticated metallurgic, chemical, and radiographic techniques:

> Archaeology is always about destruction, and so, if you really want to understand how an object is made, you often have to destroy a portion of it…And I think all museums face this issue… you have to often apply for a destructive testing permit. You have to be very careful. It's just a very difficult culture. This curatorial culture that places preservation over… They prioritize preservation over other kinds of knowledge practices (Archaeologist15).

Zoologists and archaeologists have responded to this destruction by turning their attention to collecting quality data. Quality is assessed in different ways including capturing multiple data points about a specimen, working from a good research design, and collecting contextual data. Zoologist18 discussed the importance of capturing as much information as possible about a specimen at the time of collection

when a specimen will not remain in its original condition over time, "I had a standard data sheet for tunas that tried to get most of the morphometric and meristic data because we were going to dissect the fish and turn it into a pile of bones. So you need to have all the possible information ahead of time because you weren't going to have a specimen—a standard specimen—when you got through".

Research design was another key to good data. For archaeologists, such as Archaeologist07, "good" fieldwork was difficult to achieve, but an important marker of being a member of the archaeology community, "Digging, the actual fieldwork, is really hard. It's not easy to be a 'good digger'". Yet, for zoological research there is sometimes no perfect design. Zoologist02 explained some of the difficulties he encounters in ensuring the quality of the specimens that he collects in the field:

> I work on deep water species and they're notorious for being sort of crappy in their… let's say quality…. Most folks, when they collect a fish that's, it's alive, they euthanize it, and they take the tissue sample right away. When I am collecting tissues out at sea, these nets go down for… It might take two hours for them to fish all the way down to the target depth, and then fish at the target depth for some prescribed amount of time, then another few hours to come up. The fish that I'm after could have been caught on the way down, and been dead for five, six hours.

For archaeologists, data degradation, such as a specimen or artefact breaking down over time, caused problems even when a research design was properly executed, "In any archaeological site is a unique instance of a particular culture. You really need to know the peculiar characteristics of each one. And if those are lost to some sort of degradation or deaccessioning, then you're losing valuable information. At some point, you then have to wonder whether the data you're looking at are really a representative sample anymore" (Archaeologist12).

More than collecting lots of data or good research design, collecting data in a way that would contextualize the specimen or the find was the most commonly discussed way that archaeologists and zoologists countered destruction. "And if I aim to actually rank categorically my data, I would say the first most important aspect would be the archaeological context: where my data is coming from" (Archaeologist16). Archaeologists and zoologists stressed the importance of preserving both the object and the context. Archaeologist13 identified the preservation of contextual information as important for being able to make sense of artefacts later:

> What do you archive and how do you access its context and how do you make it meaningful?… And even worse, even with the code unless you know the stratigraphy of the site that I am working on, it makes it hard for another specialist to deal with those data and then the problems that go along with all of those sort of contextual… there is the contextual extra layers upon layers of other information. Some of which only seems to exist in the heads of people that is never written down.

Archaeologists and zoologists primarily view good data collection as a means of reconstructing the data for analysis. Thus, preservation is viewed largely in terms of

reconstruction. Few of our participants explicitly linked data collection and preservation, and when they did, the association was not positive. Zoologist06 noted that he has control over data quality but not preservation. "I can't go back and make sure that the preservation process was correct. But what I can do is [know] the data I collect is accurate". Archaeologist22 commented about the lack of concern he viewed from his colleagues, "I think [repositories are] way more important than lots of people understand. Many, many, many people are only collecting data digitally anymore and don't have a clear concern for its long-term preservation".

In spite of the lack of explicit references to preservation, these quotations demonstrate the importance of data collection to preservation, and in particular, the preservation of meaning and the ability to return to the data and analyse it later. Closely related to data collection are how the data are managed, and the recordkeeping practices of the researchers in the data processing and analysis stages.

Data management and recordkeeping

Archaeologists and zoologists learn how to manage the data that they collect through academic advisors who communicate community norms as well as trial and error. Their data management and recordkeeping practices influence their attitudes toward data preservation—researchers in both disciplines view preservation of research data as being important, but are sceptical about the viability of long-term preservation for their data.

Archaeologist17 explained, "And so I'm totally open to having help in making sure that my data is preserved and is archivable. But it would depend on whether that means I get a pro-forma sheet which I'm required to use or whether it's more like I do the fieldwork that I want and then I then, sort of, map those terms to sort of existing terminology". This archaeologist expressed an attitude of openness toward producing data documentation that supports data management and preservation, but expressed uncertainty about what kind of help would be available to support that production. The cost of changing data management practices for preservation, potentially to the detriment of the research, was a consistent theme among both archaeologists and zoologists.

The costs that researchers associated with data management and recordkeeping, and in particular the higher cost of managing data when also trying to preserve that data, were a common concern among both archaeologists and zoologists. Researchers were concerned about the higher costs in terms of time and resources associated with managing large amounts of data. For Zoologist02, the shift from managing relatively small amounts of data as a student to managing large amounts of rich data as a more senior researcher brought the issue of cost to the forefront:

> And so from someone who works on dissertation that is like mine was, there is little, smaller caches of data in each little part of the dissertation work stage; each chapter required its own type of data. But the data, they weren't huge in terms of their content or… Let's say, I didn't have any storage issues. But now that I work with much more rich data and, say, hundreds of different text files, managing those, it's just… Now I just feel like I'm staring at directories all

> day and wondering where the file was, where the last time I put it… right now I'm grappling with how to maybe develop a server type system.

The issue of increased data management costs was also described by Archaeologist01, who expressed concern about managing large amounts of data. In particular, this archaeologist discussed the high cost, and higher stakes, that are present when managing large amounts of data, "The way in which everybody has to be retooled, reeducated into the latest technology every 3–5 years. That's the scariest part there as we're setting this stuff up. Like someone says. 'Okay. You've done it this way, but now we need you to jigger it just slightly this way with 10,000 items'". Archaeologist01 went on to describe a situation in which the repository managing his data switched servers and in the process lost the contextual information linking photographic images:

> I built this elaborate thing, with thousands and thousands of links, and then they simply switched servers. And they did some sort of process where they said, "We've got all the links. It's okay". And I found out that data that was stored in folders, within folders, within folders all got disconnected, and so most of our photographic links are broken.

These two examples show that data management and recordkeeping practices can be costly and difficult for researchers managing their own data, and also for researchers whose data are stored in a repository. In both cases, archaeologists and zoologists consider preservation to add to the cost of managing data in terms of time and money.

In addition to being costly, managing data for preservation is difficult and takes time away from other activities. The high cost in terms of time that archaeologists and zoologists associate with managing data for preservation is often seen as taking time away from other activities, "To be quite honest, the biggest hurdle when you're dealing with genetic data in like depositing … the information and the sequence data onto GenBank is associating that with museum specimens or locality data …. It's really kind of clunky and it really takes a lot of time to do that" (Zoologist10). And for Archaeologist12, preservation of digital data is described as a lower priority than preserving analogue data, because of the greater likelihood that analogue data will survive over a long period of time:

> But a hand-written or typed spreadsheet on acid-free paper can last for centuries. And archaeologists have intuitively recognized the long-term preservation potential in that. And that, I think, is actually driven a lot of the, not refusal, but lack of. Not a refusal to do data work, but if you only have so much time and money to allocate to putting your data on the record, then you're going to do it in the way that's going to last most certainly the longest amount of time… If you put it on a floppy disk and no one cares about it for 30 years, you know by the time they find it, the floppy disk is rotten and then it's all lost.

Similarly, Zoologist22 discussed guarding against loss of digital data by maintaining print copies of digital collections:

> The thing I'd know that is kind of sketchy, which I don't even do yet myself, is there's not really a hard copy. Like I didn't print out my database, and I don't know how many people still do that. And I'm sort of trusting that it's out there in different collections now but maybe I should… And we back it up, but I heard that people think that you should be printing out stuff and I don't know if that's… I haven't heard of anybody losing their entire collection and multiple backups all at the same time …That did give me pause the other day, I was thinking about, somebody said that what if all the servers in the US go down or something? [chuckle] I don't know if that's a possibility, but I guess it is, somehow. And so we should have a hard copy of our collections. I guess it's something to think about.

This concern about spending time and resources to manage data for preservation only to lose that data was echoed by both archaeologists and zoologists. And Archaeologist16 described an experience of actually losing data that was not properly preserved, "I wrote my MA dissertation using software like Professional Writing, PW WordStar. Now I don't have a copy because things have changed and I haven't been able to update or upgrade my stuff, it's gone. I only have a printed copy". Zoologist08 shared a similar concern and had in fact lost data that he was attempting to manage on his own, "I do have a lot of images which I've never used, which are sitting on my computer which would probably be better on a long-term repository because I have lost external hard drives in the past actually, lost lots of data". While Zoologist05 talked about both the fear of losing data and challenges in managing data, "For now, it's mostly just a fear of losing it, so backing up multiple copies. With the direction we're going with the next generation data, the genomic data, I think there'll be lots of other complications with storing it, as well as lots of ongoing work with how to handle the data when you receive all of these data. It's such a massive amount of data, how to computationally handle it? So I think that would be one of the biggest hindrances".

For both archaeologists and zoologists, preservation was seen to add to the cost of data management and recordkeeping practices. In some cases, these costs became prohibitive and the researcher was put in a position where preserving data was perceived to be detrimental to the actual research, "I guess it depends on what that system is… Because it then becomes this level of what point do the standards for preservation or archiving become sort of prohibitive of the needs of a site or a project?" (Archaeologist17).

Disciplinary norms guide the formats in which data were collected, and therefore the formats that researchers must be prepared to manage and preserve. For archaeologists working in multidisciplinary teams, this meant collecting data in a format common to the team; for zoologists, this entailed collecting the required information along with the specimen in a way that was easily exported when the specimen was turned over to a museum. Yet, the sheer number of formats used by researchers in these disciplines posed preservation problems. Archaeologist15 described the data formats he used throughout his research process:

> It's a very interdisciplinary collection method that we use … and draws on gross geological techniques and zoological techniques and architecture… So,

> the kinds of ways that I record data are first and foremost paperwork. We have notebooks in the field that are either pencil and paper version or digital format…The next step would be photography; a lot of digital photography. Both, say, publishable and stuff that's more of record keeping but doesn't have the resolution to be published. We also use high-resolution survey equipment like Total Station so we are collecting say a lot of spatial data. So, that is recorded in spreadsheet files and CAD files that we can then use to reconstruct in a CAD-like environment, GIS/CAD environments…We also catalogue all the artifacts that we find, we assign everything a unique identification number, and build databases to keep that information together. And we do a lot of what we call post-excavation processing, in which that material is processed in a field laboratory, and then it's prepared, oftentimes prepared either to be stored for the long term in the Middle East, or a portion of it is shipped to the United States for laboratory analysis, and there, a whole other level of recording takes place.

From this quotation, we not only see a variety of formats, but a mix of analogue and digital, proprietary and non-proprietary. Furthermore, preservation includes not only preservation of the pieces, but also preservation of the relationship between data residing in the different formats.

Researchers expressed scepticism that digital data could be preserved long term. When discussing the problems of data management and preservation with regard to data format and storage media over time, Archaeologist23 commented:

> When I went to high school, we had a computer, but nobody was really using them yet. It was the late 1980s. And you hear the old timers tell stories about how they still have datasets and they're on like computer cards and all that kind of crap. And I guess you start to get more and more jaded about the permanence of anything having to do with technology that's changing so rapidly.

These quotations express the importance of data management and recordkeeping for preservation and demonstrate the scepticism that researchers have about the viability of long-term preservation for their research data, particularly when those data are stored in digital rather than analogue formats.

Funding

A major external factor, which colours ideas on preservation in both communities, is funding for research. Both archaeologists and zoologists are dependent on external support for data collection; however, the infrastructure in each discipline varies and this in turn affects both how research data are preserved and perceptions about preservation. In this section, we discuss three elements of the funding infrastructure: data management and sharing mandates; funding for data sharing, management, archiving, and preservation; and granting agencies attention (or lack thereof) to preservation issues. Researchers in both archaeology and zoology expressed positive attitudes toward the concept of data management requirements but were

uncertain about the viability of such practices, in particular expressing concern about the financial sustainability of such requirements.

Data management and sharing mandates

In the USA, funding for research in archaeology and zoology comes from a variety of public and private funding sources. In particular, interview subjects mentioned the NSF, the NEH, and Mellon as major sources of funding for projects. Funding that comes from public sources such as the NSF often comes with stipulations for data management. While management is not necessarily long-term preservation, data management does affect the ability to preserve data. As described previously, archaeologists and zoologists attitudes and actions around data management are indicators of preservation.

Interview subjects spoke about those stipulations both as mandates that require compliance, and as norms that indicate the ways in which they are expected to behave as researchers. Archaeologist13 discussed the NSF requirement to create a data management plan. "I completed an NSF grant in December and… you have to have now a section that describes what you are going to do with your data…Data availability and where you're going to archive it… So you're being forced to deal with it now whereas in the past you're like, 'Well it's in my file cabinet'".

Zoologists also described the shift in norms from individual researchers managing their own data toward a model of making research data publicly available. This type of public availability of data contrasts with other models of data sharing such as providing data to individuals only upon request. Zoologist04 stated that, "increasingly, anytime you're funded by National Science Foundation, which is where my primary funding comes from, I think there's an increasing expectation, which I think is a positive thing, towards making your data publicly available". While the focus of these data management requirements is often on making data available, zoologists expressed an assumption that the repositories where they deposited their data would both provide access to and also preserve their research data, "Of course, that's a wonderful thing about GenBank, is that if I submitted all my data to GenBank, it's going to be there in perpetuity, or so one would assume" (Zoologist02).

While data management practices are changing, respondents expressed some scepticism that making their data available to others via a repository would make their data usable for others. In particular, Archaeologist09 said, "I was largely funded by NSF to do this, I always felt that I had to make it available. But it's pretty tricky on some of it. I think you have to put a lot of documents with parts of it to make it really available because I'm not always sure that people understand what they're using". This researcher expressed a common concern among researchers: that simply depositing data into a repository will not be sufficient in terms of making data accessible to other researchers, or usable over long periods of time. Zoologist23 discussed the types of contextual information that had to be preserved along with the data, and how he uses field notes to help make sense of data that he obtains from repositories, "field notes can be really, really helpful in understanding some of these data that are on the databases". These conversations shed light on the

ways in which researchers understand the importance of preserving the relationships between objects and contextual information in order to meet the spirit, but not the stipulations, of the data management requirements.

The mandates that funding agencies issue regarding data management have helped to raise the awareness about the importance of managing and, to a lesser extent, preserving research data. In some cases, as with Zoologist20, these mandates reflect long-held norms for data management within research communities, "I mean, as we continue to gather more and more data for different groups of organisms, having long-term access and long-term preservation of these data is extremely important. I mean people use data, biological data that's hundreds or more years old from museums to look at conservation changes or species diversity changes as humans expand their footprint on everything, so having that data in long-term storage is very valuable". In other cases, such as Archaeologist13, these mandates have initiated slow changes in the norms of research communities that do not have a longstanding tradition of depositing data beyond the objects or specimens into museums or repositories, "The data themselves are sort of hidden, hidden away, either by previous generations who still do things on paper, so the folks who are, I don't know, who are in their 50s and 60s, not all of them, but who are sort of closing on retirement and who run a lot of departments and a lot of field projects, a lot of their data are still in file cabinets". In either case, researchers are becoming aware of the expectation to manage data in ways that ensure its accessibility, and that one of the ways to ensure accessibility of data is to include the contextual information that users require in order to be able to make sense of that data.

Funding for data sharing, management, archiving, and preservation

Despite these generally positive opinions of data management and sharing requirements, interviewees also noted that grant funding often cannot be used for data management, archiving, or preservation. Sometimes this prohibition is described as being explicit, "you can't buy servers with NSF money or NEH money" (Archaeologist11). Others, such as Archaeologist19, described it as being implied or understood, "even though the NSF, for example, is saying come up with an archiving plan, in truth I think that the field and the peers who are reviewing are not really that interested in seeing a huge chunk of change goes to that portion of the project".

Some of the researchers that we spoke with argued that preservation should take place outside of the actual research project and should be managed by an organization other than the research team collecting the data, in order to ensure the long-term viability of that data. Archaeologist11 spoke about the fact that grant funding cannot be used to purchase equipment for data preservation, but suggested that preservation should take place at an institutional level rather than a project level, "I mean, actual excavation is a five-year project or a ten-year project and that's over … Somebody could die tomorrow, so if the data is not institutionally based, you are going to lose it". While this appears to contradict Archaeologist21′s statement that preservation will only happen when researchers are required to allocate funding to it out of their grants, both are making the point that the

responsibility for long-term preservation should not be placed on individual researchers and projects.

Not being able to use research project funding for data preservation was a common theme among our interviewees. In some cases, as with Archaeologist05, grants awarded to increase data accessibility still cannot be used for preservation. This is particularly interesting given that one way to think about preservation is as accessibility over time. Archaeologist05 discussed a specific example of a grant from a private foundation whose focus was on making data and artefacts accessible, but whose funds could not be used for preserving that data. While this project did involve digitization of resources in order to improve access, the interviewee drew a sharp distinction between access and preservation of their data, "…but in our case for example all our [funding amount] from the Mellon Foundation was really about the digital scholarship, it was not about the preservation of the original archives… our focus, our mandate was not to be restoring glass plates or reattaching torn pieces off of maps and things like that. It was to try to get the scholarship accessible" (Archaeologist05).

Researchers who are aware of both the need to preserve their data and the lack of funding that can be allocated toward preservation expressed concern about what will happen to their data in the future. Zoologist22 described a scenario in which grant funding was able to support data management during the grant period but he expressed concern that once the grant period ended there was no financial support for data preservation, "I'm glad there's all these databases. But my main concern is some of the ones that are supported by grants, what happens when the grants run out?… GenBank will always be there obviously, but who knows if EOL [Encyclopedia of Life] will always be there? Animal Diversity Web… So, hopefully, they will be, but if they're grant-dependent, what happens when the PI…you know, what happens to that data later?"

Another way in which this concern about how to fund data preservation efforts was expressed came in the form of respondents discussing concerns about repositories' ability to preserve data. Archaeoloigst04 explained that she wanted a repository to be able to demonstrate a plan for long-term preservation, "one of the concerns that I know comes up with a lot of these repositories is, what happens when the NEH or the NSF funding runs out? Who's going to take care of the collection, who's going to run it? So for using a repository, I'd want to know that there was a long-term plan for it. That's really important". This concern is distinct from earlier discussions about how to fund preservation of data in that CCU04 specifically talked about who would care for data rather than how those efforts would be funded or what equipment would be available.

Requirement of data deposit for publishing

In addition to funding agency mandates for data management, many journals also require data to be deposited in a repository. This influences attitudes around preservation by providing researchers with an incentive to deposit their data. Zoologist19 described this requirement as a reflection of long-held community norms about depositing data, "Most of the top journals are now requiring deposit of

data into Dryad upon publication of a paper. And certainly the museum community has always required that specimen-based data be tied to unique identifiers in our world's museums. That's been true for hundreds of years". For genomic data, journals require deposit into GenBank, "every journal requires that you upload it on GenBank if it's a sequence" (Zoologist27). While the requirements do not include mechanisms to ensure or enforce compliance, researchers' behaviour is shifting toward preserving their data in repositories where it will be accessible to more people over the long term. This is particularly true for researchers who want to publish in these journals in the future.

While norms for data management in zoology were described by respondents as being focused on depositing specimens into museums, the linking of data deposit to publication also reflected norms about what types of data should be shared and what should be kept private. "When I sit down and start checking my own measurements and counts on specimens then those data are not publicly available in any way, and so I wouldn't be sharing those, no. I would never… Just not be sharing until I publish to paper" (Zoologist28). In particular, Zoologist28 explained that while the specimens themselves were shared freely, any data generated based on analysis of those specimens belonged to the individual researcher and that work product need not be shared until required for publication. This information included contextual information about the specimen that in many cases was required to make sense of the specimens. In this case, the researcher described a situation in which the publication process delays long-term data management or preservation activities for some types of data because the data are not deposited until after publication.

Researchers generally expressed positive attitudes toward funding agency and publisher requirements about data management and preservation. Archaeoloigst21 made the point that until funding agencies allowed data preservation costs to be included in the grant, it was unlikely that researchers would fully comply with these requirements, "And it seems to me that one of the things we need is for granting agencies generally, not just NEH or NSF, but everybody who gives money to archaeology, has to start saying, 'Not only may you put archival preservation into your budget, but it must be there. It must be there from day one. And you must have a repository for your budget'. Because that's what will make the repositories happen, is when they know that… Or somebody knows, that there is enough money coming in from datasets that will be deposited that they can actually fund themselves".

Legal requirements

Legal requirements support preservation for both digital and analogue data, including artefacts and specimens. Both archaeologists and zoologists are required to obtain permits to collect data in the field, and museums and repositories for both disciplines require proof that the proper permits for data collection were obtained before they will accept data, including artefacts and specimens, for deposit. Researchers obtain permits that govern data collection and must then engage in the responsible and ethical data collection practices that the permits require. The practices that are required by these permits reflect behaviours that are recognized

within the respective communities as being ethical and are required as proof that the ethical norms were upheld when artefacts and specimens were collected in the field. The evidence required to demonstrate compliance with these norms comprises important contextual information that supports preservation.

Zoologist13 explained that he must provide information about the location of collection, demonstrating to the government agency granting the permits that the specimens were obtained legally, an issue that is particularly sensitive for endangered species, "if I have a permit to collect endangered species I have to make it very clear exactly where I am going collecting to the state agency that provided the permit. So that's not something I can hide, nor would I want to". Location information constitutes an important piece of contextual information that is needed in order to understand data over a long period of time. Similarly, location plays a significant role for archaeologists, although the ways in which this information is (or is not) shared with the public often looks quite different for archaeologists than for zoologists. Archaeologist03 explained that location information about artefacts is important to collect and provide to the local government, but when asked about sharing that information with the public said, "that's just not going to apply there". In this case, the researchers collected location information and provided that information to the government that granted them the data collection permit, but this particular country did not make that information publicly available. These two examples illustrate that there are different ways of meeting the goal of preservation—by making data and contextual information publicly available, or by restricting access to that same information in order to preserve sites or specimens for future use.

In addition to museums and repositories requiring permits as a condition of accepting data, some also require the researcher to agree to deposit their data into a museum or repository. For example, Zoologist04 described a situation in which he was required to make an agreement with a museum in order to obtain a permit to collect the data in the first place, "for permits, it's a requirement that you make all of your specimens—that they are going to be accessioned through somewhere".

Overall, the legal requirements that govern how artefacts and specimens are collected and managed contribute to the preservation of research data. They do so by requiring the collection of valuable contextual information that can be used to help make sense of the data over long periods of time, and by ensuring that this information is kept with the artefacts and specimens in repositories and/or museums where researchers are required to deposit their data.

Institutional infrastructure of museums/repositories

A great deal of data preservation—preservation of both physical specimens and artefacts as well as digital objects—takes place not at the level of the individual researcher but rather at the level of the museum or repository. Our respondents talked about the ways in which this orientation toward museum- and repository-based preservation of research data reflects norms within their respective communities of practice.

Zoologists expected that any specimen collected in the field would automatically be placed in a museum. For example, Zoologist03 explained that "on the zoological side the data I collect is, or it will be, museum specimen records". In this case, the respondent was talking about depositing not only the specimen itself into a museum but also the metadata about that specimen. This assumption of data deposit was expressed by several respondents from the zoology community, such as Zoologists 05, 10, 25, and 06. Zoologist06 noted, "Any specimens that I collect in the field get deposited to a museum. Along with the data, GPS coordinates, any tissues, things like that". This contextual information that researchers deposit along with their artefacts and specimens helped to ensure that the meaning of the data would be preserved over time.

While Archaeologist01 described leaving items that museums would not accept in the field, Archaeologist15 described a situation in which he retained artefacts that he collected in the field and could not find a museum to accept them: "[Government] allowed me take 99 % of what we excavate out of the country. It's one of the last [region] countries that allows this. They don't have repository facilities, and so their logic is, obviously here in the United States we do. But I don't, personally; the [name] Museum does not have space. They are not accepting collections like my material".

Zoologists also discussed the ways in which museums and repositories influenced the preservation of digital data in addition to physical specimens, a much broader spectrum of the available data than for archaeologists. Zoologist06 described the kinds of metadata that are collected and submitted to a museum along with a specimen:

> So usually we submit a lot of photographs of specimens that we catch. Obviously, we take down the locality, date, time information, whether or not we obtained tissues from the animal, and then beyond that, usually like if it is male or female, some natural history information like was it was sitting on a log, under a log, or under a rock, was it on a road, those kind of things. Usually during a collection event … you'll have to [note] what weather was outside, and the temperature, as well as like was it raining. That's usually in the field notes.

Zoologist23 discussed the ways in which museums were providing access to field notes and metadata by entering that information into a searchable database for researchers, "I know in the herpetology collection here, they're working to that where they actually go through field notes and pretty much put as much of that information from the field notes as possible into a database. So, I think that's pretty smart, but very time consuming".

While zoologists spoke often about depositing objects into museums, and data, such as genetic sequence data into GenBank, archaeologists were more likely to talk about finding ways to build their own databases or repositories to store and manage their non-object data. Archaeologist02, for example, discussed a project to create an online museum, **"**I mean, there are very few repositories. I'm working on another project with my friends at [name] University to build a sort of virtual museum for Arctic archaeology, and we're sort of thinking through those things right now. How are we going to create this visual metadata? We want to put full field notes up

online. Do we scan those as PDFs, or do we recreate them into digital format? How do we do all those sort of stuff? We were hoping we could sort of find solutions to solve those problems soon because we want to encourage people to submit their data to the project".

For both of these communities, the institutional infrastructure of museums and repositories affected the ways in which researchers thought about preservation and utilize their options for preserving data. For zoologists, the well-established norms around specimen and data deposit into museums and repositories were reflected in the attitudes about who can access data and how to gain access to data. For example, "I deposit all my collections in the institution where I currently reside. So yeah, I put those in the [state] museum. And basically, as far as sharing goes, I guess, anybody that wants to gain access to those records, they could potentially use them" (Zoologist23).

For archaeologists, the expectation that a researcher would be able to go to a museum to gain access to research data seemed far less certain. Rather than discussing the ways in which museums and repositories provided access to data, archaeologists discussed ways in which museums limited access to data for research in the interest of pursuing goals that focus on data preservation. For example, Archaeologist15 explained: "archaeology is always about destruction, and so, if you really want to understand how an object is made, you often have to destroy a portion of it, to get at its guts… And I think all museums face this issue, not… I mean, you have to often apply for a destructive testing permit. You have to be very careful. It's just a very difficult culture. This curatorial culture that places preservation over… They prioritize preservation over other kinds of knowledge practices, I suppose".

For both archaeologists and zoologists, the institutional infrastructure of museums and repositories greatly influenced preservation of physical and digital research data. In some cases, museums and repositories furthered preservation goals by aligning them with the goals of access and reuse, and in others, they furthered preservation goals by restricting access and reuse. In both cases, museums and repositories played a role in the preservation of research data.

Researchers discussed preservation as something that was someone else's responsibility. Specifically, that the goal and responsibility of repositories were to preserve data and that once a researcher has deposited her data, it was no longer her responsibility to worry about long-term preservation, "Well, I think they are extremely important for archaeologists and almost everything we do comes down to primary data and to this massive, massive complex datasets. So having databases and repositories is what I feel it's based on. I guess it's kind of integrated into our field, it's extremely important. And sometimes part of the mission statement of archaeology is to preserve our data so having them digitized and having these repositories is a way of doing that" (Archaeologist04).

Discussion

In this paper, we demonstrate how the disciplinary practices of data collection, and data management and recordkeeping, as well as external forces such as funding

agency mandates, legal requirements, and the infrastructure of museums and repositories affect archaeologists' and zoologists' attitudes and activities concerning the preservation of research data. We found that researchers generally recognized the importance of data preservation, but that their focus was on more immediate concerns of limited time in the field for data collection and the necessities of data analysis. Our participants in both disciplines remain sceptical about the feasibility of actually preserving data over long periods of time and tend not to see it as their responsibility.

Archaeologists in particular expressed their most salient preservation concern in terms of their ability to reconstruct a site after the inevitable destruction that is inherent in much of their data collection. The goal of preserving meaning was to mentally reconstruct the physical site, which was destroyed. This reconstruction process was based on the data collected—particularly contextual information to support the fullest possible understanding of the site. Interestingly, this attention to contextual information has been identified as one of the key elements for preserving the meaning of research data (Faniel and Jacobsen 2010; Zimmerman 2008). In this sense, we can think of data preservation as reconstructing the information that was destroyed during data collection.

The attention to contextual data is good news for repositories, although capturing contextual data in a standardized and usable form for preservation and subsequent use is difficult given the data management practices. Archaeologists and zoologists described idiosyncratic data management practices that depend on a number of factors, including data format, available funding, technical knowledge or expertise, and publishing schedules that affect whether and how contextual information is preserved.

We found that both archaeologists and zoologists made a distinction between data management and data preservation, perhaps reinforced by funding agency requirements, which reinforce this dichotomy. While researchers will often manage at least some portion of their own data rather than depositing that data directly into a repository, archaeologists and zoologists emphasized current use and their own sensemaking when discussing data management. For archaeologists, this meant that they developed idiosyncratic methods of recording data that addressed their particular research question. Members of both disciplines also engaged in a real calculus concerning time, effort, and resources spent on data management, which they saw as taking away from focus on the science. These behaviours work both for and against preservation. While data are valued, the focus on immediate use as opposed to future reuse makes preservation more difficult, particularly for archaeological data where there is less standardization and therefore greater cost to the museum or online repository to process the data.

While researchers generally expressed expectations that they would manage their own data in the short term, neither the zoologists nor the archaeologists saw long-term preservation as their responsibility. Both discussed instances in which they personally lost data, either through neglect or as a result of actions taken on the data. Members in both disciplines viewed museums and repositories as responsible for addressing preservation issues, such as format obsolescence, and as being best able to preserve data long-term. These votes of confidence for museums and repositories

were tempered by reservations about whether long-term preservation, particularly with regard to digital information, is feasible, even if deposited in a repository.

Archaeology and zoology are disciplines whose data collection practices involve destruction. Awareness of the destructive nature of data collection makes researchers in these disciplines sensitive to the need to preserve both the artefact/specimens and the contextual information that is required to make sense of that object. This focus on data collection as a destructive act has resulted in a greater reliance on the contextual information and data about the object and collection event. The need to preserve the data about the objects has therefore become more intense.

Conclusion

We examined the results of 49 interviews with archaeologists and zoologists and found that disciplinary practices and norms around data collection, and data management and recordkeeping emphasize reconstruction of the research site or data collection event rather than preservation per se. Research practices in these disciplines emphasized the importance of data preservation while at the same time making researchers aware of the difficulties associated with ensuring the long-term accessibility of data. We also found that external factors such as funding agency mandates, legal requirements, and the infrastructure of museums and repositories also focus attention explicitly on data and data sharing and only implicitly on preservation. The lack of funding for long-term data preservation makes preservation planning and activities difficult, even when required by funding agencies.

We found that while archaeologists and zoologists are uniquely positioned to appreciate the value of data preservation, because data collection in both disciplines often involves destruction, they are sceptical about whether preservation is possible, and that these attitudes are influenced by both internal and external factors. For future research in this area, we suggest examining attitudes toward preservation among other disciplines with destructive data collection practices, and also among disciplines whose data collection practices are not destructive for comparison.

An important factor that influences whether researchers will comply with data management and sharing mandates, or with disciplinary norms around preservation, is whether researchers see preservation as feasible or even possible. This paper has examined some of the factors that influence attitudes toward preservation, and we have found that researchers are generally aware of the need to preserve data, but are uncertain of the feasibility of preserving data long-term. This knowledge gap between recognizing the need and understanding how to preserve data suggests that repository managers can help to preserve data and support data reuse by focusing on the how rather than the why of data preservation when addressing researchers.

Acknowledgments This research was made possible by a National Leadership Grant from the Institute for Museum and Library Services, LG-06-10-0140-10: Dissemination Information Packages for Information Reuse.

References

Aitchison K (2009) Standards and guidance in archaeological archiving: the work of the archaeological archives forum and the institute for archaeologists. Grey J 5(2):67–71

Akmon DR (2014) The role of conceptions of value in data practices: a multi-case study of three small teams of ecological scientists (Dissertation). University of Michigan, Ann Arbor

Arnold SJ (2003) Too much natural history, or too little? Anim Behav 65(6):1065–1068. doi:10.1006/anbe.2003.2143

Atici L, Kansa S, Lev-Tov J, Kansa E (2012) Other people's data: a demonstration of the imperative of publishing primary data. J Archaeol Method Theory, 1–19. doi:10.1007/s10816-012-9132-9

Bartholomew GA (1986) The role of natural history in contemporary biology. Bioscience 36(5):324–329. doi:10.2307/1310237

Beagrie N, Beagrie R, Rowlands I (2009) Research data preservation and access: the views of researchers. Ariadne, (60). http://www.ariadne.ac.uk/issue60/beagrie-et-al/

Belton T (2009) The dawn of the "chaotic account": horatio hale's australia notebook and the development of anthropologists' field notes. Librar Cult Rec 44(1):138–152

Bezanson M, Stowe R, Watts SM (2013) Reducing the ecological impact of field research. Am J Primatol 75(1):1–9. doi:10.1002/ajp.22086

Bicarregui J, Gray N, Henderson R, Jones R, Lambert S, Matthews B (2013) Data management and preservation planning for big science. Int J Digit Curation 8(1):29–41. doi:10.2218/ijdc.v8i1.247

Bijker WE, Hughes TP, Pinch T (2012) The social construction of technological systems: new directions in the sociology and history of technology, Anniversary Edition. MIT Press, Cambridge

Blair A (2004) Note taking as an art of transmission. Crit Inq 31(1):85–107. doi:10.1086/ci.2004.31.issue-1

Borgman CL (2008) Data, disciplines, and scholarly publishing. Learn Publ 21(1):29–38. doi:10.1087/095315108X254476

Borgman CL, Wallis JC, Mayernik MS (2012) Who's got the data? Interdependencies in science and technology collaborations. Comput Support Coop Work 21:485–523. doi:10.1007/s10606-012-9169-z

Bowker GC, Star SL (1999) Sorting things out: classification and its consequences. MIT Press, Cambridge

Brower JE, Zar JH, von Ende CN (1998) Field and laboratory methods for general ecology. Mcgraw-Hill, Boston

Brussard PF (1982) The role of field stations in the preservation of biological diversity. Bioscience 32(5):327–330. doi:10.2307/1308849

Condron F, Richards J, Robinson D, Wise A (1999) Strategies for digital data—a survey of user needs. http://ads.ahds.ac.uk/project/strategies/. Accessed 24 July 2012

Cox AM, Pinfield S (2013) Research data management and libraries: current activities and future priorities. J Librar Inf Sci. doi:10.1177/0961000613492542

Edwards PN (2010) A vast machine: computer models, climate data, and the politics of global warming. MIT Press, Cambridge

Faniel I, Jacobsen T (2010) Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. Comput Support Coop Work (CSCW) 19(3):355–375. doi:10.1007/s10606-010-9117-8

Faniel I, Yakel E (2011) Significant properties as contextual metadata. J Libr Metadata 11(3–4):155–165

Faniel IM, Zimmerman A (2011) Beyond the data deluge: a research agenda for large-scale data sharing and reuse. Int J Digit Curation 6(1):58–69. doi:10.2218/ijdc.v6i1.172

Faniel I, Kansa E, Whitcher Kansa S, Barrera-Gomez J, Yakel E (2013) 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. ACM, New York. doi:10.1145/2467696.2467712, pp 295–304

Garrett J, Waters DJ (1996) Preserving digital information: report of the task force on archiving of digital information (No. 9781887334501 1887334505). The Commission on Preservation and Access & Research Libraries Group, Washington, p 68

Gilbert SF, Sarkar S (2000) Embracing complexity: organicism for the 21st century. Dev Dyn 219(1):1–9. doi:10.1002/1097-0177(2000)9999:9999<:AID-DVDY1036>3.0.CO;2-A

Gray J, Liu DT, Nieto-Santisteban M, Szalay A, DeWitt DJ, Heber G (2005) Scientific data management in the coming decade. SIGMOD Rec 34(4):34–41. doi:10.1145/1107499.1107503

Greene HW (2005) Organisms in nature as a central focus for biology. Trends Ecol Evol 20(1):23–27. doi:10.1016/j.tree.2004.11.005

Griffiths A (2009) The publication of research data: researcher attitudes and behaviour. Int J Digit Curation 4(1):46–56. doi:10.2218/ijdc.v4i1.77

Hodder I (1999) The archaeological process : an introduction oxford [England]. Blackwell, Malden

Hodder I, Hutson S (2003) The problem, Chapter 1. Reading the past: current approaches to interpretation in archaeology, 3rd edn. Cambridge University Press, Cambridge, pp 1–19

Hughes T (2012) The evolution of large technical systems. In: Bijker W, Hughes T, Pinch T (eds) The social construction of technological systems. MIT Press, Cambridge, pp 45–76

Ilerbaig J (2010) Specimens as records: scientific practice and recordkeeping in natural history research. Am Arch 73(2):463–482

Interis M (2011) On norms: a typology with discussion. Am J Econ Sociol 70(2):424–438

Jackson JE (1990) "I am a fieldnote": fieldnotes as a symbol of professional identity. In: Sanjeck R (ed) Fieldnotes: the makings of anthropology. Cornell University Press, Ithaca, pp 3–33

Kanfer AG, Haythornthwaite C, Bruce BC, Bowker GC, Burbules NC, Porac JF, Wade J (2000) Modeling distributed knowledge processes in next generation multidisciplinary alliances. Inf Syst Front 2(3–4):317–331. doi:10.1023/A:1026520728644

Kristiansen K (1996) The destruction of the archaeological heritage and the formation of museum collections: the case of Denmark. Learning from things. Smithsonian Institution Press, Washington, pp 89–101

Messenger PM, Fagan BM (1999) The ethics of collecting cultural property: whose culture? Whose property?. University of New Mexico Press, Alburquerque

Michener WK, Brunt JW, Helly JJ, Kirchner TB, Stafford SG (1997) Nongeospatial metadata for the ecological sciences. Ecol Appl 7(1):330–342. doi:10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

National Endowment for the Humanities Office of Digital Humanities (2013) Data management plans for NEH office of digital humanities proposals and awards. http://www.neh.gov/files/grants/data_management_plans_2013.pdf

Miles MB, Huberman AM (1994) Qualitative data analysis: an expanded sourcebook. Sage Publications, Thousand Oaks

National Science Foundation (2010) Data management & sharing frequently asked questions. National Science Foundation. Government. http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp#1. Accessed 30 Jan 2014

National Science Foundation (2013) Dissemination and sharing of research results. National Science Foundation Office of Budget, Finance, and Award Management. Government. http://www.nsf.gov/bfa/dias/policy/dmp.jsp. Accessed 25 Nov 2013

Noss RF (1996) Editorial: the naturalists are dying off. Conserv Biol 10(1):1–3

Orser CE (1996) Images of the recent past: readings in historical archaeology. Alta Mira Press, Walnut Creek

Plale B, Gannon D, Alameda J, Wilhelmson B, Hampton S, Rossi A, Droegemeier K (2005) Active management of scientific data. IEEE Internet Comput 9(1):27–34. doi:10.1109/MIC.2005.4

Ribes D, Finholt T (2009) The long now of technology infrastructure: articulating tensions in development. J Assoc Inf Syst 10(5):375–398

Rombouts J, Princic A (2010) Building a "Data Repository" for heterogenous technical research communities through collaborations (Vol. Paper 10). Presented at the International Association of Scientific and Technological University Libraries, 31st Annual Conference. http://docs.lib.purdue.edu/iatul2010/conf/day2/10

Star SL, Griesemer JR (1989) Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. Soc Stud Sci 19(3):387–420. doi:10.1177/030631289019003001

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. Trends Ecol Evol 14(8):323–327. doi:10.1016/S0169-5347(99)01637-7

United Nations Educational, Scientific, and Cultural Organizations (1970) Convention on the means of prohibiting and preventing the illicit import, export and transfer of ownership of cultural property—1970. UNESCO. http://www.unesco.org/new/en/culture/themes/illicit-traffic-of-cultural-property/1970-convention/

Van der Vleuten E (2004) Infrastructures and societal change. A view from the large technical systems field. Technol Anal Strateg Manag 16(3):395–414. doi:10.1080/0953732042000251160

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. J Wildl Manag 69(4):1419–1433. doi:10.2193/0022-541X(2005)69 [1419:NGSTFW]2.0.CO;2

Whitlock MC (2011) Data archiving in ecology and evolution: best practices. Trends Ecol Evol 26(2):61–65. doi:10.1016/j.tree.2010.11.006

Wilson EO (1989) The coming pluralization of biology and the stewardship of systematics. Bioscience 39(4):242–245. doi:10.2307/1311161

Winker K, Reed JM, Escalante P, Askins RA, Cicero C, Hough GE, Bates J (2010) The importance, effects, and ethics of bird collecting. Auk 127(3):690–695

Zimmerman AS (2008) New knowledge from old data: the role of standards in the sharing and reuse of ecological data. Sci Technol Hum Values 33(5):631–652. doi:10.1177/0162243907306704

Rebecca D. Frank is a Ph.D. Candidate at the University of Michigan School of Information. Her research interests include risk and disaster planning for digital repositories, digital preservation, and digital curation. She also holds a Masters degree from the University of Michigan School of Information, with a focus on preservation of information.

Elizabeth Yakel, Ph.D. is a Professor and Associate Dean for Research and Faculty Affairs at the University of Michigan School of Information where she teaches in the Archives and Records Management and Preservation of Information areas. Her research interests include access and use of primary sources and user information behaviour in archives and digital repositories. Yakel's research has been supported by the National Historical Publications and Records Commission, the Institute of Museum and Library Services, and the Andrew W. Mellon Foundation.

Ixchel M. Faniel is an Associate Research Scientist at OCLC, Online Computer Library Center, Inc. Dr. Faniel's research interests include the reuse of content, particularly how to capture, document, and disseminate content in a way that makes it easier to share and reuse via technology. Her research has focused on the reuse of knowledge and scientific research data within industry organizations and academic communities. Her work has been supported by the National Science Foundation and the Institute of Museum and Library Services.