

Looting Hoards of Gold and Poaching Spotted Owls: Data Confidentiality Among Archaeologists & Zoologists

Rebecca D. Frank
University of Michigan
School of Information
105 S. State St
Ann Arbor, MI 48109
frankrd@umich.edu

Adam Kriesberg
University of Maryland
College of Information
Studies
Room 405, Hornbake
Building South
College Park, MD
20742
akriesbe@umd.edu

Elizabeth Yakel
University of Michigan
School of Information
105 S. State St
Ann Arbor, MI 48109
yakel@umich.edu

Ixchel M. Faniel
OCLC Research
6565 Kilgour PI
Dublin, OH 43017
fanieli@oclc.org

ABSTRACT

Researchers in the social and health sciences are used to dealing with confidential data, and repositories in these areas have developed mechanisms to prevent unethical or illegal disclosure of this data. However, other scientific communities also collect data whose disclosure may cause harm to communities, cultures, or the environment. This paper presents results from 62 interviews and observations with archaeologists and zoologists. It focuses on how researchers' perceptions of potential harm influence attitudes about data confidentiality, and how these, in turn, influence opinions about who should be responsible for managing access to data. This is particularly problematic in archaeology when harm is not to a living individual but is targeted at a community or culture that may or may not have living representatives, and in zoology when an environment or a species may be at risk. We find that while both archaeologists and zoologists view location information as highly important and valuable in facilitating use and reuse of data, they also acknowledge that location should at times be considered confidential information since it can be used to facilitate the destruction of cultural property through looting or decimation of endangered species through poaching. While researchers in both disciplines understand the potential dangers of allowing disclosure of this information, they disagree about who should take responsibility for access decisions and conditions.

Keywords

Digital repositories, data confidentiality, access to data, data

ASIST 2015, November 6-10, 2015, St. Louis, MO, USA.

Author Retains Copyright.

sharing, data reuse, data curation

INTRODUCTION

Data curation and reuse depend on open access to data. However, many datasets contain information that for ethical or legal reasons should remain confidential. Much has been written in the social sciences and health care research about personally identifiable information and the potential for individual harm (e.g. Corti et al., 2000; Sweeney, 1997). Less has been written about confidentiality concerns in other disciplines. In this paper, we examine data confidentiality in archaeology and zoology. In these disciplines, social harms, specifically cultural and ecological concerns, drive confidentiality decisions. This has implications not only for data collection and analysis but also for data curation and reuse.

We report on the results of 62 interviews with researchers from two disciplines: archaeology and zoology. We focus on how archaeologists and zoologists make legal and ethical decisions about whether specific types of data, such as location information, should be deemed confidential and who should have the right to access certain information. We argue that the scale of potential harms relating to the sharing of sensitive information, such as location data in both fields, is larger and more distributed than in other disciplines, such as quantitative social science. In archaeology and zoology location data is particularly relevant when considering data confidentiality and it signifies larger ethical concerns regarding the responsibility that researchers have to protect the populations or cultures that they study. Since the potential harms for archaeology and zoology are primarily at the cultural/social or species/biodiversity levels, decisions about whether data should be kept confidential are less clear-cut than in a field such as quantitative social science, where the potential harms are both more immediate and more likely to have direct impact on specific living individuals.

Our study is motivated by the following research questions:

- How do perceptions of risk or potential harm influence attitudes about data confidentiality among archaeologists and zoologists?
- How do perceptions of risk or potential harm influence attitudes about who should take responsibility for managing issues of data confidentiality among archaeologists and zoologists?

LITERATURE REVIEW

In recent years, the scholarly community has paid increased attention to issues surrounding data management, sharing, and reuse. Many have identified potential benefits that could be realized through broader data sharing (e.g. Arzberger et al., 2004; Karasti et al., 2006). Yet challenges remain, particularly around incentives for sharing and the logistics for access (e.g. Borgman, 2012; Chavan and Penev 2011; Wallis et al., 2013). An additional concern among scientists is how to balance restrictions to personally identifiable information while maintaining open access to the greatest amount of data. For projects that collect human subjects data on topics, such as socio-economic status and health, the benefits to researchers are clear, but the risks for subjects are real. Some Institutional Review Boards (IRBs) have put in place a range of restrictions on the dissemination of this type of human subjects data, but more work is needed to identify solutions that work for different types of sensitive data (Hartter et al., 2013). The related risks of data breaches on cloud storage systems further demand the attention of researchers as they consider how they manage potentially sensitive data before transfer to a repository (Jahnke and Asher, 2013).

The fields of healthcare and the social sciences are emblematic examples of the current state of data curation practice around the challenges of confidentiality and data ethics. Health and social science data often represent sensitive details of living individuals, and may require additional work to ensure that participants cannot be identified in publicly available data. The Inter-university Consortium for Political and Social Research (ICPSR) maintains a guide for researchers seeking to deposit data in its digital archives. This document contains detailed instructions around data anonymization (The Inter-university Consortium for Political and Social Research, 2012). Additional examples from the research literature demonstrate the ways in which this community has worked to build disciplinary norms around confidentiality of research subjects, ranging from recommendations to include provisions about data archiving and reuse in informed consent documents to engaging in broader discussions about the ethics of data sharing and access (e.g. Corti et al., 2000; Kuula 2010; Parry and Mauthner, 2004). Repositories have also devised specific policies to transfer the risk of disclosure from the repository to the researcher, thus allowing access to confidential data but restricting the publication of results that identify individuals. More recently, repositories, such as ICPSR, have employed

technology to assist in the protection of personally identifiable data. ICPSR has established data enclaves to allow the use but not the copying or downloading of confidential data (Lyle, 2014). Specific policies around participant confidentiality are also driven by the requirements of IRBs, which seek to ensure that academic research conforms to basic ethical standards and that the benefits outweigh the risks. Health researchers also take confidentiality and participant privacy very seriously; IRBs also have a lot of influence over this research community, stemming from their legal mandate to protect human subjects (45 CFR 46, 2009). However, patient consent is not required for disclosure of health information if the data are considered de-identified. This has resulted in a major push by the health research community to develop techniques for de-identification and protection against potential re-identification (El Emam et al., 2012). Research into privacy for health data is ongoing as the tensions between privacy and research remain unresolved (e.g. Oderkirk et al., 2013).

Since the establishment of their discipline, archaeologists have struggled with the unique ethical challenges inherent in their research practices. As Alison Wylie (1996) explains, the foundation of the Society for American Archaeology (SAA) was partially motivated by the desire of academics to distance themselves from amateur archaeologists through the establishment of ethical codes of conduct. One of the main concerns in the field has long been the practice of looting historical sites and the threat this poses to both cultural heritage and future archaeological work. Archaeologists today condemn looting for its piecemeal destruction of archaeological sites, its effect on illicit antiquities markets, and its secondary effects, wherein the search for valuable objects places human remains at risk (Kersel and Chesson, 2013). The current version of the SAA “Principles of Archaeological Ethics” makes reference to protecting sites from potential looters, noting the need to “[preserve] and [protect] in situ archaeological sites... when publishing and distributing information about their nature and location” (Society for American Archaeology, 1996). While short on specifics, this principle builds upon earlier pieces of the document, which call for more widespread data sharing and publication across the field. Digital archaeology repositories, such as Open Context, also have addressed this concern with regard to research data. Their “Data Publication Guide for Contributors” explains the responsibility of individual researchers to decide about any desired restrictions on location data before depositing datasets into the repository (“Open Context,” 2015). Mark Lynott notes that the diversity of the discipline demands that each case be considered individually (Lynott, 1997), while more recently, Ian Hodder similarly argued that universal ethical principles are extremely difficult to develop and apply (Hodder, 2011). The field continues to grapple with these ethical challenges, and to seek ways to reduce or eliminate looting (Krieger, 2014).

Similarly, researchers in zoology routinely deal with issues related to confidentiality and location data in their work. The data they collect reveals the habitats of endangered species and disclosure could attract poachers and place species at further risk. As such, The Global Biodiversity Information Facility (GBIF) published a guide for sharing Sensitive Primary Species Occurrence Data. This document suggests some specific solutions for various situations around potentially sensitive data that require special procedures or protection (Chapman and Grafton, 2008). For example, the guide calls for researchers to consider the level of sensitivity for the particular dataset when determining how much to round geospatial references (i.e. to the 0.1 degree or 0.001 degree). Johnathan Fong and Gexia Qiao (2010) describe a project to map locations of an endangered species of turtle in China and argue that while this location data is valuable to researchers, it should not be made publicly available due to concerns about the safety of the animals. Harter and colleagues (2013) point to the United States Forest Service as an example of a government agency that seeks to protect research sites by not disclosing geospatial references along with its data.

Both academic disciplines under examination in this paper, archaeology and zoology, face their own data-related challenges. Yet, a common element in both fields is the importance of location or geospatial information in answering research questions. While, archaeologists and zoologists agree that sensitive data, such as location information, are encountered in the research process, perceptions of risk and harm vary among individuals in each field. We explore these differences in this paper.

METHODS

This paper draws upon data collected as part of a larger study which examined data reuse across three academic communities: quantitative social science, archaeology, and zoology. The results of this paper are primarily drawn from interviews conducted with archaeologists, and interviews and observations with zoologists.

Participant Recruitment

Both sets of interview participants, as well as the observations of zoological researchers, were recruited from a set of contacts provided to the project team by our community partners in each research area. In addition, we attended disciplinary conferences to recruit participants and then employed a snowball sampling strategy to identify additional participants, asking those we interviewed about other potential participants who they thought would be interested in participating and whose research employed data reuse.

Data Collection

Using our existing networks as well as snowball sampling, we continued data collection until the team felt that we had achieved data saturation and had elicited the major findings from our participants based on our interview protocol. In total, we conducted interviews with 22 archaeologists and 27 zoologists as well as an additional 13 observational

interviews with zoologists as they worked with specimens from collections at a natural history museum. The interviews and observations lasted between 30 minutes and one hour and participants were paid \$25 for their participation in the study. All interviews were recorded and transcribed for subsequent analysis.

Data Analysis

All interview data as well as the audio from observations were analyzed using NVivo, a qualitative data analysis software package. Project team members began with an initial codeset based on the literature on data reuse and the interview protocol. The codeset covered a wide range of topics, reflecting the semi-structured nature of the interviews. The relevant codes analyzed for the present paper included confidentiality, legal issues, ethics and values, and funding agency mandates. During analysis, the project team met regularly to discuss coding and the emergence of themes not represented in the initial codeset. For example, archaeologists spoke about discipline-specific ethical challenges often enough to warrant the creation of a new code to document this concept. Coding began in tandem, with two project team members working on the same transcript. After multiple rounds of this paired coding, Scott's pi was used to calculate inter-rater reliability. The scores on this measure of reliability for coding qualitative data were 0.73 for archaeologists, 0.74 for zoological interviews, and 0.88 for zoological observations.

Community Descriptions

We collected demographic and descriptive information about our research participants. These attributes highlight the diversity of our research participants across research experience and methods.

Of the 22 archaeologists we spoke with, 13 were faculty members or faculty-curators while only 4 were graduate students. The remaining 5 participants were research scientists and curators. Sixteen (72.7%) of the archaeologists collected their own data in addition to reusing data from a range of sources. Eleven reported reusing data from existing publications, 14 reported contacting colleagues directly to obtain data for reuse, 17 utilized archival or museum collections for reuse, while 9 used online repositories.

Of the 40 zoologists who participated in interviews or observations, 22 were either students or postdocs while 14 were faculty or faculty-curators. The remaining 4 participants were either research scientists or curators. Twenty-four (60%) of these researchers reported collecting data. In terms of data reuse, 11 participants reused data from publications and 8 obtained data directly from colleagues. Nearly every zoologist reported reusing data from archives/museum collections and online repositories, with 38 and 36 participants, respectively, indicating reuse from these sources. These attributes indicate that the archaeologists and zoologists have potentially dealt with confidential data as a data collector, a data reuser, or both.

FINDINGS

We present our findings in four sections. We begin by discussing the types of data that archaeologists and zoologists collect, focusing on the unique value of data collected in the field and the importance of location information as a significant context element that accompanies artifacts and specimens. In the second section we explore perceptions of risk or potential harm that the interviewees associated with sharing location information. In section three, we examine views about the effectiveness of data confidentiality as a way to protect sites and species, and in the fourth section we investigate interviewee attitudes about who should be responsible for controlling access to data and deciding which information to keep confidential.

The interviewees viewed location information as highly important and valuable in facilitating use and reuse of archaeological and zoological data. Yet, researchers from both disciplines also expressed the view that location information produced by researchers could be used to facilitate looting or poaching. In addition to the loss of cultural heritage or biodiversity, potential harms for living populations were recognized. Interviewees, however, questioned the effectiveness of keeping data confidential as a way to prevent looting or poaching. Likewise, their attitudes varied concerning whether and how much data should be kept confidential. Those who stressed the importance of keeping data confidential focused on restricting access to information for the general public but not other researchers. Interviewees who discussed the importance of restricting access to certain types of information largely viewed the decision-making and gatekeeping functions as belonging to others, typically either government actors or repository managers.

Data: Artifacts, Specimens, and Location Information

Data collection for both archaeologists and zoologists is a unique experience that cannot be recreated; as such data collected in these disciplines are uniquely valuable. Location information in particular is highly valued. Archaeologists use location as a key piece of information to assist in dating objects and often establishing the interaction between cultures at specific points in time. For zoologists, location information enables replication as well as understanding the changing ecology of a site over time. Yet, location data brings with it concerns about confidentiality and ethics. Specifically, both disciplines have norms around whether to keep location information confidential in order to protect heritage sites as well as endangered and threatened species. In our interviews, 8 of 22 archaeologists (36.3%) and 11 of 40 (27.5%) zoologists explicitly discussed discipline-specific ethical norms.

Archaeologists and zoologists collect a variety of data points about the physical context in which artifacts or specimens are collected. In archaeology, this may include geolocation information, stratigraphy, information about the relationships between artifacts, and the conditions under

which data were collected. In zoology, geolocation information is paired with species counts, sensor data, and information about the time and conditions under which the data were collected. For both disciplines location information is a key piece of contextual information that aids in establishing relationships between other data surrounding the artifacts or specimens themselves or the research site in general.

For example, archaeologist17 explained that “archaeological data is broken into images, spatial data, which is usually in sort of GIS or CAD plans. Sort of textual data, which is either scanned, like as scanned as an image.”

Zoologist07 similarly described his typical dataset as consisting of:

Information about the age and the location in a particular stratigraphy where you find a specimen, the location of the spot where the specimen came from, and the taxonomic information about the assessment. It's a pretty typical way to represent fossil data, what stratigraphic layer, chrono or bio or lithostratigraphic layer it's from, what it is, and where was the deposit (Zoologist07).

Interviewees from both archaeology and zoology discussed the value of collecting comprehensive datasets of which location was an integral piece. Specifically, in each field data collection is a unique event that cannot necessarily be recreated once a site has been excavated or a specimen collected. Researchers in both archaeology and zoology discussed the unique value of data, Archaeologist01 explained: “archaeological field work is kind of priceless. Even the lousiest data is data.” Zoologist07 also discussed the importance of data, focusing on the ways that large datasets can be reused:

I've had a long history of working in a lab and I realized my interest really has kind of coalesced into working with larger and larger datasets, thinking about the ethical and moral and sort of philosophical issues regarding the importance of reusable data, and building that into a rich program that really looks at global change at multiple timescales. (Zoologist07).

For both archaeology and zoology, data collection is difficult and expensive, thus the ability to reuse data is key. The availability of location information in particular can be an important factor for researchers in determining whether they can reuse data to their fullest extent. Zoologist03 explained the process of deciding whether or not to reuse data, “when it doesn't meet my needs the most obvious reasons would be there's just not enough data or it doesn't cover . . . Like geographically it doesn't cover the area I'm interested in well enough.” The access to location data is tied to the ability of researchers to be able to use and reuse data, as a marker of quality, of breadth and/or depth, and also as an important piece of information that researchers need on order to make sense of artifacts and specimens.

In both disciplines, geolocation data are considered crucial to understanding artifacts and specimens that are collected in the field. However, interviewees from both groups reported that location data are viewed as sensitive and that in some cases there are reasons to restrict access to this type of data. Researchers weigh concerns about confidentiality and ethics with regard to sharing specific types of data such as location information against the value of that data and the potential risk or harm that could occur.

Data Confidentiality: Sharing or Restricting Access to Location Information

Researchers are increasingly making their data publicly available. This trend towards more widespread access is influenced by a number of factors, such as funding agency mandates, journal or publication requirements, and disciplinary norms in research communities. For researchers in archaeology and zoology, the decision also involves weighing benefits against risks and/or potential harms. In contrast with quantitative social science, where data confidentiality risks have immediate impacts on specific living individuals represented in the data, for archaeology and zoology the potential harms are often at a broad level with impacts spread across large social groups, such as indigenous communities whose ancestors may be difficult to trace, or populations of endangered species. Concerns for these groups include loss of artifacts of importance to cultures (often endangered cultures), or loss of species and negative impacts on biodiversity. It is less common for the harms associated with archaeological or zoological data to directly impact living individuals, although not impossible. In our 62 interviews and observations, issues around confidentiality came up in 22 (35.5%).

Archaeologists were concerned about making the locations of historical sites available because of the potential for looting. The harms associated with looting are at a cultural and societal level, for example items can have significant cultural or religious value for particular indigenous groups. Archaeologist13 explained that archaeological data itself can be problematic because of considerations around preventing looters, “There’s a big problem with archaeological data, about not wanting people looting things.” Archaeologist17 also commented that preventing looting of a heritage site could be a valid reason to restrict access to information, “I guess there always could be cases where it’s appropriate to restrict access if it’s a site that is unknown or that has been subject to looting.”

Interviewees explained that in some cases sharing location information for archaeological sites is viewed as encouraging looting of those sites. Archaeologist06 explained that the process for obtaining data involved signing a form agreeing not to share location information, “they give me a form to fill out that says I’m not going to do anything bad with this like give away the site information thus encouraging looting . . . Given the amount of looting that we have to deal with, it seems to me a really reasonable

and important precaution. Site locations are an issue.” Archaeologist19 explained that “by providing any close locational data, it’s considered to be putting the site at risk.”

In addition to looting, archaeologists also expressed concerns about sharing location data for sites that are connected to living populations. Archaeologist04 described location information for sites relating to living populations as being particularly problematic with regard to depositing data into repositories, “I think actually that’s one of the reasons that people obviously are not using [repositories] . . . in the U.S. where we have these [extended] populations where we still have the living ancestors around that are highly involved with the protection of the remains, there are problems with that.” Archaeologist02 expressed a similar concern, “We have very precise locational data that we don’t want the general public to access because we’re afraid of the potential looters and hunters coming in and trying to locate some of the sites or sensitive data related to burial site that Aboriginal groups don’t want people to know the location of.” In both cases, these researchers were making decisions about keeping location information confidential. They weighed factors including protecting heritage sites from looters along and respecting the direct wishes of the communities of living ancestors who have a stake in the protection of those sites.

Zoologists discussed concerns about sharing location information because it might lead to poaching. Zoologist06 explained that releasing location information for an endangered species can be risky because it will provide a clear map for anyone wishing to track down those populations:

Yes, so there’s been some concern if you release a locality information that people will go and collect an entire hillside of the species then it can be sold either... Usually legally, actually, because the specimen hasn’t been put in the IUCN [International Union for Conservation of Nature] Red List or anything like that because it’s unknown. Like, you can go to jail for bringing over an iguana or a boa constrictor, which no one in science cares about. But because it is on the Red List, you can go and get in trouble. Whereas some species are really endangered that’s a new species that’s described that hasn’t been the Red List, you really can’t get in trouble for. So, we debated on, myself and a couple of people that were writing the paper, debated on if we should disclose that exact locality. (Zoologist06).

Zoologist05 also expressed apprehension about making location information available to the public: “I think there are certainly cases where one would be somewhat cautious with making all of that information public, for the conservation of the species.” Similarly, Zoologist18 explained that poaching would be a reason for keeping this information confidential, particularly when the data in question pertain to threatened or endangered species:

If, for example, you discovered a new species in a place that looked like it was threatened or maybe you collected a

very rare species. And this would be particularly true of some other groups of animals, some terrestrial turtles, for example. You wouldn't want to publish the locality data because somebody might go collect them all or kill them all or whatever. So for rare and endangered species, there is a good reason to withhold the exact locality data (Zoologist18).

Interviewees from both archaeology and zoology expressed concern about sharing location information, specifically with regard to providing information that could result in looting heritage sites or poaching endangered species. Other reasons for keeping location information confidential include following the wishes of living ancestor communities or protecting newly discovered species that may not have had time to go through the process to be placed on the International Union for Conservation of Nature (ICUN) Red List which contains endangered species. Researchers also discussed potential problems with making this information available to the general public, meaning outside of the community of scientists/researchers who would use this data. Given these reasons for potentially restricting access to data, the next section will explore views about whether data, such as location information should be kept confidential.

Views About the Effectiveness of Restrictions on Sensitive Data

Archaeologists' and zoologists' attitudes regarding the question of whether some information should be kept confidential varied widely. In some cases archaeologists argued for keeping location data confidential from the general public in order to protect heritage sites, but still make that information available to researchers. In discussing location information about Aboriginal burial sites, Archaeologist02 explained that, "we have to figure out how do we provide access to the locational data to researchers" while limiting use by others. Archaeologist03 made the point that even if an individual researcher wanted to share location data about a particular site, sometimes that is not possible because of legal restrictions imposed by other agencies or governments who may wish to protect those sites, "Well, personally because I am very interested in community archaeology, I don't think there's anything wrong with making all of that public. But that's just not going to apply there."

Some zoologists also argued for keeping location information confidential from the public. Zoologist05 explained that even when giving specimens to a museum there are cases when it would be appropriate to restrict access to the location data that would accompany those specimens:

In terms of new species that were clearly endangered, or in trouble or those sorts of things, I would give the museum, probably I'd certainly give them the specimens, the actual specific locality information, but I would request that based on the paper I'm publishing on it, that I'm not giving the

lat/long or the details of where they came from just so people don't go and raid them (Zoologist05).

Zoologist11 expressed a similar concern about sharing location information with the general public but explained that even if information is withheld from the public it should still be shared among scientists:

No, not from other scientists. I mean, generally it's sometimes appropriate in my field to keep data from being sort of publicly queried, if it's involving rare, endangered or sensitive species. If you have data on all the nests of spotted owls of the Pacific Northwest, yeah, that data should not be queryable by the public. But that doesn't mean that if another scientist emails you and asks you for that data, that you should not give it to them (Zoologist11).

This sentiment, that information should be withheld from the public but made available to other researchers, was expressed by both archaeologists and zoologists. Archaeologist02 described the process of managing a database and trying to limit access to locational data:

The solution they've figured out back in '95 when they originally developed the database was that we just won't provide locational data online, and that's become a real problem for us because the database is quite old . . . So we're trying to think through a way to provide general locational data but without providing the level that someone would find the site and dig it up (Archaeologist02).

Archaeologist06 expressed some uncertainty about whether withholding location information will prevent looting, but still fell on the side of keeping that information confidential, explaining that while looters will find a way regardless of whether information is easily obtainable or not, it is still a good idea to take steps to make it more difficult for them:

Looting is such a problem. And there are people out there that will use these things just to find places where they can loot. If they're... They'll do it anyway. People who are seriously into it will find some way to get access to that information but I kind of think that having some road blocks is a good idea (Archaeologist06).

Others in the archaeological and zoological communities argued for making location information widely available. Archaeologist17 argued that the availability of location information is not necessarily related to looting and that focusing on the problem of restricting access to information instills a false sense of addressing the actual problem of looting, "I kind of feel like there's a problem of looting, and rather than addressing the problem of looting you're obscuring the archaeological data, and I'm not sure that they're necessarily related. I mean the looters aren't looting just because there's archaeological data available; they are looting for completely different reasons."

Not every participant felt that withholding certain elements of research data would protect sensitive sites. Archaeologist09 took the position that sharing location information actually protected sites, “we really want people to know where the sites are because we believe that that protects them.” Zoologist03 similarly pointed out that knowing location information for particular species can be helpful in making policy decisions about managing locations with rich biodiversity or protecting endangered species. Conversations such as this necessitate sharing sensitive data beyond the research community.

And you can find those places, which is then both important for understanding the process of evolution, and also for making decisions for conservation. They're places that have got a lot that's unique and if we want to protect it, we better protect it in that particular location, because it's kind of not found elsewhere. (Zoologist03).

The view that a determined looter or poacher would not be stopped by restricting information to location data was shared by zoologists as well as archaeologists. Zoologist12 explained that “My feeling is that there’s no way to stop a determined person from finding the endangered species, whether the lat/longs are available or not. And so, I tend to just say, I just tend to publish it.”

Although researchers generally agree that location information is sensitive and that there are risks involved in making that information publicly available, views differed about whether that information should be kept confidential. Some argued that the way to protect sites or species was to restrict access to location information, while others thought that making information widely available would protect those sites. Others argued that sites were at risk regardless of whether information was shared or not and so decided to make that information available.

The question of whether data should be kept confidential then leads to another dilemma – who should make decisions about whether and how to manage access to this sensitive information?

Data Management Responsibility

For both archaeologists and zoologists, decisions about data availability are sometimes dictated by legal data collection agreements with government agencies. In these cases, researchers must disclose the location information for the artifacts or specimens that they collect, and decisions about whether that information is publicly available or not are managed by the government agency rather than the researcher.

For archaeologists, governments that grant permission to collect artifacts control the access to that information. Archaeologist03 explained, “we cannot take anything public without the Ministry's okay.” And Archaeologist06 was not entirely certain about the administrative details but did explain that data collection information, such as location had to be sent to the appropriate agencies, “I'm not

entirely sure of what the actual administration thing is, but basically when anybody does cultural resources work within the states, generally one of the requirements is that they send the basic data over to the respective state office, and it winds up being put in these places.”

Zoologist10 explained that data collection permits often require researchers to deposit specimens and/or data with repositories or museums in order to make that data available to others, including location information. “So just basically the locality information it’s collected at, when it was collected, who collected it, the landowner of the property that was collected on, you know, I share all that information. I also need to share collecting permits, you have to make copies of those, to deposit and to show that it was actually legally collected.”

Researchers in archaeology expressed a particular concern. If they made location information available and the site was then looted, they feared they might be held responsible. Archaeologist07 described a situation in which she found herself reconsidering her publication plans in light of the country’s anti-looting laws and the potential damage to the site. She questioned whether it was her responsibility to help protect a site from looters if her host country already had laws in place to safeguard archaeological sites, and whether she would be held legally responsible if the site was in some way damaged.

I hadn't occurred to me that by publishing those burials I was telling people where the site was and inviting people to come loot but that hasn't happened . . . And of course yes, there are laws in the country about not looting, and then the question is, is it my responsibility to cover up what I'm doing? (Archaeologist07).

In many cases interviewees expressed the opinion that while someone should make deliberate decisions about whether to keep data confidential, they did not want that responsibility themselves. When asked about sharing sensitive data, Zoologist19 said, “Essentially, it’s not my responsibility to police who’s doing [what] with it.” And Zoologist25 said, “I’m not really in a position to make that decision.”

Archaeologist06 also expressed a desire for someone else to handle decisions about access to sensitive information:

I'm going to pass the buck, so repositories need to figure it out...I wouldn't want to be responsible for making these choices. I mean, there are issues with information that indigenous communities are comfortable sharing, and then what archaeologists' expectations for what should be shared, and that would have to be negotiated I think in particular cases. It does, it gets very complicated. There are a lot of issues involved (Archaeologist06).

Zoologist13 described that in the process of depositing data with a repository, he included all of the information required for submission, implying that once the data have been deposited it is the responsibility of the repository to

manage access to those data, “when I upload things, you have to give all of the information that’s available.”

In the cases described above, archeologists and zoologists generally reported that decisions about whether to keep information confidential should be made by others. Specifically, that government agencies often have mandates around whether and/or how researchers can share data, and also that once data has been deposited in a repository, as is increasingly required by many funding agencies, then it is the responsibility of the repository to make decisions about who should have access to data and how that access is managed.

DISCUSSION

We reported on the results of 62 interviews, 22 with archaeologists and 40 with zoologists, to examine attitudes about data confidentiality and disclosure. Our findings indicate that archaeologists see the destruction of archaeological data through looting and the impact on living ancestors as well as ancient cultures without traceable descendants as the primary harms associated with making location information about archaeological sites widely available to the public. Zoologists view poaching endangered and threatened species as the primary harms associated with making location information public. For both of these disciplines, the primary harm is at a broad, group or community level where the risk is distributed, in contrast to quantitative social science data where the risk relates to immediate impacts on specific living individuals who are represented in the data. For quantitative social science, there is a greater consensus around keeping certain data confidential in order to protect research subjects. The absence of direct impact for specific living persons in archaeological and zoological research complicates decisions about whether to keep this location information confidential or make it available. Sometimes these groups may not have an identifiable spokesperson or may have competing leaders. The person or people with legal standing to speak for a cultural or ethnic group may be debatable and the current government may have a mixed record of maintaining the heritage from the culture in question. While disclosure of personal information in the health and social sciences can directly cause social and economic harm, it is more difficult to assign causality to the more diffuse societal harms resulting from exposing archaeological and zoological data.

Given the increasing attention that is devoted to data sharing and data reuse (e.g. Dallmeier-Tiessen et al., 2014; Zimmerman, 2008), data confidentiality issues present a particular challenge. Researchers are increasingly dependent on funding from sources that require them to share their data (e.g. National Science Foundation, 2009), or on publishing in journals that require data to be deposited for publication (e.g. Chavan & Penev, 2011). Simultaneously, researchers are still driven by disciplinary norms that place a high value on data confidentiality as an ethical best practice (e.g. Society for American

Archaeology, 1996). This tension between new data practices and traditional disciplinary ethics and norms can be seen in the disagreements among both archaeologists and zoologists about data disclosure: whether, when, and to whom. While interviewees overwhelmingly agreed about the potential risks of looting and poaching, they did not agree about how to go about preventing these harms. They generally agreed that some other organization, such as a repository, should be responsible for making decisions about which data to keep confidential.

Implications for Data Curation and Repository Management

Many of the archaeologists and zoologists interviewed suggested that data repositories should be responsible for ensuring the confidentiality of, and managing access to, sensitive data. Along with this recommendation is the implication that repository staff rather than the researchers should make decisions about whether and/or which data to keep confidential. Repositories generally exist at the end of the chain of custody for archaeological and zoological research data. Decisions about confidentiality may be dictated by any number of groups using diverse means (e.g. government regulations, data collection permits, funding agency mandates), or by the researchers themselves before data is deposited in a repository. However, as the custodians who are responsible for providing access to the data, it falls to repositories to figure out how to implement the requirements around restricting access to data.

This raises several issues for repositories. First, repositories must adhere to multiple and sometimes conflicting laws, rules, and ethical frameworks. Second, for online repositories in particular, sensitive data may be able to be inferred from other data points as is the case in the social sciences. Third, for museums that often face budgetary constraints the identification of confidential data in past accessions may be difficult. Finally, developing systems, protocols and regimes that enable monitored reuse can be difficult and costly

Given the complicated landscape of data collection for archaeology and zoology (e.g. overlapping requirements from funding agencies providing financial support, government offices granting permits for data collection, scholarly journals requiring data deposit for publication, professional societies outlining ethical considerations to protect vulnerable sites, groups, or species), the decisions that repositories must make about data confidentiality are difficult. Our literature review identified codes of ethics for researchers in archaeology and zoology, but no guidelines for repositories that collect, preserve, and provide access to these types of data. Given the data management landscape described by our participants, repositories should develop more stringent policies around confidential data and protected locations. In this way, repositories can enforce the best practices established by individual disciplines and protect confidential or otherwise sensitive data from being made widely available.

In many cases, repositories may also be limited by technological capabilities that conflict with their goals or obligations. For example, systems may not allow the repository to restrict access to parts of the data that need to remain confidential. Archaeologist06 explained that his repository was trying to find a way to provide secure access to location information, but had not yet managed to accomplish it. Rather, users were required to email repository staff and ask for special permission to get location data, “So we’re trying to think through a way to provide general locational data but without providing the level that someone would find the site and dig it up.”

Online repositories face particular problems in maintaining data confidentiality. For example, a zoologist may find an endangered species during the same data collection event in which he or she collects other species. Masking the location of the endangered species may not be sufficient if the entire data location event can be reconstructed by the name of the researcher, date of data collection, and location of the more common species collected. This type of data triangulation that is characteristic of the social sciences also applies here. In the social sciences, the solutions proposed have included data enclaves (Lyle 2014) and metadata masking extreme values and precise distributions (Lagoze 2014). These tactics might also work for archaeology and zoology.

While this paper discussed current data collection events, archaeological repositories and museums curate artifacts and specimens collected in the past. Unlike recent accessions, which have warnings about confidentiality, previous accessions may not. Species collected in the past may have been collected when they were not endangered, therefore repositories need to establish security mechanisms that take the bi-directionality between not sensitive and sensitive into account and allow the status of particular types of data to change over time.

Taken together these findings complicate the role of the repository as gatekeeper, asking repositories to consider not only whether they *can* provide access to data, but also whether they *should*. These findings suggest that the disciplines providing data to repositories believe that in some cases the answer to the latter question is no.

CONCLUSION

This study found that both archaeologists and zoologists understand the potential harms or risks associated with sharing location information. Some favor keeping the data confidential as the potential for looting and poaching remains a real and present threat to cultures and living ancestral groups. Researchers would generally prefer to restrict access to their data from the general public but maintain open data for colleagues. Given these potential harms, researchers in archaeology and zoology view decisions about who should have access to data and how those decisions are made as complicated and would prefer that some other organization, such as a repository, take that responsibility.

ACKNOWLEDGMENTS

The DIPIR Project 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.” We also thank other members of the DIPIR team who participated in data collection and/or analysis efforts, including Jessica Schaengold, Julianna Barrera-Gomez, Morgan Daniels, and Ayoung Yoon.

REFERENCES

- About Open Context: Publishing. (2015). Retrieved April 29, 2015, from <http://opencontext.org/about/publishing>
- Arzberger, P. W., Schroeder, P., Beaulieu, A., Bowker, G. C., Casey, K., Laaksonen, L., ... Wouters, P. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*, 3(29), 135–152.
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. <http://doi.org/10.1002/asi.22634>
- Chapman, A.D., & Grafton, O. (2008). Guide to Best Practices for Generalising Primary Species-Occurrence Data (p. 27). Copenhagen: Global Biodiversity Information Facility. Retrieved from http://www.gbif.org/orc/?doc_id=1233
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12 (Suppl 15), S2. <http://doi.org/10.1186/1471-2105-12-S15-S2>
- Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/viewArticle/1024>
- Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., ... Wilson, M. (2014). Enabling Sharing and Reuse of Scientific Data. *New Review of Information Networking*, 19(1), 16–43. <http://doi.org/10.1080/13614576.2014.883936>
- El Emam, K., Arbuckle, L., Koru, G., Eze, B., Gaudette, L., Neri, E., ... Gluck, J. (2012). De-identification Methods for Open Health Data: The Case of the Heritage Health Prize Claims Dataset. *Journal of Medical Internet Research*, 14(1), e33. <http://doi.org/10.2196/jmir.2001>
- Fong, J. J., & Qiao, G. (2010). New localities of endangered Chinese turtles from museum specimens and the practical and ethical challenges using and reporting natural history collection data. *Zootaxa*, 2393, 59–68.
- Hartter, J., Ryan, S. J., MacKenzie, C. A., Parker, J. N., & Strasser, C. A. (2013). Spatially Explicit Data: Stewardship and Ethical Challenges in Science. *PLoS*

- Biol, 11(9), e1001634.
<http://doi.org/10.1371/journal.pbio.1001634>
- Hodder, I. (2011). Is a Shared Past Possible? The Ethics and Practice of Archaeology in the Twenty-First Century. In K. Okamura & A. Matsuda (Eds.), *New Perspectives in Global Public Archaeology* (pp. 19–28). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-0341-8_2
- Inter-university Consortium for Political and Social Research (ICPSR). (2015). About ICPSR. Retrieved April 19, 2015, from <http://www.icpsr.umich.edu/icpsrweb/content/membership/about.html>
- Jahnke, L. M., & Asher, A. (2013). Dilemmas of Digital Stewardship: Research Ethics and the Problems of Data Sharing. *Research Data Management*, 80.
- Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work (CSCW)*, 15(4), 321–358. <http://doi.org/10.1007/s10606-006-9023-2>
- Kersel, M. M., & Chesson, M. S. (2013). Looting Matters. Early Bronze Age Cemeteries of Jordan's southeast Dead Sea Plain in the Past and Present. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1059&context=morag_kersel
- Krieger, W. H. (2014). Marketing Archaeology. *Ethical Theory and Moral Practice*, 17(5), 923–939. <http://doi.org/10.1007/s10677-014-9497-9>
- Kuula, A. (2010). Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly*, 34(3), 12–17.
- Lagoze, C., Block, W. C., Williams, J., Abowd, J., & Vilhuber, L. (2013). Data Management of Confidential Data. *International Journal of Digital Curation*, 8(1), 265–278. <http://doi.org/10.2218/ijdc.v8i1.259>
- Lyle, J. (2014). OpenICPSR: OpenICPSR. *Bulletin of the American Society for Information Science and Technology*, 40(5), 55–56. <http://doi.org/10.1002/bult.2014.1720400514>
- Lynott, M. J. (1997). Ethical principles and archaeological practice: development of an ethics policy. *American Antiquity*, 589–599.
- Oderkirk, J., Ronchi, E., & Klazinga, N. (2013). International comparisons of health system performance among OECD countries: Opportunities and data privacy protection challenges. *Health Policy*, 112(1), 9–18. <http://doi.org/10.1016/j.healthpol.2013.06.006>
- Parry, O., & Mauthner, N. S. (2004). Whose Data are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology*, 38(1), 139–152. <http://doi.org/10.1177/0038038504039366>
- Protection of Human Subjects (2009). Retrieved from <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html>
- Society for American Archaeology. (1996). Principles of Archaeological Ethics. Retrieved from <http://www.saa.org/AbouttheSociety/PrinciplesofArchaeologicalEthics/tabid/203/Default.aspx>
- Sweeney, L. (1997). Guaranteeing anonymity when sharing medical data, the Datafly system. *Proceedings, Journal of the American Medical Informatics Association. (AMIA)*. Retrieved from <http://dataprivacylab.org/datafly/paper4.pdf>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, 8(7), e67332. <http://doi.org/10.1371/journal.pone.0067332>
- Wylie, A. (1996). Ethical dilemmas in archaeological practice: Looting, repatriation, stewardship, and the (trans) formation of disciplinary identity. *Perspectives on Science: Historical Philosophical, Social*, 4(2), 154–194.
- Zimmerman, A. S. (2008). New Knowledge from Old Data. *Science, Technology & Human Values*, 33(5), 631–652. <http://doi.org/10.1177/0162243907306704>