YEARBOOK OF BIOLOGICAL ANTHROPOLOGY ARTICLE

WILEY

Data sharing in biological anthropology

Connie J. Mulligan[1,2] | Doug M. Boyer[3] | Trudy R. Turner[4,5] | Eric Delson[6,7] |
William R. Leonard[8]

[1]Department of Anthropology, University of Florida, Gainesville, Florida, USA

[2]University of Florida Genetics Institute, University of Florida, Gainesville, Florida, USA

[3]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA

[4]Department of Anthropology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA

[5]Department of Genetics, University of the Free State, Bloemfontein, South Africa

[6]Department of Anthropology, Lehman College, City University of New York, Bronx, New York, USA

[7]New York Consortium in Evolutionary Primatology, New York, New York, USA

[8]Department of Anthropology, Northwestern University, Evanston, Illinois, USA

Correspondence
Connie J. Mulligan, Department of Anthropology, University of Florida, Gainesville, Florida, USA.
Email: cmulligan@ufl.edu

Abstract

Open data sharing democratizes science by making data more equitably available throughout the world. Furthermore, open data sharing improves the reproducibility and quality of research and enables new collaborations powered by the freely available data. Open data are defined as data that can be freely used, reused, and redistributed by anyone. For an interdisciplinary field like biological anthropology, data sharing is critical since one person cannot easily collect data across the domains relevant to our field. The goal of this paper is to encourage broader data sharing in our discipline by exploring the state of data sharing in the field of biological anthropology. Our paper is divided into four parts: the first section describes the benefits, challenges, and emerging solutions to open data sharing; the second section presents the results of our data archiving and sharing survey that was completed by over 700 researchers; the third section presents personal experiences of data sharing by the authors; and the fourth section discusses the strengths of different types of data repositories and provides a list of recommended data repositories.

KEYWORDS
data archiving, data sharing, repositories

1 | INTRODUCTION

We believe that more open data sharing is important for the advancement of the field of biological anthropology. Data in our field are collected from individuals and physical objects that are often temporally or geographically limited and may not be able to be collected again. Originally, the ideas of data sharing were focused on the digital characteristics of data and were formalized in the FAIR principles that specify shared data should be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (Wilkinson et al., 2016). As concerns about confidentiality, participant anonymity, data sovereignty, power differentials between researchers and participant communities, and historical context emerged, the discussion expanded from the digital characteristics of the data to also consider the people involved in the research and the purpose of the proposed research. These concerns led to the complementary CARE principles of data governance that recognize the importance of **C**ollective benefit, **A**uthority to control, **R**esponsibility, and **E**thics (GIDA, 2019). Simply put, data sharing should be "as open as possible and as closed as necessary" (H2020Programme, 2020), although researchers may disagree on how that philosophy should be interpreted and implemented. Thus, while we aspire for data to be openly shared when possible and appropriate, we recognize that certain restrictions are often necessary and valid. Acknowledging the validity of caveats expressed by CARE, we believe data sharing in biological anthropology can become much more open than it currently is.

In this paper, we focus on data and metadata. We recognize that sharing of samples and materials, as well as software and technology, are also important topics, but we consider them beyond the scope of the paper. We define "data" to include facts or pieces of information,

 wileyonlinelibrary.com/journal/ajpa

including raw data and summary statistics, as well as digital manifestations of information, such as scanned images, and related databases, that is, any information about a sample or specimen. We define "metadata" to be additional information necessary to understand, analyze, interpret, and replicate findings, including documentary, behavioral, demographic, and historical data as well as experimental protocols, custom software programs, scripts, and annotated code for statistical analyses, that is, information about the data. We differentiate between "archiving"—depositing digital data in an online database or repository that can persist without maintenance by the researcher who deposited the data—and "sharing"—providing some level of access (e.g., open access, restricted access) to the data that have been archived. *We hope the information in this paper will help to further foster a culture of data sharing in biological anthropology in which the worthy aims of open science are fully realized and a future is built in which "there will be so much data available that ownership will no longer be an issue"* (Hipsley & Sherratt, 2019).

The authors of this paper are members of the American Association for Biological Anthropologists (AABA)[1] ad hoc committee on Data Archiving and Sharing. In 2019, we organized a workshop on data sharing in biological anthropology that was funded by the National Science Foundation (NSF). A diverse group of 40 participants representing a range of sub-disciplines attended the workshop. Two days of discussions contributed greatly to a shared understanding of the benefits, challenges, and concerns about open data sharing. The discussions also supported the main goal of the workshop, which was to formalize a set of guiding principles and best practices to increase and normalize data sharing in biological anthropology.

The results of the workshop were published in the *American Journal of Physical Anthropology* (AJPA) and briefly outlined guiding principles and best practices for data sharing in biological anthropology (Turner & Mulligan, 2019). The *AJPA* article invited members of the biological anthropology community to submit comments through Letters to the Editor (LTE) of *AJPA*; four LTEs were published that supported the importance of more open and standardized data sharing in biological anthropology, but highlighted challenges that needed to be addressed, such as preserving participant anonymity, the importance of engaging all relevant stakeholders, and the need to provide more detail on how to openly share data (Elton, 2020; Leigh, 2020; McDade, 2020; Wagner, 2020). A response to the letters to the editor was published with input from participants of the 2019 workshop and briefly addressed the challenges to open data sharing and proposed some solutions (Boyer, Jahnke, et al., 2020).

The current manuscript is influenced by the intellectual contributions of the workshop participants and LTE writers. Here we provide a more thorough examination of the benefits, challenges, and emerging solutions to more open data sharing in the field of biological anthropology. Specifically, we provide a detailed discussion of the benefits to open data sharing, such as a more collaborative, equitable, dynamic science, as well as the challenges, such as issues of participant confidentiality and anonymity, the need for stronger collaborations with participant communities, and other stake holders, and the importance of creating a culture of data sharing (Section 2). We also present results from a survey on data archiving and sharing that was sent to researchers from different biological fields to query their attitudes and practices with data archiving and sharing (Section 3). Finally, we present several personal stories of experiences with data sharing (Section 4) and a guide and recommendations to choosing a data repository (Section 5).

2 | BENEFITS, CHALLENGES, AND SOLUTIONS TO MORE OPEN DATA SHARING IN BIOLOGICAL ANTHROPOLOGY

The goal of data sharing is for all scholars to archive (preserve) and share (make available) their data in such a way that anyone can find and reuse those data without knowing a priori that the data exist and without the need to inform or ask permission from the researcher who generated the data. The movement toward open sharing of data has been building for decades, but often only in specific disciplines (Birney et al., 2009; Delson et al., 2007; Gewin, 2016; Parr & Cummings, 2005; Pienta et al., 2010; Rausher et al., 2010).

The underlying philosophy of open data sharing is that data and samples collected by scholars are part of our global human heritage and should be shared with other scholars and with the public (Mons, 2020; O'Doherty et al., 2021; Phillips et al., 2020; Turner & Mulligan, 2019; Vitelli & Colwell-Chanthaphonh, 2006).

Data sharing has the short-term goal to promote the replicability, reproducibility, and correction of research and the long-term goal to transform existing scientific fields and create new fields of research. M*ore open data sharing requires a shift from a competitive view of science that only recognizes individuals for the scholarly impact of the papers they author to a more collaborative one that also recognizes individuals for the impact of the data they make available.* This culture shift will incentivize researchers to share their data more openly and will help align researcher career interests with the greater good of science and society (DigitalScience et al., 2020).

Increases in data sharing have been driven by the digitization of information in a form that is easily shared (Hipsley & Sherratt, 2019). The largest and most transformative data sharing initiative was the Human Genome Project, begun in 1990 and completed in 2003. During the Human Genome Project, raw DNA sequence data were uploaded within 24 h of their generation to GenBank, a repository where anyone in the world could easily and freely access the data (NCBI, 2021; NHGRI, 2020). Most recently, the UK Biobank released 200,000 genome sequences of participants in the long-running British health study, along with anonymized medical information (Kaiser, 2021). This open sharing of genetic data has resulted in hundreds of millions of publicly available DNA sequences, transformed the generation and analysis of genetic data, and driven the creation of new fields like personalized medicine, conservation genetics, paleogenomics as well as the ongoing response to the COVID-19 pandemic.

As of today, there seems to be a certain inevitability to more open data sharing as scholarly journals, funding agencies, and professional

organizations move towards greater expectations of data sharing. Journals began encouraging the sharing of data presented in published articles over a decade ago. Historically, very few social science data sets have been archived and disseminated through public repositories since informal data sharing has traditionally been much more common (Pienta et al., 2010). However, that situation is changing as more journals implement data policies with a range of data sharing expectations. Data sharing policies in journals vary by field with anthropology journals falling in the middle of the social sciences along with political science and psychology (58% of anthropology journals have a data policy versus 74% in economics, 60% in political science and psychology, 46% in sociology, and 18% in history) (Crosas et al., 2018). Anthropology journals are less likely to have a strict data policy that requires data sharing than political science or psychology (10% in anthropology vs. 30% in political science and 22% in psychology). Interestingly, there is an association between a journal's prominence (measured as ranking or impact factor) and the existence of a data policy (Crosas et al., 2018). In 2019, the *American Journal of Biological Anthropology*, the flagship journal for biological anthropology, began requiring a Data Availability Statement for all publications and currently "expects data sharing" though a move to "requires data sharing" may be in the future.

One of the biggest challenges to the paradigm shift needed to move from a competitive view of science to a more collaborative culture is to demonstrate to individual researchers the personal and societal benefits of such a shift. Some of the benefits to individual researchers and to the research community are increased publications, better citation impact, and an accelerated pace of research, as shown in several studies. Researchers who reviewed over 50,000 journal articles reported that articles that included a data availability statement and a link to a data repository were associated with an increase in citation impact up to 25% (Colavizza et al., 2020). Based on a survey of over 7000 social science projects funded by the NSF and the National Institutes of Health (NIH) over the past 40 years, researchers evaluated the effect of no data sharing, informal data sharing, and archived data sharing (i.e. data sharing through an archive); they found in comparison to no data sharing, that informal and archived data sharing resulted in an increase of 2.31 and 2.42, respectively, more publications by members of the original research team, even after controlling for demographics, faculty status, scientific discipline, number of federal grants awarded, and institutional home of the principal investigator (Pienta et al., 2010). Using a number of secondary publications (in which no member of the original research team was a coauthor) as a measure of data reuse and resultant downstream research productivity, Pienta et al. (2010) also found that informal and archived data sharing resulted in 4.78 and 6.81 more secondary publications, respectively, relative to no data sharing. Furthermore, in fields with more established data sharing cultures, like computational fields, data sharing has accelerated the pace of research as measured by more collaborations, larger research networks, increased number of publications, the reduced time between original and follow-up research, an influx of enthusiastic young researchers, and 'previously unheard-of vertical and horizontal "collaboration mobility"' (Plis & Calhoun, 2021).

Despite the demonstrated benefits of open data sharing, there are significant challenges to expanding data sharing in biological anthropology. Some of these challenges include issues of indigenous data sovereignty and including participant communities in decision-making; repatriation of samples as well as restricted access to physical collections and image data by museums; confidentiality and anonymity when dealing with potentially identifying data; and the lack of a strong culture of data sharing in biological anthropology. The need to meaningfully engage with marginalized, indigenous, Aboriginal, and BIPOC communities is critical in order to address their concerns and incorporate their goals, but also to ensure they are not left behind in the move to more open data sharing (Mc Cartney et al., 2022). Protecting vulnerable populations, particularly those that have been abused by the research community in the past, is of paramount importance (Lipphardt et al., 2021; Normile, 2021; Thomas & Quinn, 1991). Although some scholars view the data they have collected as theirs to use or share, many participant indigenous communities believe that data are part of their group's identity, property, and heritage (Nicholas, 2014). Some indigenous communities have endorsed the CARE Principles for Indigenous Data Governance (see above) that reinforce their right to engage in the decision-making process in accord with indigenous values and collective interests (GIDA, 2019). Recent publications that provide guidelines for ethical ancient DNA research expand the stakeholder communities by stressing the need to engage descendent communities and highlight data sharing as a particular issue of concern (Tsosie et al., 2021; Wagner et al., 2020). Ultimately, greater discussion and collaboration between scholars and indigenous communities support the CARE principles and will lead to richer and more holistic research (e.g. Austin et al., 2021).

Another key issue is that many valuable specimens and images are held by museums. Historically and currently, access to museum specimens for research and imaging has been easier for researchers affiliated with the museum, as well as members of well-known labs (Davies et al., 2017; Hipsley & Sherratt, 2019). This is somewhat unavoidable given the museum mission as stewards of their collections, but more open sharing of digital versions of specimens provides an opportunity for more equitable access to museum holdings for research and education. However, museums still have restrictions on third-party data sharing, such as the need to cite the museum and the specimen identifier. Even these light restrictions can create problems since most museums are not equipped to make third-party digital holdings discoverable or facilitate data access on request. The 3D data repository MorphoSource.org has helped alleviate these blockages by allowing researchers to upload data and museums to manage access to those data. MorphoSource has thus enabled massive amounts of museum data to be mobilized by providing a FAIR-principled platform that effectively models the rights and roles of key stakeholders.

A more difficult challenge faced by museums and anthropology research communities is the call to repatriate collections of physical samples to descendant communities. Many museums have already begun, or are in the process of repatriating such collections. For example, the Morton collection at the University of Pennsylvania Museum of Archaeology and Anthropology, which contains almost 1000 crania

collected by Dr. Samuel G. Morton during the first half of the 19th century, represents an excellent example of the challenges faced with museums. Morton (1849) used these crania to support racist notions of biological variation by looking for cranial differences between the races and ranking each race's intellectual abilities (see Geller, 2020 for a historical account). In the early 2000s, computerized tomography (CT) scans of these skulls and thousands of others at the museum were made available online allowing more than 100 researchers access and leading to more than 70 publications. However, in the summer of 2020, the long history of racial injustice in the United States boiled over into protests after the killing of George Perry Floyd and the museum put the entire physical and digital collection in storage and halted research. The museum is now in the process of hiring an anthropologist of color to direct repatriation efforts, and the hope is that improved engagement with descendant communities will expand ideas of what research is and lead to better, more holistic research. "People think about repatriation as something that's going to empty out museum shelves, but in reality, it fills the museum back up with these relationships and connections" says Dorothy Lippert, an archaeologist and tribal liaison at the National Museum of Natural History and a citizen of the Choctaw Nation (Wade, 2021).

There are also concerns that data could lead to the identification of study participants, rendering promises of confidentiality and anonymity meaningless. In theory, having multiple and different data elements could allow for the identification of a study participant or a non-participating individual although it is unclear how often such a situation occurs. Genetic data are often cited as an example of potentially identifiable data. The advent of low cost, high-throughput DNA sequencing platforms and broad sharing of genetic data through data repositories and public ancestry websites have made genetic data accessible on a scale unimaginable a couple of decades ago (Bonomi et al., 2020). For example, in 2005, a 15-year-old boy was able to identify his anonymous sperm donor father by sequencing his Y chromosome through an online genealogy DNA-testing service and using two websites that enabled him to identify two close DNA matches with a shared surname and another website that identified his father as a man with the same surname and the birthdate and place of the sperm donor (BBCNews, 2005). Since then, there have been many stories of people identifying relatives, and police identifying perpetrators of crimes, through DNA testing of samples and evidence in combination with ancestry websites and national DNA databases. However, these stories represent a fraction of the hundreds of millions of human DNA sequences available in online repositories. Genetic identification of a specific person would typically require genome-scale data or hundreds of forensically-informative genetic markers as well as similar data from that individual or a close relative, which is not true for the majority of research participants. Nevertheless, researchers should make clear to study participants who are donating DNA samples that their anonymity cannot be guaranteed.

Finally, there is not yet a strong culture of data sharing in biological anthropology and this must be acknowledged and changed in order to realize the full benefits of open data sharing. The most popular means of sharing data, for example, *"data available upon request" statements and providing data in supplemental material or on personal websites, are not considered data sharing according to the FAIR principles* for the following reasons: the data are not Findable without knowing a priori that the data exist and the possibility exists that the files could disappear in the future, the data are frequently not Accessible without contact or permission from the author, and the data are often not Interoperable or Reusable because of incompatible formatting or a lack of interpretable metadata (Wilkinson et al., 2016). Furthermore, authors frequently do not respond to data requests despite claiming that data are available upon request. Although limited research has been done to investigate the effectiveness of "data available upon request" statements, the consensus is that such statements are an ineffective means to share data and result in significant delays in obtaining data (Langille et al., 2018; Tedersoo et al., 2021; Vines et al., 2013). Vines et al. (2013) compared four means of data sharing and reported that contacting authors of "data available upon request" statements resulted in data being sent in only ∼59% of cases and often required multiple emails and weeks of waiting (although all seven requested data sets from *BMC Evolutionary Biology* arrived within 2 weeks). Tedersoo et al. (2021) sent data requests to corresponding authors (and co-authors) of 875 papers published in *Nature* and *Science* from 2000 to 2019 representing nine different disciplines and reported that data requests were fulfilled only 39.4% of the time, with a range of 27.9%–56.1% among disciplines. Interestingly, social science authors declined to share data (as opposed to ignoring the request) more often than authors from the other eight disciplines.

Low levels of data sharing likely impact the quality of research since the data are not freely available for replication or future studies. Based on a study of 49 papers published in two major psychology journals, Wicherts et al. (2011) reported that a reluctance to share data was associated with weaker support for results and a higher prevalence of statistical errors. A tendency to report only the best results and poor availability and documentation of data may lead to more error and bias in published studies, suggesting that increased data sharing will improve the reproducibility of results and the robustness of research in general (Destro Bisol et al., 2014). Creating a culture of more open data sharing will likely require moving beyond voluntary action to data sharing requirements. Vines et al. (2013) reported that journals with mandated data archiving policies improved the odds of finding the data online almost 1000-fold compared to journals with no data policy. A study of genetic data sets extracted from over 500 papers published between 2008 and 2011 reported that 77% of papers provided data immediately (through supplemental material or online repositories), 78% shared data after emails sent to authors of papers where data available upon request was explicitly stated, and 83% shared data when emails were sent to all authors of papers with no information on data availability; this relatively high level of data sharing shows the effectiveness of an established culture of data sharing culture in the field of genetics where even emails to authors resulted in high levels of data sharing (Destro Bisol et al., 2014).

The big challenge in encouraging more researchers to share their data more openly is to address their concerns, which include lack of

30 | WILEY—YEARBOOK OF BIOLOGICAL ANTHROPOLOGY The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

recognition for sharing data, apprehensions about how the data will be used (e.g. being "scooped," misrepresented, or errors found), and a perceived loss of power through sharing (see Section 3 for the results from our survey). Concerns about being "scooped" are often cited, although stories of actual scooping by researchers other than collaborators (i.e., those who do not already have access to the project data) are scarce. As Elton points out (2020), the data are not the intellectual property, the ideas are, and chances are slim that two people will have the same idea when it comes to a specific data set. A combined carrot-and-stick approach to encourage more open data sharing may be most effective. As mentioned above, open data sharing is associated with many benefits, including an increased number of publications and enhanced citation impact (Colavizza et al., 2020; Pienta et al., 2010). There are several movements to further increase recognition and reward for scholars who share their data, and more broadly those who share their research efforts and time. Marwick and Birch (2018) have proposed a scholarly citation standard for databases, comparable to that for published articles, that ensures authors receive appropriate credit. Journals like *PLOS ONE* are encouraging researchers to submit their protocols for publication in recognition of the effort necessary to develop protocols and to increase broader sharing of research methods and transparency in research (Protocols—PLOS). ORCID iDs are unique digital identifiers that follow researchers' accomplishments throughout their careers, even with name changes, by recognizing publications, preprints, databases, grants, peer reviews, and other achievements. There is also a movement to include shared data sets in the H-index that is often used to measure the impact of an author's scholarly record (DigitalScience et al., 2020). Finally, faculty can encourage universities to give credit for data sharing by including this information in their curricula vitae, annual activity reports, and tenure-and-promotion packets.

Journals and funding agencies are becoming more assertive in creating and enforcing strict data-sharing policies that conform to the FAIR principles. The *Proceedings of the National Academy of Sciences* recently banned a researcher from publishing in the journal for 3 years after he refused to share a mutant algal strain in violation of the journal's data and materials sharing policy (Berenbaum, 2021; Offord, 2021). Since a journal's data-sharing policies take effect after the study is complete and ready for publication, data sharing policies at funding agencies are potentially more impactful because they require the researcher to consider data sharing before the project begins and are more likely to ensure that data sharing is integral to the success of the project. Data management plans are now required for all grant proposals to NSF and NIH and encourage researchers to resolve data management and sharing issues before the project has begun. Discussions with stakeholders (e.g. study participants, indigenous communities, government officials, and museums) before a project has begun show respect for stakeholder input and are more likely to be productive and result in innovative agreements about data sharing. Since granting agencies require data management and sharing policies, costs for these activities should be written into the grant budget. Ultimately, federal funding agencies have clearly worded data and materials sharing policies that will increasingly be enforced so it

behooves researchers to prepare for this change and embrace more open data sharing. For example, NSF's policy states that

> Investigators are expected to 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 grants. Grantees are expected to encourage and facilitate such sharing. (NSF, 2020).

Progress is being made in addressing the confidentiality and privacy issues associated with sharing data. Several creative solutions are in development that allow the analysis of shared data while protecting identifying information, as well as the integration of computational tools with sharing platforms that allow analyses without access to the raw data. The Dataverse Project (dataverse.org) is one platform that is finding the middle ground between fully open and fully closed data sets (Crosas et al., 2018). For instance, metadata that are unrestricted are published so the data set is findable through a public repository, but restricted data are only accessed by authorized users with an approved use agreement. Developers are also integrating tools to create differentially private releases that introduce a minimum amount of noise to preserve individuals' privacy so that researchers can conduct certain analyses without accessing the raw data. COINSTAC (coinstac.org) is a platform that creates decentralized data sets but does not move the data from the sites that own them and has integrated analytic tools that allow researchers to launch computations on a custom decentralized data set from anywhere in the world (Plis & Calhoun, 2021). The development of innovative data-sharing platforms that grapple with issues of confidentiality, privacy, and limited access to samples, data, and analytic tools is an emerging, interdisciplinary field with calls for researchers and developers from around the world to join the cause (Plis & Calhoun, 2021).

As data sharing becomes more widespread, more resources are becoming available. University libraries are getting involved by offering workshops, one-on-one consultations, and online resources for developing data management plans, creating data spreadsheets for archiving and sharing, identifying appropriate data repositories, providing assistance to upload data to repositories, and sometimes paying fees for institutional access to online repositories. In addition to a veritable smorgasbord of online data repositories (Section 5), there are also a growing number of resources that evaluate different repositories (https://www.nature.com/sdata/policies/repositories; Devriendt et al., 2021; Pauli et al., 2017).

3 | RESULTS OF DATA ARCHIVING AND SHARING SURVEY

3.1 | Characteristics of survey respondents

We developed a survey to query researchers on their attitudes and practices with data archiving and sharing across different anthropological and biological fields (full survey report at https://ql.tc/4reI82). A

primary goal of the survey was to better understand the potential bene-
fits and challenges that motivate and limit broader data archiving and
sharing. In reading some of the responses to the survey, a secondary
benefit of the survey may have been to educate researchers, possibly
young researchers primarily, about the importance of data archiving
and sharing.

The survey questions were developed by the authors based on our
experiences with data sharing, data repositories, and research and pub-
lishing in biological anthropology. The survey was administered through
Qualtrics at http://bit.ly/DataShareSurvey (note: survey is closed and
no longer accessible from this link). The link to the survey was distrib-
uted via listserv, website, newsletter, and/or Twitter account to 25 pro-
fessional societies and research communities in anthropology, biology,
evolution, ecology and related disciplines, and department websites (-
Table S1). We surmised from write-in answers about where respondents
accessed the survey that the link to the survey was further distributed
by respondents to other websites, departmental listservs, and col-
leagues. The goal was to distribute the survey as widely as possible, and
we specifically included disciplines outside of anthropology that had
more experience with open data sharing, in order to get a broad range
of responses. Nevertheless, the sample of respondents was not a ran-
dom sample of researchers, but was biased toward people who are likely
to respond to surveys and the distribution sites were biased towards
those known to the authors. These limitations are likely to have
influenced our data set of responses in several ways. First, we purposely
included researchers outside of anthropology in order to collect a broad
range of responses. It is possible that people who felt strongly in favor
of open data sharing were more likely to respond than those who held
negative views on open data sharing. However, the reverse could also
be true. Second, although many questions allowed write-in responses,
the majority of questions were in a multiple-choice format meaning that
other responses were not possible. Ultimately, the results from a survey
are only as good as the survey itself. We believe that the number and
range of responses indicate we were successful in reaching a broad
audience and receiving diverse responses to the questions.

The survey was live from May 23 to August 2, 2021, and the last
response was received on July 23, 2021. We received 722 complete
responses. The majority of respondents accessed the survey through
the website for the American Association of Physical/Biological
Anthropologists and were tenured full professors or graduate students
(Table 1). Respondents came from a variety of fields with the highest
representation from physical/biological anthropology (13%), evolu-
tionary biology (12%), paleoanthropology (10%), archaeology (8%), and
bioarchaeology and skeletal biology (6% each). Twenty-one percent of
respondents identified as a member of an under-represented or mar-
ginalized group with Racial identity and Sex reported as the most com-
mon under-represented or marginalized identities.

3.2 | Using data shared by other researchers

*Almost 80% of respondents (530 respondents, 78.3%) reported that they
had benefited from other researchers sharing their data.* In answer to the

TABLE 1 Characteristics of survey respondents

Society membership (>30 respondents/category)	Percent[a]	#
American association of physical/biological anthropologists	29%	221
European society for the study of human evolution	20%	157
American society of ichthyologists and herpetologists	11%	87
Human biology association	7.0%	54
American society of primatologists	4.6%	35
Morphmet	4.4%	34
American association of anthropological genetics	4.0%	31
Career stage or occupation (>30 respondents/category)		
Full professor (tenured)	19%	133
Graduate student	17%	119
Associate professor (tenured)	13%	90
Assistant professor	12%	80
Postdoctoral fellow	12%	80
Emeritus professor	4.5%	31
Field of expertise (>30 respondents/category)		
Physical (biological) anthropology	13%	204
Evolutionary biology	12%	186
Paleoanthropology	10%	155
Archaeology	7.8%	120
Bioarchaeology	6.3%	97
Skeletal biology	6.1%	94
Human biology	5.8%	90
Primate behavior and ecology	5.8%	90
Genetics/genomics	5.7%	88
Functional anatomy/biomechanics	5.7%	84
Paleontology	4.8%	74
Phylogenetics	4.1%	63
Biogeography	2.8%	43
Member of an under-represented or marginalized group (21% of respondents)		
Racial identity	26%	57
Sex	25%	55
Sexuality	15%	34
Socioeconomic	15%	34
Gender identity	8.2%	18

[a]Respondents were able to choose more than one category and
percentages are based on the total number of responses to each question
rather than the number of total respondents.

question "How did you benefit?", the top three responses indicated
clear benefits to the respondents' research, namely that research was
improved, performed more quickly and less expensively, or would not

32 | WILEY—**YEARBOOK OF BIOLOGICAL ANTHROPOLOGY** The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

TABLE 2 Benefits from using data shared by other researchers (530 respondents)

Response	Percent[a]	#
Improved an ongoing study through data resources made accessible to me	32%	379
I was able to do a research project much more quickly and less expensively (without having to travel or pay for as much data collection)	21%	247
I was able to design and execute a study that would not have been possible otherwise	21%	241
Improved my teaching/outreach through data resources made accessible to me	15%	177
I was able to design and execute all or part of a thesis with shared data	8.5%	99
Other (write-in)	2%	23

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.

have been possible otherwise (total of 74%; Table 2). *Additional write-in answers mentioned an improved ability to do data analyses and specifically included: being able to personally conduct analyses with shared data; learn new skills with shared software; understand the authors' code and learn analyses by reproducing their code; learn data management techniques; verify results and correct mistakes; provide new evidence for existing theories; improve a publication; and have a better understanding of the research.* Two selected statements capture some of the goals of open shared data: "My student was able to execute all of a thesis with shared data" and "My disability limits my travel to field sites; shared data allows me to still ask questions."

In terms of how researchers accessed the shared data, responses were evenly split between online repositories (332 responses, 46%) and direct sharing with the author (327 responses, 45%). Respondents were asked to specify the online repositories they had accessed and the NCBI databases, MorphoSource, and Dryad were each mentioned ~50 times (Table 3). In total, almost 100 repositories were mentioned, demonstrating that researchers are accessing data from a wide variety of sources and providing a valuable summary of the types of shared

TABLE 3 Online repositories used by survey respondents to access data shared by other researchers (530 respondents)

Repository	# mentions
NCBI repositories, such as GenBank/SRA/dbGAP/GEO	51
MorphoSource	50
Dryad	47
GitHub	15
Figshare	14
Zenodo	14
Digital Morphology Museum/KUPRI (Kyoto University's Primate Research Institute)	8
GBIF (Global Biodiversity Information Facility)	6
Digimorph; European Nucleotide Archive (ENA); ResearchGate; OSF (Open Science Framework)	5
Academia.edu; PRIMO	4
ADS (Astrophysics Data System); EMBL (European Molecular Biology Laboratory)/Ensembl; Mendeley; NHANES (from CDC); NOW (New and Old Worlds paleontology database); PalaeoDB; Wellcome Osteological Research Database	3
BOLD (Barcode of Life Data System); DHS (Demographic and Health Surveys); Google Scholar; ICPSR (Inter-university Consortium for Political and Social Research); MoveBank; Neotoma Paleoecology Database; NESPOS (Pleistocene People and Places, formerly Neanderthal Studies Professional Online Service); NOAA National Centers for Environmental Information; SketchFab; VertNet; WorldClim	2
AAOF (American Association of Orthodontists Foundation) Legacy Collection; African fossils; AddHealth (National Longitudinal Study of Adolescent to Adult Health); AHOB (Ancient Human Occupation of Britain) Database; Allen Brain Atlas; Animal Diversity Web; ArchNet.org; Arctos; Aves 3D; Bakeng se Afrika/University of Pretoria; BioBank Japan; CARTA; Dataverse Project; DDBJ (DNA Data Bank of Japan); Earth System Grid Federation; eLife; European Mammal Foundation; ESRF (European Synchrotron Radiation Facility); FaceBase; Federal government; HAPMAP; HCHS/SOL (Hispanic Community Health Study/Study of Latinos); Human Fossil Record; iDigBio; iNaturalist; IPUMS (Integrated Public Use Microdata Series); IsoBank; IUCN (International Union for Conservation of Nature); JSTOR; Love Consortium; KEGG (Kyoto Encyclopedia of Genes and Genomes); mGAP (Macaque Genotype and Phenotype Resource); MIDUS (MIDlife in the United States); Morphobank; MorphoMuseuM; NatureServe; NCD-RisC (Non-Communicable Diseases - Risk Factor Collaboration); NHM (Natural History Museum) Data Portal; NICHD DASH (Data And Specimen Hub); NMDID (New Mexico Decedent Image Database); NSHAP (National Social Life; Health and Aging Project); ProteomeXchange; OBIS (Ocean Biodiversity Information System); Open Context; Personal Genome Project; ORSA (Open Research Scan Archive); panTHERIA; Pfam (Protein families); PsychENCODE; PubMed; Qiita; Road Database; ScienceDirect; TAPS (Tsimane' Amazonian Panel Study); tDAR (the Digital Archaeological Record); TNT (The Neandertal Tools); UK Biobank; UK Data Archive; UniProt; USGS National Inventory of Dams; World Bank; ZFIN (ZebraFish Information Network); https://c14.arch.ox.ac.uk; https://wonder.cdc.gov/	1

data that are available to researchers. Respondents were also able to write in other ways in which they had accessed data and these 61 responses primarily mentioned journal supplementary materials and lab/author websites, but also included non-online museum databases or collections, local or departmental repositories or databases, data sharing through collaborators, hard-copy dissertations, and state government data requests.

The 147 respondents (21.7%) who reported they had not benefited from data sharing by other researchers identified two main barriers—difficulties in re-using data (because the data collection and reporting methods are complex and poorly standardized) and the lack of a data-sharing culture in their field (Table 4). Lack of repositories was the third most frequent barrier, but lack of knowledge about appropriate repositories may be the real barrier as this was the most common write-in response to a later question about the challenges that limit data sharing (Figure 2). Stakeholders who prevent data sharing were also cited as a reason that limits the benefits from data sharing. Respondents were able to write in other reasons they had not been able to benefit from openly shared data. The vast majority of write-in responses (74%, 42/57 responses) indicated that respondents did not need other's data to conduct their research, had never tried to access other's data, or were too busy working with their own data to be interested in others' data. Some of these responses, for example, "Simply haven't pursued this option yet" or "I am confident that if I did need this [others' raw data], I'd likely be able to get it," suggest an openness to data sharing even among the 21.7% of respondents who reported they had not benefited from data sharing by other researchers. The remaining write-in responses were generally mentioned only once and included: use of data that are published in papers or supplemental materials (in contrast, other respondents interpreted data shared in papers as openly shared data); reluctance to share data by respondents and other researchers in the field;

shared data are not useful or sufficiently annotated to be useful; the need to generate original data for a degree; the difficulty of anonymizing human subject data; work with communities with data sovereignty and access agreements; the obligation to include others as co-authors if you use their data; concern that shared data might be false; and the mechanisms for sharing are still incipient.

3.3 | Sharing one's own data using an online repository

When asked "Have you archived and/or shared data using an online repository before?" more than half of respondents (57%, 369 respondents) answered in the affirmative. When asked what motivated them to archive and share their data, three of the top four answers indicated an altruistic view of research in which data sharing was done to help others, satisfy personal or professional standards, or to "pay it forward" after having benefited from others sharing their data (total of 55%, Table 5). Two strongly endorsed responses (Required to share data by journals or funding agencies, total of 26%) suggest that journal and funder requirements to make data publicly available are effective at expanding data sharing. The desire to increase the impact of one's research, for example, increased data citations, was endorsed by 12% of respondents. Low endorsement of responses that indicate expectations or encouragement of data sharing by stakeholders or employers (total of 5.3%) strengthens the view that there is not yet a culture of data sharing. Additional write-in responses included: "it's the right thing to do"; transparency and reproducibility; create new collaborations; foster scientific advancement; encouraged but not required by a journal; part of project design; potential benefit to study population; data preservation; expected/encouraged by co-authors; and part of a thesis. One respondent disputed that data are the property of researchers and stated that our survey should have engaged more with the impacts of colonialism and white supremacy.

When asked how study respondents had benefited from sharing their own data, three of the top four responses indicated benefits specific to the respondent, that is, increased citations, invitations to collaborate, and providing evidence of data sharing for a grant proposal (total of 51%, Table 6). The other response in the top four stated that another researcher was able to confirm and build on the shared data, supporting the idea that data sharing leads to an increased pace of research. The vast majority of write-in responses indicated that the respondents had not received any benefits from sharing their data, but some noted that they had not received any benefits "yet," or any benefits "except good feelings" indicating a generally positive attitude toward data sharing regardless of benefits. Only one respondent was rueful about spending much money to collect the shared data and having gotten nothing from it. Other write-in responses included: helps my own data analysis and retrieval; positive feedback from other researchers; acknowledgment in papers; asked to contribute a paper; an error was noted and corrected; expand my professional network; being a good citizen; personal satisfaction; happy to share; and "it's just the correct thing to do." One respondent stated "I put years

TABLE 4 Reasons that respondents were not able to benefit from data sharing by others (147 respondents)

Response	Percent[a]	#
The data collection and reporting methods are complex and poorly standardized meaning it is difficult to re-use data between studies	28%	56
There is not a culture of sharing data in my field (it is not required or expected)	22%	43
There are no repositories capable of storing my field's data in a useful way	11%	21
Stakeholders other than the researchers who collected the original data act to prevent data sharing in my field (e.g., museum administrators, descendent populations)	8.6%	17
Other (write-in)	30%	60

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.

TABLE 5 Motivation to archive and share data (369 respondents)

Response	Percent[a]	#
I wanted to benefit others by sharing my data	20%	244
Satisfy my own personal/professional standards	19%	237
Required to by journals in which I publish my research	17%	211
I have benefited from others sharing data and wanted to pay it forward	16%	191
I hoped to increase impact of my data (e.g., to generate data citations)	12%	152
Required to by a foundation supporting my research (e.g., NSF)	9.1%	112
Expected/encouraged by other stakeholders (e.g., museum policy)	2.9%	36
Expected/encouraged by my employer (e.g., for promotion/tenure)	2.4%	30
Other	1.5%	18

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.

TABLE 6 Benefits experienced from sharing own data (369 respondents)

Response	Percent[a]	#
My paper was cited specifically for the shared data it included	20%	101
An independent researcher was able to confirm and build upon results of my study because I shared the data	18%	92
I was invited to collaborate on a paper using data I shared	17%	84
I was able to provide evidence of my past data sharing in a new grant proposal	14%	72
The data were used for outreach and education and I was able to report that as impact of my data	8.4%	42
Evidence of my data sharing was remarked upon positively in my performance reviews/promotion assessments	5.4%	27
I was able to report more publications using my data to the grant sponsor who funded the data collection than if I hadn't shared my data	4.0%	20
Other (write-in)	12%	62

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.

of effort into gathering data. Nice to see it being used by others. This increases its impact. I don't care about credit." A word cloud based on both the multiple-choice and write-in answers to the question about the benefits of data sharing highlights new and increased results, papers, grants, collaborations, and invitations constitute the primary benefits of data sharing (Figure 1).

Respondents who indicated they had archived and/or shared their data using an online repository were asked to list the repositories they had used. A wealth of databases and websites were mentioned, although not all were data repositories that meet FAIR guidelines (e.g., journal supplementary information, manuscript repositories like ResearchGate, or personal websites; Table 7). The top three repositories were Dryad, NCBI databases, and MorphoSource, representing ∼40% of the online data repositories mentioned. These repositories were also the top three sites mentioned for accessing data (Table 3), suggesting that these repositories are easy to use for accessing, archiving, and sharing data sets. Over 80 repositories were used by respondents to archive and share data, again demonstrating that researchers are archiving and sharing data through a wide variety of sources and providing a valuable summary of the repositories that are available for sharing data. The sheer number of repositories available creates a challenge for researchers to determine where best to share their data and we provide guidance on trusted repositories in Section 5 of this paper.

3.4 | Breakdown of data sharing by field of expertise and professional society

Survey responses were also analyzed according to the respondent's field of expertise and professional society (Table 8). The first result to

note is that respondents from all fields reported high levels of having benefited from others sharing their data, ranging from 66% in primate behavior and ecology to 98% in phylogenetics. Majority support for the benefits of data sharing was also shown in the professional societies, ranging from 57% in the Human Biology Association to 97% in the American Association of Anthropological Genetics. The other result to note is that all fields and societies showed more individuals who reported they had benefited from others sharing their data than those who reported having shared their own data. Sometimes the gap between benefiting from shared data and sharing one's own data was quite large, such as the 48% and 36% differences seen in bioarchaeology and archaeology, respectively. The 31% difference seen in physical (biological) anthropology respondents was the third-largest gap among research fields and the 28% difference seen in AABA respondents was the largest gap among societies—both of these numbers suggest there is not a strong data sharing culture in biological anthropology. The gap may also reflect a lack of data to be shared because of not having completed a research project, as reported by 19% of respondents who had never shared their data (Table 10). The high number of respondents who have not completed a research project may reflect the high percentage of graduate student respondents; *the increase in benefiting from others' shared data relative to sharing one's own data suggests that junior faculty who have not completed a research project may benefit preferentially from more open data sharing.* Benefiting from other's shared data and sharing one's own data are highly correlated and the rankings for both sets of responses are

FIGURE 1 A word cloud displaying multiple-choice and write-in answers to describe the benefits of data sharing. Font size represents the number of times a word was mentioned. Non-descriptive words, for example, "and" and "are," and verbs were filtered out and answers indicating no benefit, for example, "none" or "not sure," were removed. WordItOut. com was used to generate the word cloud

nearly the same with the exception of archaeology and bio-archaeology. In general, phylogenetics, biogeography, and genetics/genomics are the fields with the highest levels of data sharing while Bioarchaeology, Archaeology, and Human biology are the fields with the lowest levels of data sharing.

It is interesting to speculate why researchers from some fields are more or less likely to share data. However, the sample sizes of reasons for not sharing one's own data (Table 10) were too small for formal analysis once responses were divided by field of expertise. While these ideas could be more thoroughly ascertained with explicit models reasoned with Directed Acyclic Graphs (DAGs), we did not have the luxury of the time required to clean and organize the data for this kind of analysis. We have made the raw data available in hopes that other researchers will delve deeper into these data (Data Availability Statement).

3.5 | Challenges to data sharing

Of the 369 respondents who had archived or shared their data, 14% (47 respondents) reported they had experienced negative consequences of data archiving or sharing. The top four responses indicated that data were used without acknowledging the author, data were misused, and research was scooped (total of 77%, Table 9). These negative experiences should be further explored in order to better understand how they occurred and how to avoid them in the future. Fears about data being misused or scooped are often cited as a reason for not sharing data, but anecdotal evidence suggests that data misuse or scooping is often done by collaborators, which presumably would occur regardless of open data sharing. Fourteen write-in responses reflected similar concerns as those in the multiple-

choice answers and included: anxiety about being scooped; data used without permission or attribution; unpublished data were published without consent; data sharing takes time and resources; paper was rejected for already being "published"; collaboration or authorship was not offered in return for data sharing; other researchers refused to share data; data were stolen; pushback from other study members; study population had misgivings; and undermining indigenous sovereignty. Although only 14 write-in responses were given, these negative consequences are varied and deserve additional study to fully understand the underlying reasons so these consequences can be minimized in the future.

In order to accomplish broader data sharing, it is critical to understand why researchers may be reluctant to share their data through public online repositories. Thus, we asked two questions about why respondents had not archived or shared data using an online repository. The first question, "Why have you never archived/shared your data?" was asked of respondents who reported they had never archived or shared data (277/729 respondents) and the most common response was that they had not yet completed a research project to the point of having data to archive or share (19%, 92 responses) (Table 10). The lack of a completed research project may reflect the high percentage of graduate students who completed the survey (119/729 respondents). The other most frequent responses included a lack of awareness of useful repositories, a lack of a culture of data sharing, and other stakeholders who did not allow data sharing. The second question queried researchers who had shared data in the past, but had chosen not to share a particular data set (150 respondents). Focusing on the 10 responses that were common between the two questions, four responses were more frequent for the second question: other stakeholders prevented data sharing; fear of being scooped; giving

36 | WILEY—YEARBOOK OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

TABLE 7 Online repositories used by survey respondents to archive and share their data (369 respondents)

Repository	# mentions
Dryad	82
NCBI repositories, such as Genbank/dbGAP/SRA/GEO/TOPMed	61
MorphoSource	41
Figshare	35
github	34
Zenodo	29
OSF (Open Science Framework)	20
University data repositories	16
iDigBio	12
MoveBank; NOW (New and Old Worlds paleontology database); PRIMO	11
ENA (European Nucleotide Archive); GBIF (Global Biodiversity Information Facility); Mendeley	9
University and museum online databases	9
Qiita; tDAR (the Digital Archaeological Record)	4
BOLD (Barcode of Life Data System); Dataverse Project; MorphoBank; NESPOS (Pleistocene People and Places, formerly Neanderthal Studies Professional Online Service); Open Context	3
ADS (Astrophysics Data System); EGA (European Genome-Phenome Archive); EMBL (European Molecular Biology Laboratory)-EBI (European Bioinformatics Institute)/Ensembl; government surveys, e.g. http://srtm.usgs.gov; ROAD Database	2
AHOB (Ancient Human Occupation of Britain) Database; Alexandria Archive; Arctos; Ariadne; Atlas of Living Australia; CRAN (Comprehensive R Archive Network); DAHP (Digital Archives of Human Paleobiology); DigiMorph; digital.csic.es/; Edmond; eMammal; FaceBase; FishesofTexas.org; human-fossil-record.org; ICPSR (Inter-university Consortium for Political and Social Research); IFWIS; IsoBank; KNB Data Repository; KUPRI (Kyoto University's Primate Research Institute); MiDAS Global (Global database of microbes); morphDBase; MorphoMuseuM; NEODAT; Neotoma Paleoecology Database; NHANES (from CDC); Paleobiology Database; PaleoCore; Primate-radiograph.com; ProteomeXchange Consortium; Sistema de Biodiversidad Colombia (SiB-Colombia); SNIF (Symphion Network Instrumentation Framework): TNT (The Neandertal Tools); TreeBASE; TreeShare; UK Data Archive; VertNet; XMAPortal; 3Dtheque; https://www.kiglobalhealth.org; https://primatelocomotion.org/; http://anthropologicaldata.free.fr	1
Manuscript repositories	
ResearchGate	5
Academia.edu	3
bioRxiv	2
Not FAIR online repositories, but mentioned in survey	
Journal websites/supplementary information	29
Personal/laboratory research website	6
Google Drive, Squarespace, Microsoft OneDrive	4

first priority for the data to students; and too time-consuming. Thus, the first question queries actual and theoretical reasons for not sharing data while the second question queries actual reasons for not sharing a particular data set. The reasons given for not sharing data align well with previous studies (e.g., Devriendt et al., 2021) and focus on the realities of working with multiple stakeholders who have varying views on data sharing, fears of being scooped and priorities given to students, lack of awareness of appropriate repositories, and the time and expense of using online repositories. It seems that researchers are not generally spiteful and only a handful cited other researchers who had not shared data in the past as their reason for not sharing data. There were also more than 100 write-in responses to both questions that often reinforce, or more fully explain, the reasons listed in

Table 10 and these responses can be seen in the full report available at https://ql.tc/4reI82.

Our survey also included an optional question with write-in responses that asked respondents to list the challenges that have limited their data sharing or archiving. A word cloud emphasizes some key challenges (Figure 2). *The challenge mentioned most often focused on knowledge—the lack of knowledge, expertise, and established best practices for repositories, guidance of the best repositories to use, how to prepare data for archiving, and how to discuss data sharing with collaborators.* The AABA ad hoc committee on data access and sharing has published two articles on guiding principles and best practices for data sharing in biological anthropology that addresses some of these issues (Boyer, Jahnke, et al., 2020; Turner & Mulligan, 2019). Other often-mentioned challenges

TABLE 8 Breakdown of data sharing by field of expertise and professional society

Field of expertise (>15 respondents/category)	Sample (#)	Benefited from other researchers sharing data[a]	Shared own data[a]
Phylogenetics	63	62 (98%)	48 (76%)
Biogeography	43	40 (93%)	34 (79%)
Genetics/genomics	88	79 (90%)	65 (74%)
Paleontology	74	66 (89%)	50 (68%)
Evolutionary biology	185	160 (86%)	130 (70%)
Bioarchaeology	97	81 (84%)	35 (36%)
Physical (biological) anthropology	204	170 (83%)	106 (52%)
Functional anatomy/biomechanics	84	69 (82%)	51 (61%)
Paleoanthropology	155	123 (79%)	86 (55%)
Skeletal biology	94	74 (79%)	51 (54%)
Archaeology	120	90 (75%)	47 (39%)
Human biology	90	64 (71%)	45 (50%)
Primate behavior and ecology	88	58 (66%)	49 (56%)
Professional society (>30 respondents/category)			
American association of anthropological genetics	31	30 (97%)	21 (68%)
Morphmet	34	31 (91%)	25 (74%)
American association of physical/biological anthropologists	219	168 (77%)	107 (49%)
European society of the study of human evolution	157	117 (75%)	83 (53%)
American society of ichthyologists and herpetologists	87	62 (71%)	55 (63%)
Human biology association	54	31 (57%)	23 (43%)

[a]Respondents were able to choose more than one category and percentages are based on the number of responses for each field or society. Data are sorted by the third column.

TABLE 9 Negative consequences of data archiving/sharing (47 respondents)

Response	Percent[a]	#
People used the data without acknowledging/citing me or other sources	28%	23
People misused and/or misrepresented my data	22%	18
My research was scooped	20%	17
My student's research was scooped	7.2%	6
People criticized me for perceived errors	4.8%	4
Another stakeholder became angry with me	2.4%	2
Other (write-in)	16%	13

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.

included: the cost and time needed to clean, prepare and upload data to a repository; the need for guidelines to standardize data formats for usable, high-quality data and the need for rules or restrictions on data usage; restrictions enforced by museums. *Many*

responses described issues of confidentiality, privacy, trust, commercial use, and past abuse of marginalized, indigenous, Aboriginal, and BIPOC communities. There were calls to make data sharing mandatory as well as calls to not strictly enforce data sharing. There were questions about how to share data that are part of a larger, ongoing research project (in this case, openly sharing the data from a specific publication would be a possible compromise). *Other challenges included: conflict between collaborators or researchers and other stakeholders; concerns about giving data away for free; privacy issues when working with humans and vulnerable field sites; issues with availability and standardization of, and ability to include, metadata; making sure code data are clean enough to be used by other; predicting which data might be useful in the future; behavioral and ethnographic data are particularly complex data to standardize and share; data sets that are too small to be worth sharing and data sets that are so large they are difficult to share.*

3.6 | Conclusions

We believe that expanded data sharing using online repositories will benefit researchers and the field of biological anthropology as well as

38 | WILEY— YEARBOOK OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

TABLE 10 Reasons for not archiving or sharing data

Response	Why have you never archived/shared your data? (277 respondents)		Reason for not sharing a particular data set although researcher had shared data in the past (150 respondents)	
	Percent[a]	#	Percent[a]	#
I have not yet completed a research project or scientific manuscript to the point of having to plan for data archiving or sharing yet	19%	92	Not a possible response	
There is not a culture of data archiving/sharing in my field (it is not required or expected)	10%	50	Not a possible response	
Other stakeholders did not allow me to archive/share data (e.g., collaborator, advisor, museum administrator, descendant population)	9.4%	46	24%	64
I could have been scooped on current or planned future research if I did	7.2%	35	17%	30
I planned to give first priority of reuse to my students (or future students)	4.5%	22	11%	30
Too time-consuming	5.7%	28	11%	29
I am not aware of any repositories capable of storing my data in a useful way	13%	64	8.6%	23
I am not aware of any repositories that allow my data to be archived with the appropriate restrictions/embargos on access (write in the restrictions/embargos of known repositories).	7.6%	37	5.6%	15
Too expensive to pay repository fees	5.1%	25	4.5%	12
Not relevant to verifying the study's conclusion	3.3%	16	3.0%	8
Other researchers denied my requests for data in the past and it did not seem fair that they would get access to my data if I shared it openly	0.82%	4	0.75%	2
Other (write-in)	14%	69	15%	39

[a]Respondents were able to choose more than one category and percentages are based on the total number of responses to each question rather than the number of total respondents.
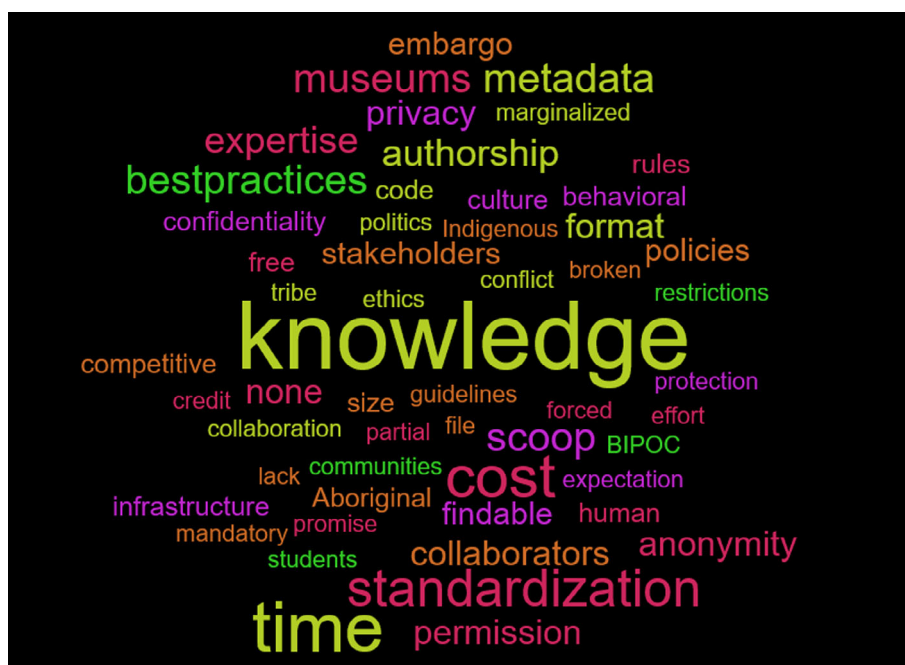
FIGURE 2 A word cloud displaying write-in answers to describe the challenges that limit data sharing. Font size represents the number of times a word was mentioned. Non-descriptive words, for example, "and" and "are," and verbs were filtered out and only words mentioned more than once were displayed. WordItOut.com was used to generate the word cloud

TABLE 11 Recommended trusted, open-source, community-led data repositories

Repository	Description	Fees/costs	Size limits	Information on open data sharing principles
Dryad digital repository	Maintains data sets from diverse fields and accommodates multiple data types with options for restricted access	$120 for first 20 GB, $50 for each additional 10 GB	None stated	view FAIRsharing entry
Figshare	Maintains data sets from diverse fields and accommodates multiple data types with options for restricted access	100 GB free per Scientific Data manuscript. Additional fees apply for larger data sets.	1 TB per data set	view FAIRsharing entry
Harvard Dataverse	Maintains data sets from diverse fields and accommodates multiple data types, innovating new approaches to confidentiality and privacy concerns	Contact repository for data sets over 1 TB	2.5 GB per file, 1 TB per data set	view re3data entry
Inter-university Consortium for Political and Social Research - ICPSR	Maintains >250,000 files of research in the social and behavioral sciences, hosts 21 specialized collections of data on education, aging, criminal justice, substance abuse, terrorism, and other fields			view re3data entry
MorphoSource 3D data repository	Maintain image data representing research objects. Heavy specialization for 3D modalities and formats.	Free for small data deposits, contributors must arrange to pay deposit fee for large scale use (>500Gb)	No strict limits. Different upload methods for files greater than ~10Gb.	view re3data entry
Open science framework	Maintains data sets from diverse fields and accommodates multiple data types with options for restricted access	Free	5 GB per file, multiple files can be uploaded	view FAIRsharing entry
Science Data Bank	Maintains data sets from diverse fields and accommodates multiple data types with options for restricted access	Free	8 GB per file, no limit to data set size	view FAIRsharing entry
tDAR (Digital Archaeological Record)	Maintains archaeological data sets—digital documents, data sets, images—while maintaining confidentiality of legally protected information	$10 for 1 file/10 MB $100 for 10 files/100 MB; $500 for 100 files/1 GB	None stated	view FAIRsharing entry
Zenodo	Maintains data sets from diverse fields and accommodates multiple data types with options for anonymized data or restricted access	Donations towards sustainability encouraged	50 GB per data set	view re3data entry

Source: Adapted from https://www.nature.com/sdata/policies/repositories.

TABLE 12 Recommended NIH genetic repositories (see more NIH repositories at ncbi.nlm.nih.gov)

GenBank	DNA sequences	Free	None stated	view FAIRsharing entry
dbGaP	The database of Genotypes and Phenotypes (dbGaP) for research on interactions between genotypes and phenotypes in humans	Free	None stated	view FAIRsharing entry
dbSNP	Simple genetic polymorphisms, for example, single nucleotide polymorphisms—cannot accept sensitive human subject data	Free	None stated	view FAIRsharing entry
dbVar	Structural variations, for example, insertion/deletions—cannot accept sensitive human subject data	Free	None stated	view FAIRsharing entry
Gene Expression Omnibus (GEO)	Gene expression array and sequence data	Free	None stated	view FAIRsharing entry
Sequence Read Archive (SRA)	High throughput sequencing data	Free	None stated	view FAIRsharing entry

the broader scientific community and general public. Our survey queried over 200 biological anthropologists and 500 scientists in other biological and social sciences. The fact that almost 80% of respondents reported they had benefited from other researchers sharing their data (Table 2) indicates there are clear advantages to open data sharing. However, the fact that fewer than 60% of respondents reported sharing their own data using an online repository suggests there is room for improvement. It is worth noting that the leading reasons for sharing data noted by respondents who did share their own data were altruistic reasons rather than the increased research impact that is also associated with sharing one's data (Table 5). Forty-seven respondents (14% of those who had shared data) reported negative experiences (Table 9), so it is important to find ways to ensure that everyone benefits from data sharing. The challenges that limit data sharing are diverse, but many focus on a lack of knowledge about which repository to use and how to use repositories (Figure 2). In Section 5, we explain some of the important issues surrounding repositories and we provide a list of trusted data repositories. Finally, the personal stories presented in Section 4 may be helpful as they detail how we have navigated the ever-changing waters of data sharing throughout our careers.

4 | PERSONAL EXPERIENCES WITH DATA SHARING

4.1 | Doug Boyer, Associate Professor, Department of Evolutionary Anthropology, Duke University

I obtained my PhD from Stony Brook University in 2009. During the closed-door defense, my external committee member, Philip Gingerich (a former director of the University of Michigan, Museum of Paleontology) came to my rescue, defending a methodological approach I had used by explaining to the rest of the committee that I had been academically "trained at Michigan." Which is to say, I got my undergraduate degree there. While I certainly hope that some of the lessons from my 6 years at Stony Brook stuck, I do tend to agree with Gingerich in certain respects. As an undergraduate, I was mentored by Gingerich and other

scholars in the University of Michigan Museum of Paleontology (UMMP) including Daniel Fisher, Gregg Gunnell, William Sanders, and Jonathan Bloch. At UMMP (1997–2003), I was repeatedly reminded to "let the data speak" and to avoid pushing narratives in place of objectively defensible statements. What I saw at UMMP was a research culture committed to describing and documenting the physical evidence that illuminates the biology of extinct and extant species. I believe that categorizing the 672 publications of Phil Gingerich as "descriptive" or "theoretical" would reveal the vast majority of papers to be in the first category, even though Gingerich's theoretical work has been monumentally impactful. In 2019, Gingerich published a book unifying observations of evolutionary rate across time scales Rates of Evolution: A Quantitative Synthesis and deposited the data supporting it (compiled over decades) in Dryad.

In any case, from the time I started doing research, I have felt that unless the data used to conduct analyses were available in association with corresponding publications, there was little separating the conclusions of that research from subjective opinion. I have made consistent efforts to always provide every number that I have analyzed or drawn inferences from.

When my research data reflected measurements of physical specimens, my urge was to publish every raw measurement. When my data were qualitative scorings of character states for a species, my urge was to provide a list of all specimens used to determine the score, and possibly photographs. When my data reflected measurements from 3D renderings of specimens, I felt compelled to share both the 3D renderings and the measurements.

It was when I started relying on 3D models of specimens for my research that I noticed a gap in sharing practices. In 2008, there were already data-sharing resources and strong expectations for sharing specialized genomic data, but nothing equivalent existed for 3D data representing biological specimens. A few platforms had 3D data, but they were not open for public deposits. There were not even many generic data archiving platforms. Supplementary data sections for journal articles were just emerging. Still, I had the feeling that the meaningfulness of my published research would be greatly diminished if the digital models supporting my research were not accessible.

To make a long story short, eventually (in 2013), I launched MorphoSource.org (Boyer et al., 2014; Boyer et al., 2016; Boyer,

Winchester, & Silverton, 2020) to help fill the infrastructural void I perceived for 3D data archiving. As alluded to earlier in the paper, MorphoSource is a digital repository specialized for 3D data representing physical objects (primarily from museum collections) that is available as a data archiving/sharing solution for researchers and other subject experts. Anyone can search, visualize, and in many cases, download data from the site. As discussed above, our survey shows MorphoSource to be making a notable impact on the field both in terms of data access (Table 3) and data archiving (Table 7). There were a large number of challenges that had to be solved in order for MorphoSource to "work" politically and technically, and many of the "solutions" are still in progress. The other important point to make about MorphoSource is that it was not created in vacuum according to unilateral whims. The plans for the platform were initiated in Jukka Jernvall's research group in 2009–2010. We originally thought it would represent a massive upgrade to a platform called MorphoBrowser hosted by the University of Helsinki. This initial work was put on indefinite hold for a variety of reasons. After joining NYCEP via Brooklyn College, City University of New York (CUNY), I joined Eric Delson, Gregg Gunnell (then at Duke), Meng Jin (American Museum of Natural History) and Jonathan Bloch (University of Florida) to propose a NSF Thematic Collection Network (TCN) focused on generating 3D scans of fossil primates (Delson also mentions this effort below). NSF's TCN program required its applicants to present their own data hosting solutions. Since no adequately specialized 3D repository existed at that time, we dug out the plans from my work with Jernvall's group and submitted an internal CUNY proposal for an independent platform based on that previous work. The proposal to CUNY was awarded in 2012, but the NSF TCN primate scanning proposal was declined. In the meantime, I moved to Duke University and was permitted to pursue the 3D platform using Duke resources as part of my startup negotiations. Even with the platform in progress, the NSF proposal was declined again and we decided not to pursue it further. But the 3D data-sharing platform was initiated and it has grown under Duke University's watch. Finally, in 2017, NSF made an award to Duke to help improve the resource and expand its data holdings. This award allowed me to bring Dr. Julie Winchester on as product manager and lead developer. Julie had a strong vision for the platform because she was involved in the original discussions about MorphoSource in 2009–2010 as a member of the Jernvall lab and because of her experience working with and building tools to analyze 3D data. She is now essentially the co-director of the platform. A key mechanism of "improving" the MorphoSouce was to engage intensively with diverse communities of 3D data producers and consumers, and to better align the platform with established and emerging standards for preservation of 3D data. This engagement included: regular meetings and discussions with an advisory board, scores of users, and scanning experts; formal participation in a variety of working groups (e.g., the Community Standards for 3D Data Preservation group, IIIF 3D community working group, TDWG task group for Enhancement of Audubon Core Vocabulary for 3D data); and hosting/participating in workshops to understand community needs. At the same time that we began the project to update and improve the MorphoSource platform, NSF funded a proposal for a completely different TCN led by David

Blackburn at the Florida Museum of Natural History called "oVert." The improved and updated version of MorphoSource launched in February 2021 and "oVert" has contributed 15,000 scans of ~10,000 specimens.

At the time of writing, MorphoSource has 14,800 users and 1600 data contributors. Its holdings include 152,000 data files representing objects sourced from 620 museum collections and imaged at 250 facilities. Dozens of museums use MorphoSource to actively manage data representing their objects. Fileset DOIs are minted upon request for publication purposes and MorphoSource has been cited in almost 1100 scholarly publications for data archiving or access.

Some reasons that MorphoSource has been successful are that it supports access for outreach and education as well as research, and caters aspects of its platform specifically to those use cases. Another reason for MorphoSource's success is that (counter intuitively) it allows data contributors and museum collection managers to put a variety of access restrictions on shared data. This allows data to be discoverable and accessible even when it cannot be shared with full open access. Finally, MorphoSource tracks/records details on who downloads data for what purpose by requiring a brief survey when users download data. Usage statistics are then made available to museums and data contributors. These usage reports incentivize data sharing for all stakeholders by allowing them to demonstrate the value/demand of their specimens and data.

My personal experience with data sharing of 3D scans has helped me think about how sharing obligations and risks vary according to certain qualities of the data. Though there are many kinds of risks associated with sharing data, in this discussion I mean to refer to the risk of being "scooped" or losing priority to publish a research idea supported by one's data. However, hopefully, it goes without saying by this point in the article, that consideration of other kinds of risks may affect decisions about data shareability, as well (e.g., according to the CARE principles referenced earlier). Characteristics of data suggest different levels of obligation or risk for sharing. Data with lower associated risk, also come with a greater obligation for more open sharing. In my personal experience, the risk of being scooped from early sharing of CT scans and other 3D models (even on a massive scale) seems extremely small. The suggestion by Elton (2020) mentioned earlier seems borne out in my experiences. Furthermore, I believe researchers have a strong obligation to share data derived from public resources (a sentiment reflected in Delson's section below as well). Thus, when CT scans and other 3D models represent specimens from natural history collections, there is a strong obligation to make those data accessible to other researchers (in ways that do not violate museum specific restrictions, of course). Though researchers experience an opportunity cost from the time and money it takes to scan a specimen, they also impose a cost on other researchers by preventing them from accessing the specimen. The handling of the specimen during scanning also degrades that specimen for future researchers. If researchers withhold data they have collected, degradation of the specimen is compounded because other researchers must access the physical specimen again and degrade it further instead of measuring or studying a freely shared scan.

42 | WILEY **YEARBOOK OF BIOLOGICAL ANTHROPOLOGY** The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

Obligations aside, I have never had my in-progress or planned research scooped due to upfront sharing of 3D data sets that were collected by me or my collaborators. That said, the scan data sets I tend to share up front tend to be comparative data sets as opposed to those representing newly discovered or unique fossil specimens that are under study. Samples that represent comparative data sets present the least risk and strongest obligation. In these situations, I believe in making data discoverable and potentially accessible while the imaging work is in progress. I have taken this approach with data collection for my NSF CAREER grant (8499 files representing 291 specimens from the Smithsonian; and 3609 files representing 129 specimens from AMNH Mammalogy).

When a sample is a unique specimen that is the focus of a research project, there is more risk to making the image accessible before the study is published, and I believe it is more acceptable to hold back access longer even when the specimen represents a public resource. This is more often a characteristic of fossil specimens. Even so, plans for restricting data access in this way must be coordinated with museums, which are increasingly developing their own policies about allowable embargo periods on data that represent their specimens.

Comparing the qualities of 3D scans to another data type such as gene sequences strengthens the argument that there is a strong obligation to share 3D data as soon as possible. 3D data more often represent public resources than genetic data and are even less "derivative" than genomic data. Thus, more opportunity cost to the field is imposed by withholding these data, and less intellectual investment is foregone by sharing them.

Ideals of data sharing aside, I believe (as the director of a data repository) that long-term accessibility and interpretability (i.e., maintaining FAIRness) of shared data is the biggest challenge. A key strategy for any web service to increase its sustainability, is to increase its user base. A seemingly obvious way for a data archiving and sharing platform to increase its user base is to increase the variety of data types it supports. This pulls archives towards becoming more generic. While increasing the size and diversity of the user community improves long-term potential of the data files themselves, it unfortunately tends to work against strong adherence to FAIR principles in many respects. FAIRness is especially difficult to maintain in a generic repository for data that are highly specialized. 3D imagery represents such a specialized data type for which there is typically extremely nuanced information on provenance that needs to be maintained in a standardized way and updated over time, and for which there are complex networks of data stakeholders preventing full open access sharing of the data. In contrast, platforms with deeper support for specific data types (and thus better potential for effective discovery and long-term reuse according to FAIR principles), necessarily have smaller communities and more difficulty with sustainability. On top of that, general concerns about availability of global (or at least American) infrastructure that can continue to be tapped to support such efforts grow as economic functioning and supply chains are disrupted by surging pandemics, climate change-accelerated natural disasters, malicious technology, and political unrest. Nevertheless, I feel confident the necessary solutions for robustly sustaining strong access to a diversity of deeply specialized data types can emerge in the coming decades as research communities and funding agencies commit more fully to data archiving and sharing. One approach might be for agencies to support collaborations between larger-scale generic repositories (with demonstrated business models and sustainability) and smaller specialized repositories, that is, design funding programs to allow smaller specialized repositories to transfer their workflows, governance, and holdings to larger ones.

4.2 | Eric Delson, Professor, Department of Anthropology, Lehman College, City University of New York

I have been involved with efforts in data sharing and access for over 20 years. I am most familiar with work in paleoanthropology (broadly speaking), but as we all know, the geneticists got there first with Genbank starting in 1982. In the early 2000s, a groundswell of action began in paleoanthropology to make new fossils more widely available to researchers not affiliated with the discovery team, especially if public funds supported the field research. One of the first published efforts was Gerhard Weber's, 2001 (Weber, 2001) call for *Glasnost* involving open access archiving of 3D models of new fossils within a fixed time after recovery or initial announcement. Additional calls and comments appeared soon afterward (Dalton, 2004; Gibbons, 2002; Tattersall & Schwartz, 2002) and the concept was taken up in comparative biology more broadly (e.g., Parr & Cummings, 2005). Those who found the fossils reasonably responded that they had put in huge effort to obtain funding and search for new fossils, but others argued that public funding led to a responsibility to ensure early access.

At the same time, several individual researchers began developing "silo" archives to make their own data available to a wider audience. Building on a computer-science grant from the NSF, the late Leslie Marcus and I started work in 2000 on a planned online open-access database to provide access to metrical and 3D data that I and my colleagues had collected on living and fossil cercopithecids (and eventually other primates). PRIMO (the NYCEP PRImate Morphometric Online database, http://primo.nycep.org) eventually went live in 2007 at a meeting I organized (see below), and by 2010 it was a solid framework with both craniodental measurements and 3D coordinate (Microscribe) data, the latter mainly from Steve Frost's work (Frost et al., 2003). Over 200 colleagues around the world have since requested a login, and at least 20 publications have used PRIMO data (e.g., Adams et al., 2010; Alba et al., 2014; Alba et al., 2015; Alba et al., 2016; Alba et al., 2018; Alba et al., 2019; Alba et al., 2021; Ferràndez-Cañadell et al., 2014; Geraads & de Bonis, 2020; Gilbert et al., 2014; Gilbert et al., 2015; Hlusko & Mahaney, 2009; Koufos, 2019; Koufos et al., 2016; Pugh & Gilbert, 2018; Radović et al., 2019; Rossie et al., 2013; Singleton et al., 2016; Takai et al., 2015; Tran, 2014).

What makes PRIMO unique is that most of the data are unpublished and raw, so that users can do what they wish without limitation. There are excellent search options, both taxonomic and by

element and variable, thanks to the work of several talented web programmers. I hope to release an expanded version and data set in 2022, and I will release the underlying code and organization once it is stable, in case anyone wants to use it as a basis to build their own database. My view has always been that I collected data in order to work on research problems, and I have supervised a number of student volunteers and postdocs who have contributed data while working with me and afterward. PRIMO is a way to make these data freely available as part of my contribution to better science.

As a result of the growing interest in open data, I worked with Will Harcourt-Smith, Steve Frost, and Chris Norris to obtain funding from the NSF and Wenner-Gren to organize a workshop at the American Museum of Natural History (AMNH) in April 2007 to discuss data access and data-sharing in the restricted area of paleoanthropology (Delson et al., 2007). Few major conclusions were reached unanimously, but the idea of access to data was supported without any significant pressure to provide it. We agreed to set up a web "portal" through which a potential user could access several different online databases from a single site, building on the success of The Paleontology Portal (http://paleoportal.org/), which facilitates searches of the paleontology collections of 12 US museums. The Paleoanthropology Portal (http://paleoanthportal.org/; http://paleoanthportal.org/index_fr.php in French and http://paleoanthportal.org/index_de.php in German) provides links to a number of United States and international web sites related to paleoanthropology. However, it has not been updated in some time, and some of the links are no longer active.

After the 2007 meeting, I met with the organizers of the first two sites below to see if it would be possible to make our databases interoperable; however, the fields we were using were so different that no connection was thought possible. In May 2009, at the initiative of Carol Ember of HRAF (Human Relations Area Files, cultural information for education and research), a workshop was held at the NSF (with NSF and Wenner-Gren support) to spread the concepts of data sharing, and especially data preservation (or archiving) across anthropology. The *AnthroDataDPA* (Anthropology Data Preservation and Access) workshop spanned all four fields of anthropology; a full report is posted at http://anthrodatadpa.org/addpa/. I represented biological anthropology at the workshop and learned more about database methods. This initiative did not proceed further despite requests for additional funding.

Numerous colleagues have produced working databases which provide data of various types, in paleoanthropology or related fields. These include:

1. Human Origins Database (organized by Bernard Wood and Adam Gordon, https://www.humanoriginsdatabase.org/) includes original data on skeletal elements present for the chimpanzee and gorilla collections at the Powell-Cotton Museum and also published measurements and skeletal element information from (Wood, 1991).

2. Smithsonian Institution National Museum of Natural History "What does it mean to be human" website and database (https://humanorigins.si.edu/evidence/3d-collection) is mostly a public education site but includes access to 3D models of extant primates and (sometimes) to scans of casts of fossil human specimens and archaeological artifacts.

3. Human Fossil Teeth (http://anthropologicaldata.free.fr/) includes published (and some unpublished) data on dental remains of *Homo* species, a few australopiths, and great apes from a variety of sources, downloadable in Excel files.

4. NESPOS (Pleistocene People and Places, formerly Neanderthal Studies Professional Online System, https://www.nespos.org/display/openspace/Home) includes information on Pleistocene sites, cave paintings, and downloadable 3D scans (mainly CT) of fossil and extant hominins, as well as links to scans of extant primates from various collections, especially the Digital Morphology Museum of Kyoto University Primate Research Institute (separately accessible at http://dmm.pri.kyoto-u.ac.jp/dmm/WebGallery/index.html).

5. AHOB (the Ancient Human Occupation of Britain) (https://www.ahobproject.org/) is a series of related projects analyzing the earliest human presence on the island, and the website provides a variety of data resulting from this work, as well as links to many publications.

6. Outside paleoanthropology, one of the most successful related web data sources is NOW (New and Old Worlds, originally Neogene of the Old World) where https://nowdatabase.org/ describes the project, while the database itself is located at https://nowdatabase.luomus.fi/). NOW offers extensive information about the mammalian fauna of the Cenozoic and the sites from which fossils are recovered.

I can mention one last initiative that initially failed, but ultimately led to great success. In 2010, NSF announced the Advancing Digitization of Biological Collections program. A group of us at the City University of New York (CUNY), University of Florida, and the AMNH proposed to borrow and CT scan tens of thousands of original specimens of fossil primates from numerous United States museums and make the data available as an online resource. We tried three times, but the reviewers said that we were providing data on too few specimens. They wanted millions of specimens, but the approved projects mainly scanned catalogue cards and made the temporal and geographical data available, while we wanted to provide access to actual specimens. We gave up, but one member of our CUNY group moved to Duke, and there Doug Boyer made it happen when he built MorphoSource (see his story above). Dreams do come true on occasion.

4.3 | William Leonard, Professor, Department of Anthropology, Northwestern University

My experiences with data sharing have been largely through our longitudinal research in lowland Bolivia, The Tsimane' Amazonian Panel Study (TAPS; Tsimane' Amazonian Panel Study| The Heller School at Brandeis University). The TAPS project was initiated in 2002 with the objectives of this research being to better understand the influences of both ecological factors and lifestyle change (e.g., market integration) on the biology and health of an indigenous tropical rainforest population. This project has collected longitudinal waves of survey

and biometric/health data in 13 Tsimane' communities ($n = 633$ adults; 820 children) along the Maniqui River in the Beni Department of Bolivia.

The project has been a collaboration of researchers in biological and cultural anthropology, economics, and ecology and agricultural science. Each of the collaborators has brought a distinct yet complementary set of skills to the project. In terms of human biology and health, TAPS researchers have looked at a range of issues, including the assessment of (a) physical growth and nutritional status using anthropometric methods, (b) household food consumption and energy availability, (c) biomarkers of health (e.g., hemoglobin, lipids, CRP), and (d) parasite loads.

In 2008, we began making the TAPS data publicly available. We announced the availability of the first 5 years of TAPS data in a short communication in the journal *Economic & Human Biology*, which outlined the broad objectives of the research and how to request access to the data (Leonard & Godoy, 2008). In 2015, we released the second batch of data (through 2010), again with a paper in *Economics & Human Biology* (Leonard et al., 2015). This second paper discussed some of major findings of the TAPS research and the challenges with conducting a long-term, longitudinal study in a remote field setting. The paper also provided a complete list of the TAPS publications to that point, and the sources of funding.

Through 2019, all requests for the TAPS data were submitted directly to my colleague Ricardo Godoy. Each researcher completed a short form, providing some background information and agreeing to the terms of use. Between 2008 and 2019, the data were sent to 116 different researchers from around the world. In addition to the data set itself, the researchers were provided with a full data dictionary and paper providing a historical overview of the study.

The majority of the data requests came from students – 50% from graduate students (either Master's or PhD), 8% from undergraduates (often working on theses), and another 8% from postdoctoral fellows. Faculty at academic institutions made up 28% of the requests, and the remaining 6% were from other professionals – independent researchers, those working in industry. Geographically, the majority of the requests came from North America (71%), with 23% coming from Europe and 3% each from South America and Asia.

Reflective of the nature of the TAPS research program, these data have attracted a broad, interdisciplinary audience spanning the biological and social sciences. Economists (41%) and Anthropologists (21%) comprise majority of the total requests. Other fields represented included: international development, ecology, public health/medicine, sociology, psychology, computer science, and public policy.

In 2020, we made the full set of TAPS data available through the Inter-university Consortium for Political and Social Research (ICPSR) sponsored by the Institute of Social Research at the University of Michigan. The site provides a full codebook with detailed definitions of all the variables along with background material on the study itself (Godoy & Leonard, 2020; Tsimane' Amazonian Panel Study, El Beni, Bolivia, 2002–2010 [umich.edu]). Since August of 2020, six researchers have accessed the TAPS data.

While all the senior members of the TAPS research team contributed to making the data accessible, the project was directed and overseen by Ricardo Godoy of the Heller School of Social Policy and Management at Brandeis University. A challenge with the project was cleaning and effectively documenting a complex data set containing information from different hierarchical levels (i.e., community, family, individual). Over the span of the research, there has been considerable movement of families across the villages in sample, making the linking of individuals to communities an important consideration. Additionally, in standardizing various biological and health measures (e.g., anthropometric measures of nutritional status), care was taken to ensure that the most recently and widely recognized references were being used.

In 2014, the TAPS team submitted a proposal to the Cultural Anthropology Program of NSF to fully curate TAPS data. The proposal was rejected, with one critique being that curation should have been part of the regular field research project. This underscores one of the ongoing challenges with making high quality data from field-based studies more broadly accessible. Our experience has shown the importance of spending the extra time needed to make these data useful to third parties unfamiliar with the research. Consequently, key issues for promoting greater data accessibility have to do with changing the incentives, and perhaps realizing that extra funds should be allocated to curating data.

4.4 | Ben Marwick, Associate Professor, Department of Anthropology, University of Washington

The types of data my research uses fall into three rough categories that might be labeled "found," "borrowed," and "collected." Each of these data types have different contexts and considerations for making data openly available. "Found" data are like found objects, or objects that are considered art although they are not normally considered as such. Found data are not normally considered research data, at least by most archaeologists. These data include word counts of journal articles for text analysis and the number of edits made to Wikipedia pages about archaeological sites. I have used these data in my published research to explore questions about digital cultural heritage and the history of archaeology. Access to these data is relatively straightforward, thanks to researcher-friendly policies from non-profit organizations like JSTOR and Wikipedia, and acquiring and using the data has few restrictions. Sharing the data derived from these sources is also uncomplicated, thanks in part to a distinctive and radical culture of openness and accessibility that is fundamental to these organizations. This is very satisfying as it simplifies the process of making the quantitative aspects of these research projects reproducible, which is important to me. Social media data, such as Twitter and Facebook posts, may also be considered as found data, but these are more complex to collect and share because social media services are private companies that protect their data with elaborate terms of service that researchers need to agree to in order to access the data. Publishing

and sharing the data derived from social media services requires thoughtful efforts to de-identify the text to protect the identities of the authors since it is typically not practical to seek permission to use text from thousands or tens of thousands of people. Furthermore, a user can switch their account from public to private at any time so we cannot assume that a public tweet will always be so.

My work with "borrowed" data refers to collaborating on more traditional archaeological science projects that combine archaeological data from multiple researchers into a final publication. Although I will directly "collect" data to contribute to the project myself, I also often work with data contributed by peer collaborators as we bring the final report to completion. Often my collaborators are comfortable with freely sharing their data, both with me and with the wider research community, and are willing for their data to be openly shared on a repository at the time of publication. Sometimes, however, for a variety of reasons, collaborators prefer not to make their data available at the time of publication. In these cases when the data do not end up being shared beyond the group of collaborators, I refer to the data as borrowed. The use of the term "borrowed" conveys the sense of a temporary use for a discrete purpose, and then when that purpose is fulfilled (e.g., the paper is published), the data are returned to the control of the researcher who collected them. Some of the reasons that my collaborators have for not sharing the data publicly are reasonable and compelling, such as the high cost of long-term storage of unusually large volumes of data, or spatial data that will reveal the location of sensitive archaeological sites. Sometimes community members and Indigenous collaborators indicate that some parts of our research data cannot be made publicly available. I have had this experience with data from human remains and rock art, and it is usually an easy request to satisfy, in part because those data do not relate to the primary objectives of my research. I also believe it is important to prioritize the relationship that makes the research possible in the first place, even if some compromises around data openness need to be made.

Although there are many good reasons for not sharing data, there are some reasons that I find not at all compelling, such as concern that another researcher will find errors in the analysis or that it is too complicated to prepare the files for others to understand. When I hear these from collaborators I will try to have a conversation to ensure I correctly understand my colleagues' concerns. I also want to share my perspective that research data are a common good rather than private property, and that part of our duty is to make it easily available to other researchers. I know the unpleasant sensation of having others find mistakes in my work, both during peer review or post-publication, and can understand the desire to avoid it. Most modern data repositories, such as Figshare, Zenodo, and OSF, provide versioning so that anyone can see earlier versions of the data and identify if changes have been made. This means corrections can be done very easily and transparently. This is important because if we do not enable the possibility for finding errors by sharing our data, then we cannot call archaeology a science because there is no opportunity for self-correcting. I think it's important that we do the best we can in our research, and accept the possibility that someone might find an unintended error in our work, and relish the opportunity to promptly

correct the record. Nevertheless, I do not force my colleagues to do something with their data that they are not comfortable with because I value relationships more than getting my way on sharing data.

I have mixed feelings about a few reasons for not sharing that I sometimes hear from collaborators. For example, I do feel sympathetic when a junior colleague, such as a PhD student or recent graduate, is concerned about gatekeeping senior researchers who might take a data set that is publicly shared by the student with the intention of finding a flaw, critiquing their work, or quickly publishing analyses that the student was planning to publish. I have experienced this with one paper that was published in a high-profile journal where the first author was a PhD student who was anxious that people might scour the stone artifact data for errors. On one hand, flaw-finding and publication of valid critique play an important role for advancing our discipline and deterring sloppy research, especially when it is done tactfully and mindful of reputational impacts. On the other hand, I wonder whether the precarity of the early career stages that students and junior colleagues are navigating requires special consideration, to insulate them from ancient scholarly vendettas and rivalries that mean that not every critique is primarily motivated by a wholesome desire to benefit the research community. Furthermore, when many well-known and senior scholars in archaeology are not in the habit of openly depositing their research data on repositories to accompany their publications, junior scholars may struggle to see value in sharing data themselves. I have found that good compromise here is an embargo arrangement, where the data are released a year after the paper has been published, and hopefully by then the predators' thirst for young blood has diminished.

As data sharing becomes more normal and widespread, I expect we'll encounter increasingly challenging questions about what and when to share and not share. I'm confident that the technical solutions to most of these challenges are already freely available to us. The hard work we must be ready to tackle to make our research data openly available is in negotiating expectations and risk, communicating effectively, maintaining collegial relationships, and finding shared priorities that are supported by open data practices.

4.5 | Connie J. Mulligan, Professor, Department of Anthropology, University of Florida

For my research, I use molecular genetic data to investigate human evolution and health. At the beginning of my career, I used genetic data to reconstruct the evolutionary histories related to peopling of the Americas and migration out of Africa. Over the past decade, my research has evolved to combine genetic and psychosocial data in order to better understand complex phenotypes like response to stress and hypertension, as well as related racial disparities. These projects all require me to explain complicated genetic research to populations around the world and ask potential participants to contribute biological samples for DNA as well as report details about their lives. My goal has always been to find common ground between my desire to pursue research and the wide range of requests and

46 | WILEY— YEARBOOK OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

requirements by the participant communities. Initially, my focus was on customizing compensation to study participants by asking community leaders and key participants how to best compensate study participants. These conversations occurred before the research began and almost always led to solutions I would never have developed on my own, such as visits by dentists and outdoor feasts of rice, beans, and ham for all members of specific indigenous communities in Panamá; English dictionaries for college students in Eritrea; genetic ancestry reports upon request for African American participants in Tallahassee; and health care for infants and small business loans to mothers in the Democratic Republic of Congo (DRC).

The key part of these collaborations was the involvement of communities and participants in conversations about the project both before and during the research. These days, I focus discussions on the samples and data that will be collected during the project and how these data can be shared with researchers who are not involved in the project. In the past, I have asked members of our advisory board (for our project in Tallahassee) or the local directors of the project (for our project in the DRC), based on their knowledge of and connections to our study participants, to approve the data I would like to share in a planned publication. My requests to share data have typically been agreed to and have often led to new discussions about future publications or initiatives. For future projects, I will also include an explicit option in the informed consent form for participants to allow their data to be shared. To share data, I used supplementary information in the early years but have generally used public repositories since the 1990s. In terms of repositories, my research group typically uses GenBank and other NCBI databases (www.ncbi.nlm.nih.gov) as the standard site for genetic data and Dataverse (dataverse.org) as a site that allows us to submit both genetic and psychosocial data. In this way, the genetic data are submitted to two different repositories in order to make the data maximally available to both geneticists and non-geneticists. When my lab shares data through these repositories, we do so in the hope that our data will be used in future publications, most likely without our knowledge prior to publication, but with no expectation that we will be included as authors on any of the publications. I believe this philosophy is justified in that the majority of my group's most highly cited papers include publicly available data. Although many people worry about being scooped by making their data publicly available, I have never had that experience. Furthermore, the people I have spoken to who were scooped had their data misused by their own collaborators who had access to the data through the collaboration and did not need to access the data through a public repository – to me, this is a not a cautionary tale on data sharing, but on how to pick your collaborators.

As a geneticist, my data originally consisted only of DNA sequence data and individual genetic variants. These data are sensitive since genetic data potentially can identify a specific individual if the data are sufficiently unique and comprehensive. As my research has moved into biocultural investigations, we have incorporated psychosocial measures such as self-reports of traumatic exposures, unfair treatment, and depressive symptoms. These data are sensitive since study participants share painful and personal experiences directly with

the interviewer and, by extension, with the entire research team. When sharing these data, my philosophy is to share as much data as is necessary to replicate the study while also protecting the identity and privacy of our study participants. The primary goal is to ensure that no single individual can be identified and connected to specific responses in the data. Typically, my research group generates data on several genes or a subset of genetic variants that are insufficient to identify specific individuals (generally genome sequences from the study participant and at least one relative are necessary to specifically identify an individual).

I had one experience in which data generated and shared by my lab was used by another lab in a new method that ended up showing we had made a mistake. The other research group used our DNA methylation (DNAm) data set from Congolese mothers, placentas, and umbilical cord blood (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE54399) to verify a method they had developed to identify maternal contamination of umbilical cord blood samples (Morin et al., 2017). In the course of testing their method, they determined that two of our umbilical cord blood samples were either contaminated with maternal venous blood or were actually maternal venous blood samples. The principal investigator contacted me prior to publication of their paper to inform me of the two contaminated or misidentified samples. My lab had already determined that one sample was mislabeled but the other sample was a surprise to us. We immediately deleted the sample from the NCBI GEO database. Even though no one enjoys being wrong, we survived and suffered no further negative repercussions. Furthermore, even though our study is clearly identified in Morin et al. (2017) as having two contaminated samples, I doubt readers paid much attention to that part of the paper because detection and correction of errors is a normal part of science.

My research group has also reused, or 'mined', molecular genetic data from public repositories, like NCBI. In 2008, we published a three-stage colonization model for peopling of the Americas based on our analysis of 77 previously published Native American DNA sequences (Kitchen et al., 2008). We proposed that a recent, rapid expansion into the Americas was preceded by a long period of population stability in greater Beringia by the proto-Amerind population after divergence from their ancestral Asian population. Soon after, Fagundes et al. (2008) published a reanalysis of those 77 DNA sequences and reported that we had incorrectly included nine non-Native American DNA sequences. When these DNA sequences were removed from the analysis, the evidence for an early (~40,000 years ago) population expansion of the proto-Amerind population disappeared. We published a correction in which we stated that "Fagundes et al. are correct in this assessment, i.e. five sequences were reclassified as Asian after their publication as Native American sequences [5] and four sequences were mistakenly included in our original study." (Mulligan et al., 2008). In our second paper, in addition to removing the nine misidentified DNA sequences, we also added more Native American sequences and reported that the long period of population stability during which New World-specific genetic variants evolved occurred after divergence from the ancestral population and before entry to the New World, thus supporting our proposal of a

three-stage colonization model. Both papers are well-cited—currently 119 citations for the first publication and 89 citations for the second publication. Even though having a publication called out for an incorrect analysis is one of a researcher's worst nightmares, I feel that we made the best of a bad situation since we were able to publish an improved and expanded analysis. Most importantly, the experience did not derail my career or that of my then-graduate student, who was first author on the publication. Open sharing of data allows the self-correction that is integral to good science.

I have also shared DNA samples that I collected with researchers outside of our research team as well as received DNA samples collected by other researchers. Sharing of samples (as opposed to data) almost always creates a collaboration in which co-authorship is promised for resulting publications. The difference between sharing samples versus data are that shared samples result in new data that are customized to your research project whereas shared data must be used as they are and may not be exactly the data you would generate if you had the ability to do so. Samples that other researchers have shared with me have resulted in completely new collaborations that now comprise my three main research projects, namely those in Tallahassee FL, the Democratic Republic of Congo, and Syria/Jordan. It is no exaggeration to say that those projects, and my current research program, would not exist without the culture of collaboration that led researchers to share their biological samples and include me in their projects. I feel that I personally have benefited, and the fields of genetics and anthropology will benefit by broad, open, and collegial sharing of data and samples.

4.6 | Trudy Turner, Professor, Department of Anthropology, University of Wisconsin-Milwaukee

My interest in data sharing derives from multiple sources—writing on ethics in biological anthropology, collecting both genetic and morphological data on wild primate populations, and serving as a Program Officer at the National Science Foundation. In the late 1990s and early 2000s, I organized a NSF-sponsored workshop on Anthropological Genetics and Ethics which led to a symposium at the AABA meetings on ethics in the field in general. The workshop and symposium led to the volume *Biological Anthropology and Ethics: From Repatriation to Genetic Identity* (Turner, 2005a), which had a commentary on data sharing (Turner, 2005b). I became much more aware of the disparities in the way subfields shared data during my tenure as a Program Officer at NSF in 2005–2006. Three years later I was approached by Dr. Nelson Freimer and Dr. Anna Jasinska at the University of California Los Angeles to participate in a project that would allow me to actualize my interest in data and resource sharing.

Nelson and Jasinska were initially interested in genetic variation in savanna monkeys (genus *Chlorocebus*) and the relationship to SIV infection. The International Vervet Research Consortium, which they led, was launching a project designed to sample broadly across the genus range as well as intensively in select populations. The project was initially funded by NIH to be a resource for future research. All materials collected would become part of the resource and would be made available to future researchers. What this meant practically was that if an animal was trapped and released to acquire a sample, that sample would be available for future research without needing to return to the wild to re-trap and re-sample vulnerable animal populations. It took nearly 2 years to ensure this project could be accomplished. Ethics clearance had to obtain from the United States and foreign universities and multiple wildlife organizations in several countries. In some countries, this involved specific permits and, in some cases, these were bioprospecting permits since cell lines were to be established to ensure a continued source of genetic materials. Community consultation was required if there were local populations who had beliefs about the animals. In all, the work was reviewed and permits were granted from IACUCs in the US, local universities or wildlife organizations, ethics committees in Africa, the Caribbean, as well as the United States Fish and Wildlife Service and CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora). Memoranda of Understanding were signed that required a return to the country of origin for discussion if the materials were to be used by anyone outside academia.

Work on the relationship among various species of *Chlorocebus* and variation in SIV infection (Coetzer et al., 2018; Svardal et al., 2017) provided the initial set of research output—the samples have also been used to examine gene sequence variation, gene expression, and functionality, and epigenetic modification. The samples have provided annotated genomic databases for comparison with other organisms (for a detailed discussion see Jasinska et al., 2013; Jasinska, 2020). In addition to work on genetics and genomics, research has focused on taxonomy, morphological, color, hormonal, and microbiome variation as well as differences in parasitic infection rates (Amato et al., 2015; Cramer et al., 2013; Fourie et al., 2015; Gaetano et al., 2014; Svardal et al., 2017; Turner et al., 2018; Turner et al., 2019).

Funding for the resource was provided through a NIH division on repositories that no longer exists. While continued funding was not available, the samples remain with UCLA.

I have learned an enormous amount from this collaboration and experience. This project was designed to provide a shared resource. It required enormous planning – both in terms of logistics and permissions—to actualize. At the end, samples were put into a resource repository and measurement data were made available through a web-based repository. Even though there are differences between resource sharing and data sharing, there are also similarities—there needs to be a place or facility to store the information. Resources require an actual physical place and data require a web-based location, yet they both require resources and maintenance. Both require permissions for collection and a way to ensure that sharing can take place without difficulty. There needs to be a responsible party to ensure ongoing ethical considerations. But most importantly, if the project is planned with the idea of sharing in mind, it will be easier to share when resources or data are available. I do not know what questions may arise about this widely distributed animal population in the future, but resources are available to address that question.

48 | WILEY—YEARBOOK OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

MULLIGAN ET AL.

5 | WHICH DATA REPOSITORY SHOULD I USE?

5.1 | Data repositories and important distinctions with respect to FAIR principles for data sharing

Our survey showed that lack of knowledge about which data repository to use is a major challenge that limits data sharing (Figure 2). Below we provide a list of recommended data repositories (Tables 11 and 12). Hesitancy about which repository to use, and more generally how to use data repositories, may be driven by uncertainty about how to share data according to FAIR principles and how openly to share data. One thing that researchers should know about "FAIRness" of data sharing is that FAIRness is mostly not about how researchers personally manage the accessibility of their data. Instead, FAIRness is primarily a quality of data repositories and one that researchers should evaluate in the data archiving and sharing platforms available to them. Furthermore, it is important to realize that no data-sharing platform can claim to maximize FAIRness for every scientific field or every kind of data. In particular, researchers with highly specialized data formats should recognize that even if a platform claims to preserve their file formats, this does not always mean the human-readability of their data will be maintained. Researchers in fields with complex data are well aware of this issue and, as a result, may be less enthusiastic about investing a lot of effort in data sharing. In other words, why share data if you know that the person who downloads it will not be able to interpret it without consulting you or your colleagues? It may be better to make the data available upon request in order to facilitate discussion and "on-boarding" with the would-be "re-user." In this way, problems of "data complexity" can be inter-dependent with lack of a "sharing culture." Of course, in the long-term, private sharing is a recipe for data loss. Instead, what is needed for many researchers is more specialized repositories or better-articulated standards that allow them to archive their complex data or describe them in a way that ensures their interpretability into the future. The challenge of data complexity may explain why the proliferation of options for data sharing has so far failed to catalyze a data-sharing revolution in certain scientific fields. In other words, we should not necessarily assume that lack of a data-sharing culture reflects cultural intransigence. It could instead reflect remaining gaps in infrastructure and standards.

Data complexity is a real concern: researchers should not be lulled into thinking that zipping up all the documents and data of their study and depositing them in a project-based data repository will necessarily ensure the reproducibility of their study or the re-usability of their data in the long-run. In this sense, generalist data repositories like Zenodo, Figshare, Dryad, or OSF usually cannot meet FAIR standards for a particular type of specialized data for a particular field of research as well as a repository that is specialized for that data type or designed for a particular field of research. The need for more specialized repositories is reflected in the plethora of genetic databases that have been created at NCBI since Genbank was created in 1982—go to ncbi.nlm.nih.gov to see the 39 databases currently available.

The notion that specialized repositories can store data more "FAIR-ly" than generalist archives is supported by our survey results. We asked respondents to list repositories in which they deposited data (Table 7) and from which they had accessed data (Table 3). In Table 7 we see that the most-used repository for depositing data is the generalist Dryad. Dryad was mentioned 82 times while second place NCBI repositories were mentioned 61 times and third place MorphoSource was mentioned 41 times. In contrast, for accessing data, NCBI and MorphoSource were essentially tied for first place with 51 and 50 mentions, respectively, and Dryad was close behind with 47 mentions. Going back to data "deposits" in Table 7, if we look at the fourth resource on the list, Figshare, it was listed 35 times, which is not too far behind MorphoSource at 41. However, when we compare MorphoSource and Figshare for access rates (Table 3), users listed MorphoSource more than three times as frequently as Figshare. These results show that researchers are more likely to deposit data in generalist repositories (probably for several reasons), but are more likely to access data from specialized repositories. Researchers are likely choosing generalist repositories because they are faster and/or easier to use than specialized repositories or because there is no specialized repository for their particular data. If researchers were rewarded professionally for data sharing via their data's impact (e.g., reuse rates), then the results of this study would be a clear incentive to seek out more specialized repositories when archiving and sharing data.

An ideal model for FAIR data sharing is one in which the researcher begins by schematically outlining the categories of data included in their study, and then seeks appropriately specialized repositories for each category. As an example, a researcher who publishes a study evaluating the phylogenetic significance of new fossil material and ancient DNA for an extinct subfossil lemur or Pleistocene hominid would likely have at least three important kinds of data, a phylogenetic character matrix used to explicitly re-assess relationships of the fossil, 3D scans of the new fossil showing the newly available morphology (perhaps structured light and/or CT scans), and ancient DNA sequences extracted from the fossil. A best practice data sharing approach might be to deposit the character matrix in MorphoBank, a repository designed to preserve phylogenetic taxon-character matrices, deposit the 3D scan data in MorphoSource, a previously mentioned repository for 3D data representing museum objects, and deposit the ancient DNA sequence in Genbank, a NCBI repository for DNA sequences from all organisms. The researcher would then cite DOIs from all three repositories in the data availability statement of their publication.

A logistical drawback of this model is the implied need for many differently specialized repositories. Funding and sustainability are difficult and it is hard to imagine dozens of organizationally independent, narrowly specialized repositories being able to last very long with small user bases reflecting the small size of the scientific subfields they support. Yet researchers are not the only ones needing deeper support for documenting data. Natural history and biodiversity collections (e.g., museums) are increasingly pushing for a model of data organization referred to as the "extended

specimen network." The core of the network is the physical specimen (or observation) itself. Data related to that specimen are made accessible through differently specialized, informatic resources assigned to either primary, secondary, or tertiary extension layers (Thiers et al., 2019), where each extension layer adds a level of abstraction. The network is formed by the linkages between these differently specialized resources, all of which contain information that can be related back to the physical specimen or observation. Assigning an informatic resource to an extension layer helps clarify its role in the network and exposes potential mechanisms of interoperation with other resources. If multiple groups of stakeholders recognize the need for improving FAIRness for a diversity of scientific fields and data types, that may improve the potential for finding solutions. Instead of envisioning many organizationally independent repositories, a conglomerate platform that has separate solutions for a variety of subfields could perhaps exist more sustainably. Such a mega-repository would need a mechanism for prioritizing, researching, building, supporting, and regularly updating the different specialties that are hosted. This would require engaging with sufficiently inclusive communities of practice and getting broad buy-in for the decisions for each field and data type. In other words, a mega-repository that was sufficiently effective in supporting a critical breadth of fields and data types with a critical depth of specialized support would be difficult to realize. However, a large generalist repository like Figshare could go beyond linking to specialized repositories and begin acquiring specialized and their data. Such an approach might be more sustainable. Phoenix Bioinformatics is a non-profit corporation that was originally started to support TAIR (The Arabidopsis Information Resource), a curated database for plant genome information (PhoenixBioinformatics, 2021). Phoenix Bioinformatics has actually begun to acquire repositories, helping them continue to support their user communities while also taking advantage of economies of scale and a common, tested business model for joint sustainability.

The other challenge to enhanced data sharing is that researchers are often uncertain about whether or how they can ethically and legally share their data because of other stakeholders. They may perceive that they are unable to share their data while respecting the ethical and legal restrictions imposed by other stakeholders. Alternatively, they may feel unqualified to judge whether they have ethical and legal obligations to other stakeholders to keep data private. This issue is another place where specialized repositories are important. Stakeholder communities are different for different data types and fields and the dimensions of sharing restrictions are accordingly different in specialized ways. This is one area where MorphoSource provides critical support. Since museums rarely allow fully open sharing of 3D data sets collected by researchers, a platform for sharing 3D data has to allow for certain kinds of restrictions on data openness if those data sets are to be made publicly accessible. On top of that, many museums differ in details of their sharing policies. Thus, MorphoSource had to carefully model the axes of variation in sharing

policies reflected by museums across the globe in order to create an adequate set of standardized tools and options. These tools allow researchers to tailor the sharing settings of each data set to the policies of the relevant museum.

5.2 | Recommended data repositories

Given the caveats listed above, we believe it is important to provide a list of recommended data repositories, which are trusted, non-profit, open-source, community-led data repositories with no or minimal fees. A trusted repository is one with a mission "to provide reliable, long-term access to managed digital resources to its designated community, now and in the future" (ResearchLibrariesGroup, 2002). The repositories listed below are a mix of generalist and specialized repositories (Table 11) and NIH molecular genetic repositories (Table 12) that maintain data sets from the biological and social sciences. All of these repositories have made arrangements to maintain all submitted files in perpetuity with no additional effort required on the part of the researcher. Access to the databases is generally unrestricted without contact or permission from the researcher required although some repositories have options for anonymized data or restricted access, as noted below. Repositories that are not recommended for scholarly research data and are not included below include Academia.edu, Research Gate, and Mendeley, which are for-profit companies that may leverage the data sets in their repositories for profit in a manner that is inconsistent with the trusted, open-source repositories listed below, and GitHub, which is an excellent platform for developing and maintaining software and annotated code, but is not intended as a digital data repository.

AUTHOR CONTRIBUTIONS

Connie J. Mulligan: Conceptualization, data curation, funding acquisition, methodology, visualization, writing – original draft, writing-review & editing. Doug M. Boyer: Conceptualization, methodology, visualization, writing – original draft, writing-review & editing. Trudy R. Turner: Conceptualization, funding acquisition, writing – original draft, writing-review & editing. Eric Delson: Conceptualization, writing – original draft, writing-review & editing. William R. Leonard: Conceptualization, writing – original draft, writing-review & editing.

ENDNOTE

[1] The American Association of Physical Anthropologists (AAPA) recently changed its name to American Association of Biological Anthropologists (AABA), and its journal similarly changed from the *American Journal of Physical Anthropology* to *American Journal of Biological Anthropology*. In order to avoid constant repetition of these points, we use AABA/AJBA throughout the article except in the formal references cited.

DATA AVAILABILITY STATEMENT

The full report of all data from the survey are available at https://ql. tc/4reI82. Raw data are available at Dryad doi:10.5061/ dryad.5x69p8d40.

ORCID

Connie J. Mulligan https://orcid.org/0000-0002-4360-2402

REFERENCES

Adams, J. W., Herries, A. I., Hemingway, J., Kegley, A. D., Kgasi, L., Hopley, P., Reade, H., Potze, S., & Thackeray, J. F. (2010). Initial fossil discoveries from Hoogland, a new Pliocene primate-bearing karstic system in Gauteng Province, South Africa. *Journal of Human Evolution*, *59*(6), 685–691. https://doi.org/10.1016/j.jhevol.2010.07.021

Alba, D. M., Daura, J., Sanz, M., Santos, E., Yagüe, A. S., Delson, E., & Zilhão, J. (2019). New macaque remains from the middle Pleistocene of Gruta da Aroeira (Almonda karst system), Portugal. *Journal of Human Evolution*, *131*, 40–47.

Alba, D. M., Delson, E., Carnevale, G., Colombero, S., Delfino, M., Giuntelli, P., Pavia, G., & Pavia, G. (2014). First joint record of Mesopithecus and cf. Macaca in the Miocene of Europe. *Journal of Human Evolution*, *67*, 1–18. https://doi.org/10.1016/j.jhevol.2013.11.001

Alba, D. M., Delson, E., Morales, J., Montoya, P., & Romero, G. (2018). Macaque remains from the early Pliocene of the Iberian Peninsula. *Journal of Human Evolution*, *123*, 141–147. https://doi.org/10.1016/j.jhevol.2018.07.005

Alba, D. M., Madurell-Malapeira, J., Delson, E., Vinuesa, V., Susanna, I., Espigares, M. P., Ros-Montoya, S., & Martínez-Navarro, B. (2016). First record of macaques from the early Pleistocene of Incarcal (NE Iberian Peninsula). *Journal of Human Evolution*, *96*, 139–144. https://doi.org/10.1016/j.jhevol.2016.05.005

Alba, D. M., Montoya, P., Pina, M., Rook, L., Abella, J., Morales, J., & Delson, E. (2015). First record of Mesopithecus (Cercopithecidae, Colobinae) from the Miocene of the Iberian Peninsula. *Journal of Human Evolution*, *88*, 1–14. https://doi.org/10.1016/j.jhevol.2015.08.003

Alba, D. M., Rodríguez-Hidalgo, A., Aouraghe, H., van der Made, J., Oujaa, A., Haddoumi, H., Saladié, P., Aissa, A. M., Marín, J., Farkouch, M., Lorenzo, C., Bengamra, S., Delson, E., Chacón, M. G., & Sala-Ramos, R. (2021). New macaque fossil remains from Morocco. *Journal of Human Evolution*, *153*, 102951. https://doi.org/10.1016/j.jhevol.2021.102951

Amato, K. R., Yeoman, C. J., Cerda, G., Schmitt, A. C., Cramer, J. D., Miller, M. E., Gomez, A., Turner, T. R., Wilson, B. A., Stumpf, R. M., Nelson, K. E., White, B. A., Knight, R., & Leigh, S. R. (2015). Variable responses of human and non-human primate gut microbiomes to a Western diet. *Microbiome*, *3*, 53. https://doi.org/10.1186/s40168-015-0120-7

Austin, C. C., Bernier, A., Bezuidenhout, L., Bicarregui, J., Biro, T., Cambon-Thomsen, A., Carroll, S. R., Cournia, Z., Dabrowski, P. W., Diallo, G., Duflot, T., Garcia, L., Gesing, S., Gonzalez-Beltran, A., Gururaj, A., Harrower, N., Lin, D., Medeiros, C., Méndez, E., ... Research Data Alliance. (2020). Fostering global data sharing: Highlighting the recommendations of the research data alliance COVID-19 working group. *Wellcome Open Research*, *5*, 267–267. https://doi.org/10.12688/wellcomeopenres.16378.2

BBCNews. (2005). Boy tracks his sperm donor father. BBC News - Health. Retrieved from http://news.bbc.co.uk/2/hi/health/4400778.stm

Berenbaum, M. R. (2021). Editorial Expression of Concern: New class of transcription factors controls flagellar assembly by recruiting RNA polymerase II in *Chlamydomonas*. *Proceedings of the National Academy of Sciences*, *118*(24), e2108930118. https://doi.org/10.1073/pnas.2108930118

Birney, E., Hudson, T. J., Green, E. D., Gunter, C., Eddy, S., Rogers, J., Harris, J. R., Ehrlich, S. D., Apweiler, R., Austin, C. P., Berglund, L., Bobrow, M., Bountra, C., Brookes, A. J., Cambon-Thomsen, A., Carter, N. P., Chisholm, R. L., Contreras, J. L., Cooke, R. M., ... Yu, J. (2009). Prepublication data sharing. *Nature*, *461*(7261), 168–170. https://doi.org/10.1038/461168a

Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics*, *52*(7), 646–654. https://doi.org/10.1038/s41588-020-0651-0

Boyer, D. M., Gunnell, G. F., Kaufman, S., & McGeary, T. (2016). Morpho-Source - Archiving and sharing digital specimen data. *The Paleontological Society Papers*, *22*, 157–181.

Boyer, D. M., Jahnke, L. M., Mulligan, C. J., & Turner, T. (2020). Response to letters to the editor concerning AJPA commentary on "data sharing in biological anthropology: Guiding principles and best practices". *American Journal of Physical Anthropology*, *172*(3), 344–346. https://doi.org/10.1002/ajpa.24065

Boyer, D. M., Kaufman, S., Gunnell, G. F., Rosenberger, A. L., & Delson, E. (2014). Managing 3D digital data sets of morphology: MorphoSource is a new project-based data archiving and distribution tool. *American Journal of Physical Anthropology*, *153*(s58), 84A.

Boyer, D. M., Winchester, J. M., & Silverton, E. (2020). MorphoSource: Creating a 3D web repository capable of complex workflows and providing novel viewing experiences. *Europeana Pro Technology*, *14*, 3D.

Coetzer, W. G., Turner, T. R., Schmitt, C. A., & Grobler, J. P. (2018). Adaptive genetic variation at three loci in south African vervet monkeys (Chlorocebus pygerythrus) and the role of selection within primates. *PeerJ*, *6*, e4953. https://doi.org/10.7717/peerj.4953

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS One*, *15*(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

Cramer, J. D., Gaetano, T., Gray, J. P., Grobler, P., Lorenz, J. G., Freimer, N. B., Schmitt, C. A., & Turner, T. R. (2013). Variation in scrotal color among widely distributed vervet monkey populations (Chlorocebus aethiops pygerythrus and Chlorocebus aethiops sabaeus). *American Journal of Primatology*, *75*(7), 752–762. https://doi.org/10.1002/ajp.22156

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). Data policies of highly-ranked social science journals. *SocArXiv, March 30*. doi:https://doi.org/10.31235/osf.io/9h7ay

Dalton, R. (2004). Anthropologists rocked by fossil access row. *Nature*, *428*(6986), 881. https://doi.org/10.1038/428881a

Davies, T. G., Rahman, I. A., Lautenschlager, S., Cunningham, J. A., Asher, R. J., Barrett, P. M., Bates, K. T., Bengtson, S., Benson, R. B. J., Boyer, D. M., Braga, J., Bright, J. A., Claessens, L. P. A. M., Cox, P. G., Dong, X. P., Evans, A. R., Falkingham, P. L., Friedman, M., Garwood, R. J., ... Donoghue, P. C. J. (2017). Open data and digital morphology. *Proceedings of the Biological Sciences*, *284*(1852), 20170194. https://doi.org/10.1098/rspb.2017.0194

Delson, E., Harcourt-Smith, W. E. H., Frost, S. R., & Norris, C. A. (2007). Databases, data access, and data sharing in Paleoanthropology: First steps. *Evolutionary Anthropology*, *16*, 161–163.

Destro Bisol, G., Anagnostou, P., Capocasa, M., Bencivelli, S., Cerroni, A., Contreras, J., Enke, N., Fantini, B., Greco, P., Heeney, C., Luzi, D., Manghi, P., Mascalzoni, D., Molloy, J., Parenti, F., Wicherts, J., & Boulton, G. (2014). Perspectives on Open Science and scientific data sharing: An interdisciplinary workshop. *Journal of Anthropological Sciences*, *92*, 179–200. https://doi.org/10.4436/jass.92006

Devriendt, T., Borry, P., & Shabani, M. (2021). Factors that influence data sharing through data sharing platforms: A qualitative study on the views and experiences of cohort holders and platform developers. *PLoS One*, *16*(7), e0254202. https://doi.org/10.1371/journal.pone.0254202

DigitalScience, Hahnel, M., McIntosh, L. D., Hyndman, A., Baynes, G., Crosas, M., Brian NosekBrian Nosek, Kathleen Shearer, Mariette van SelmMariette van Selm, Greg Goodey Nature Research. (2020). The state of open data 2020. Digital Science, Report. https://doi.org/10.6084/m9.figshare.13227875.v2

Elton, S. (2020). My new Year's resolution for 2020 is to try harder at sharing data. *American Journal of Physical Anthropology*, *172*(3), 337–338. https://doi.org/10.1002/ajpa.24071

Fagundes, N. J. R., Kanitz, R., & Bonatto, S. L. (2008). A Reevaluation of the native American MtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One*, *3*(9), e3157. https://doi.org/10.1371/journal.pone.0003157

Ferràndez-Cañadell, C., Ribot, F., & Gibert, L. (2014). New fossil teeth of Theropithecus oswaldi (Cercopithecoidea) from the early Pleistocene at Cueva Victoria (SE Spain). *Journal of Human Evolution*, *74*, 55–66. https://doi.org/10.1016/j.jhevol.2014.02.020

Fourie, N. H., Turner, T. R., Brown, J. L., Pampush, J. D., Lorenz, J. G., & Bernstein, R. M. (2015). Variation in vervet (*Chlorocebus aethiops*) hair cortisol concentrations reflects ecological disturbance by humans. *Primates*, *56*(4), 365–373. https://doi.org/10.1007/s10329-015-0486-y

Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P., & Delson, E. (2003). Cranial allometry, phylogeography, and systematics of large-bodied papionins (primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, *275*(2), 1048–1072. https://doi.org/10.1002/ar.a.10112

Gaetano, T. J., Danzy, J., Mtshali, M. S., Theron, N., Schmitt, C. A., Grobler, J. P., Freimer, N., & Turner, T. R. (2014). Mapping correlates of parasitism in wild south African vervet monkeys (*Chlorocebus aethiops*). *South African Journal of Wildlife Research*, *44*(1), 56–70.

Geller, P. L. (2020). Building nation, becoming object: The bio-politics of the Samuel G. Morton crania collection. *Historical Archaeology*, *54*, 52–70. https://doi.org/10.1007/s41636-019-00218-3

Geraads, D., & de Bonis, L. (2020). First record of Theropithecus (Cercopithecidae) from the Republic of Djibouti. *Journal of Human Evolution*, *138*, 102686. https://doi.org/10.1016/j.jhevol.2019.102686

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a

Gibbons, A. (2002). Paleoanthropology. Glasnost for hominids: Seeking access to fossils. *Science*, *297*(5586), 1464–1468. https://doi.org/10.1126/science.297.5586.1464

GIDA. (2019). *CARE principles for indigenous data governance*. Retrieved from https://www.gida-global.org/care https://static1.squarespace.com/static/5d3799de845604000199cd24/t/5da9f4479ecab221ce848fb2/1571419335217/CARE+Principles_One+Pagers+FINAL_Oct_17_2019.pdf

Gilbert, C. C., Bibi, F., Hill, A., & Beech, M. J. (2014). Early guenon from the late Miocene Baynunah formation, Abu Dhabi, with implications for cercopithecoid biogeography and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10119–10124. https://doi.org/10.1073/pnas.1323888111

Gilbert, C. C., Steininger, C. M., Kibii, J. M., & Berger, L. R. (2015). Papio Cranium from the hominin-bearing site of Malapa: Implications for the evolution of modern baboon cranial morphology and south African Plio-Pleistocene biochronology. *PLoS One*, *10*(8), e0133361. https://doi.org/10.1371/journal.pone.0133361

Godoy, R. A., & Leonard, W. R. (2020). *Tsimane' Amazonian Panel Study* (pp. 2002–2010). El Beni.

H2020Programme. (2020). *Guidelines on FAIR Data Management in Horizon 2020*.

Hipsley, C. A., & Sherratt, E. (2019). Psychology, not technology, is our biggest challenge to open digital morphology data. *Scientific Data*, *6*(1), 41. https://doi.org/10.1038/s41597-019-0047-0

Hlusko, L. J., & Mahaney, M. C. (2009). Quantitative genetics, pleiotropy, and morphological integration in the dentition of Papio hamadryas. *Evolutionary Biology*, *36*(1), 5–18. https://doi.org/10.1007/s11692-008-9048-1

Jasinska, A. J. (2020). Resources for functional genomic studies of health and development in nonhuman primates. *American Journal of Physical Anthropology*, *171*(Suppl 70), 174–194. https://doi.org/10.1002/ajpa.24051

Jasinska, A. J., Schmitt, C. A., Service, S. K., Cantor, R. M., Dewar, K., Jentsch, J. D., Kaplan, J. R., Turner, T. R., Warren, W. C., Weinstock, G. M., Woods, R. P., & Freimer, N. B. (2013). Systems biology of the vervet monkey. *ILAR Journal*, *54*(2), 122–143. https://doi.org/10.1093/ilar/ilt049

Kaiser, J. (2021). 200,000 whole genomes made available for biomedical studies. *Science*, *374*, 1036.

Kitchen, A., Miyamoto, M. M., & Mulligan, C. J. (2008). A three-stage colonization model for the peopling of the Americas. *PLoS One*, *3*(2), e1596. https://doi.org/10.1371/journal.pone.0001596

Koufos, G. D. (2019). Late Turolian Mesopithecus (Mammalia: Cercopithecidae) from Axios Valley (Macedonia, Greece): Earliest presence of M. monspessulanus in Europe. *Comptes Rendus Palevol*, *18*, 1057–1072.

Koufos, G. D., de Bonis, L., & Kugiumtzis, D. (2016). New material of the hominoid Ouranopithecus macedoniensis from the late Miocene of the Axios Valley (Macedonia, Greece) with some remarks on its sexual dimorphism. *Folia Primatol (Basel)*, *87*(2), 94–122. https://doi.org/10.1159/000446433

Langille, M. G. I., Ravel, J., & Fricke, W. F. (2018). "available upon request": Not good enough for microbiome data! *Microbiome*, *6*(1), 8. https://doi.org/10.1186/s40168-017-0394-z

Leigh, S. R. (2020). Comment on data sharing in biological anthropology. *American Journal of Physical Anthropology*, *172*(3), 339. https://doi.org/10.1002/ajpa.24073

Leonard, W. R., & Godoy, R. (2008). Tsimane' Amazonian panel study (TAPS): The first 5 years (2002-2006) of socioeconomic, demographic, and anthropometric data available to the public. *Economics and Human Biology*, *6*(2), 299–301. https://doi.org/10.1016/j.ehb.2008.04.001

Leonard, W. R., Reyes-García, V., Tanner, S., Rosinger, A., Schultz, A., Vadez, V., Zhang, R., & Godoy, R. (2015). The Tsimane' Amazonian panel study (TAPS): Nine years (2002-2010) of annual data available to the public. *Economics and Human Biology*, *19*, 51–61. https://doi.org/10.1016/j.ehb.2015.07.004

Lipphardt, V., Surdu, M., Ellebrecht, N., Pfaffelhuber, P., Wienroth, M., & Rappold, G. A. (2021). Europe's Roma people are vulnerable to poor practice in genetics. *Nature*, *599*, 368–371.

Marwick, B., & Birch, S. E. P. (2018). A standard for the scholarly citation of archaeological data as an incentive to data sharing. *Advances in Archaeological Practice*, *6*(2), 125–143. https://doi.org/10.1017/aap.2018.3

Mc Cartney, A. M., Anderson, J., Liggins, L., Hudson, M. L., Anderson, M. Z., TeAika, B., Geary, J., Cook-Deegan, R., Patel, H. R., & Phillippy, A. M. (2022). Balancing openness with indigenous data sovereignty: An opportunity to leave no one behind in the journey to sequence all of life. *Proceedings of the National Academy of Sciences*, *119*(4), e2115860119. https://doi.org/10.1073/pnas.2115860119

McDade, T. W. (2020). Letter to the editor in support of data sharing, with caveats. *American Journal of Physical Anthropology*, 172(3), 340. https://doi.org/10.1002/ajpa.24072

Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. *Nature*, 578, 491.

Morin, A. M., Gatev, E., McEwen, L. M., MacIsaac, J. L., Lin, D. T. S., Koen, N., Czamara, D., Räikkönen, K., Zar, H. J., Koenen, K., Stein, D. J., Kobor, M. S., & Jones, M. J. (2017). Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. *Clinical Epigenetics*, 9(1), 75. https://doi.org/10.1186/s13148-017-0370-2

Morton, S. G. (1849). *Catalogue of skulls ofMan and the inferior animals, in the collection of Samuel George Morton*. Merrihew & Thompson.

Mulligan, C. J., Kitchen, A., & Miyamoto, M. M. (2008). Updated three-stage model for the peopling of the Americas. *PLoS One*, 3(9), e3199.

NCBI. (2021, 04/26/2021). GenBank overview.

NHGRI. (2020). The Human Genome Project.

Nicholas, G. P. (2014). Indigenous cultural heritage in the age of technological reproducibility: Towards a postcolonial ethic of the public domain. In R. Coombe, D. Wershler, & M. Zeilinger (Eds.), *Dynamic fair dealing: Creating Canadian culture online* (pp. 213–224). University of Toronto Press.

Normile, D. (2021). Genetics papers from China face ethical scrutiny. *Science*, 373, 727–728.

NSF. (2020). Proposal & Award Policies & Procedures Guide (Section D.4. a). *Chapter XI - Other post award requirements and considerations*. Retrieved from https://www.nsf.gov/bfa/dias/policy/dmp.jsp

O'Doherty, K. C., Shabani, M., Dove, E. S., Bentzen, H. B., Borry, P., Burgess, M. M., Chalmers, D., de Vries, J., Eckstein, L., Fullerton, S. M., Juengst, E., Kato, K., Kaye, J., Knoppers, B. M., Koenig, B. A., Manson, S. M., McGrail, K. M., McGuire, A. L., Meslin, E. M., … Burke, W. (2021). Toward better governance of human genomic data. *Nature Genetics*, 53(1), 2–8. https://doi.org/10.1038/s41588-020-00742-6

Offord, C. (2021). Researcher sanctioned by PNAS for not sharing alga. *The Scientist*.

Parr, C. S., & Cummings, M. P. (2005). Data sharing in ecology and evolution. *Trends in Ecology & Evolution*, 20(7), 362–363. https://doi.org/10.1016/j.tree.2005.04.023

Pauli, J. N., Newsome, S. D., Cook, J. A., Harrod, C., Steffan, S. A., Baker, C. J. O., Ben-David, M., Bloom, D., Bowen, G. J., Cerling, T. E., Cicero, C., Cook, C., Dohm, M., Dharampal, P. S., Graves, G., Gropp, R., Hobson, K. A., Jordan, C., MacFadden, B., … Hayden, B. (2017). Opinion: Why we need a centralized repository for isotopic data. *Proceedings of the National Academy of Sciences*, 114(12), 2997–3001. https://doi.org/10.1073/pnas.1701742114

Phillips, M., Molnár-Gábor, F., Korbel, J. O., Thorogood, A., Joly, Y., Chalmers, D., Townend, D., & Knoppers, B. M. (2020). Genomics: Data sharing needs an international code of conduct. *Nature*, 578, 31–33.

PhoenixBioinformatics. (2021). Phoenix Bioinformatics. Supporting science is our mission. Retrieved from https://www.phoenixbioinformatics.org/

Pienta, A. M., Alter, G., & Lyle, J. (2010). The enduring value of social science research: The use and reuse of primary research data. *Electronic document*, http://deepblue.lib.umich.edu/handle/2027.42/78307.

Plis, S., & Calhoun, V. (2021). Opinion: Toward better data sharing. The Scientist, March.

Pugh, K. D., & Gilbert, C. C. (2018). Phylogenetic relationships of living and fossil African papionins: Combined evidence from morphology and molecules. *Journal of Human Evolution*, 123, 35–51. https://doi.org/10.1016/j.jhevol.2018.06.002

Radović, P., Lindal, J., Marković, Z., Alaburić, S., & Roksandic, M. (2019). First record of a fossil monkey (primates, Cercopithecidae) from the late Pliocene of Serbia. *Journal of Human Evolution*, 137, 102681. https://doi.org/10.1016/j.jhevol.2019.102681

Rausher, M. D., McPeek, M. A., Moore, A. J., Rieseberg, L., & Whitlock, M. C. (2010). Data archiving. *Evolution*, 64(3), 603–604. https://doi.org/10.1111/j.1558-5646.2009.00940.x

ResearchLibrariesGroup. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Retrieved from Mountain View, CA: www.rlg.org

Rossie, J. B., Gilbert, C. C., & Hill, A. (2013). Early cercopithecid monkeys from the Tugen Hills, Kenya. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5818–5822. https://doi.org/10.1073/pnas.1213691110

Singleton, M., Gilbert, C. C., Frost, S. R., & Seitelman, B. C. (2016). Comparative morphometric analysis of a juvenile papionin (primates: Cercopithecidae) from Kromdraai a. *Annals of the Ditsong National Museum of Natural History*, 6, 1–17.

Svardal, H., Jasinska, A. J., Apetrei, C., Coppola, G., Huang, Y., Schmitt, C. A., Jacquelin, B., Ramensky, V., Müller-Trutwin, M., Antonio, M., Weinstock, G., Grobler, J. P., Dewar, K., Wilson, R. K., Turner, T. R., Warren, W. C., Freimer, N. B., & Nordborg, M. (2017). Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nature Genetics*, 49(12), 1705–1713. https://doi.org/10.1038/ng.3980

Takai, M., Thaung-Htike, Zin-Maung-Maung-Thein, Soe, A. N., Maung, M., Tsubamoto, T., Egi, N., Nishimura, T. D., & Nishioka, Y. (2015). First discovery of colobine fossils from the late Miocene/early Pliocene in Central Myanmar. *Journal of Human Evolution*, 84, 1–15. https://doi.org/10.1016/j.jhevol.2015.04.003

Tattersall, I., & Schwartz, J. H. (2002). Is paleoanthropology science? Naming new fossils and control of access to them. *The Anatomical Record*, 269(6), 239–241. https://doi.org/10.1002/ar.10187

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 192. https://doi.org/10.1038/s41597-021-00981-0

Thiers, B., Monfils, A., Zaspel, J., Ellwood, E., Bentley, A., Levan, K., Bates, J., Jennings, D., Contreras, D., Lagomarsino, L., Mabee, P., Ford, L., Guralnick, R., Gropp, R., Revelez, M., Cobb, N., Lendemer, J., Seltmann, K. & Aime, M. C. (2019). Extending U.S. Biodiversity Collections to Promote Research and Education. Retrieved from https://bcon.aibs.org/2019/04/04/bcon-report-extending-u-s-biodiversity-collections-to-promote-research-and-education/

Thomas, S. B., & Quinn, S. C. (1991). The Tuskegee syphilis study, 1932 to 1972: Implications for HIV education and AIDS risk education programs in the black community. *American Journal of Public Health*, 81(11), 1498–1505. https://doi.org/10.2105/ajph.81.11.1498

Tran, L. A. (2014). The role of ecological opportunity in shaping disparate diversification trajectories in a bicontinental primate radiation. *Proceedings of the Biological Sciences*, 281(1781), 20131979. https://doi.org/10.1098/rspb.2013.1979

Tsosie, K. S., Bader, A. C., Fox, K., Bolnick, D. A., Garrison, N. A., & Smith, R. W. A. (2021). Ancient-DNA researchers write their own rules. *Nature*, 600(7887), 37. https://doi.org/10.1038/d41586-021-03542-y

Turner, T. R. (2005a). *Biological Anthropology and ethics: From repatriation to genetic identity*. State University of New York Press.

Turner, T. R. (2005b). Commentary: Data access and sharing of information. In T. R. Turner (Ed.), *Biological Anthropology and ethics: From repatriation to genetic identity*. State University of New York Press.

Turner, T. R., & Mulligan, C. J. (2019). Data sharing in biological anthropology: Guiding principles and best practices. *American Journal of Physical Anthropology*, 170(1), 3–4. https://doi.org/10.1002/ajpa.23909

Turner, T. R., Schmitt, C. A., & Cramer, J. D. (2019). *Savanna monkeys: The genus Chlorocebus*. Cambridge University Press.

Turner, T. R., Schmitt, C. A., Cramer, J. D., Lorenz, J., Grobler, J. P., Jolly, C. J., & Freimer, N. B. (2018). Morphological variation in the genus Chlorocebus: Ecogeographic and anthropogenically mediated variation in body mass, postcranial morphology, and growth. *American Journal of Physical Anthropology*, 166(3), 682–707. https://doi.org/10.1002/ajpa.23459

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J. S., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, 27(4), 1304–1308. https://doi.org/10.1096/fj.12-218164

Vitelli, K. D., & Colwell-Chanthaphonh, J. S. (2006). *Archaeological ethics*. Rowan AltaMira.

Wade, L. (2021). The ghosts in the museum. *Science*, 373(6551), 148–152. https://doi.org/10.1126/science.373.6551.148

Wagner, J. K. (2020). Concerning data sharing in biological anthropology. *American Journal of Physical Anthropology*, 172(3), 341–343. https://doi.org/10.1002/ajpa.24074

Wagner, J. K., Colwell, C., Claw, K. G., Stone, A. C., Bolnick, D. A., Hawks, J., Brothers, K. B., & Garrison, N.'. A. (2020). Fostering responsible research on ancient DNA. *American Journal of Human Genetics*, 107(2), 183–195. https://doi.org/10.1016/j.ajhg.2020.06.017

Weber, G. W. (2001). Virtual anthropology (VA): A call for glasnost in paleoanthropology. *The Anatomical Record*, 265(4), 193–201. https://doi.org/10.1002/ar.1153

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One*, 6(11), e26828. https://doi.org/10.1371/journal.pone.0026828

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wood, B. A. (1991). *Koobi fora research project. Volume 4: Hominid cranial remains (Vol. 4)*. Oxford University Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mulligan, C. J., Boyer, D. M., Turner, T. R., Delson, E., & Leonard, W. R. (2022). Data sharing in biological anthropology. *Yearbook Biological Anthropology*, 178(Suppl. 74), 26–53. https://doi.org/10.1002/ajpa.24499