A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing

Ben Marwick and Suzanne E. Pilaar Birch

Data are the building blocks of research, and access to data is necessary to fully understand and extend the work of others (Fienberg, Martin, and Straf 1985). Yet archaeological data are often not available for a variety of reasons, including a perceived lack of credit for making data available. The goal of this article is to present a standard for the scholarly citation of archaeological data, akin to the way we cite publications. The need for this has been motivated by Huggett's (2017) findings of low levels of data citation and reuse in archaeology and Clarke's (2015) analysis that digital data preservation and management systems are underdeveloped in archaeology, despite our obligations as stewards of cultural heritage. A standard for data citation is important because it gives credit to the authors or compilers of the data and so accords them visibility and recognition for their scholarly contribution in a way that is valued broadly in the discipline. A citation standard is also important for tracking the use of data and generating metrics about the

ABSTRACT

How do archaeologists share their research data, if at all? We review what data are, according to current influential definitions, and previous work on the benefits, costs, and norms of data sharing in the sciences broadly. To understand data sharing in archaeology, we present the results of three pilot studies: requests for data by e-mail, review of data availability in published articles, and analysis of archaeological datasets deposited in repositories. We find that archaeologists are often willing to share but that discipline-wide sharing is patchy and ad hoc. Legislation and mandates are effective at increasing data sharing, but editorial policies at journals lack adequate enforcement. Although most of the data available at repositories are licensed to enable flexible reuse, only a small proportion of the data are stored in structured formats for easy reuse. We present some suggestions for improving the state of date sharing in archaeology; among these is a standard for citing datasets to ensure that researchers making their data publicly available receive appropriate credit.

¿Cómo comparten los arqueólogos sus datos de investigación, si lo hacen? Revisamos qué datos son, de acuerdo con las actuales definiciones influyentes, y el trabajo previo sobre los beneficios, costos y estándares de intercambio de datos en las ciencias en general. Para comprender el intercambio de datos en arqueología, presentamos los resultados de tres estudios piloto: solicitudes de datos por correo electrónico; revisión de la disponibilidad de datos en artículos publicados y análisis de conjuntos de datos arqueológicos depositados en repositorios. Encontramos que los arqueólogos están a menudo dispuestos a compartir, pero el intercambio de toda la disciplina es desigual y ad hoc. La legislación y los mandatos son eficaces para aumentar el intercambio de datos, pero las políticas editoriales de las revistas carecen de aplicación adecuada. Aunque la mayoría de los datos disponibles en repositorios están autorizados para permitir la reutilización flexible, sólo una pequeña proporción de los datos se almacena en formatos estructurados para su fácil reutilización. Presentamos algunas sugerencias para mejorar el estado de la fecha de intercambio en arqueología, incluyendo un estándar para citar conjuntos de datos para asegurar que los investigadores que hacen sus datos disponibles públicamente reciben crédito apropiado.

Advances in Archaeological Practice 6(2), 2018, pp. 125–143

DOI:10.1017/aap.2018.3

impact of a dataset on the discipline. Understanding the use of resources such as datasets is important for allocating resources within archaeology and identifying novel research directions.

Archaeologists work with data in research, commercial, government, and other contexts. A full investigation of data in each of these contexts is beyond the scope of this article, so we limit this study to what we are most familiar with, the research context. However, many of the practical details we discuss here will also be relevant to archaeologists working in other contexts. We begin by defining what we mean by "data" and reviewing previous work on researchers' attitudes to sharing data. We motivate data sharing and citing behaviors by explaining why data sharing is beneficial to archaeology as a discipline and to individual researchers. Then we survey some of the data sharing and citing norms that have emerged in other research areas. From these norms we derive a standard suitable for the citation of archaeological data. Finally, we present some basic instructions on how to share archaeological data.

WHAT ARE DATA?

Defining what data are is an ongoing research problem for information scientists (Borgman 2012; Leonelli 2016). Table 1 summarizes some of the definitions debated in the philosophy of science literature. A more concrete starting point is this definition from the National Academies of Sciences:

> The term "data" as used in this document is meant to be broadly inclusive. In addition to digital manifestations of literature (including text, sound, still images, moving images, models, games, or simulations), it refers as well to forms of data and databases that generally require the assistance of computational machinery and software in order to be useful, such as various types of laboratory data including spectrographic, genomic sequencing, and electron microscopy data; observational data, such as remote sensing, geospatial, and socioeconomic data; and other forms of data either generated or compiled, by humans or machines [Uhlir and Cohen 2011, cited in Borgman 2015 p. 1061].

This is a useful definition because it emphasizes data as digital manifestations and so suitable for transmission, storage, and retrieval using the internet. Physical objects (such as samples, collections, hard copies of notes, etc.) are data also, but they are out of the scope of this discussion because of the different challenges and requirements for access, archiving, and so on. The categories shown in Table 2 are useful as a rough guide to what data can be, although this table is not comprehensive and does not represent the full diversity of data used by archaeologists. Ultimately, data are whatever the community decides, as the National Science Foundation notes in its commentary on data management plans: "'Data' are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. This includes original data, but also 'metadata' (e.g. experimental protocols, code written for statistical analyses, etc.)" (2010).

TABLE 1. Selection of Definitions of Research Data.

Definition of Data	Source
Any portable thing that is treated as potential evidence for one or more claims about phenomena	Leonelli 2015
A lack of uniformity in the real world	Floridi 2011
Traces or signals that have been manipulated so that they can be stored and retrieved	Rheinberger 2011
Anchors of knowledge claims that are stable across different contexts	Latour 1999
Marks or traces produced by human interactions with research instruments	Hacking 1992
Evidence for the identification and characterization of phenomena	Bogen and Woodward 1988

TABLE 2. Selection of Types of Data Encountered by Archaeologists.

Type of Data	Description
Observational	Includes site descriptions and artifact measurements, which may be associated with specific places and times, or multiple places and times, as well as ethnoarchaeological observations
Computational	Includes the results of simulations and computer models, such as agent-based models or virtual reality models, hypothesis tests, and regression models
Experimental	Includes results from laboratory procedures, such as instrumental analyses and chemical reactions to determine artifact composition, or field experiments such as knapping, trampling, and taphonomic studies
Records	Includes government, business, public, and private documents

These definitions in policy are consistent with the relational account of data in philosophy of science scholarship, which holds that what data are is dependent on who uses it, how, and for what purpose (Daston and Galison 1992; Leonelli 2015). How can these definitions and philosophical discussions inform our practice? They tell us that what count as data are emergent from our professional interactions. Data are the materials we request when we do peer review of manuscripts and grant proposals or when we make an informal request for more information by e-mail. They are not exclusively objective archaeometric records but, rather, include interpretive and derivative results. In short, data are anything another person needs to understand our work better.

To translate this into a more applied context, our practice of data sharing is to make available the computer files (e.g., containing text, raw structured data, images, etc.) that we analyze or otherwise use to create the figures, tables, and quantitative results

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

that we include in our reports and publications. These include our derived results and secondary data, such as taxonomic and typological classifications. This selection omits materials that could also be considered data according to the definitions noted above (e.g., field notes), but we feel that our selection represents the most information-dense and meaningful items that are useful for our community of interest. These are also the items that have been most frequently requested from us by others in our professional interactions. Our citation standard is inclusive and imposes no constraints on what data can be, but we propose a few minimal qualities for how data should be organized that will help ground our discussion.

WHAT ARE THE ETHICS OF DATA SHARING?

A full treatment of the ethical issues of data sharing is beyond the scope of this article, but we can highlight some points specifically relevant to the question of when *not* to share archaeological data. Much of the discussion on ethics and data sharing generally relates to appropriate attribution (assigning credit for the data) in data reuse (Duke and Porter 2013). The default position for archaeological data, like data from many other sciences, is that they are part of the global human patrimony and should be accessible to other scholars and the public (Vitelli and Colwell-Chanthaphonh 2006; Zimmerman et al. 2003). However, archaeologists, like other researchers working with local and indigenous communities, governments, and international collaborators, generate research data that are intellectual property with many stakeholders. With diverse stakeholders, archaeological data can become entangled in economics, nationalism, cultural politics, and identity (Nicholas and Wylie 2009). Often, one of these stakeholders is an indigenous community, and archaeologists may find differences between Western and indigenous conceptions of what data are and how they may be generated, used, shared, and "owned" (Harding et al. 2012).

One of these differences is that Western researchers often have individual autonomy in sharing data, and once data are shared, they are free for all to use. In contrast, within many indigenous communities, data are part of the group's identity and property, akin to tangible heritage (Nicholas 2014), with certain members entrusted to keep that data on behalf of the group, and sometimes it may be inappropriate for these people to share data with other members of the group or beyond the group (Tsosie 2007). Differences between Western and indigenous concepts of data are not only cultural—substantial inequalities exist in access to resources for converting archaeological data into economic and political benefit (e.g., graduate degrees, employment, publications, grants, promotions). On one hand, making archaeological data openly available benefits indigenous communities because it removes some of the financial and technical barriers to accessing that data (e.g., subscription fees to journals). On the other hand, this may be little more than a useless courtesy (Hymes 1972), because simply making data openly available does not give indigenous people access to the contexts where data can be used to generate economic and political benefits. Thus, open data have different values for Western and indigenous communities, and if archaeologists want to make a compelling case to their indigenous partners for data sharing, they should also make a commitment to reducing some of the inequalities in making use of the data, for example, by providing some training in data analysis or adapting the data to a context that is valued by the indigenous community.

In short, there is more to the ethics of data sharing than avoiding risk of damage. Redacting the locations of archaeological sites to prevent looting is an obvious case of preparing data for sharing to reduce the risk of damage to the sites (and in some jurisdictions, mandated by law). However, because of cultural, economic, and political differences between Western archaeologists and the communities they work with, archaeologists cannot consider their own risk assessment to be comprehensive and exhaustive (Nicholas and Wylie 2012). It is crucial for archaeologists to discuss data sharing plans during negotiations with representatives authorized by the indigenous peoples whose cultural heritage is the subject of investigation. This negotiation process is important because only indigenous peoples themselves can identify potential adverse outcomes to data sharing, and they can do this only if they understand the proposed research and anticipated results. The concrete outcome of this process is often a memorandum of understanding between the indigenous community and the archaeologist's employer that documents the outcome of negotiations about data sharing. If an archaeologist's indigenous partners have reasons not to share data in public, then the archaeologist should respect that position and note in publication that the data are restricted at the request of the indigenous collaborators.

WHAT ARE THE BENEFITS OF DATA SHARING TO THE INDIVIDUAL RESEARCHER?

While public and community benefits and costs are well rehearsed and widely agreed on (Tables 3 and 4), they offer no, or only weak, tangible incentives to the individual researcher (Rowhani-Farid, Allen, and Barnett 2017). Previous work has identified citation advantages as a tangible benefit to individual researchers who share data with publications, although the magnitude of the advantage varies greatly (see Table 5 for a summary). Increased publication productivity also appeared to be associated with data sharing behaviors. Across more than 7,000 National Science Foundation and National Institute of the Humanities awards, Pienta, Alter, and Lyle (2010) report that research projects with archived data produced a median of 10 publications, vs. only five for projects without archived data.

WHAT DO RESEARCHERS THINK ABOUT SHARING DATA?

Sharing of data is a quality that is fundamental to the definition of science (cf. Merton's [1942] communalism norm, where intellectual property rights are normally given up in exchange for recognition and esteem) and to data both as an abstract quality (cf. the definitions in Table 1 that mention data as "portable," "stable across different domains") and as a concrete detail to be managed with policy. For example, for over a decade the

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

TABLE 3. Summary of Public and Community Costs of Data Sharing.

#	Reasons Not to Share
1.	The time it takes to clean up and document data for release
2.	The possibility that your data may be used without citation
3.	Legal barriers, such as copyright
4.	Time to verify privacy or other administrative data concerns
5.	The potential loss of future publications using these data
6.	Competitors may get an advantage
7.	Dealing with questions from users about the data
8.	Technical limitations, i.e., Web platform space constraints
9.	Intense competition in the topic
10.	Investment of large amount of work building the dataset
11.	Insufficient perceived reward, such as promotion or subsequent citation
12.	Effort in documenting
13.	Concerns for priority, including control of results and sources
14.	Intellectual property issues

Source: From Borgman 2012; Stodden 2010.

TABLE 4. Summary of Public and Community Benefits of Data Sharing.

#	Reasons to Share
1.	Encouraging scientific advancement
2.	Being a good community member
3.	Potential to encourage others to work on the problem
4.	Encouraging sharing and having others share with you
5.	The potential to set a standard for the field
6.	Improvement in the caliber of research
7.	Increase in publicity, track metrics of impact
8.	Opportunity to get feedback on your work
9.	Potential for finding collaborators
10.	Normalizing understanding in a field
11.	To reproduce or to verify research
12.	To make the results of publicly funded research available to the public
13.	To enable others to ask new questions of extant data
14.	To advance the state of research and innovation

Source: From Borgman 2012; Stodden 2010.

National Science Foundation has had a data sharing policy that states:

> 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 [2016].

TABLE 5. Summary of Individual Benefits to Data Sharing.

Research Area	Citation Advantage (%)	Source
Genetics	9	Piwowar and Vision 2013
Astronomy	20	Henneken and Accomazzi 2011
Astrophysics	28	Dorch et al. 2015
Paleoceanography	35	Sears 2011
Peace research	100	Gleditsch et al. 2003

Surveys of researchers across different domains repeatedly show that people value the ideal of data sharing (Kratz and Strasser 2014, 2015). However, data sharing is not widespread in many disciplines, raising questions about why researchers do or do not share their data. Three recent studies have attempted to answer these questions, in the search for strategies to encourage sharing.

Stodden (2010) surveyed registrants at a top machine learning conference (the Neural Information Processing Systems conference) to identify the factors that affect researchers' decisions to reveal data (Table 4). Stodden found that communitarian factors influence scientists to share their data but private incentives determine when scientists choose not to share their code and data. In the top three reasons not to share, we see a lack of incentives to invest time in data sharing (however, there are many other time-consuming tasks in research, and they get done because clear reward structures are in place), a lack of faith in data citation (which we hope this article will help with), and a lack of an efficient copyright process. Stodden's (2009a, 2009b) "Reproducible Research Standard," summarized in Table 6, is a convenient and robust copyright schema that is in wide use for many typical research projects.

Other surveys have found similar results. Tenopir and colleagues (2011) received 1,329 responses to a survey conducted by the research team of the National Science Foundation–funded DataONE project. They found that when researchers did not share data it was due to reasons that included insufficient time (54%), a lack of funding (40%), having no place to put the data (24%), a lack of standards (20%), and the fact that their sponsor did not require it (17%). A remarkable observation by Tenopir and colleagues is that while 75% of respondents said that they shared data, only 36% of those respondents said that their data are easy to access, while only 6% said that they make all their data available. This suggests that most data sharing occurs in private directly between individual researchers, rather than in public via open repositories. A similar focus on private, peer-to-peer sharing is evident in interviews conducted by Wallis and colleagues (2013) with 43 researchers at the Center for Embedded Networked Sensing. They found that the most commonly cited conditions for sharing data were (1) first rights to publish results, (2) proper attribution to the data source, (3) familiarity between sharer and recipient, (4) funding agency expectations, and (5) the amount of effort required to share.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

TABLE 6. Summary of Stodden's "Reproducible Research Standard."

Content	Examples	License(s)	Comments
Media	Texts of reports, journal article manuscripts, figures, tables	CC-BY	Creative Commons Attribution: License permits free reuse only with appropriate attribution to give credit to the original author. Allows commercial reuse (e.g., in cultural resource management reports). Author permission not required for reuse.
Code	R or Python scripts, packages, NetLogo codes	MIT, BSD, GPL	Berkeley Software Distribution, GNU General Public License: License permits free reuse with minimal restrictions on the use and redistribution of code. Gives no warranty to other users of the code. Allows the code to be incorporated into proprietary products (e.g., cultural resource management software). Author permission not required for reuse.
Data	Spreadsheets, database files, geographic information systems files, images, plain text files	CC-0	Public Domain: License places data in the public domain by waiving all of the author's rights to the work worldwide under copyright law. Author permission not required for reuse.

Note: This standard realigns legal rights with scientific norms by freeing scientific work from copying and reuse restrictions, with attribution. It encourages researchers to release their products with standardized instructions for reuse and attribution.

Source: Stodden 2009a.

WHAT ARE THE CURRENT NORMS OF DATA SHARING?

The surveys above show that data sharing practices tend to be a private activity, but where public sharing does occur, what forms does it take, and how do these vary across different fields of research? Wouters and Haak (2017) surveyed 1,200 researchers about data sharing behaviors and found a spectrum from "intensive data-sharing fields" to "restricted data-sharing fields." Intensive data-sharing fields often depend on access to global databases, for example, human genetics, while restricted data-sharing fields depend mostly on data collected by the individual researcher or his or her immediate collaborators, such as digital humanities. These differences in disciplinary norms have implications for data ownership (intensive: collectively owned, willing to share; restrictive: owned by the individual, reluctant to share) and the effort required to share (intensive: low effort because data are prepared for sharing as an integral part of the research process; restrictive: high effort because data are prepared for sharing after the research, or apart from it, in response to a request from another researcher).

This intensive-restricted spectrum also maps onto Sawyer's (2008) analysis of data sharing in data-rich fields compared with data-poor fields. Data-rich fields, such as astronomy, oceanography, and ecology, have three common characteristics: Pooling and sharing data resources is expected; the cumulative nature of work that builds on common data and common methods reduces the likelihood of alternative data sources and other methods being widespread; and the number of theoretical choices is limited because there is enough data to adequately test competing theories. Data-poor fields have three common characteristics: Data are rare, or hard to get, or both; the types of data available often dictate the methods used; and theories are relatively easy to generate and more difficult to validate. In data-poor fields, researchers often invest substantial time and resources on gathering their own data. The accumulation of data, and the quality of the data collected, is often a key determinant of the status of a researcher, leading to hoarding of data as an asset to be traded for professional benefit.

Archaeologists can be found all over this spectrum because of the diversity of archaeological research. However, our observations of data sharing in archaeology, and the nature of archaeological research generally, indicate that it is, for the most part, currently a restricted data-sharing and data-poor field. There are few standardized and widely used methods for sharing and citing data, a challenge for archaeological data sharing that we hope this article will help to address.

Even though there is a wide spectrum of data sharing behaviors, overall rates of data sharing by individual researchers are very low in many fields. Wicherts and colleagues (2006) e-mailed authors of 141 empirical articles recently published by American Psychological Association (APA) journals to request a copy of the original data reported in the article. They received 64 of the 249 datasets they requested, resulting in a data sharing rate of about 26%, despite all the authors of the articles in their study being bound by the APA's Ethical Principles, which stated that authors must "not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims" (APA 2001:396). Vanpaemel and colleagues (2015) similarly requested data for 394 articles, published in all issues of four APA journals in 2012, and found that 38% of authors responded with their data. They conclude that although the data sharing rate increased as compared with the study by Wicherts and colleagues, their findings are "worrisome" and indicate a "poor availability of data" (Vanpaemel et al. 2015: 3).

Savage and Vickers (2009) requested data from the authors of 10 papers published in PLOS journals and received only one reply. Although this is a very small sample, it is notable because PLOS journals require authors to share data as a condition of publi-

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

cation: "PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception" (*PLOS* Editorial and Publishing Policies 2017); and yet only one out of the 10 authors actually complied with this requirement. Alsheikh-Ali and colleagues (2011) inspected 500 articles published in high-impact science journals in 2009 and found that only 47 essays (9%) deposited full primary raw data online. Similarly, Womack (2015) inspected 4,370 articles published in high-impact science journals in 2014 and found that 13% of articles with original data published had made the data available to others.

The absence of a standardized and sustainable approach to data sharing is starkly demonstrated by Vines and colleagues (2014) study of data availability over time in biology publications. They examined the availability of data from 516 studies of plants or animal morphology between two and 22 years old. Vines et al. found that the odds of a dataset referred to in a publication being available fell by 17% per year. They found that the key factors responsible for the unavailability of data were that the data were lost or on inaccessible storage media.

DO ARCHAEOLOGISTS SHARE DATA?

Here we present the results of three small pilot studies aimed at understanding how archaeologists share their data. First, we e-mailed a sample of authors of research publications to request access to the data behind their publication. Second, we surveyed a sample of journal articles to see what data were available. Third, we analyzed information about archaeological data files in online repositories tracked by DataCite (an organization that provides persistent identifiers, or digital object identifiers [DOIs], for research data). The small sample sizes of the first two of these pilot studies limit how generalizable the results are, but as the first of their kind for archaeology, they are valuable for proposing hypotheses for larger-scale investigations.

Reproducibility and Open-Source Materials for This Study

To enable reuse of our materials and improve reproducibility and transparency according to the principles outlined in Marwick 2017, we include the entire R code used for all the analysis and visualizations contained in this article in our compendium at http://doi.org/10.17605/OSF.IO/KSRUZ. Also in this version-controlled compendium are the raw data for all the tests reported here. All of the figures and quantitative results presented here can be independently reproduced with the code and data in this repository. In our compendium our code is released under the MIT license, our data as CC-0, and our figures as CC-BY, to enable maximum reuse (for more details, see Marwick 2017).

E-mail Requests for Data in a Published Article

Our sample focused on *Journal of Archaeological Science* articles on experiments with stone artifacts that were published during 2009 and 2015. We focused on stone artifact experiments because all research materials in these projects are generated by the researcher, so there are no other stakeholders to consider when determining how to share the data: the researcher has complete control over the data. There are also no risks of damage to people or property in sharing these data because they do not include archaeological sites or human subjects. Because we contacted authors of essays at the intersection of (1) containing eminently sharable data and (2) our bona fide research interests (i.e., stone artifacts for BM), this is a targeted, nonrandom sample that may not be representative of the entire archaeological community. In contacting the authors, we requested the datasets for use in a graduate seminar with the aim of reproducing the published results and further exploring the data (this stage is still in progress). We promised to contact the authors if we discovered anything new and not to share the data further without their permission. The data included with this article are anonymized so the specific authors we contacted are not identifiable.

We sent out e-mails to the first authors of 23 stone artifact experiment articles and received 15 replies, resulting in a 70% response rate. We received five responses that included data files, giving an overall sharing rate of 20%. Our small sample size limits the robustness of our analysis but points to some observations that may be worth following up in a larger study. We found no clear relationship of data sharing and date of publication. We do not have any insights into why authors of older essays declined to share, but three of the six responses for articles published in 2014–2015 said that they declined to share data because they intended to use these data in a future publication. Other noteworthy reasons provided by respondents for declining to share include the fact that the authors were about to defend their thesis or about to go on vacation. We found no effect of the author's student status at time of publication on data sharing behaviors, with roughly equal proportions of students and nonstudents sharing and not sharing (Figure 1).

Similarly there are no clear patterns of data sharing among different journals. It is remarkable that the authors of the *PLOS ONE* article we contacted declined to share because that refusal violates the policy of that journal at the time of publication of the essay (early 2014). Our finding is not unique—Chambers (2017) examined 50 *PLOS ONE* articles on brain images published in November 2014–May 2015 and found that only 38% of the essays' authors had archived their data in an open repository. Similarly, Nuijten and colleagues (2017) found that of 462 *PLOS ONE* articles that promised data availability, 29% of these did not have any available data. Unfulfilled promises to share data are not unique to *PLOS ONE* (Kidwell et al. 2016), but these studies highlight the need to resolve ambiguity among authors about what count as data and for editors and peer reviewers to more critically assess whether authors have made their data openly available.

We see similar rates of sharing in response to e-mail requests among archaeologists to what have been described in other disciplines (Table 7; Tenopir et al. 2015; Vanpaemel et al. 2015). However, our results may overestimate the true rate because our sample did not include datasets that authors might not be able or willing to share because they contain sensitive location or cultural information. Analysis of the qualitative responses we received suggests two key reasons why authors are reluctant to share: they are fearful that they will lose the opportunity to produce further publications from those data (a typical concern of researchers in data-poor fields) and that their data are not

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

TABLE 7. Summary of a Selection of Previous Studies of Data Sharing in Various Fields.

Subject Area	Requested Datasets	Available Datasets	Source Journals	Study
Cancer	64	42 (66%)	Various	Kyzas et al. 2005
Medicine, cancer	85	41 (48%)	Various	Piwowar et al. 2007
Psychology	141	64 (45%)	*Journal of Personality and Social Psychology*, *Developmental Psychology*, *Journal of Consulting and Clinical Psychology*, *Journal of Experimental Psychology: Learning, Memory, and Cognition*	Wicherts et al. 2006
Economics	15	6 (40%)	*American Economic Review*, *International Journal of Industrial Organization*, *Journal of International Economics*	McCullough and Vinod 2005
Economics	193	69 (36%)	*Journal of Money, Credit and Banking*	McCullough et al. 2006
Medicine	10	1 (10%)	*PLOS Medicine*, *PLOS Clinical Trials*	Savage and Vickers 2009
Medicine	29	1 (3%)	*British Medical Journal*	Reidpath and Allotey 2002
Epidemiology	69	0 (0%)	*American Journal of Epidemiology*, *Journal of the American Medical Association*	Peng et al. 2006

Note: In the study by Peng, Dominici, and Zege (2006), the authors did not attempt to contact other researchers but note the availability of data as described in the published article.

Source: From Spencer 2010.

in a form suitable for sharing and the task of organizing those data is not a high priority. These reasons not to share have also been documented in other research areas, noted above. The absence of patterns in the date of publication, student status of the author, and other variables (Figure 1) may relate more to our small sample size than to a real absence of relationships in data sharing behavior.

Survey of Articles Published in the *Journal of Archaeological Science*

In addition to our e-mail survey, we also conducted a pilot study of a random sample of 48 articles published during February–May 2017 in the *Journal of Archaeological Science* to investigate data sharing behaviors. In this sample we found openly available raw data for 18 essays (53%), even though only seven articles (21%) include a data availability statement. For essays where data are available, nine (50%) have data contained in supplementary files published with the article, six (33%) have all the raw data in tables in the text of the essays, and three (17%) have data in an online repository. The most frequently shared type of data is compositional (e.g., element concentrations; $n = 9$, 28%), followed by radiocarbon and luminescence age data (where the dates are the primary object of analysis; $n = 5$, 16%) and DNA sequences ($n = 4$, 12%). Generally, compositional data are presented in tables in the text, dating data are presented in supplementary files, and DNA data are in repositories (as is typical for DNA data; Figure 2). Data shared in supplementary files and online repositories were most often in Microsoft Excel format ($n = 38$, 38% of the files in supplementary files and repositories), followed by CSV (aka comma-separated values, an open-source plain text spreadsheet format; $n = 4$, 31%), and tables embedded in Microsoft Word or PDF documents ($n = 4$, 31%). When a scripting language such as R, Python, MATLAB, or OxCal was used as the analysis software for an article ($n = 7$, 54% of the essays would identify the software), only two also provided the script with the article.

The *Journal of Archaeological Science* has had a "data disclosure" policy since at least 2013 that states that "all data relating to the article must be made available in Supplementary files or deposited in external repositories and linked to within the article" (2013). The data availability rate of 53% reported in our pilot study reflects either weak enforcement of this policy or an incomplete understanding among editors, reviewers, and authors of how to interpret and implement this policy. The prominence of compositional and dating data among the shared data types suggests a least effort strategy, with authors sharing data that do not require extensive cleaning and tidying (because the data are generated by instruments rather than entered by hand, such as artifact attribute data). The size of compositional and dating datasets means that they mostly fit easily into tables in the text or supplementary files and do not require much effort to organize (e.g., very few metadata are required for others to use these data). The small proportion of articles in our sample using or sharing a scripting language suggests that, unlike authors in other fields (Eglen et al. 2017), archaeologists do not appear to recognize code for data analysis as a primary research object to share with the research community. Code written by the researcher for even the simplest operations should be shared along with data because code communicates decisions made about the data analysis that can help others understand, analyze, and reuse the data.

Datasets in Repositories Tracked by DataCite

Our third investigation of archaeologists' data sharing behaviors is an analysis of all records ($n = 118{,}311$, as of August 2017) of archaeological data files in online repositories tracked by DataCite. DataCite is the primary (but not the only) organization

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

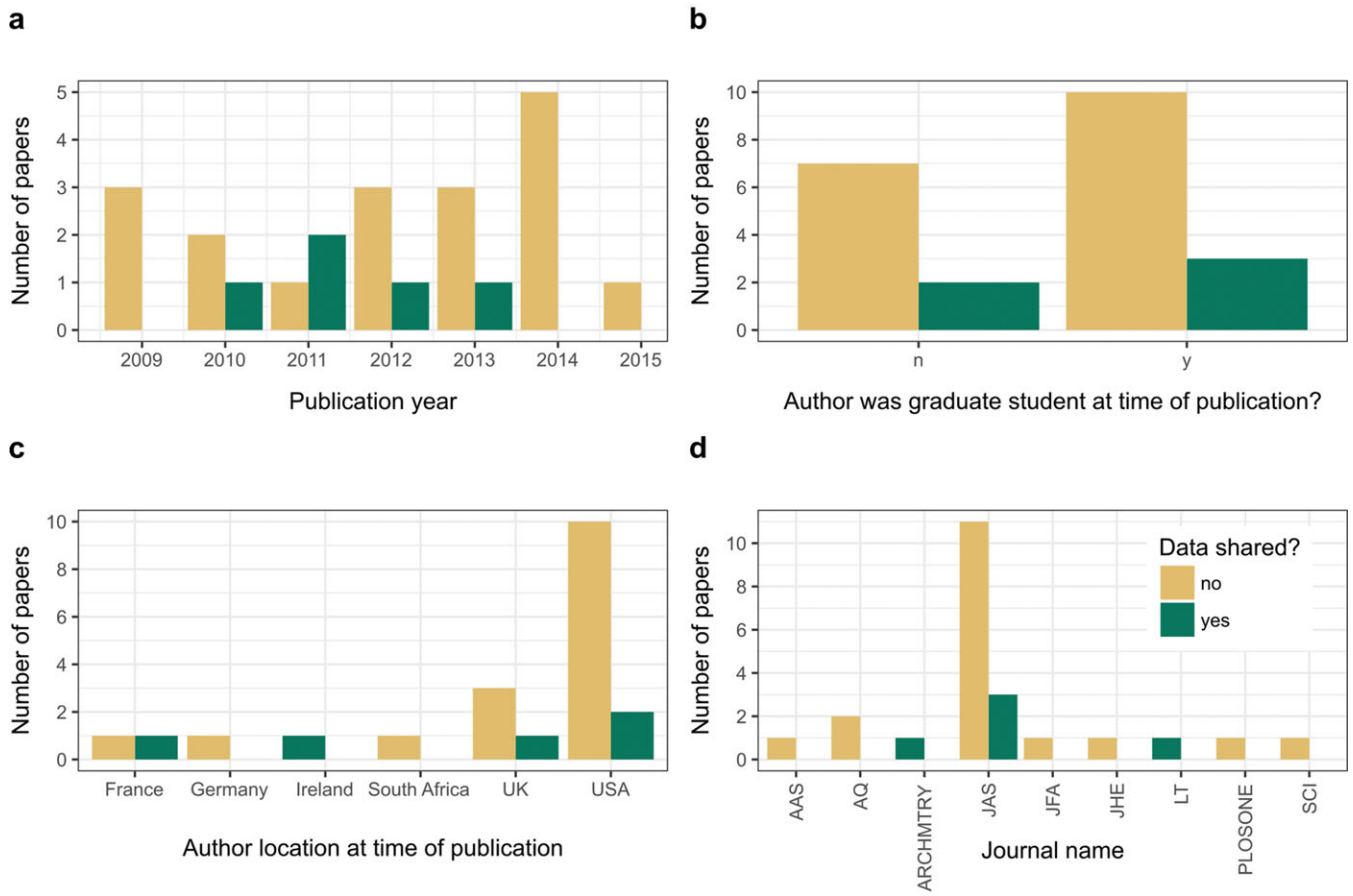

FIGURE 1. Summary of results of e-mail requests for data: (a) responses by year of journal article publication; (b) student status of first author; (c) country location of first author; (d) journal. AAS = *Archaeological and Anthropological Sciences*; AQ = *American Antiquity*; ARCHMTRY = *Archaeometry*; JAS = *Journal of Archaeological Science*; JFA = *Journal of Field Archaeology*; JHE = *Journal of Human Evolution*; LT = *Lithic Technology*; PLOSONE = *PLOS ONE*; SCI = *Science*.

that provides persistent identifiers, or DOIs, for research data. Although DataCite is an international organization, our results are highly concentrated in just a few locations in Europe. The majority of records come from the Archaeology Data Service (ADS) in the United Kingdom, followed by the Dutch Data Archiving and Networked Services (DANS-EASY; Figure 3). Although the earliest data date to 1640, file deposition started and steadily increased during 1990–2008, as personal computers became more affordable and data became easier to digitize. There was a substantial increase in 2009, when the DANS-EASY repository saw an influx of records from Dutch cultural resource management firms (Figure 4a). This can be explained as a result of Dutch archaeology legislation, the Wet op de archeologische monumentenzorg, in 2007 that formally obligated archaeologists in the Netherlands to archive their data via DANS within two years of completion of their projects (Keers et al. 2011). Another spike occurred in 2012, when tDAR: The Digital Archaeological Record had a large increase in deposits, likely due to it waiving its deposit fees for that year (Ellison 2012). After 2009 the volume of deposits declined in many repositories, with the exception of generic repositories such as Figshare, Zenodo, and ResearchGate, which continued to grow (Figure 4b).

Of the 75,632 records with copyright information, the majority are released under the ADS conditions of use ($n = 43,661$, 58%), which is broadly similar to the CC-BY-NC (Moore and Richards 2015; Figure 5). The next most frequently used condition is the EU Open Access license ($n = 22,198$, 29%), similar to the CC-BY license. Creative Commons licenses such as CC-BY, CC-0, and others are used by less than 3% of records.

Many of the records are gray literature such as unpublished reports (Figure 6), produced by commercial consultants or research projects. Of the 66,121 (56%) records that include information about the file types in the deposit, 86% ($n = 56,935$) contain only PDFs. Structured data files such as spreadsheets, databases, and shapefiles are found either alone or together with other file types (typically PDFs) in 7% ($n = 4,481$) of records. The highest proportion of records with structured data files is found in the DANS-EASY repository ($n = 4,304$, 14%), followed by the ADS with 0.4% ($n = 174$). Note that file format information is not present for all records in this sample, and some of the smaller repositories do not supply this information to DataCite at all, so it is possible that the true proportion of structured files in these records is higher than observed here.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

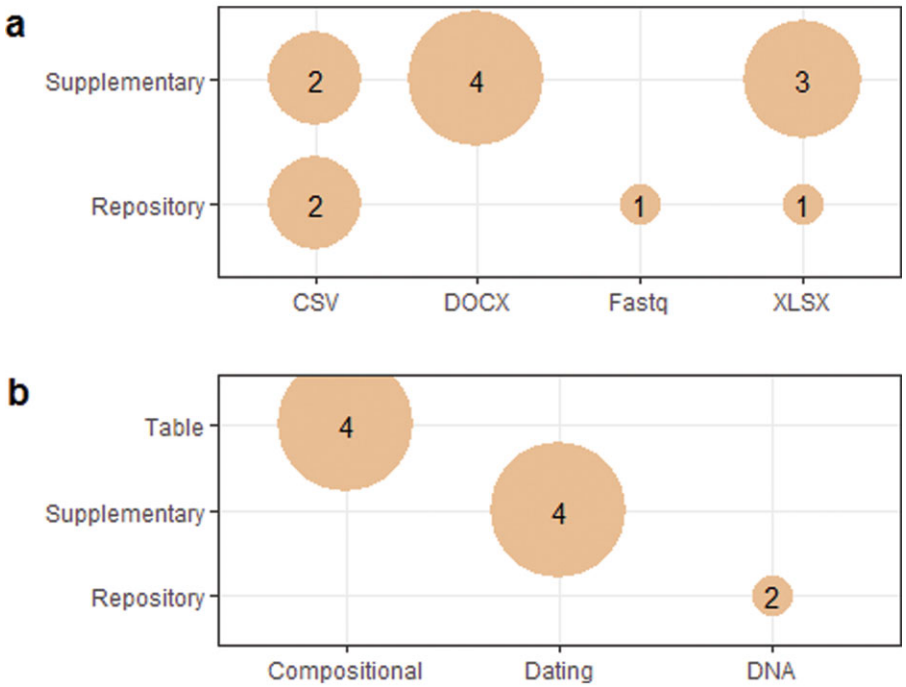

FIGURE 2. Summary of relationships among file types, data types, and data locations: (a) file types and data locations; (b) data types and data locations.

From this pilot study of repository data we see that policy mandates in the United Kingdom and the Netherlands have been effective in populating repositories with archaeological data. With the 2012 spike in records deposited in tDAR we also see that archaeologists are sensitive to the price of sharing data. Nearly all records are made available under conditions that permit others to use the data while requiring that the original author of the data receive credit in reuse. In most cases this is because the repository hosting the data requires the author to choose one of these permissive licenses as a condition of making the deposit. However, although many items are *able* to be reused because of the licenses, the ease of reuse of these items is low. Only a small fraction of the items deposited are machine-readable structured data files. The majority of items are PDFs, which require extensive manual handling to extract tabular data for convenient reuse and combination with other datasets.

We can get some insights into the impact of repository data by querying the Crossref (2017) Event Data service, a freely accessible DOI citation dataset produced by periodic scanning of a variety of online sources. Crossref is an official DOI registration agency, focused on issuing DOIs for scholarly materials such as journal articles, books, e-books, and so on. Another source for data citation data is the Thomson Reuters Data Citation Index (Pavlech 2016), but this resource is not free, and we did not have access to a subscription. Of the 118,311 DOIs in our sample, we found a total of 366 citations of 80 unique DOIs in the Crossref events data. The majority of these citations were found on Wikipedia (Figure 7), followed by other data repositories (i.e., cross-citations from one dataset to another in the same repository). The largest source of citations is the Zenodo repository, followed by articles in the journal *Medieval Archaeology*, which appear here because Volumes 1–50 are hosted by the ADS. According to the Crossref data, citation of archaeological data in repositories is a recent phenomenon, with the majority in our sample occurring over the last 12 months. We found only one example of a dataset DOI cited in a journal article (McCoy and Codlin 2016). This may be an underestimate, given Huggett's (2017) finding that 56 of 476 archaeology data sources in the Thomson Reuters Data Citation Index have citations in Web of Science databases. However, Huggett notes that the Thomson Reuters citation data are unreliable, for example, with incomplete citations and citations that predate the cited dataset. Huggett also shows an example of archaeologists citing datasets but omitting the DOI (e.g., Nugent and Williams 2012), so the citation is not captured by services such as Crossref or Thomson Reuters. These malformed citations add to the difficulty of accurately measuring data citation in archaeology (cf. Belter 2014).

HOW TO SHARE ARCHAEOLOGICAL DATA?

In the pilot studies reported above we see three common approaches to sharing data: sending files privately by e-mail, following a personal request; including data as supplementary files submitted for publication with a journal article; and sharing data by depositing files in a trustworthy, public, DOI-issuing repository. The third method, via a public repository, is by far the most widely recommended method both specifically by archaeologists (Kansa and Kansa 2013; Kansa et al. 2011; Kintigh 2006) and in other sciences (see Cranston et al. 2014; Mounce 2014;

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

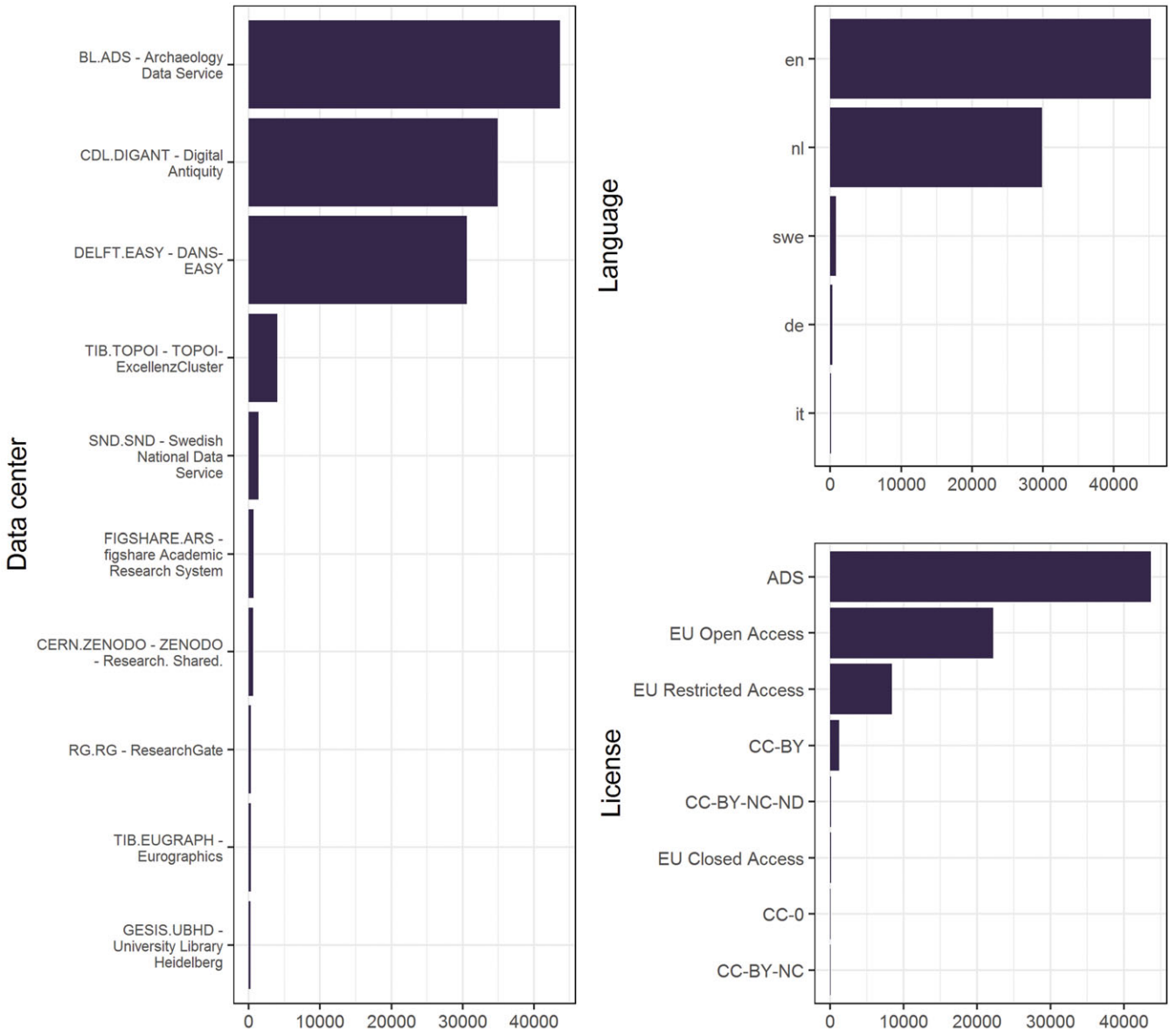

FIGURE 3. Distributions of data centers, languages, and licenses (see text for details) among archaeology datasets known to DataCite. Horizontal axis is the number of digital object identifiers. DANS = Data Archiving and Networked Services; en = English; nl = Dutch; swe = Swedish; de = German; it = Italian.

Penev et al. 2017; Thessen and Patterson 2011; Whitlock 2011). Lists of trustworthy repositories are maintained by the editors of the journal *Scientific Data* (Nature: *Scientific Data* Editors 2017) and http://www.re3data.org, a global registry of research data repositories (Pampel et al. 2013). University librarians are another resource for advice on choosing repositories and preparing data for sharing. There are many guidelines and conventions from other disciplines that are relevant to sharing archaeological data (Inter-university Consortium for Political and Social Research 2012; Strasser et al. 2012; Wilkinson et al. 2016). Table 8 summarizes some of these, and recommendations specifically written for archaeologists can be found in Archaeology Data Service/Digital Antiquity (2011) and Archaeological Resources in Cultural Heritage: A European Standard (2014). Many of these guidelines may create new technical and practical challenges for archaeologists who are not accustomed to sharing data, and many archaeologists do not want to dedicate resources to sharing data. However, we should not let the perfect be the enemy of the good. In our own work we try to follow these guidelines with no dedicated resources for data sharing and with an incomplete knowledge of some of the technical details of repository operations. While we recognize that our data sharing practices fall short of all the guidelines, we still see value in sharing data imperfectly, rather than not at all.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

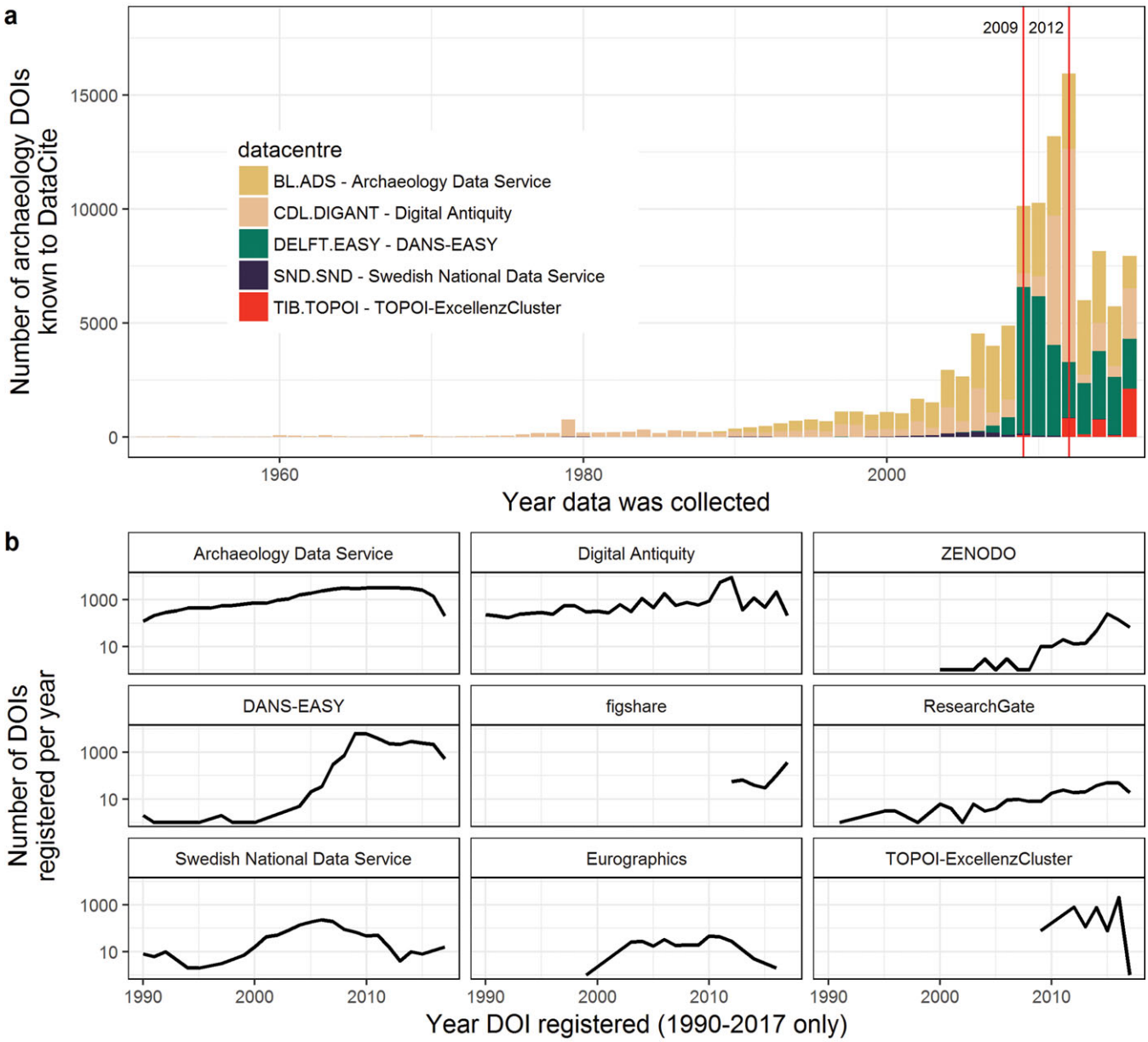

FIGURE 4. Change over time in archaeology datasets known to DataCite. (a) Counts of datasets per year by the major data centers; red lines indicate major increases in the number of datasets in repositories. Note that the date on the horizontal axis indicates when the data were collected, not when they were deposited in the data center. (b) Temporal trends in the number of digital object identifiers (DOIs) at individual data centers. DANS = Data Archiving and Networked Services.

Data sharing is a relatively new behavior for most archaeologists, so it is important to clarify what does not count as data sharing. Making data available on a website that is not a trustworthy, persistent-identifier-issuing repository is not data sharing. For example, data files hosted on a personal faculty web page or the Web server of a research project is problematic because these files could disappear at any time, the data can change in untracked ways, and published links to the files are broken when these websites are reorganized, making the files inaccessible.

Sharing data via supplementary files attached to a journal article is problematic because they are inaccessible to nonsubscribers for paywalled journals, because publishers often alter the data file formats and file names during the production process, and because publishers own those data and the original authors may need to request permission to reuse their own data. For example, machine-readable formats such as CSV are sometimes presented as PDF or Microsoft Word files in supplementary files. The modification of supplementary file names during the article pro-

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

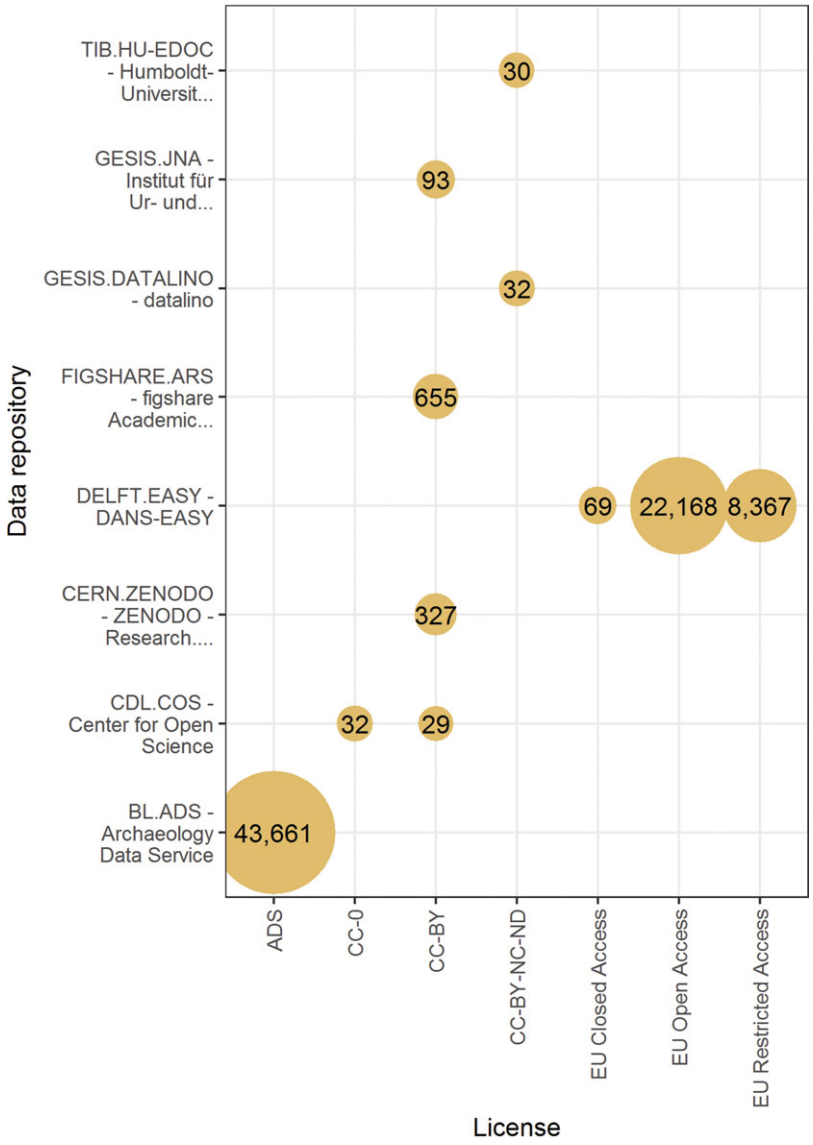

FIGURE 5. Summary of licenses used by data centers with archaeological data. ADS = Archaeology Data Service; DANS = Data Archiving and Networked Services.

duction process can break scripts that attempt to read files by referring to file names as they were prior to the publisher handling the files.

Finally, authors who write in their publications that "the data are available by request" are not sharing data. Not only are the data not actually available, but there is no way to enforce that the authors will actually honor any requests for their data. A simple strategy for avoiding these pseudo-sharing behaviors is to deposit data files in a trustworthy data repository and use persistent links to the files in reports and manuscripts. In our own work we have used figshare.com, zenodo.org, osf.io, neotomadb.org, opencontext.org, and our university repositories. However, there are many other options, and we recommend seeking advice from librarians and funding agencies to help with finding the most suitable service.

HOW TO CITE ARCHAEOLOGICAL DATA?

The Joint Declaration of Data Citation Principles is the current standard in many disciplines for data citation within any research object (Data Citation Synthesis Group 2014). The result of these principles is a citation formation as follows (the exact order of the items may vary slightly across different referencing styles):

> Creator (PublicationYear). Title. Version. Publisher. Identifier

The "PublicationYear," "Title," and "Publisher" are familiar from traditional citation formats, except that the publisher may not

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

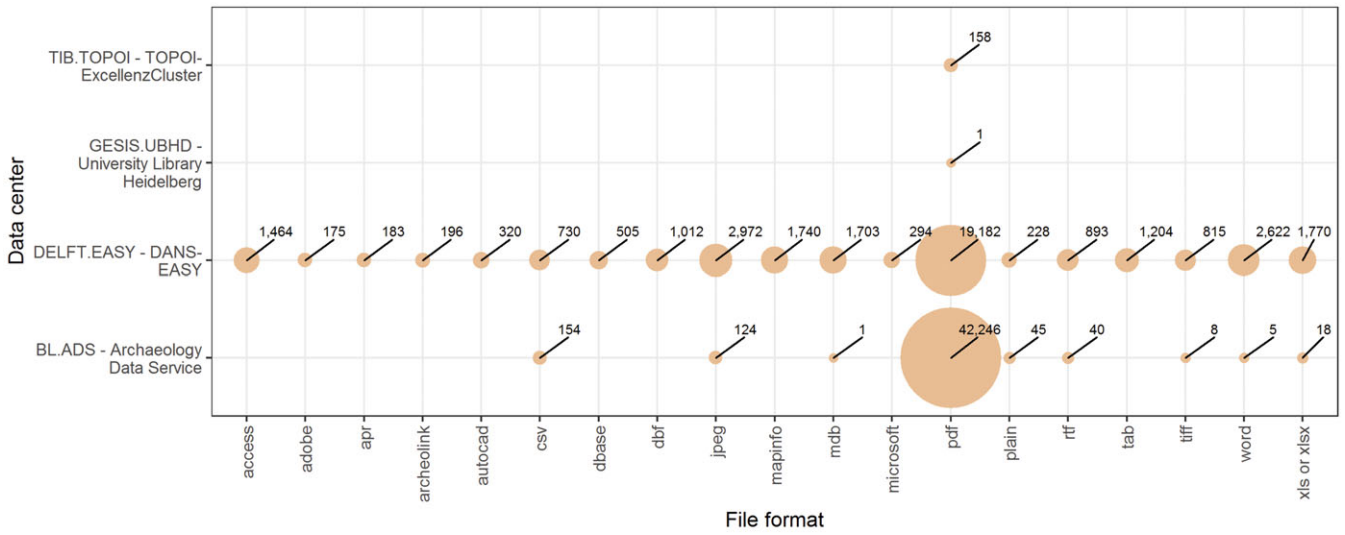

FIGURE 6. Relationships between data centers and file formats. The 10 most frequently found file formats are shown, in the data centers with these files from the top 10 data centers by number of digital object identifiers. Open file formats for structured data, such as CSV, are not common. DANS = Data Archiving and Networked Services.

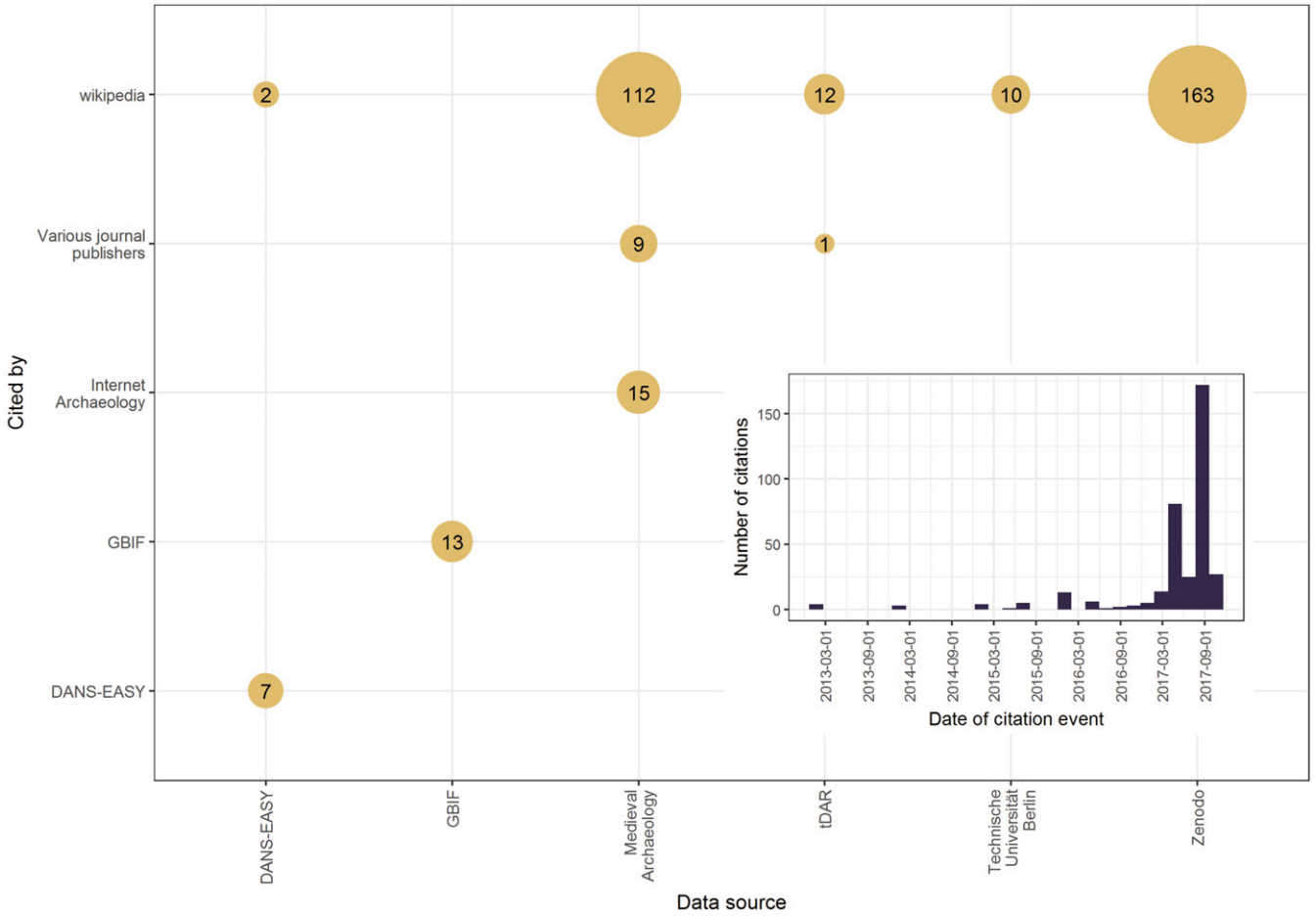

FIGURE 7. Citation of archaeological data digital object identifiers tracked by the Crossref Event Data service. Inset shows the distribution of these citations over time. DANS = Data Archiving and Networked Services, GBIF = Global Biodiversity Information Facility.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

TABLE 8. Nine Key Things to Do before Sharing Your Data.

Rule	Details
Anticipate how your data will be used	Before collecting data, decide what formats and metadata schema will be used; how and where the data will be archived, embargoed, and released; and what software others are likely to use to work with your data.
Keep raw data raw	Data should be as unprocessed as possible, within practical limitations, to facilitate future analysis with new methods. The line between raw and derived data is not often clear, so the general idea is to provide data that will enable others to confirm your analytic results, assess the validity of your statistical output, or directly compare findings across other studies.
Store data in open formats	Data should be in a format that is easily readable by computers, without being dependent on specific software. Examples of open data formats include CSV for tabular data, PNG for images, and KML (or other Open Geospatial Consortium format) for spatial data. Examples of closed formats include Photoshop documents (PSD) for bitmap images and Microsoft Excel (XLS) for tabular data. While day-to-day work may require closed formats, the files that are archived should be in open formats.
Data should be structured for analysis	Data should be structured so they are easy to analyze with a computer and easy to combine with other datasets. For tabular data, each variable or measurement should be a single column, each observation or record should be a single row, and each type of observational unit should be a table. Each cell should contain a single value and not include units, comments, or derivative values (such as percentage) in the same cell as the value. Be consistent with capitalization, abbreviations, and naming conventions, and avoid exotic characters such as accents that may not be readable on other systems. Use a good null value such as a blank cell or "NA." Do not use spaces in column names.
Data should be uniquely identifiable	Data used in a scholarly publication should be archived online at a location that can be uniquely identified by a persistent URL or a globally unique identifier/universally unique identifier. Types of unchanging URLs include digital object identifiers, Archival Resource Keys, and persistent URLs. Individual files should have a distinct name, which includes a version identifier. This identifier could be the date in the ISO 8601 standard (YYYY-MM-DD, i.e., from largest time unit to smallest).
Provide relevant metadata	Include information about the what, when, where, and how of the data collected. For tabular data, this should include descriptions of the column names, units of measurement, warnings about missing or inconsistent data, and information to check that the data are complete (e.g., the number of rows and columns that were in the table when it was created).
Adopt the proper privacy protocols	Take care not to include data that might expose people, places, or objects to risk of harm or loss. Negotiating with stakeholders such as indigenous groups is essential to avoid harm in data sharing. This might include omitting or degrading location data (e.g., so that site locations are not revealed to the public) and removing or anonymizing personally identifiable information. Methods for ensuring privacy can be complex and change rapidly, so seeking expert help may be necessary.
Use a trustworthy repository	Data should be easy to find and available for a long time. Archive your data in a major well-established repository that provides persistent URLs and has a commitment to sustainability (e.g., participates in the CLOCKSS archive). Do not use your personal website or a project website, which may change or disappear.
Use an open license	Data should have an explicit and well-established license to let others know exactly what they can and cannot do with the data you shared. One of the most widely used and frequently recommended licenses is CC-0 or public domain. This places no restrictions on data use, promoting greater engagement with the data.

Source: Hart et al. 2016; White et al. 2013.

be a traditional book or journal publisher but, rather, a dedicated data repository organization. "Creator" is equivalent to the author or primary collector and arranger of the data (e.g., Open Context or tDAR). The "Version" and "Identifier" fields are unique to data citation and can take a variety of forms. The identifier is a special type of persistent internet address. The identifier must have two important qualities: the address contains some metadata (such as the identity of the publisher, but not necessarily in a human-readable format), and it remains fixed to a specific file or bundle of files over the lifetime of those objects, even if their location on the internet changes. There are several varieties of these in use, with DOIs currently the most common form. When a data publisher changes the internet address of a dataset, it is responsible for informing the DOI-issuing authority so that the DOI resolves to the new location of the files.

A DOI consists of three parts:

- the proxy, usually https://doi.org but sometimes also https://hdl.handle.net/ or just DOI
- the prefix, usually a string of numbers that is unique to the organization publishing the DOI. For example, the tDAR prefix is 10.6067
- the suffix, a random-looking string of any combination of characters and numbers that uniquely identify the file or bundle of files; this is also known as a local identifier (local ID) or an

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

"accession," an identifier that, by itself, is only guaranteed to be locally unique within the database or source (McMurry et al. 2017)

A DOI alone is not sufficient for citing a dataset, because, as with many things on the internet, the files can change over time. For example, we might upload data files to the repository when we submit a paper for peer review and then get a DOI and cite that in our manuscript so the peer reviewers can see them (some services provide a private URL that can be shared with reviewers but is not publicly accessible or discoverable by a search engine). Then when we resubmit our manuscript after peer review, we might update the repository to include additional data files relating to additional analyses requested by the reviewers. We still have the same DOI in our manuscript for this dataset, but the files have changed because of the revisions we did in response to the peer reviewers' comments. Most data repositories will keep each version of the files that you upload, so in this scenario there will be at least two versions of the files in our data repository. Even if we do not follow this exact sequence of updating the files, we may still change them in the future, after our article has been published. Because of this potential to change the files stored at the DOI, it is necessary to identify specific file versions in our citation of the data. A simple method to record the version is to include the date that the data files were accessed, as when citing any other resource found on the internet. Some repositories also provide DOIs that have a version number suffix such as ".v2," ".2," "/2." However, this is not formally a functionality of the DOI system, so including the date accessed is a more robust and general method of indicating the version. Technological solutions to issues of transience, permanence, and granularity in data citation are still in development.

When citing data in the text of a manuscript or report, the ideal practice is to cite the dataset and the traditional publication that first described it, like so (Penev et al. 2017): "This paper uses data from the [name] data repository at http://doi.org/xxx/xxx (Author YYYY), first described in Author YYYY." For example: "This paper uses data from the Open Science Framework data repository at http://doi.org/10.17605/OSF.IO/32A87 (Marwick 2017), first described in Marwick et al. 2017." Citing both the dataset and the publication helps to resolve ambiguity about the source of the data and gives context to the data. Furthermore, the authorship of the data files may differ from that of the paper describing the data, and it is important that both authors (or sets of authors) receive credit for their work through the normal mechanisms of citation recognition. In cases where authors have published their data and then make an intellectual contribution to a separate study led by others that uses the published data, then the data author should also be a coauthor of the separate study. If other researchers use published data in their publication without any input from the data author, then citation of the data is all that is required, and adding the data author as a coauthor is not necessary.

In the paper in which these data are first described, the citation statement should be included in the body of the paper, in the "Material and Methods" section. In addition, the formal data reference should be included in the paper's reference list, using the recommended journal's reference format. For example, here is a statement that would appear in the paper in which the data are first presented: "The data underpinning the analysis reported in this paper were deposited in the Open Science Framework Repository at http://doi.org/10.17605/OSF.IO/32A87 (Marwick 2017)." The reference list of that paper would include Marwick 2017 as

Marwick, B. (2017). Dataset and R code for "Movement of lithics by trampling: An experiment in the Madjedbebe sediments," http://doi.org/10.17605/OSF.IO/32A87, Open Science Framework, Accessed 7 Sept 2017.

SUMMARY AND CONCLUSION

The status of research data is slowly changing in many fields from a gift culture where data are privately held commodities traded for coauthorship, invitations, and other status tokens (Hagstrom 1982; Wallis et al. 2013) to part of the public commons, accessible to anyone with an internet connection. This change has been fastest in data-rich "big science" fields, such as astronomy, physics, and oceanography, that use expensive instruments that generate data shared with large teams of collaborators to produce new discoveries (Reichman and Uhlir 2003). However, in "smaller" data-poor sciences such as archaeology, where collaborative groups are modest, data sharing is still mostly ad hoc and disaggregated. Our pilot studies, while small and limited in many ways, support this statement, with an overall sharing rate of 20% in responses to private requests for data and 53% of sampled journal articles with openly available data, in a wide variety of formats and locations. We found that legislation requiring data sharing resulted in a substantial increase in the rate of data deposition and that archaeologists are sensitive to repository fees. In examining archaeological data deposited in repositories, we found that most of the contents are PDF files of reports and only a very small fraction of the data files are structured data such as spreadsheets. Archaeologists still have a long way to go in making their data easier to reuse by sharing data in plain text structured data formats such as CSV. We found that citation of datasets, although problematic to measure, appears to be almost nonexistent in archaeological literature. Table 8 summarizes nine key tasks archaeologists should do with their data to optimize their discovery and reuse.

Among the many challenges of increasing the rate of data sharing in archaeology and other "small" sciences is that data sharing is unfunded, unrewarded, and only rarely required. Who is responsible for changing this? There is no simple answer, but we believe that the task must be shared between individuals and institutions, with individuals pushing for institutional change. The challenge of funding can be tackled by practical action from members of professional societies. For example, the Society for American Archaeology should communicate to funding bodies, such as the National Science Foundation, about the importance of allocating funds specifically for data sharing and providing reliable information on how to choose and use data repositories. The challenge of requiring data sharing can also be addressed by funding bodies and federal agencies, through more rigorous evaluation and enforcement of data management plans during the review process. However, changes in the policies of professional societies and other agencies are primarily responses to changes in norms in the community of researchers. In practice, individual researchers are responsible for

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

communicating the value of data sharing to professional societies and funding agencies (e.g., by holding peers to high standards of data sharing during grant reviews). Changes in individual practice will demonstrate a change in norms and motivate our institutions to formalize these new data sharing values through policies and regulations.

One area where changes in individual practice may be especially effective is among journal editors and editorial boards. For example, editors of archaeology journals should adopt the Transparency and Openness Protocols of the Center for Open Science (Nosek et al. 2015). Editors should then educate their editorial board members, peer reviewers, and contributors about their obligations regarding these protocols. Editors should only accept papers that, minimally, provide the raw data used to produce each table and visualization in the manuscript. Changes in journal-specific submission guidelines can result in substantial changes in authors' research practices (Giofrè et al. 2017; Macleod et al. 2017). For example, if archaeology journals adopt incentives such as badging (Kidwell et al. 2016), and require as a condition of publication that submitted articles include in the text a properly formatted citation to the data analyzed in the article (or if the data cannot be shared, an explanation why the data are not available), authors would quickly adapt to this new requirement to ensure that their articles will be published.

Individual researchers can also contribute to changing the norms on data sharing via several practical actions suitable for typical day-to-day work:

- When preparing research for publication, we can upload data to a trustworthy repository and cite the data in the manuscript.
- When writing peer reviews of journal articles, we can tell editors that we require data to be available for papers that we read and that the manuscripts we review must include citations of datasets analyzed in them. We can write in our peer reviews that data need to be available for us to make a proper assessment of the manuscript.
- When reviewing grant applications, we can critically evaluate data management plans and report on the applicants' past record on data sharing. We can offer helpful suggestions about how to improve data availability (e.g., Table 8).
- When teaching students, we can instruct them how to prepare data for sharing, and how to find and use publicly available datasets to give students experience analyzing and writing about shared data, and cultivate the expectation among students that data sharing is a normal part of archaeological research.
- When evaluating candidates for hiring and promotion, we can ensure that commitments to data sharing are valued.

The challenge of data sharing as an unrewarded behavior was the primary motivation for this article. By specifying a standard for the scholarly citation of archaeological data we provide a method for giving credit to the authors or compilers of data. This rewards those authors with visibility and recognition for their effort with citations, a currency that is valued broadly in the discipline. The data citation standard outlined above is easy for humans and search engines to understand and is highly functional, with a full internet address for the data contained in the citation. The standard can accommodate the addition of further information required by different citation styles and publishers' requirements. The data citation standard is not dependent on any specific choices of software, archive, data producer, publisher, or author and is easy to adapt into existing writing and research workflows. There are still many unresolved challenges in data sharing, such as how to manage high volumes of data generated by photogrammetry. However, we are confident that solutions to many of these challenges will emerge as data sharing is increasingly valued by the research community.

Acknowledgments

This article emerged out of discussions at the 2017 Society for American Archaeology meeting, specifically the forum "Beyond Data Management: A Conversation about 'Digital Data Realities,'" organized by Sarah Kansa and Eric Kansa, and the Society for American Archaeology Publications Committee. Thanks go to Canan Çakirlar for insights into the history of archaeological data sharing in the Netherlands and Stefano Costa for many helpful suggestions. BM was supported by an Australian Research Council Future Fellowship (FT140100101). Thanks go to Stefanie Crabtree and the other four peer reviewers for their constructive suggestions.

Data Availability Statement

This article uses data from the Open Science Framework repository at http://doi.org/10.17605/OSF.IO/KSRUZ. (Marwick and Pilaar Birch 2018), first described in this article. See the methods section of the text for more details. A capsule of text, data and code to reproduce this article is also available in Code Ocean, a computational reproducibility platform at https://doi.org/10.24433/CO.ca12b3f0-55a2-4eba-9687-168c8281e535.

REFERENCES CITED

Alsheikh-Ali, Alawi A., Waqas Qureshi, Mouaz H. Al-Mallah, and John P. A. Ioannidis
2011 Public Availability of Published Research Data in High-Impact Journals. *PLOS ONE* 6(9): e24357. DOI:10.1371/journal.pone.0024357, accessed September 7, 2017.

American Psychological Association
2001 Ethical Standards for the Reporting and Publishing of Scientific Information. In *Publication Manual of the American Psychological Association*, 5th ed., pp. 387–396. American Psychological Association, Washington, DC.

Archaeological Resources in Cultural Heritage: A European Standard
2014 *The Standard and Guide to Best Practice in Archaeological Archiving in Europe*. Electronic document, http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=ARCHES%20Introduction, accessed September 7, 2017.

Archaeology Data Service/Digital Antiquity
2011 Guides to Good Practice. Electronic document, http://guides.archaeologydataservice.ac.uk/, accessed September 7, 2017.

Belter, Christopher W.
2014 Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. *PLoS One* 9(3): e92590. https://doi.org/10.1371/journal.pone.0092590

Bogen, James, and James Woodward
1998 Saving the Phenomena. *Philosophical Review* 97(3):303–352.

Borgman, Christine L.
2012 The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* 63(6):1059–1078.
2015 *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, Cambridge, Massachusetts.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

Chambers, Chris
2017 *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton University Press, Princeton, New Jersey.

Clarke, Mary E.
2015 The Digital Dilemma: Preservation and the Digital Archaeological Record. *Advances in Archaeological Practice* 3(4):313–330.

Cranston, Karen, Luke J. Harmon, Maureen A. O'Leary, and Curtis Lisle
2014 Best Practices for Data Sharing in Phylogenetic Research. *PLOS Currents Tree of Life* (blog), June 19. DOI:10.1371/currents.tol.bf01eff4a6b60ca4825c69293dc59645, accessed September 7, 2017.

Crossref
2017 Event Data User Guide, Version 1.0.0—Beta. Electronic document, https://www.eventdata.crossref.org/guide/, accessed November 7, 2017.

Daston, Lorraine, and Peter Galison
1992 The Image of Objectivity. *Representations* 40:81–128.

Data Citation Synthesis Group
2014 Joint Declaration of Data Citation Principles. In *FORCE11; 2014*, edited by M. Martone. Electronic document, https://doi.org/10.25490/a97f-egyk, accessed September 7, 2017.

Dorch, Bertil F., Thea M. Drachen, and Ole Ellegaard
2015 The Data Sharing Advantage in Astrophysics. *Proceedings of the International Astronomical Union* 11(A29A):172–175.

Duke, Clifford S., and John H. Porter
2013 The Ethics of Data Sharing and Reuse in Biology. *BioScience* 63(6):483–489.

Ellison, Leigh A.
2012 Just 20 Days Left!! Upload Your Archaeological Information FREE to tDAR!! *tDAR: The Digital Archaeological Record* (blog), December 12. Electronic document, https://www.tdar.org/news/2012/12/just-20-days-left-upload-your-archaeological-information-free-to-tdar/, accessed September 7, 2017.

Eglen, Stephen J., Ben Marwick, Yaroslav O. Halchenko, Michael Hanke, Shoaib Sufi, Padraig Gleeson, and Jean-Baptist Poline
2017 Toward Standard Practices for Sharing Computer Code and Programs in Neuroscience. *Nature Neuroscience* 20(6):770–773.

Fienberg, Stephen E., Margaret E. Martin, and Miron L. Straf
1985 *Sharing Research Data*. National Academy Press, Washington, DC.

Floridi, Luciano
2011 *The Philosophy of Information*. Oxford University Press, London.

Giofrè, David, Geoff Cumming, Luca Fresc, Ingrid Boedker, and Patrizio Tressoldi
2017 The Influence of Journal Submission Guidelines on Authors' Reporting of Statistics and Use of Open Research Practices. *PLOS ONE* 12(4). DOI:10.1371/journal.pone.0175583, accessed September 7, 2017.

Gleditsch, Nils P., Claire Metelits, and Håvard Strand
2003 Posting Your Data: Will You Be Scooped or Will You Be Famous? *International Studies Perspectives* 4(1):89–97.

Hacking, Ian
1992 The Self-Vindication of the Laboratory Sciences. In *Science as Practice and Culture*, edited by Andrew Pickering, pp. 29–64. University of Chicago Press, Chicago.

Hagstrom, Warren O.
1982 Gift Giving as an Organising Principle in Science. In *Science in Context: Readings in the Sociology of Science*, edited by B. Barnes and D. Edge, pp. 21–34. Open University Press, Milton Keynes, United Kingdom.

Harding, Anna, Barbara Harper, Dave Stone, Catherine O'Neill, Patricia Berger, Stuart Harris, and Jamie Donatuto
2012 Conducting Research with Tribal Communities: Sovereignty, Ethics, and Data-Sharing Issues. *Environmental Health Perspectives* 120(1):6–10.

Hart, Edmund M., Pauline Barmby, David LeBauer, François Michonneau, Sarah Mount, Patrick Mulrooney, and Jeffrey W. Hollister
2016 Ten Simple Rules for Digital Data Storage. *PLOS Computational Biology* 12(10): e1005097. DOI:10.1371/journal.pcbi.1005097, accessed September 7, 2017.

Henneken, Edwin A., and Alberto Accomazzi
2011 Linking to Data—Effect on Citation Rates in Astronomy. Electronic document, https://arxiv.org/abs/1111.3618, accessed September 17, 2017.

Huggett, Jeremy
2017 Citing Data Reuse. *Introspective Digital Archaeology* (blog), May 23. Electronic document, https://introspectivedigitalarchaeology.wordpress.com/2017/05/23/citing-data-reuse/, accessed September 7, 2017.

Hymes, Dell
1972 *Reinventing Anthropology*. Pantheon, New York.

Inter-university Consortium for Political and Social Research
2012 *Guide to Social Science Data Preparation and Archiving: Best Practice throughout the Data Life Cycle*, 5th ed. Institute for Social Research, University of Michigan, Ann Arbor.

Journal of Archaeological Science
2013 Guide for Authors. *Journal of Archaeological Science*. Electronic document, https://web.archive.org/web/20130125222940; https://www.elsevier.com/journals/journal-of-archaeological-science/0305-4403/guide-for-authors, accessed September 7, 2017.

Kansa, Eric C., and Sarah W. Kansa
2013 Open Archaeology: We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication. *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 1(1):88–97.

Kansa, Eric C., Sarah W. Kansa, and Ethan Watrall (editors)
2011 *Archaeology 2.0: New Approaches to Communication and Collaboration*. Cotsen Digital Archaeology Series, University of California, Los Angeles. Electronic document, https://escholarship.org/uc/item/1r6137tb, accessed November 7, 2017.

Keers, Geurt, Hans van derReijden, and Hans van Rossum
2011 *Planning Archaeology: A Synthesis of the Thematic Sub-reports*. RIGO Research en Advies, Amsterdam.

Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, and Brian Nosek
2016 Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14(5): e1002456.

Kintigh, Keith
2006 The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3):567–578.

Kratz, John, and Carly Strasser
2014 Data Publication Consensus and Controversies. *F1000Research*. Electronic document, https://f1000research.com/articles/3-94/v1, accessed September 7, 2017.
2015 Researcher Perspectives on Publication and Peer Review of Data. *PLOS ONE* 10(2): e0117619. DOI:10.1371/journal.pone.0117619, accessed September 7, 2017.

Kyzas, Panayiotis A., Konstantinos T. Loizou, and John P. A. Ioannidis
2005 Selective Reporting Biases in Cancer Prognostic Factor Studies. *Journal of the National Cancer Institute* 97:1043–1055.

Latour, Bruno
1999 *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press, Cambridge, Massachusetts.

Leonelli, Sabina
2015 What Counts as Scientific Data? A Relational Framework. *Philosophy of Science* 82(5):810–821.
2016 The Philosophy of Data. In *The Routledge Handbook of Philosophy of Information*, edited by L. Floridi. Routledge, New York 191–202.

Macleod, Malcolm R., The NPQIP Collaborative group
2017 Findings of a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution. bioRxiv 187245; doi: https://doi.org/10.1101/187245, accessed March 12, 2018.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

McCoy, Mark D., and Maria C. Codlin
2016 The Influence of Religious Authority in Everyday Life: A Landscape-Scale Study of Domestic Architecture and Religious Law in Ancient Hawai'i. *World Archaeology* 48(3):411–430.
McCullough, Bruce D., Kerry A. McGeary, and Teresa D. Harrison
2006 Lessons from the *JMCB* Archive. *Journal of Money, Credit, and Banking* 38(4):1093–1107.
McCullough, Bruce D., and Hrishikesh D. Vinod
2005 Comment: Econometrics and Software. *Journal of Economic Perspectives* 17(1):223–224.
McMurry, Julie A., Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, and Helen Parkinson
2017 Identifiers for the 21st Century: How to Design, Provision, and Reuse Persistent Identifiers to Maximize Utility and Impact of Life Science Data. *PLOS Biology* 15(6): e2001414. DOI:10.1371/journal.pbio.2001414, accessed September 7, 2017.
Marwick, Ben
2017 Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 24(2):424–450.
Marwick, Ben, Elspeth Hayes, Chris Clarkson, and Richard Fullagar
2017 Movement of Lithics by Trampling: An Experiment in the Madjedbebe Sediments, Northern Australia. *Journal of Archaeological Science* 79:73–85.
Marwick, B. and S. Pilaar Birch
2018 Dataset and R code for "A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing," http://doi.org/10.17605/OSF.IO/KSRUZ, Open Science Framework, Accessed 11 March 2018.
Merton, Robert K.
1942 The Normative Structure of Science. In *The Sociology of Science: Theoretical and Empirical Investigations*, edited by Robert K. Merton. University of Chicago Press, Chicago 267–278.
Moore, Ray, and Julian Richards
2015 Here Today, Gone Tomorrow: Open Access, Open Data and Digital Preservation. In *Open Source Archaeology: Ethics and Practice*, edited by Andrew T. Wilson and Ben Edwards. De Gruyter, the Netherlands 30–43.
Mounce, Ross
2014 Open Data and Palaeontology. In *Issues in Open Research Data*, edited by S. A. Moore, pp. 151–164. Ubiquity Press, London.
National Science Foundation
2010 Data Management for NSF SBE Directorate Proposals and Awards. Electronic document, https://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf, accessed 15 March 2018.
2016 Dissemination and Sharing of Research Results. Electronic document, https://www.nsf.gov/bfa/dias/policy/dmp.jsp, accessed September 7, 2017.
Nature: Scientific Data Editors
2017 Recommended Data Repositories. Electronic document, https://www.nature.com/sdata/policies/repositories, accessed September 13, 2017.
Nicholas, George P.
2014 Indigenous Cultural Heritage in the Age of Technological Reproducibility: Towards a Postcolonial Ethic of the Public Domain. In *Dynamic Fair Dealing: Creating Canadian Culture Online*, edited by Rosemary Coombe, Darren Wershler, and Martin Zeilinger, pp. 213–224. University of Toronto Press, Toronto.
Nicholas, George P., and Alison Wylie
2009 Archaeological Finds: Legacies of Appropriation, Modes of Response. In *The Ethics of Cultural Appropriation*, edited by James O. Young and Conrad G. Brunk, pp. 11–54. Wiley-Blackwell, Oxford.
2012 "Do Not Do unto Others . . . ": Cultural Misrecognition and the Harms of Appropriation in an Open-Source World. In *Appropriating the Past: Philosophical Perspectives on Archaeological Practice*, edited by Robin Coningham and Geoffrey Scarre, pp. 195–221. Cambridge University Press, Cambridge.
Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, and T. Yarkoni
2015 Promoting an Open Research Culture: Author Guidelines for Journals Could Help to Promote Transparency, Openness, and Reproducibility. *Science* 348:1422–1425.
Nugent, Ruth, and Howard Williams
2012 Sighted Surfaces: Ocular Agency in Early Anglo-Saxon Cremation Burials. In *Encountering Imagery: Materialities, Perceptions, Relations*, edited by Ing-Marie Back Danielsson, Fredrik Fahlander, and Ylva Sjöstrand, pp. 187–208. Stockholm University, Stockholm.
Nuijten, Michèle B., Jeroen Borghuis, Coosje Veldkamp, Linda Alvarez, Marvel van Assen, and Jelte M. Wicherts
2017 Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology. *Collabra: Psychology* 3(1): 31. DOI:10.1525/collabra.102, accessed September 7, 2017.
Pampel, Heinz, Paul Vierkant, Frank Scholze, Roland Bertelmann, Maxi Kindling, Jens Klump, and Uwe Dierolf
2013 Making Research Data Repositories Visible: The re3data.org Registry. *PLOS ONE* 8(11): e78080. DOI:10.1371/journal.pone.0078080, accessed September 7, 2017.
Pavlech, Laura L.
2016 Data Citation Index. *Journal of the Medical Library Association* 104(1):88–90.
Penev, Lyubomir, Daniel Mietchen, Vishwas Shravan Chavan, Gregor Hagedorn, Vincent Stuart Smith, David Shotton, Éamonn Ó Tuama, Viktor Senderov, Teodor Georgiev, Pavel Stoev, Quentin John Groom, David Remsen, and Scott C. Edmunds
2017 Strategies and Guidelines for Scholarly Publishing of Biodiversity Data. *Research Ideas and Outcomes* 3: e12431. DOI:10.3897/rio.3.e12431, accessed September 7, 2017.
Peng, Roger D., Francesca Dominici, and Scott L. Zege
2006 Reproducible Epidemiological Research. *American Journal of Epidemiology* 163:783–789.
Pienta, Amy M., George Alter, and Jared A. Lyle
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, accessed September 27, 2017.
Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma
2007 Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE* 2(3): e308.
Piwowar, Heather A., and Todd J. Vision
2013 Data Reuse and the Open Data Citation Advantage. *PeerJ* 1: e175. DOI:10.7717/peerj.175, accessed September 7, 2017.
PLOS Editorial and Publishing Policies
2017 Sharing of Materials, Methods and Data. *PLOS Medicine*. Electronic document, http://journals.plos.org/plosone/s/data-availability, accessed March 12, 2018.
Reichman, James H., and Paul F. Uhlir
2003 A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment. *Law and Contemporary Problems* 66(1/2):315–462.
Reidpath, Daniel D., and Pascale A. Allotey
2002 Data Sharing in Medical Research: An Empirical Investigation. *Bioethics* 15(2):125–134.
Rheinberger, H. J.
2011 Infra-experimentality: from traces to data, from data to patterning facts. *History of Science*, 49(3), 337–348.
Rowhani-Farid, Anisa, Michelle Allen, and Adrian G. Barnett
2017 What Incentives Increase Data Sharing in Health and Medical Research? A Systematic Review. *Research Integrity and Peer Review* 2(4). DOI:10.1186/s41073-017-0028-9, accessed September 7, 2017.
Savage, Caroline J., and Andrew J. Vickers
2009 Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLOS ONE* 4(9): e7078. DOI:10.1371/journal.pone.0007078, accessed September 7, 2017.

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press

Sawyer, Steve
2008 Data Wealth, Data Poverty, Science and Cyberinfrastructure. *Prometheus* 26(4):355–371.

Sears, Jonathan R. L.
2011 Data Sharing Effect on Article Citation Rate in Paleoceanography. Paper presented at the fall meeting of the American Geophysical Union. Electronic document, https://doi.org/10.6084/M9.FIGSHARE.1222998.V1, accessed September 7, 2017.

Spencer, Hilary
2010 Thoughts on the Sharing of Data and Research Materials and the Role of Journal Policies. Electronic document, http://web.stanford.edu/~vcs/Conferences/RoundtableNov212009/ThoughtPieces, accessed September 7, 2017.

Stodden, Victoria
2009a Enabling Reproducible Research: Open Licensing for Scientific Innovation. *International Journal of Communications Law and Policy*. Electronic document, https://ssrn.com/abstract=1362040, accessed September 7, 2017.
2009b The Legal Framework for Reproducible Scientific Research: Licensing and Copyright. *Computing in Science and Engineering* 11(1):35–40.
2010 The Scientific Method in Practice: Reproducibility in the Computational Sciences. Columbia University Academic Commons. Electronic document, http://hdl.handle.net/10022/AC:P:11417, accessed September 7, 2017.

Strasser, Carly, Robert Cook, William Michener, and Amber Budden
2012 Primer on Data Management: What You Always Wanted to Know. In *DataONE Best Practices Primer*. Electronic document, https://doi.org/10.5060/D2251G48, accessed September 7, 2017.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev U. Aydinoglu, Lei Wu, and Mike Frame
2011 Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE* 6(6): e21101. DOI:10.1371/journal.pone.0021101, accessed September 7, 2017.

Tenopir, Carol, Elizabeth Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, and Kristina Dorsett
2015 Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* 10(8): e0134826. DOI:10.1371/journal.pone.0134826, accessed September 7, 2017.

Thessen, Anne E., and David J. Patterson
2011 Data Issues in the Life Sciences. *ZooKeys* 150:15–51. DOI:10.3897/zookeys.150.1766, accessed September 7, 2017.

Tsosie, Rebecca
2007 Cultural Challenges to Biotechnology: Native American Genetic Resources and the Concept of Cultural Harm. *Journal of Law, Medicine and Ethics* 35(3):396–411.

Uhlir, Paul F., and Daniel Cohen
2011 Internal document. Board on Research Data and Information, Policy and Global Affairs Division, National Academy of Sciences, Washington, DC, March 18.

Vanpaemel, Wolf, Maarten Vermorgen, Leen Deriemaecker, and Gert Storms
2015 Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra* 1(1): 3. DOI:10.1525/collabra.13, accessed September 7, 2017.

Vines, Timothy H., Arianne Y. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, and Diana J. Rennison
2014 The Availability of Research Data Declines Rapidly with Article Age. *Current Biology* 24(1):94–97.

Vitelli, Karen D., and John S. Colwell-Chanthaphonh (editors)
2006 *Archaeological Ethics*. Rowman AltaMira, New York.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman
2013 If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE* 8(7): e67332. DOI:10.1371/journal.pone.0067332, accessed September 7, 2017.

White, Ethan, Elita Baldridge, Zachary T. Brym, Kenneth Locey, Daniel McGlinn, and Sarah Supp
2013 Nine Simple Ways to Make It Easier to (Re)Use Your Data. *Ideas in Ecology and Evolution* 6(2):1–10. DOI:10.4033/iee.2013.6b.6.f, accessed September 7, 2017.

Whitlock, Michael C.
2011 Data Archiving in Ecology and Evolution: Best Practices. *Trends in Ecology and Evolution* 26(2):61–65.

Wicherts, Jelte M., Denny Borsboom, Judith Kats, and Dylan Molenaar
2006 The Poor Availability of Psychological Research Data for Reanalysis. *American Psychologist* 61(7):726–728.

Wilkinson, Mark D., Michel Dumontier, Isbrand J. Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, and Barend Mons
2016 The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3: 160018. DOI:10.1038/sdata.2016.18, accessed September 7, 2017.

Womack, Ryan P.
2015 Research Data in Core Journals in Biology, Chemistry, Mathematics, and Physics. *PLOS ONE* 10(12): e0143460. DOI:10.1371/journal.pone.0143460, accessed September 7, 2017.

Wouters, Paul, and Wouter Haak
2017 Open Data Report: The Researcher Perspective. Electronic document, https://www.elsevier.com/about/open-science/research-data/open-data-report, accessed October 11, 2017.

Zimmerman, Larry J., Karen Vitelli, and Julie Hollowell-Zimmer (editors)
2003 *Ethical Issues in Archaeology*. Rowman AltaMira, New York.

AUTHORS INFORMATION

Ben Marwick ■ University of Washington, Seattle, WA 98115, USA (bmarwick@uw.edu, corresponding author)

Suzanne E. Pilaar Birch ■ University of Georgia, Athens, GA 30602, USA

https://doi.org/10.1017/aap.2018.3 Published online by Cambridge University Press