



Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Essay

Forgotten treasures: the fate of data in animal behaviour studies



Daniel S. Caetano^{a, *}, Anita Aisenberg^{b, 1}

^a Department of Biological Sciences, University of Idaho, Moscow, ID, U.S.A.

^b Laboratorio de Etología, Ecología y Evolución, Instituto de Investigaciones Biológicas Clemente Estable, Avenida Italia, Montevideo, Uruguay

ARTICLE INFO

Article history:

Received 2 May 2014

Initial acceptance 9 July 2014

Final acceptance 26 August 2014

Available online 15 October 2014

MS. number: AE-14-00366R

Keywords:

data reuse
data sharing
database
ontology
science policy

The majority of publications across many disciplines in biology do not make their data available in repositories. On the other hand, there are several advantages associated with archiving data in open access repositories, and technological resources are available to do so. To date, no study has estimated the frequency of data from animal behaviour publications made available in digital repositories or supplementary materials. To evaluate how much data from those studies are available, we surveyed publications from two well-known journals in the field. Since journal policy is an important factor influencing the availability of data from publications, we list the data policy of each journal that publishes animal behaviour research. We found that only a small proportion of the articles published in 2013 made even part of their data available and that journals do not require data to be archived prior to or after publication. If not deposited in repositories, data supporting those findings will most likely be lost from lack of usage, inability to access obsolete storage devices or even the death of the authors. Although it is difficult to appreciate the relevance of data for future studies at the time of a research article's publication, such data may inspire fruitful opportunities that we cannot afford to lose. We discuss the benefits of making data available, review resources for data archiving and provide practical guidelines. We hope that raising awareness about this problem will help foster a data-sharing culture among animal behaviourists.

© 2014 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

The internet has facilitated scientific communication, and one would expect such advances not only to enhance access to articles but also to enhance access to the data that support them (Costello, 2009). Because of the absence of a data-sharing culture among researchers in the biological sciences, data availability of published articles does not follow this expectation (Heidorn, 2008; Magee, May, & Moore, 2014; Reichman, Jones, & Schildhauer, 2011; Vines, Albert, et al., 2013; Vines, Andrew, et al., 2013; Wolkovich, Regetz, & O'Connor, 2012). Many researchers have reservations about sharing their data in repositories (Costello, 2009; Parr & Cummings, 2005; Pryor, 2009; Roche et al., 2014; Smith, 2009; Tenopir et al., 2011; Wolkovich et al., 2012), and this seems to explain the general lack of data available for publications (Hartter, Ryan, MacKenzie, Parker, & Strasser, 2013; Zamir, 2013; but see Wallis, Rolando, & Borgman, 2013). The main reasons why authors opt to avoid storing data in digital repositories are related to concerns about limited time and appropriate tools to prepare and upload data sets, the potential for data misuse (Whitlock, 2011) and lack of personal benefits (Wolkovich et al., 2012). While access to

digital storage space and limitations to exchange data sets were technical issues in the 1980s and early 1990s, initiatives to promote open science and reproducibility over the last decade have led to increased availability of suitable resources to help manage, archive and share data. Furthermore, there seem to be far more advantages than disadvantages associated with archiving data from publications (Costello, 2009; Koslow, 2002; Piwowar & Vision, 2013; Wolkovich et al., 2012).

Another important factor responsible for the proportion of studies that make their data accessible to peers is journal policy (Vines, 2014). A strict and clear data policy from journals can make a significant improvement in data availability. A series of influential journals in ecology and evolutionary biology adopted the Joint Data Archiving Policy (JDAP: <http://datadryad.org/pages/jdap>), which is a standard journal policy that requires data archiving in a public repository as condition for publication. Journals such as *Evolution* and those published by the *Public Library of Science (PLOS)* have recently improved their data policy by asking authors to report where the data supporting their findings are deposited (Bloom, Ganley, & Winker, 2014; Fairbairn, 2011). Both are examples of requirements known to result in an increase in data available from publications (Magee et al., 2014; Vines, Andrew, et al., 2013).

The frequency of data archiving has been estimated for ecological (Hampton et al., 2013), evolutionary (Drew et al., 2013;

* Correspondence: D. S. Caetano, Department of Biological Sciences, University of Idaho, Moscow, ID 83844, U.S.A.

E-mail address: caetanods1@gmail.com (D. S. Caetano).

¹ E-mail address: anita.aisenberg@gmail.com (A. Aisenberg).

Magee et al., 2014; Stoltzfus et al., 2012; Vines, Andrew, et al., 2013), health sciences (Chan et al., 2014) and molecular studies (Noor, Zimmerman, & Teeter, 2006; Piwowar, 2011). In general, only a small proportion of publications have their data in repositories and authors fail to release their data sets upon direct request. While most of the published surveys are focused on data recorded in standard formats, which facilitate collating data from different studies (e.g. molecular sequences), animal behaviour research produces a wider diversity of data, such as images, ethograms, video and audio recordings. To our knowledge, no study has surveyed the frequency of data from animal behaviour publications made available in digital repositories or supplementary materials. Herein we estimate this proportion for publications from two well-known journals in the field and argue in favour of a data-sharing culture among animal behaviourists.

QUANTIFYING ANIMAL BEHAVIOUR DATA AVAILABILITY

We randomly selected and reviewed one-third of the articles published during 2013 in *Animal Behaviour* (AB, 103 out of 308 articles) and *Behavioral Ecology* (BE, 54 out of 161 articles). We chose to sample from these journals because we recognize them to be among the most influential journals in the field. We searched for database indications (hyperlinks and/or references) in the methods, results and acknowledgments sections of each publication. We recorded whether at least part of the raw data was made available in tables or supplementary material or stored in a digital repository. Summary statistics, parameter estimates, results from tests of significance and effect sizes are the main information needed to understand the findings and conclusions of a scientific publication and to perform meta-analyses (Deeks, Higgins, & Altman, 2008; and see examples in Bell, Hankison, & Laskowski, 2009; Stankowich & Blumstein, 2005). However, we did not include those in this survey, since they cannot be considered as data and do not allow for reproduction of the findings.

Our survey shows that a small proportion of the analysed articles from *Animal Behaviour* (13%) and *Behavioral Ecology* (7%) made at least some portion of their data available (see S1_sampled_articles.csv in <http://dx.doi.org/10.6084/m9.figshare.1003857>). Molecular sequence data have a strong potential for reuse, and archiving this type of data in specialized repositories such as GenBank (Benson et al., 2014) has become common practice (but see Noor et al., 2006). In our survey only two sampled articles used molecular sequence data and both were archived in GenBank. Although our sample is restricted to 1 year, our results are similar to a survey of environmental biology publications over a 5-year period that reported only 8% of articles (sequence data excluded) made their data available (Hampton et al., 2013). Furthermore, Vines, Andrew, et al. (2013) showed a comparable proportion of 7.3% of data available for articles published over a 2-year period in journals that do not have data policy or only recommend, but do not mandate, data to be made available. While Hampton et al. (2013) and the present survey included studies that shared some or all of their data, Vines, Andrew, et al. (2013) were more specific and only sampled articles that used the same method (population genetics analysis with the STRUCTURE software) and made all data available. The proportion of shared data was similar among studies despite differences in the criteria used in each survey. Interestingly, our results did not differ from surveys of research areas that produce data in standard formats, especially genotypes (Vines, Andrew, et al., 2013). This suggests that the likelihood that a data set is made available may be independent of the facility for aggregation of data from different studies.

DATA POLICY OF ANIMAL BEHAVIOUR JOURNALS

The similarity in the proportion of data deposition from such different types of data also suggests that, when not mandated by journals, data availability may be a function of the authors' choice to share their data. This reinforces the notion that journals have an important role in improving the frequency at which data from publications are made accessible (Vines, 2014; Vines, Andrew, et al., 2013).

To estimate how common data policies are with regard to studies of animal behaviour, we reviewed the instructions for authors from all journals classified under the 'behavioral sciences' category of the Journal of Citation Reports database (ISI Web of Science). Although all journals accept supplementary data from a variety of media formats (e.g. sound, video and photos), less than half (34%, $N = 49$) explicitly encourage authors to store data in digital repositories or supplementary files, and requirements for data archiving prior to or after publication are nonexistent (see S2_journals_policies.csv in <http://dx.doi.org/10.6084/m9.figshare.1003857>). None of the animal behaviour journals asks authors to state where the data are made available or adopts a strict data policy (such as the JDAP).

The low frequency of data available for animal behaviour studies may be both a result of the lack of a sharing culture among researchers and the absence of data policy from journals. Even when journals require authors to share their data upon request by the scientific community, only a portion of the authors comply (Alsheikh-Ali, Qureshi, Al-Mallah, & Ioannidis, 2011; Noor et al., 2006; Savage & Vickers, 2009; Vines, Andrew, et al., 2013; Wicherts, Borsboom, Kats, & Molenaar, 2006). Stated reasons why authors do not provide their data sets range from inability to access files stored in obsolete media, data loss (Vines, Andrew, et al., 2013) or even supposed lack of prior knowledge about journal policy (Savage & Vickers, 2009). Unfortunately, data sets not stored in a repository rapidly tend to be lost over time (80% of the data is likely to be lost within 20 years; Vines, Albert, et al., 2013).

THE COST OF LOSING DATA

Failure to store data from animal behaviour studies comes at a big cost. The majority of studies result from the observation of a cohort of individuals in a specific point in time and space (Taborsky, 2010; also see Heidorn, 2008). Behavioural plasticity, geographical variation and environmental fluctuations make the reproducibility of such studies challenging (see discussions in Bissell, 2013; Heidorn, 2008; Reichman et al., 2011). As a result, specific behavioural data not made available in repositories are likely going to be lost. Making data available in online repositories prevents this loss, since data sets accessible to public reuse are more likely to survive in the long term (Gibney, 2013). On the other hand, data stored on private hard drives or local repositories are often lost from disuse (Heidorn, 2008; Wolkovich et al., 2012). Researchers, funding agencies and institutions are more prone to be concerned with large data sets resulting from collaborations and/or associated with long-term projects (Heidorn, 2008). However, the majority of published studies, especially in the animal behaviour sciences, produce smaller data sets because of characteristics of the study system or experimental design. Each data set represents a spatial, temporal or population replicate of importance to future studies, but scientists often fail to recognize such potential at the time of publication (Wolkovich et al., 2012). The heterogeneity of animal behaviour data and potential difficulties to reproduce findings are the main reasons why losing this type of data is of particular concern. If data management plans are based solely on compliance to journal or funding requirements (Costello, 2009), a significant amount of data may

Download English Version:

<https://daneshyari.com/en/article/2416320>

Download Persian Version:

<https://daneshyari.com/article/2416320>

[Daneshyari.com](https://daneshyari.com)