Animal Behaviour 98 (2014) 1–5

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
Published online
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.

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

* 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).

http://dx.doi.org/10.1016/j.anbehav.2014.09.025
0003-3472/

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; 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 always be lost (Savage & Vickers, 2009). Proper data stewardship and sharing is good scientific practice and should therefore not be viewed simply as a mandatory requirement to fulfil (Bolukbasi et al., 2013; Costello, 2009; Piwowar & Vision, 2013).

ALTRUISTIC OR SELFISH BEHAVIOUR? NEITHER ONE, NOR THE OTHER

The archiving of data in digital repositories and metadata management is the responsibility of the authors (see discussion in Roche et al., 2014). At first inspection, this practice seems to be an altruistic behaviour, beneficial to the community with no individual return (Smith, 2009). Contrary to this perception, there are benefits associated with data sharing both at the individual and community level (Costello, 2009; Piwowar & Vision, 2013). Articles with data publicly available receive more citations and are more visible to internet searches (Piwowar & Vision, 2013; Piwowar, Day, & Fridsma, 2007). Data sets from articles are also likely to be indexed by search databases such as DataCite (http://datacite.org) and Data Citation Index (ISI Web of Science, http://wokinfo.com/products_tools/multidisciplinary/dci/), which contribute to research discoverability. In addition, data availability provides transparency to publications. Lack of transparency makes publications vulnerable to acts of scientific misconduct that can damage the credibility of individuals, institutes and funding agencies (Costello, 2009; Couzin, 2006). At the individual level, transparency may increase citations because peers have more confidence in the results reported by the authors (Costello, 2009; Piwowar & Vision, 2013). Therefore, both the scientific community and individual researchers may be rewarded by the establishment of a data-sharing culture in animal behaviour sciences: use of data sets without proper citation would be discouraged; researchers would receive recognition for reused products; and articles would become more reproducible (Costello, Michener, Gahegan, Zhang, & Bourne, 2013; Piwowar, 2013; Wolkovich et al., 2012).

GUIDELINES TO ANIMAL BEHAVIOUR DATA SHARING AND ARCHIVING

Animal behaviour studies often record data in myriad media formats, including images, videos and audio recordings. Although the unique characteristics of such media make archival in a standardized format difficult, there are several digital repositories capable of storing such heterogeneous data sets (Table 1). Most of the repositories do not charge for data deposition. Those repositories also offer a limited amount of private storage space and usually unlimited storage space for released data sets. Scientists may use private storage space to archive data while conducting research. Beyond serving as a reliable backup, this practice improves data management efficiency as data are uploaded and organized prior to publication. Among repositories listed in Table 1, Dryad releases data sets in the public domain under a Creative Commons Zero licence (CC0) and figshare uses CC0 for data and Creative Commons Attribution (CC-BY) for media files. The public domain licence (CC0) waives all legal requirements to attribution of rights to the authors, whereas the attribution licence (CC-BY) requires citation of the original authors, reproduction of copyright notices present in the work and acknowledgment of modifications made to the original work. Both Zenodo and the Macaulay Library offer flexible licence options (see Table 1). Below we list some

Table 1
List of repositories suitable for archiving animal behaviour data

Repository	Link	Access to data	Embargo period	Cost	Licence	File format	File size
Dryad	http://datadryad.org/	Open-access	For selected journals or upon request	Associated fees, with waivers for developing countries	CC0	Any kind	10 GB per data package; additional fees for bigger packages
EthoSearch	http://www.ethosearch.org/	Open-access	None	None	Not stated	Ethograms; media files to help define behaviour	Not stated
figshare	http://figshare.com	Open-access; private storage	None	None	CC0 for data sets and CC-BY for media	Any kind	No limit
Macaulay Library	http://macaulaylibrary.org/	Free download for researchers	Yes; flexible release date policy	None	Flexible copyright agreement	Video and audio recordings	No limit; consult representative
Zenodo	https://zenodo.org/	Open-access; private storage	Yes; release date set by the authors	None	Author chose among Creative Commons licences	Any kind	2 GB per data file; larger files upon request

practical guidelines relevant to researchers preparing data sets to share via digital repositories (also refer to Whitlock, 2011).

(1) Record all metadata. Metadata are information that describe data collection, defines categorizations, specifies data structure and contains everything needed for another researcher to understand and reuse the data. A data set with insufficient metadata can be impossible to reuse (see discussion in Rüegg et al., 2014). Well-constructed metadata allow other researchers to understand thoroughly how the data were collected and, as a result, may facilitate collaboration in future projects.
(2) Implement extensive use of repositories. It is often the case that authors publish only a portion of the data generated by research projects within their articles. As an alternative, authors can ensure that any data not directly related to the published results are available in repositories and assigned to DOIs. These data can be cited as soon as they are made available, foster collaborations, and bring visibility to young scientists. Availability of additional data is of special relevance to animal behaviour studies in which observations of 'rare' behaviours are often not reported. New findings may potentially be uncovered by comparative studies of rare behaviours (Peretti, 2013). However, those initiatives are made impossible because of the lack of accessibility to the data.
(3) Deposit supplementary information in repositories. Many journals provide the option for authors to include supplementary data that are made available in the digital version of the publication. However, repositories increase data discoverability, are more reliable for long-term storage and make it possible to make data sets available under open access even when authors transfer copyright ownership to publishers.
(4) Keep an eye on copyright licences. Data sets and media in publicly accessible repositories are usually shared under the Creative Commons attribution licences (usually CC-BY) or in the public domain (CC0). The Creative Commons organization (https://creativecommons.org/) has extensive information on the different versions of the attribution family of licences. Poisot, Mounce, and Gravel (2013) also provide an interesting discussion on the topic.
(5) Embargo periods to release data. One common concern of sharing data is that a third party could publish findings based on the data set before the original authors. However, data repositories have optional embargo that prevent release of data sets for a specified period after publication of the research. One year seems a reasonable period to assure the 'right of first use' to authors, but it is possible to request longer periods (see discussion in Roche et al., 2014). Specific surveys are needed to estimate reasonable embargo periods for animal behaviour data sets.

FUTURE RECOMMENDATIONS

Databases such as the Global Biodiversity Information Facility (GBIF; www.gbif.org) and GenBank (Benson et al., 2014) archive data under standardized formats, which facilitate reuse and data curation. Geographical distribution coordinates and molecular sequences can be easily aggregated from independent sources. In contrast, animal behaviour is often described in free text format that is richly detailed but lacks standardization, which creates challenges to implement repositories with automated query algorithms (see discussion in Taylor & Knight, 2003). The use of ontologies (i.e. controlled vocabularies) is a means to make behavioural descriptions more comparable (Hoehndorf et al., 2014; Kalueff et al., 2013; Martins, 2004; Midford, 2004). Ontologies are already being used for model systems such as mice to make associations between genes and behaviour (Hoehndorf et al., 2014). Likewise, the Animal Behavior Ontology (ABO; http://pages.iu.edu/~emartins/Ethosource/EthoData/ethodata/; see also Martins, 2004) is under development and will be a source of vocabularies for animal behaviour in general.

Heterogeneity in animal behaviour is not exclusive to descriptions but is also present in the type of data generated. Data sets are more likely to be discoverable and useful if they are included in a broader context (i.e. connected among studies; Parsons et al., 2011). Interoperability can be achieved by consistent description of data objects and data types making use of a structured metadata populated with an ontology vocabulary (Parsons et al., 2011; Reichman et al., 2011). One example of structured metadata is the Ecological Metadata Language (EML), which is mainly applied to ecology and was based on work done by the Ecological Society of America and associated efforts (Higgins, Berkley, & Jones, 2002; Michener, Brunt, Helly, Kirchner, & Stafford, 1997). These examples point to the path towards which animal behaviour, as a discipline, should head. However, success of both metadata standards and ontologies depend upon their usage and collaborative development by the scientific community.

Some authors do not share data because of the effort needed to organize data sets and manage descriptive metadata. Therefore, tools to facilitate data stewardship and mitigate the effort needed to submit data sets are imperative for the establishment of a data-sharing culture (Molloy, 2011). Direct connection with databases could allow researchers to update data from the field alongside basic information such as date, time, geo-reference and weather conditions. The use of software to store data in databases through direct links and animal behaviour ontologies would make data organization and metadata management automatic, thus requiring no further effort from researchers. After they are stored, data sets can be linked to an online portal such as EthoSource (Martins, 2004) for easy access and reuse (see possible models of centralization in Rani & Buckley, 2012). Finally, a sharing culture among researchers would allow access not only to data from animal behaviour but also from other disciplines such as taxonomy, ecology and systematics. This would increase dramatically the availability of information and promote collaboration among areas.

CONCLUSIONS

The frequency with which data from published articles in animal behaviour is made available is extremely low and, as a result, most of the data supporting publications are likely to be lost quickly. Furthermore, the majority of journals do not have any data policy or recommendations for authors to make their data available after publication. While both resources for data sharing and reasons for doing so exist, interest from the scientific community is needed. We hope that raising awareness of data sharing, the mitigation of effort related to data stewardship and adoption of clear data policies (with preference to the JDAP) by journals, will help to foster a data-sharing culture among animal behaviourists.

Acknowledgments

We thank T. H. Kawamoto and F. Machado for fruitful discussions on the topic. B. A. Buzatto, L. E. Costa-Schmidt, A. Espíndola, D. Jochimsen, B. S. Medeiros, D. J. Machado, M. W. Pennell, A. V. Peretti, E. A. Santos, M. Simó, M. Singer and J. C. Uyeda for comments on early versions of the manuscript. We also thank T. H. Vines and an anonymous referee for suggestions to improve this manuscript. D.S.C. is supported by a fellowship from Coordenação de

Aperfeiçoamento de Pessoal de Nível Superior (CAPES – 1093/12-6). A.A. is supported by Programa Desarrollo de Ciencias Básicos (PEDECIBA), UdelaR, and Sistema Nacional de Investigadores (SNI), Agencia Nacional de Investigación e Innovación (ANII).

References

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLoS One, 6*, e24357.

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour, 77*, 771–783.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2014). GenBank. *Nucleic Acids Research, 42*, D32–D37.

Bissell, M. (2013). Reproducibility: the risks of the replication drive. *Nature, 503*, 333–334.

Bloom, T., Ganley, E., & Winker, M. (2014). Data access for the open access literature: *PLoS*'s data policy. *PLoS Biology, 12*, e1001797.

Bolukbasi, B., Berente, N., Cutcher-Gershenfeld, J., Dechurch, L., Flint, C., Haberman, M., et al. (2013). Open data: crediting a culture of cooperation. *Science, 342*, 1041–1042. http://dx.doi.org/10.1126/science.342.6162.1041-b.

Chan, A. W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P. C., et al. (2014). Increasing value and reducing waste: addressing inaccessible research. *Lancet, 383*, 257–266.

Costello, M. J. (2009). Motivating online publication of data. *BioScience, 59*, 418–427.

Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z.-Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution, 28*, 454–461.

Couzin, J. (2006). Truth and consequences. *Science, 313*, 1222–1226.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 243–296). Chichester, U.K.: J. Wiley.

Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., et al. (2013). Lost branches on the tree of life. *PLoS Biology, 11*, e1001636.

Fairbairn, D. J. (2011). The advent of mandatory data archiving. *Evolution, 65*, 1–2.

Gibney, E. (2013). LHC plans for open data future. *Nature, 503*, 447. http://dx.doi.org/10.1038/503447a.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., et al. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment, 11*, 156–162.

Hartter, J., Ryan, S. J., MacKenzie, C. A., Parker, J. N., & Strasser, C. A. (2013). Spatially explicit data: stewardship and ethical challenges in science. *PLoS Biology, 11*, e1001634.

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends, 57*, 280–299.

Higgins, D., Berkley, C., & Jones, M. B. (2002). Managing heterogeneous ecological data using Morpho. In *Proceedings of the14th International Conference on Scientific and Statistical Database Management, Edinburgh, U.K.* (pp. 69–76). Washington, D.C.: IEEE Computer Society. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1029707.

Hoehndorf, R., Hancock, J. M., Hardy, N. W., Mallon, A.-M., Schofield, P. N., & Gkoutos, G. V. (2014). Analyzing gene expression data in mice with the Neuro Behavior Ontology. *Mammalian Genome, 25*, 32–40.

Kalueff, A. V., Gebhardt, M., Stewart, A. M., Cachat, J. M., Brimmer, M., Chawla, J. S., et al. (2013). Towards a comprehensive catalog of zebrafish behavior 1.0 and beyond. *Zebrafish, 10*, 70–86.

Koslow, S. H. (2002). Sharing primary data: a threat or asset to discovery? *Nature Reviews Neuroscience, 3*, 311–313.

Magee, A. F., May, M. R., & Moore, B. R. (2014). *The dawn of open access to phylogenetic data. arXiv:1405.6623 [cs, q-bio]*. Retrieved from http://arxiv.org/abs/1405.6623.

Martins, E. P. (2004). EthoSource: storing, sharing, and combining behavioral data. *BioScience, 54*, 886–887. http://dx.doi.org/10.1641/0006-3568(2004)054[0886:ESSACB]2.0.CO;2.

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., & Stafford, S. G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications, 7*, 330–342.

Midford, P. E. (2004). Ontologies for behavior. *Bioinformatics, 20*, 3700–3701. http://dx.doi.org/10.1093/bioinformatics/bth433.

Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *PLoS Biology, 9*, e1001195. http://dx.doi.org/10.1371/journal.pbio.1001195.

Noor, M. A. F., Zimmerman, K. J., & Teeter, K. C. (2006). Data sharing: how much doesn't get submitted to GenBank? *PLoS Biology, 4*, e228. http://dx.doi.org/10.1371/journal.pbio.0040228.

Parr, C. S., & Cummings, M. P. (2005). Data sharing in ecology and evolution. *Trends in Ecology & Evolution, 20*, 362–363.

Parsons, M. A., Godøy, Ø., LeDrew, E., de Bruin, T. F., Danis, B., Tomlinson, S., et al. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science, 37*, 555–569.

Peretti, A. V. (2013). Sexual selection in Neotropical species: rules and exceptions. In R. Macedo, & G. Machado (Eds.), *Sexual selection: Perspectives and models from the Neotropics* (pp. 33–52). Amsterdam, The Netherlands: Elsevier.

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS One, 6*, e18657.

Piwowar, H. A. (2013). Altmetrics: value all research products. *Nature, 493*, 159.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One, 2*, e308.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ, 1*, e175.

Poisot, T., Mounce, R., & Gravel, D. (2013). Moving toward a sustainable ecological science: don't let data go to waste! *Ideas in Ecology and Evolution, 6*, 11–19.

Pryor, G. (2009). Multi-scale data sharing in the life sciences: some lessons for policy makers. *International Journal of Digital Curation, 4*, 71–82.

Rani, M., & Buckley, B. S. (2012). Systematic archiving and access to health research data: rationale, current status and way forward. *Bulletin of the World Health Organization, 90*, 932–939.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science, 331*, 703–705.

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., et al. (2014). Troubleshooting public data archiving: suggestions to increase participation. *PLoS Biology, 12*, e1001779.

Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G. J., Felzer, B. S., McIntyre, N. E., et al. (2014). Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment, 12*, 24–30.

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One, 4*, e7078.

Smith, V. S. (2009). Data publication: towards a database of everything. *BMC Research Notes, 2*, 113. http://dx.doi.org/10.1186/1756-0500-2-113. http://www.biomedcentral.com/1756-0500/2/113.

Stankowich, T., & Blumstein, D. T. (2005). Fear in animals: a meta-analysis and review of risk assessment. *Proceedings of the Royal Society B: Biological Sciences, 272*, 2627–2634.

Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E., Kumar, S., et al. (2012). Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes, 5*, 574. http://dx.doi.org/10.1186/1756-0500-5-574. http://www.biomedcentral.com/1756-0500/5/574.

Taborsky, M. (2010). Sample size in the study of behaviour. *Ethology, 116*, 185–202.

Taylor, A. R., & Knight, R. L. (2003). Behavioral responses of wildlife to human activity: terminology and methods. *Wildlife Society Bulletin, 31*, 1263–1271.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS One, 6*, e21101.

Vines, T. H. (2014). Journals must boost data sharing. *Nature, 508*, 44.

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., et al. (2013a). The availability of research data declines rapidly with article age. *Current Biology, 24*, 94–97.

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., et al. (2013b). Mandated data archiving greatly improves access to research data. *FASEB Journal, 27*, 1304–1308.

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One, 8*, e67332.

Whitlock, M. C. (2011). Data archiving in ecology and evolution: best practices. *Trends in Ecology & Evolution, 26*, 61–65.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*, 726–728.

Wolkovich, E. M., Regetz, J., & O'Connor, M. I. (2012). Advances in global change research require open science by individual researchers. *Global Change Biology, 18*, 2102–2110.

Zamir, D. (2013). Where have all the crop phenotypes gone? *PLoS Biology, 11*, e1001595.