

PERSPECTIVE

Gauging the Purported Costs of Public Data Archiving for Long-Term Population Studies

Simon Robin Evans*

Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

* simon.evans@ieu.uzh.ch

Abstract

It was recently proposed that long-term population studies be exempted from the expectation that authors publicly archive the primary data underlying published articles. Such studies are valuable to many areas of ecological and evolutionary biological research, and multiple risks to their viability were anticipated as a result of public data archiving (PDA), ultimately all stemming from independent reuse of archived data. However, empirical assessment was missing, making it difficult to determine whether such fears are realistic. I addressed this by surveying data packages from long-term population studies archived in the Dryad Digital Repository. I found no evidence that PDA results in reuse of data by independent parties, suggesting the purported costs of PDA for long-term population studies have been overstated.



OPEN ACCESS

Citation: Evans SR (2016) Gauging the Purported Costs of Public Data Archiving for Long-Term Population Studies. *PLoS Biol* 14(4): e1002432. doi:10.1371/journal.pbio.1002432

Published: April 8, 2016

Copyright: © 2016 Simon Robin Evans. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the Swiss National Science Foundation (31003A_159462; <http://p3.snf.ch/project-159462>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

Abbreviations: JDAP, Joint Data Archiving Policy; PDA, public data archiving; PI, principal investigator.

Data are the foundation of the scientific method, yet individual scientists are evaluated via novel analyses of data, generating a potential conflict of interest between a research field and its individual participants that is manifested in the debate over access to the primary data underpinning published studies [1–5]. This is a chronic issue but has become more acute with the growing expectation that researchers publish the primary data underlying research reports (i.e., public data archiving [PDA]). Studies show that articles publishing their primary data are more reliable and accrue more citations [6,7], but a recent opinion piece by Mills et al. [2] highlighted the particular concerns felt by some principal investigators (PIs) of long-term population studies regarding PDA, arguing that unique aspects of such studies render them unsuitable for PDA. The "potential costs to science" identified by Mills et al. [2] as arising from PDA are as follows:

- Publication of flawed research resulting from a "lack of understanding" by independent researchers conducting analyses of archived data
- Time demands placed on the PIs of long-term population studies arising from the need to correct such errors via, e.g., published rebuttals
- Reduced opportunities for researchers to obtain the skills needed for field-based data collection because equivalent long-term population studies will be rendered redundant

- Reduced number of collaborations
- Inefficiencies resulting from repeated assessment of a hypothesis using a single dataset

Each "potential cost" is ultimately predicated on the supposition that reuse of archived long-term population data is common, yet the extent to which this is true was not evaluated. To assess the prevalence of independent reuse of archived data—and thereby examine whether the negative consequences of PDA presented by Mills et al. [2] may be realised—I surveyed datasets from long-term population studies archived in the Dryad Digital Repository (hereafter, Dryad). Dryad is an online service that hosts data from a broad range of scientific disciplines, but its content is dominated by submissions associated with ecological and evolutionary biological research [8]. I examined all the Dryad packages associated with studies from four journals featuring ecological or evolutionary research: *The American Naturalist*, *Evolution*, *Journal of Evolutionary Biology*, and *Proceedings of the Royal Society B: Biological Sciences* (the latter referred to hereafter as *Proceedings B*). These four journals together represent 23.3% of Dryad's contributed packages (as of early February 2016). Mills et al. [2] refer to short- versus long-term studies but do not provide a definition of this dichotomy. However, the shortest study represented by their survey lasted for 5 years, so I used this as the minimum time span for inclusion in my survey. This cut-off seems reasonable, as it will generally exclude studies resulting from single projects, such that included datasets likely relate to studies resulting from a sustained commitment on the part of researchers—although one included package contains data gathered via "citizen science" [9], and two others contain data derived from archived human population records [10,11]. However, as these datasets cover extended time spans and were used to address ecological questions [12–14], they were retained in my survey sample. Following Mills et al. [2], my focus was on population studies conducted in natural (or seminatural) settings, so captive populations were excluded. Because I was assessing the reuse of archived data, I excluded packages published by Dryad after 2013: authors can typically opt to impose a 1-year embargo, and articles based on archived data will themselves take some time to be written and published.

Of the 1,264 archived data packages linked to one of the four journals and published on the Dryad website before 2014, 72 were identified as meeting the selection criteria. This sample represents a diverse range of taxa (Fig 1) and is comparable to the 73 studies surveyed by Mills et al. [2], although my methodology permits individual populations to be represented more than once, since the survey was conducted at the level of published articles (S1 Table). Of these 72 data packages, five had long-term embargoes remaining active (three packages with 5-year embargoes [15–17]; two packages with 10-year embargoes [18,19]). For two of these [17,19], the time span of the study could not be estimated because this information is not provided in the associated articles [20,21]. For a third package [22], the archived data indicated 10 years were represented (dummy coding was used to disguise factor level identities, including for year), yet the text of the associated paper suggests data collection covered a considerably greater time span [23]. However, since the study period is not stated in the text, I followed the archived data [22] in assuming data collection spanned a 10-year period. The distribution of study time spans is shown in Fig 2.

For each year from 2000 to 2004, these four journals contributed no more than a single data package to Dryad between them. However, around the time that the Joint Data Archiving Policy (JDAP; [24]) was adopted by three of these, we see a surge in PDA by ecologists and evolutionary biologists (Fig 3), such that in 2015 these four journals were collectively represented by 709 data packages. Of course, Mills et al. [2] argue against mandatory archiving of primary data for long-term studies in particular. For this subset of articles published in these four journals, the same pattern is observed: prior to adoption of the JDAP, only two data packages associated with long-term studies had been archived in Dryad, but following the implementation of

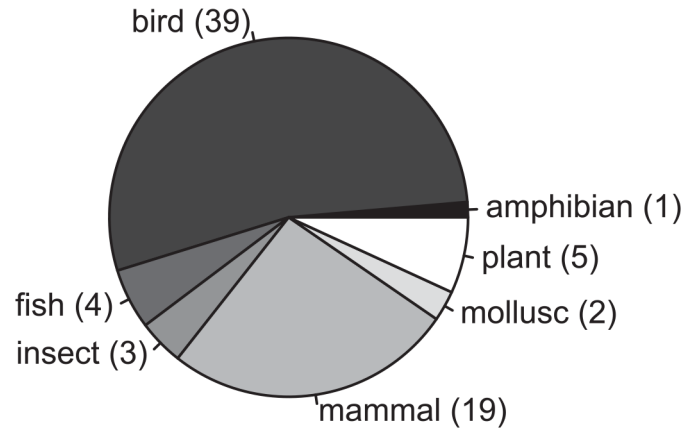


Fig 1. Taxonomic representation of the 72 data packages included in the survey. The number of packages for each taxon is given in parentheses (note: one data package included data describing both insects and plants [9], while other data packages represented multiple species within a single taxonomic category).

doi:10.1371/journal.pbio.1002432.g001

the JDAP as a condition of publication in *The American Naturalist*, *Evolution*, and *Journal of Evolutionary Biology*, there is a rapid increase in the number of data packages being archived, despite the continuing availability of alternative venues should authors wish to avoid the purported costs of PDA as Mills et al. [2] contend. As the editorial policy of *Proceedings B* has shifted towards an increasingly strong emphasis on PDA (it is now mandatory), there has similarly been an increase in the representation of articles from this journal in Dryad, both overall (Fig 3) and for long-term studies in particular (Fig 4). These observations suggest that authors rarely chose to publicly archive their data prior to the adoption of PDA policies by journals and that uptake of PDA spread rapidly once it became a prerequisite for publication. In this respect, researchers using long-term population studies are no different to those in other scientific fields, despite the assertion by Mills et al. [2] that they are a special case owing to the complexity of their data. In reality, researchers in many other scientific disciplines also seek to identify relationships within complex systems. Within neuroscience, for example, near-identical objections to PDA were raised at the turn of the century [25], while archiving of genetic and

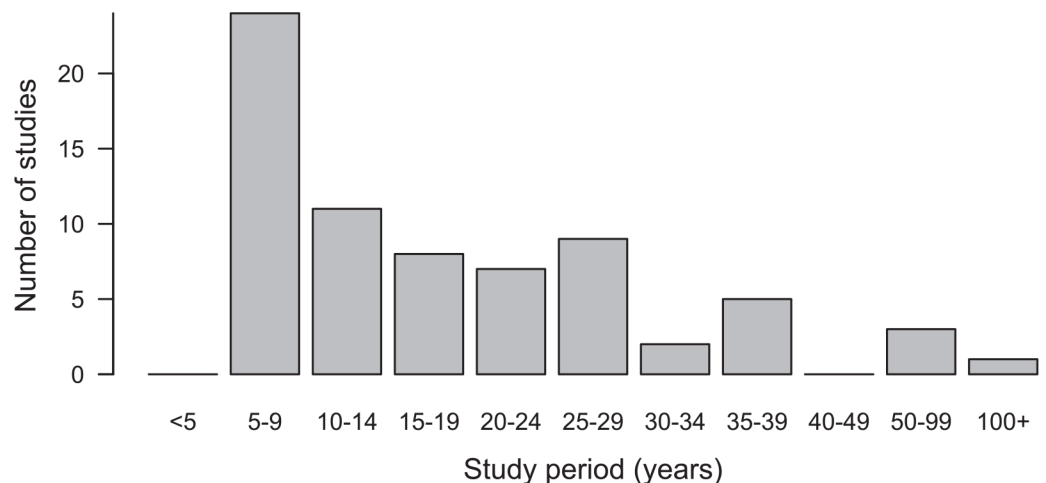


Fig 2. The study periods of the 70 data packages included in the survey for which this could be calculated.

doi:10.1371/journal.pbio.1002432.g002

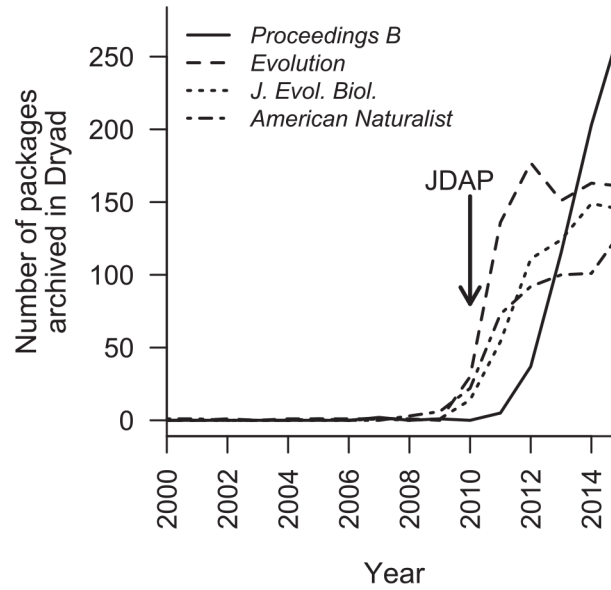


Fig 3. Total number of data packages archived in the Dryad Digital Repository each year for four leading journals within ecology and evolutionary biology. Arrow indicates when the Joint Data Archiving Policy (JDAP) was adopted by *Evolution*, *Journal of Evolutionary Biology*, and *The American Naturalist*. Note that because data packages are assigned a publication date by Dryad prior to journal publication (even if an embargo is imposed), some data packages will have been published in the year preceding the journal publication of their associated article.

doi:10.1371/journal.pbio.1002432.g003

protein sequences by molecular biologists has yielded huge advances but was similarly resisted until revised journal policies stimulated a change in culture [1,26].

A primary concern raised by opponents of PDA is that sharing their data will see them “scooped” by independent researchers [6,8,27–30]. To quantify this risk for researchers maintaining long-term population studies, I used the Web of Science (wok.mimas.ac.uk) to search for citations of each data package (as of November 2015). For the 67 Dryad packages that were

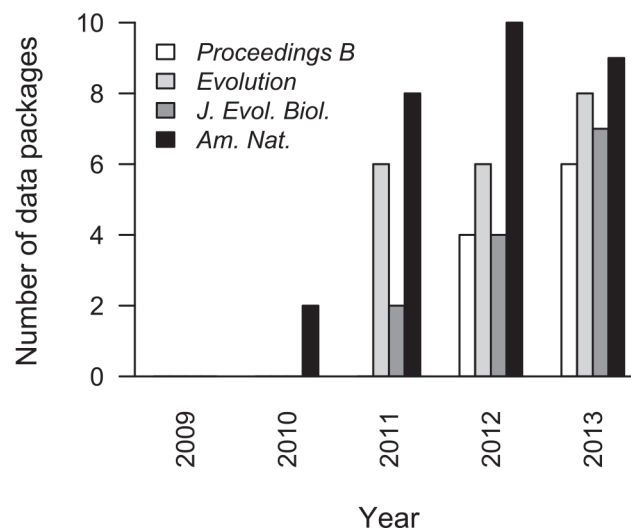


Fig 4. Publication dates of the 72 data packages from long-term study populations that were included in the survey.

doi:10.1371/journal.pbio.1002432.g004

publicly accessible, none were cited by any article other than that from which it was derived. However, archived data could conceivably have been reused without the data package being cited, so I examined all journal articles that cited the study report associated with each data package (median citation count: 9; range: 0–58). Although derived metrics from the main articles were occasionally included in quantitative reviews [31,32] or formal meta-analyses [33], I again found no examples of the archived data being reused by independent researchers. As a third approach, I emailed the corresponding author(s) listed for each article, to ask if they were themselves aware of any examples. The replies I received ($n = 35$) confirmed that there were no known cases of long-term population data being independently reused in published articles. The apparent concern of some senior researchers that PDA will see them "collect data for 30 years just to be scooped" [30] thus lacks empirical support. It should also be noted that providing primary data upon request precedes PDA as a condition of acceptance for most major scientific journals [8]. PDA merely serves to ensure that authors meet this established commitment, a step made necessary by the failure rate that is otherwise observed, even after the recent revolution in communications technology [34–36]. As my survey shows, in practice the risk of being scooped is a monster under the bed: empirical assessment fails to justify the level of concern expressed. While long-term population studies are unquestionably a highly valuable resource for ecologists [2,37–39] and will likely continue to face funding challenges [37–39], there is no empirical support for the contention of Mills et al. [2] that PDA threatens their viability, although this situation may deserve reassessment in the future if the adoption of PDA increases within ecology and evolutionary biology. Nonetheless, in the absence of assessments over longer time frames (an inevitable result of the historical reluctance to adopt PDA), my survey results raise doubts over the validity of arguments favouring extended embargoes for archived data [29,40], and particularly the suggestion that multidecadal embargoes should be facilitated for long-term studies [2,41].

Authors frequently assert that unique aspects of their long-term study render it especially well suited to addressing particular issues. Such claims contradict the suggestion that studies will become redundant if PDA becomes the norm [2] while simultaneously highlighting the necessity of making primary data available for meaningful evaluation of results. For research articles relying on data collected over several decades, independent replication is clearly impractical, such that reproducibility (the ability for a third party to replicate the results exactly [42]) is rendered all the more crucial. Besides permitting independent validation of the original results, PDA allows assessment of the hypotheses using alternative analytical methods (large datasets facilitate multiple analytical routes to test a single biological hypothesis, which likely contributes to poor reproducibility [43]) and reassessment if flaws in the original methodology later emerge [44]. Although I was not attempting to use archived data to replicate published results, and thus did not assess the contents of each package in detail, at least six packages [10,45–49] failed to provide the primary data underlying their associated articles, including a quantitative genetic study [50] for which only pedigree information was archived [47]. This limits exploration of alternative statistical approaches to the focal biological hypothesis and impedes future applications of the data that may be unforeseeable by the original investigators (a classic example being Bumpus' [51] dataset describing house sparrow survival [52]), but it seems to be a reality of PDA within ecology and evolution at present [53].

The "solutions" proffered by Mills et al. [2] are, in reality, alternatives to PDA that would serve to maintain the status quo with respect to data accessibility for published studies (i.e., subject to consent from the PI). This is a situation that is widely recognised to be failing with respect to the availability of studies' primary data [34–36,54]. Indeed, for 19% (13 of 67 non-embargoed studies) of the articles represented in my survey, the correspondence email addresses were no longer active, highlighting how rapidly access to long-term primary data can be

passively lost. It is unsurprising, then, that 95% of scientists in evolution and ecology are reportedly in favour of PDA [1]. Yet, having highlighted the value and irreplaceability of data describing long-term population studies, Mills et al. [2] reject PDA in favour of allowing PIs to maintain postpublication control of primary data, going so far as to discuss the possibility of data being copyrighted. Such an attitude risks inviting public ire, since asserting private ownership ignores the public funding that likely enabled data collection, and is at odds with a Royal Society report urging scientists to "shift away from a research culture where data is viewed as a private preserve" [55]. I contend that primary data would better be considered as an intrinsic component of a published article, alongside the report appearing in the pages of a journal that presents the data's interpretation. In this way, an article would move closer to being a self-contained product of research that is fully accessible and assessable. For issues that can only be addressed using data covering an extended time span [2,37–39], excusing long-term studies from the expectation of publishing primary data would potentially render the PIs as unaccountable gatekeepers of scientific consensus. PDA encourages an alternative to this and facilitates a change in the treatment of published studies, from the system of preservation (in which a study's contribution is fixed) that has been the historical convention, towards a conservation approach (in which support for hypotheses can be reassessed and updated) [56]. Given the fundamentally dynamic nature of science, harnessing the storage potential enabled by the Information Age to ensure a study's contribution can be further developed or refined in the future seems logical and would benefit both the individual authors (through enhanced citations and reputation) and the wider scientific community.

The comparison Mills et al. [2] draw between PIs and pharmaceutical companies in terms of how their data are treated is inappropriate: whereas the latter bear the financial cost of developing a drug, a field study's costs are typically covered by the public purse, such that the personal risks of a failed project are largely limited to opportunity costs. It is inconsistent to highlight funding challenges [2,37] while simultaneously acting to inhibit maximum value for money being derived from funded studies. Several of the studies represented in the survey by Mills et al. [2] comfortably exceed a 50-year time span, highlighting the possibility that current PIs are inheritors rather than initiators of long-term studies. In such a situation, arguments favouring the rights of the PI to maintain control of postpublication access to primary data are weakened still further, given that the data may be the result of someone else's efforts. Indeed, given the undoubted value of long-term studies for ecological and evolutionary research [2,37,39], many of Mills et al.'s [2] survey respondents will presumably hope to see these studies continue after their own retirement. Rather than owners of datasets, then, perhaps PIs of long-term studies might better be considered as custodians, such that—to adapt the slogan of a Swiss watchmaker—"you never really own a long-term population study; you merely look after it for the next generation."

Supporting Information

S1 Table. Details of the 72 data packages (and their associated articles) included in the survey.

(TXT)

S2 Table. The number of data packages archived in the Dryad Digital Repository each year from 2000 to 2015 for four leading journals within ecology and evolutionary biology (*The American Naturalist*, *Proceedings of the Royal Society B: Biological Sciences*, *Evolution*, and *Journal of Evolutionary Biology*).

(TXT)

Acknowledgments

I am grateful to all who engaged with me in discussing the pros and cons of PDA, particularly E. Postma and E. Cole for comments on an earlier version of the manuscript.

References

1. Whitlock MC, McPeck MA, Rausher MD, Rieseberg L, Moore AJ. Data archiving. *American Naturalist*. 2010; 175(2):145–6. doi: [10.1086/650340](https://doi.org/10.1086/650340) PMID: [20073990](https://pubmed.ncbi.nlm.nih.gov/20073990/)
2. Mills JA, Teplitsky C, Arroyo B, Charmantier A, Becker PH, Birkhead TR, et al. Archiving primary data: solutions for long-term studies. *Trends in Ecology & Evolution*. 2015; 30(10):581–9. doi: [10.1016/j.tree.2015.07.006](https://doi.org/10.1016/j.tree.2015.07.006)
3. Hannay T. A new kind of science? *Nature Physics*. 2011; 7(10):742. doi: [10.1038/nphys2109](https://doi.org/10.1038/nphys2109)
4. Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J. Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*. 2008; 17:167–78. doi: [10.1111/j.1365-294X.2007.03413.x](https://doi.org/10.1111/j.1365-294X.2007.03413.x) PMID: [18173499](https://pubmed.ncbi.nlm.nih.gov/18173499/)
5. Longo DL, Drazen JM. Data Sharing. *The New England Journal of Medicine*. 2016; 374(3):276–7. doi: [10.1056/NEJMe1516564](https://doi.org/10.1056/NEJMe1516564) PMID: [26789876](https://pubmed.ncbi.nlm.nih.gov/26789876/)
6. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. *PeerJ*. 2013; 1:e175. doi: [10.7717/peerj.175](https://doi.org/10.7717/peerj.175) PMID: [24109559](https://pubmed.ncbi.nlm.nih.gov/24109559/)
7. Wicherts JM, Bakker M, Molenaar D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*. 2011; 6(11):e26828. doi: [10.1371/journal.pone.0026828](https://doi.org/10.1371/journal.pone.0026828) PMID: [22073203](https://pubmed.ncbi.nlm.nih.gov/22073203/)
8. Kenall A, Harold S, Foote C. An open future for ecological and evolutionary data? *BMC Ecology*. 2014; 14:10. doi: [10.1186/1472-6785-14-10](https://doi.org/10.1186/1472-6785-14-10) PMID: [24690219](https://pubmed.ncbi.nlm.nih.gov/24690219/)
9. Phillimore AB, Stålhandske S, Smithers RJ, Rodolphe B. Data from: Dissecting the contributions of plasticity and local adaptation to the phenology of a butterfly and its host plants. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.733d9>.
10. Bolund E, Bouwhuis S, Pettay JE, Lummaa V. Data from: Divergent selection on, but no genetic conflict over, female and male timing and rate of reproduction in a human population. 2013. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.4h071>.
11. Bürkli A, Postma E. Data from: Genetic constraints underlying human reproductive timing in a premodern Swiss village. 2013. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3d2m3>.
12. Bolund E, Bouwhuis S, Pettay JE, Lummaa V. Divergent selection on, but no genetic conflict over, female and male timing and rate of reproduction in a human population. *Proceedings of the Royal Society B-Biological sciences*. 2013; 280(1772):20132002. doi: [10.1098/rspb.2013.2002](https://doi.org/10.1098/rspb.2013.2002)
13. Bürkli A, Postma E. Genetic constraints underlying human reproductive timing in a premodern Swiss village. *Evolution*. 2013; 68(2):526–37. Dryad Digital Repository. doi: [10.1111/evo.12287](https://doi.org/10.1111/evo.12287) PMID: [24117466](https://pubmed.ncbi.nlm.nih.gov/24117466/)
14. Phillimore AB, Stålhandske S, Smithers RJ, Bernard R. Dissecting the contributions of plasticity and local adaptation to the phenology of a butterfly and its host plants. *American Naturalist*. 2012; 180(5):655–70. doi: [10.1086/667893](https://doi.org/10.1086/667893) PMID: [23070325](https://pubmed.ncbi.nlm.nih.gov/23070325/)
15. Bischof R, Loe LE, Meisingset EL, Zimmermann B, Moorter BV, Mysterud A. Data from: A migratory northern ungulate in the pursuit of spring: jumping or surfing the green wave? 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3hr2c>.
16. Lebigre C, Arcese P, Sardell RJ, Keller LF, Reid JM. Data from: Extra-pair paternity and the variance in male fitness in song sparrows (*Melospiza melodia*). 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.6sb403hc>.
17. Stopher KV, Walling CA, Morris A, Guinness FE, Clutton-Brock TH, Pemberton JM, et al. Data from: Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.jf04r362>.
18. Morrissey MB, Parker DJ, Korsten P, Pemberton JM, Kruuk LEB, Wilson AJ. Data from: The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.td6tm4f6>.
19. Morrissey MB, Walling CA, Wilson AJ, Pemberton JM, Clutton-Brock TH, Kruuk LEB. Data from: Genetic analysis of life-history constraint and evolution in a wild ungulate population. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3ff10120>.

20. Morrissey MB, Walling CA, Wilson AJ, Pemberton JM, Clutton-Brock TH, Kruuk LEB. Genetic analysis of life-history constraint and evolution in a wild ungulate population. *American Naturalist*. 2012; 179(4): E97–E114. doi: [10.1086/664686](https://doi.org/10.1086/664686) PMID: [22437186](https://pubmed.ncbi.nlm.nih.gov/22437186/)
21. Stopher KV, Walling CA, Morris A, Guinness FE, Clutton-Brock TH, Pemberton JM, et al. Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *Evolution*. 2012; 66(8):2411–26. doi: [10.1111/j.1558-5646.2012.01620.x](https://doi.org/10.1111/j.1558-5646.2012.01620.x) PMID: [22834741](https://pubmed.ncbi.nlm.nih.gov/22834741/)
22. Husby A, Schielzeth H, Forstmeier W, Gustafsson L, Qvarnström A. Data from: Sex chromosome linked genetic variance and the evolution of sexual dimorphism of quantitative traits. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.451n7>.
23. Husby A, Schielzeth H, Forstmeier W, Gustafsson L, Qvarnström A. Sex chromosome linked genetic variance and the evolution of sexual dimorphism of quantitative traits. *Evolution*. 2013; 67(3):609–19. doi: [10.1111/j.1558-5646.2012.01806.x](https://doi.org/10.1111/j.1558-5646.2012.01806.x) PMID: [23461313](https://pubmed.ncbi.nlm.nih.gov/23461313/)
24. Dryad Digital Repository. Joint Data Archiving Policy (JDAP) 2014. <http://datadryad.org/pages/jdap>.
25. Koslow SH. Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*. 2000; 3(9):863–5. PMID: [10966615](https://pubmed.ncbi.nlm.nih.gov/10966615/)
26. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. *Frontiers in Ecology and the Environment*. 2013; 11(3):156–62. doi: [10.1890/120103](https://doi.org/10.1890/120103)
27. Caetano DS, Aisenberg A. Forgotten treasures: the fate of data in animal behaviour studies. *Animal Behaviour*. 2014; 98:1–5. doi: [10.1016/j.anbehav.2014.09.025](https://doi.org/10.1016/j.anbehav.2014.09.025)
28. Costello MJ. Motivating online publication of data. *BioScience*. 2009; 59(5):418–27. doi: [10.1525/bio.2009.59.5.9](https://doi.org/10.1525/bio.2009.59.5.9)
29. Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, et al. Troubleshooting public data archiving: suggestions to increase participation. *PLoS Biol*. 2014; 12(1):e1001779. doi: [10.1371/journal.pbio.1001779.g001](https://doi.org/10.1371/journal.pbio.1001779.g001) PMID: [24492920](https://pubmed.ncbi.nlm.nih.gov/24492920/)
30. Shawkey MD. 2015. <https://twitter.com/mdshawkey/status/649620776367296513>.
31. Brommer JE. Variation in plasticity of personality traits implies that the ranking of personality measures changes between environmental contexts: calculating the cross-environmental correlation. *Behavioral Ecology and Sociobiology*. 2013; 67(10):1709–18. doi: [10.1007/s00265-013-1603-9](https://doi.org/10.1007/s00265-013-1603-9)
32. Villellas J, Doak DF, Garcia MB, Morris WF. Demographic compensation among populations: what is it, how does it arise and what are its implications? *Ecology Letters*. 2015; 18:1139–52. doi: [10.1111/ele.12505](https://doi.org/10.1111/ele.12505)
33. Miller JM, Coltman DW. Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta-analysis. *Molecular Ecology*. 2014; 23(8):1899–909. doi: [10.1111/mec.12707](https://doi.org/10.1111/mec.12707) PMID: [24581039](https://pubmed.ncbi.nlm.nih.gov/24581039/)
34. Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, et al. Mandated data archiving greatly improves access to research data. *The FASEB Journal*. 2013; 27:1304–8. doi: [10.1096/fj.12-218164](https://doi.org/10.1096/fj.12-218164) PMID: [23288929](https://pubmed.ncbi.nlm.nih.gov/23288929/)
35. Magee AF, May MR, Moore BR. The dawn of open access to phylogenetic data. *PLoS ONE*. 2014; 9(10):e110268. doi: [10.1371/journal.pone.0110268.g001](https://doi.org/10.1371/journal.pone.0110268.g001) PMID: [25343725](https://pubmed.ncbi.nlm.nih.gov/25343725/)
36. Wicherts JM, Borsboom D, Kats J, Molenaar D. The poor availability of psychological research data for reanalysis. *American Psychologist*. 2006; 61(7):726–8. doi: [10.1037/0003-066X.61.7.726](https://doi.org/10.1037/0003-066X.61.7.726) PMID: [17032082](https://pubmed.ncbi.nlm.nih.gov/17032082/)
37. Birkhead T. Stormy outlook for long-term ecology studies. *Nature*. 2014; 514:405. doi: [10.1038/514405a](https://doi.org/10.1038/514405a) PMID: [25341754](https://pubmed.ncbi.nlm.nih.gov/25341754/)
38. Clutton-Brock TH, Sheldon BC. The seven ages of *Pan*. *Science*. 2010; 327:1207–8. doi: [10.1126/science.1187796](https://doi.org/10.1126/science.1187796) PMID: [20203037](https://pubmed.ncbi.nlm.nih.gov/20203037/)
39. Clutton-Brock T, Sheldon BC. Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology and Evolution*. 2010; 25(10):562–73. doi: [10.1016/j.tree.2010.08.002](https://doi.org/10.1016/j.tree.2010.08.002) PMID: [20828863](https://pubmed.ncbi.nlm.nih.gov/20828863/)
40. Whitlock MC, Bronstein JL, Bruna EM, Ellison AM, Fox CW, McPeck MA, et al. A balanced data archiving policy for long-term studies. *Trends in Ecology & Evolution*. 2016; 31(2):84–5. doi: [10.1016/j.tree.2015.12.001](https://doi.org/10.1016/j.tree.2015.12.001)
41. Mills JA, Teplitsky C, Arroyo B, Charmantier A, Becker PH, Birkhead TR, et al. Solutions for archiving data in long-term studies: a reply to Whitlock *et al*. *Trends in ecology & evolution*. 2016; 31(2):85–7. doi: [10.1016/j.tree.2015.12.004](https://doi.org/10.1016/j.tree.2015.12.004)
42. Cassey P, Blackburn DC. Reproducibility and repeatability in ecology. *BioScience*. 2006; 56(12):958–9.

43. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015; 349(6251):943. doi: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)
44. Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB. The misuse of BLUP in ecology and evolution. *American Naturalist*. 2010; 175(1):116–25. doi: [10.1086/648604](https://doi.org/10.1086/648604) PMID: [19922262](https://pubmed.ncbi.nlm.nih.gov/19922262/)
45. Cheng Y, Martin TE. Data from: Nest predation risk and growth strategies of passerine species: grow fast or develop traits to escape risk? 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.33d6c>.
46. Husby A, Hille SM, Visser ME. Data from: Testing mechanisms of Bergmann's rule: phenotypic but no genetic change in body size in three passerine bird populations. 2011. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.8989>.
47. Liedvogel M, Cornwallis CK, Sheldon BC. Data from: Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. 2012. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.6n3d6109>.
48. Manser A, Lindholm AK, König B, Bagheri HC. Data from: Polyandry and the decrease of a selfish genetic element in a wild house mouse population. 2011. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.9069>.
49. Susi H, Laine A. Data from: Pathogen life-history trade-offs revealed in allopatry. 2013. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.dm88t>.
50. Liedvogel M, Cornwallis CK, Sheldon BC. Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Journal of Evolutionary Biology*. 2012; 25:813–23. doi: [10.1111/j.1420-9101.2012.02480.x](https://doi.org/10.1111/j.1420-9101.2012.02480.x) PMID: [22409177](https://pubmed.ncbi.nlm.nih.gov/22409177/)
51. Bumpus HC. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. (A fourth contribution to the study of variation). *Biol Lectures: Woods Hole Marine Biological Laboratory*. 1899:209–26.
52. Whitlock MC. Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution*. 2011; 26(2):61–5. doi: [10.1016/j.tree.2010.11.006](https://doi.org/10.1016/j.tree.2010.11.006) PMID: [21159406](https://pubmed.ncbi.nlm.nih.gov/21159406/)
53. Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol*. 2015; 13(11):e1002295. doi: [10.1371/journal.pbio.1002295.t001](https://doi.org/10.1371/journal.pbio.1002295.t001) PMID: [26556502](https://pubmed.ncbi.nlm.nih.gov/26556502/)
54. Vines TH, Albert AY, Andrew RL, Débarre F, Bock DG, Franklin MT, et al. The availability of research data declines rapidly with article age. *Current Biology*. 2014; 24(1):94–7. doi: [10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014) PMID: [24361065](https://pubmed.ncbi.nlm.nih.gov/24361065/)
55. The Royal Society. Science as an open enterprise. 2012. https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.
56. Shelford VE. Conservation versus preservation. *Science*. 1933; 77(2005):535.