

Article (refereed) - postprint

Mills, James A.; Teplitsky, Céline; Arroyo, Beatriz; Charmantier, Anne; Becker, Peter H.; Birkhead, Tim R.; Bize, Pierre; Blumstein, Daniel T.; Bonenfant, Christophe; Boutin, Stan; Bushuev, Andrey; Cam, Emmanuelle; Cockburn, Andrew; Côté, Steeve D.; Coulson, John C.; Daunt, Francis; Dingemanse, Niels J.; Doligez, Blandine; Drummond, Hugh; Espie, Richard H.M.; Festa-Bianchet, Marco; Frentiu, Francesca; Fitzpatrick, John W.; Furness, Robert W.; Garant, Dany; Gauthier, Gilles; Grant, Peter R.; Griesser, Michael; Gustafsson, Lars; Hansson, Bengt; Harris, Michael P.; Jiguet, Frédéric; Kjellander, Petter; Korpimäki, Erkki; Krebs, Charles J.; Lens, Luc; Linnell, John D.C.; Low, Matthew; McAdam, Andrew; Margalida, Antoni; Merilä, Juha; Møller, Anders P.; Nakagawa, Shinichi; Nilsson, Jan-Åke; Nisbet, Ian C.T.; van Noordwijk, Arie J.; Oro, Daniel; Pärt, Tomas; Pelletier, Fanie; Potti, Jaime; Pujol, Benoit; Réale, Denis; Rockwell, Robert F.; Ropert-Coudert, Yan; Roulin, Alexandre; Sederger, James S.; Swenson, Jon E.; Thébaud, Christophe; Visser, Marcel E.; Wanless, Sarah; Westneat, David F.; Wilson, Alastair J.; Zedrosser, Andreas. 2015. **Archiving primary data: solutions for long-term studies.**

© 2015 [Name]

This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>



This version available <http://nora.nerc.ac.uk/512555/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

NOTICE: this is the author's version of a work that was accepted for publication in *Trends in Ecology & Evolution*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Trends in Ecology & Evolution*, 30 (10). 581-589. [10.1016/j.tree.2015.07.006](https://doi.org/10.1016/j.tree.2015.07.006)

www.elsevier.com/

Contact CEH NORA team at

noraceh@ceh.ac.uk

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Archiving primary data: solutions for long-term studies

James A. Mills^{1,2}, Céline Teplitsky^{1,3}, Beatriz Arroyo⁴, Anne Charmantier⁵, Peter. H. Becker⁶, Tim R. Birkhead⁷, Pierre Bize⁸, Daniel T. Blumstein⁹, Christophe Bonenfant¹⁰, Stan Boutin¹¹, Andrey Bushuev¹², Emmanuelle Cam¹³, Andrew Cockburn¹⁴, Steeve D. Côté¹⁵, John C. Coulson¹⁶, Francis Daunt¹⁷, Niels J. Dingemanse^{18,19}, Blandine Doligez¹⁰, Hugh Drummond²⁰, Richard H. M. Espie²¹, Marco Festa-Bianchet²², Francesca Frentiu²³, John W. Fitzpatrick²⁴, Robert W. Furness²⁵, Dany Garant²², Gilles Gauthier¹⁵, Peter R. Grant²⁶, Michael Griesser²⁷, Lars Gustafsson²⁸, Bengt Hansson²⁹, Michael P. Harris¹⁷, Frédéric Jiguet³⁰, Petter Kjellander³¹, Erkki Korpimäki³², Charles J. Krebs³³, Luc Lens³⁴, John D.C. Linnell³⁵, Matthew Low³⁶, Andrew McAdam³⁷, Antoni Margalida³⁸, Juha Merilä³⁹, Anders P. Møller⁴⁰, Shinichi Nakagawa⁴¹, Jan-Åke Nilsson²⁸, Ian C. T. Nisbet⁴², Arie J. van Noordwijk⁴³, Daniel Oro⁴⁴, Tomas Pärt³⁶, Fanie Pelletier²², Jaime Potti⁴⁵, Benoit Pujol¹³, Denis Réale⁴⁶, Robert F. Rockwell⁴⁷, Yan Ropert-Coudert⁴⁸, Alexandre Roulin⁴⁹, James S. Sedinger⁵⁰, Jon E. Swenson⁵¹, Christophe Thébaud¹³, Marcel E. Visser⁴³, Sarah Wanless¹⁷, David F. Westneat⁵², Alastair J. Wilson⁵³, Andreas Zedrosser⁵⁴

- 1 JAM & CT contributed equally
- 2 10527A Skyline Drive, Corning, NY 14830, USA
- 3 Département Ecologie et Gestion de la Biodiversité. UMR 7204 CNRS/MNHN/UPMC, Muséum National d'Histoire Naturelle, Paris, France
- 4 Instituto de Investigacion en Recursos Cinegeticos (IREC) (CSIC-UCLM-JCCM), Ronda de Toledo s/n, 13005 Ciudad, Real, Spain
- 5 Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, Campus CNRS, 1919 Route de Mende, 34293 Montpellier, cedex 5, France
- 6 Institute of Avian Research, "Vogelwarte Helgoland", An der Vogelwarte 21 D26386 Wilhelmshaven, Germany
- 7 Department of Animal and Plant Sciences, The University of Sheffield, UK
- 8 Institute of Biological and Environmental Sciences, University of Aberdeen, UK
- 9 Department of Ecology and Evolutionary Biology, University of California, 621 Young Drive South, Los Angeles, CA 90095-1606, USA
- 10 CNRS, Université Lyon 1, Université de Lyon, UMR 5558, Laboratoire Biométrie et Biologie Évolutive, 43 boulevard du 11 Novembre 1918, F-69622 Villeurbanne, Cedex, France
- 11 Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9
- 12 Department of Vertebrate Zoology, Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1/12, 119234 Moscow, Russia
- 13 UMR 5174 EDB Laboratoire Évolution et Diversité Biologique, CNRS, ENFA, Université Toulouse 3 Paul Sabatier, 31062 Toulouse Cedex 9, France
- 14 Department of Evolution, Ecology & Genetics, Research School of Biology, The Australian National University, Canberra, ACT, Australia
- 15 Département de biologie & Centre d'études nordiques, Université Laval, 1045 avenue de la Médecine, Quebec (QC), G1V 0A6, Canada
- 16 29 St Mary's Close, Shincliffe, Durham, DH1 2ND, UK
- 17 Centre for Ecology & Hydrology, Bush Estate, Penicuik, EH26 0QB UK

- 46 18 Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-
47 Martinsried, Germany
- 48 19 Evolutionary Ecology of Variation Research Group, Max Planck Institute for Ornithology,
49 Seewiesen, Germany
- 50 20 Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de
51 México, A.P. 70-275, México D.F. 04510, México
- 52 21 Technical Resource Branch, Saskatchewan Ministry of Environment, 3211 Albert Street, Regina,
53 Saskatchewan, Canada S4S 5W6
- 54 22 Département de biologie, Université de Sherbrooke, 2500 Boulevard de L'Université,
55 Sherbrooke, Québec, Canada J1K 2R1
- 56 23 School of Biomedical Sciences and Institute of Health and Biomedical Innovation, Queensland
57 University of Technology, Kelvin Grove QLD 4059 Australia
- 58 24 Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
- 59 25 Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, U.K.
- 60 26 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544-
61 1003, USA
- 62 27 Anthropological Institute and Museum, University of Zürich, Zürich, Switzerland
- 63 28 Department of Animal Ecology, Evolutionary Biology Center, Uppsala University, Uppsala,
64 Sweden
- 65 29 Department of Biology, Lund University, Ecology Building, 223 62, Lund, Sweden
- 66 30 CESCO,UMR7204 Sorbonne Universités-MNHN-CNRS-UPMC, CP51, 55 Rue Buffon, 75005 Paris,
67 France
- 68 31 Grimso Wildlife Research Station, Department of Ecology, Swedish University of Agricultural
69 Sciences (SLU) SE-73091, Riddarhyttan, Sweden
- 70 32 Section of Ecology, Department of Biology, University of Turku, FI-20014 Turku, Finland
- 71 33 Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4,
72 Canada
- 73 34 Terrestrial Ecology Unit, Department of Biology, Ghent University, Ledeganckstraat 35, B-9000
74 Gent, Belgium
- 75 35 Norwegian Institute for Nature Research, P.O.Box 5685 Sluppen, N-7485 Trondheim, Norway
- 76 36 Department of Ecology, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden
- 77 37 Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada, N1G 2W1
- 78 38 Faculty of Life Sciences and Engineering, University of Lleida, E-25198 Lleida, Spain
- 79 39 Ecological Genetics Research Unit, Department of Biosciences, PO Box 65 (Biocenter 3,
80 Viikinkaari 1), FIN-00014 University of Helsinki, Finland
- 81 40 Laboratoire Ecologie, Systématique et Evolution, Equipe Diversité, Ecologie et Evolution
82 Microbiennes, Bâtiment 362, 91405 Orsay Cedex, France
- 83 41 Evolution & Ecology Research Centre and School of Biological, Earth and Environmental
84 Sciences, University of New South Wales, Sydney, Australia
- 85 42 I. C. T. Nisbet & Company, 150 Alder Lane, North Falmouth, MA 02556, USA
- 86 43 Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O.Box 50,
87 6700 AB Wageningen, The Netherlands
- 88 44 Institut Mediterrani d'Estudis Avançats IMEDEA (CSIC-UIB), Miquel Marques 21, 07190 Esporles,
89 Mallorca, Spain
- 90 45 Departamento de Ecología Evolutiva, Estación Biológica de Doñana-CSIC, Av. Américo Vespucio
91 s/n, 41092 Seville, Spain
- 92 46 Département des Sciences Biologiques, Université du Québec A Montréal, CP 8888-succursale
93 centre-ville, Montreal, Québec, H3C3P8

- 94 47 Vertebrate Zoology, American Museum of Natural History, New York, NY 10024 USA
95 48 Institut Pluridisciplinaire Hubert Curien, CNRS UMR7178, 23 rue Becquerel 67087 Strasbourg,
96 France
97 49 Department of Ecology and Evolution, University of Lausanne, Switzerland.
98 50 Department of Natural Resources and Environmental Science, University of Nevada Reno, Reno
99 NV 89512, USA
100 51 Department of Ecology and Natural Resource Management, Norwegian University of Life
101 Sciences, P.O.Box 5003, NO-1432 Ås, Norway and Norwegian Institute for Nature Research,
102 P.O.Box 5685 Sluppen, N-7485 Trondheim, Norway
103 52 Department of Biology, Center for Ecology, Evolution, and Behavior, University of Kentucky,
104 Lexington, KY, U.S.A
105 53 Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of
106 Exeter, Cornwall Campus, Penryn TR10 9EZ, UK
107 54 Faculty of Arts and Sciences, Department of Environmental & Health Studies, Telemark
108 University College, N-3800 Bø i Telemark, Norway

109

110 To whom correspondence can be addressed: dmills@stny.rr.com or teplitsky@mnhn.fr

111 **The recent trend for journals to require open access to primary data included in publications has been**
112 **embraced by many biologists, but has caused apprehension amongst researchers engaged in long-**
113 **term ecological and evolutionary studies. A worldwide survey of 73 principal investigators (PIs) with**
114 **long term studies revealed positive attitudes towards sharing data with the agreement or**
115 **involvement of the PI and 93% of PIs have historically shared data. Only 8% were in favor of**
116 **uncontrolled, open access to primary data while 63% expressed serious concern. Here we present**
117 **their viewpoint on an issue that can have non trivial scientific consequences. We discuss potential**
118 **costs of public data archiving and provide possible solutions to meet the needs of journals and**
119 **researchers.**

120

121 Keywords: Public data archiving, long term studies

122

123 **Long-term data sharing**

124 Several funding agencies, international regulatory bodies and many major ecological and evolutionary
125 journals now require raw or primary data to be deposited in a permanent open access archive, such as
126 Dryad or TreeBASE, as a condition for funding or publication. The data must be in sufficient detail to
127 allow the analyses in the paper to be replicated. The rationale for open archiving is that archived data
128 are available to posterity when studies are completed, for error checking, for use in new studies, or for
129 future meta-analysis [1]. In addition it has been argued that the policy would benefit data providers by
130 increasing their citation index through citations by papers with new analyses [1, 2].

131 Although it is claimed that over 95% of scientists in evolution and ecology believe data should be
132 publicly archived [1], mandatory public data archiving (PDA) is raising many issues in the scientific
133 community as evidenced by debates on websites, in blogs and publications [2-10]. Here we focus on the

134 perspective from long-term individual-based studies of wild populations that often span several
135 decades.

136 Short and long-term ecological studies differ in several important aspects. For example, in the former,
137 data tend to be collected over a short period of time for one or two papers and once published the data
138 in these papers become less valuable to the collector and can be more useful to others with different
139 perspectives or analytical skills. In contrast, in studies that have followed individuals over their lifetimes,
140 a lot of crucial information is assessed from derived metrics (e.g. survival, lifetime reproductive success)
141 that can only be estimated after many years of fieldwork. Therefore, much value can remain in the
142 primary data even after some of the initial questions are answered-

143 Long-term studies are rare and have great scientific value since many important questions in ecology
144 and evolutionary biology can only be answered from the life histories of recognizable individuals [11]. A
145 detailed analysis of the importance of individual-based studies has been documented elsewhere [11],
146 but a few examples are given in Box 1.

147 While group discussions and blog posts on PDA related issues have been flourishing, little is formally
148 known and published about the position and concerns of people collecting long-term data. To fill this
149 gap, a survey was conducted to learn their perspectives, and if current data requirements were
150 perceived as problematic, to identify potential alternative data-sharing policies that could be acceptable
151 to the journals, the scientific community and the Principal investigators.

152

153 **The survey**

154 To obtain the opinions of scientists with individual-based longitudinal data, a worldwide survey was sent
155 to 146 PIs of long-term research projects. Responses were received from 73 PIs working on 59 bird
156 studies, 13 mammalian studies and 1 plant study. The 92 projects (some PIs have several projects) range
157 in duration from 5 to 68 years (Figure 1), with 55 percent collecting data for more than 30 years. Thirty-
158 five percent of researchers were required to archive data used in a publication by their current funding
159 agency and 19% by their institution. Eight researchers were required to deposit data by both; therefore
160 59% were not required to archive their data. There was diversity of opinion among PIs about data
161 archiving, but some strong points of consensus emerged. This paper synthesizes the views of all
162 respondents, many of whom have made important contributions to ecology and evolution.

163 The survey revealed that virtually all PIs were in favor of data sharing with the agreement or
164 involvement of the PI. Historically, 93% of the respondents have shared their data when asked and 80%
165 have collaborated in meta-analyses. In the 1960-70s publications using longitudinal data often involved
166 only one or two authors. However, over the past two decades studies have become more complex and
167 collaborative, with studies commonly involving collaboration among biologists with expertise in a variety
168 of disciplines.

169 Overall, 63% of PIs were against PDA as currently required. This contrasts with a previous survey of
170 ecological and evolutionary biologists that reported that 95% were in favor of PDA [1]. Among the 36%
171 of respondents in favor of open access data archiving in this survey, only six (8% of 73) were in favor of
172 unconditional data archiving. The reasons given by PIs in favor of PDA were similar to those advocated
173 by the archiving journals. In contrast, 91% of PIs supported data sharing when clear rules for data access

174 were in place. These rules could include i) co-authorship or at least acknowledgment, depending on the
175 level of PI involvement; ii) no overlap with current projects, particularly projects conducted by students
176 or postdoctoral fellows; and iii) an agreement that the data go no further than the person to whom it is
177 entrusted.

178

179 *General concerns about PDA*

180 The main issues about archiving were centered on what data would be archived and to whom access
181 would be given, as detailed below. However, these concerns are so strong that 41% of respondents said
182 that they have avoided publishing in journals that require data be deposited in open access archives.
183 Furthermore, 53% intend to avoid publishing in them in the future and for those who published a major
184 paper involving long-term data early in their careers, 63% indicated that they would not have submitted
185 it to any journal that required data archiving. Avoiding publishing in a high impact journal can have
186 major consequences in terms of career advancement and could potentially reduce the prospects of
187 obtaining future financial support; therefore the decision would not be taken lightly.

188 In discussions among the survey respondents, it was suggested that the design and data collection of a
189 long-term study is research infrastructure that is the foundation of the publications which form the
190 lifework and careers of researchers and the PhD students and postdoctoral fellows who work on these
191 programs. The analogy can be made to experimental infrastructures which involve the construction of
192 an apparatus that takes years, or sometimes decades, and requires numerous grant applications,
193 institutional support, and deferred publication effort, all of which involve significant risk, but potentially
194 have profound scientific value, both pure and applied. Developing the infrastructure is a necessary pre-
195 requisite for project completion. In this case it would not be reasonable for other scientists to have
196 immediate access to the fruits of the inventor's labors. Furthermore, compulsory and unrestricted open
197 access to the apparatus would provide a strong disincentive to making the initial infrastructural
198 investment. The same case can be made for long-term ecological studies.

199

200 **Specific concerns from long-term researchers about PDA**

201 Several concerns about the costs of PDA for researchers and the scientific community were addressed
202 previously [5]. Here we add the perspective of PIs with long-term studies. Three major concerns were
203 identified during the survey.

204 *Potential costs to science*

205 *Flawed science:* A major cost would be flawed science resulting from a lack of understanding of the
206 database or the biological system. Open access to long-term data might not allow for a full
207 understanding of all the subtle contexts, nuances and issues involved in the biological system and the
208 structure of the database from which the long-term data are collected. It has been argued that if
209 method sections are sufficiently detailed, misunderstanding the system should not be a major source of
210 error [54]. However, not all of the complexities of the biological system can be detailed in a method
211 section without making a paper unwieldy. Hence, without the PI's involvement, crucial contextual
212 information is likely to be lost under open access, leading to the potential for erroneous assumptions
213 and interpretations which could add to the growing retraction rate in scientific journals [55]. For
214 example, although it was not included as a question, three respondents of the survey indicated that on

215 four occasions their data have been misinterpreted in publications, and once published, errors or
216 misinterpretations are hard to remove.

217 *More time spent on redundant activities:* A potential cost would be simultaneous testing of the same
218 idea on the data. In some cases, hypotheses might have been already investigated but not published by
219 the PIs because they were inconclusive. In addition, the cost of monitoring publications that used PDA
220 and writing replies would be borne by the researcher with long-term data and not the scientific
221 community. These do not seem to be a productive use of research investment.

222

223 *Fewer long-term studies:* Open access archiving could reduce the incentives for carrying out long-term
224 studies and would likely result in researchers suspending ongoing studies and declining to undertake
225 new ones. This is predicted by the producer-scrouter game theory [56] where the producer spends
226 time and energy to develop a resource but is unable to monopolize it, thereby creating opportunities for
227 the resource to be exploited by scrounger(s). Over time as the scrounger strategy increases, the
228 resource decreases. In theory, the fitness of the producer and the scrounger decreases, because at some
229 point there are no more resources to scrounge since no more resources are being produced [57,58].
230

231 *Less collaboration:* New collaborations are extremely valuable to make the most of the data but
232 comparative analyses and meta-analysis among long-term studies would likely suffer because PIs might
233 decline to participate if they are required to archive their data.

234

235 *Research funding*

236 Several financial issues have been overlooked by advocates of PDA. Archiving mutualizes the benefits,
237 but not the costs of long-term studies, because there is no cost to the person accessing the data. This
238 might be a sustainable model when recurrent funding is available, but not when funding is granted on a
239 per project basis. Also, PDA could incur some new costs for long-term studies since Dryad, for example,
240 has required extra payment for large data sets. Researchers with scarce funding might not be able to
241 absorb this additional cost. Maintaining constant funding is a critical issue for long-term studies to avoid
242 fatal gaps in the data [11, 59], contrasting once again with short-term studies that can be restarted at a
243 later time. Long-term studies of all durations experienced difficulties with funding (Figure 2) as only 33%
244 were fully funded in all years with the remainder having funding gaps varying in duration from 1 to 19
245 years (Figure 2). To maintain funding, PIs with long-term projects have to keep identifying new uses of
246 the data to obtain short-term funding because recurrent funding is essentially nonexistent [11].
247 Therefore, PDA could lead to a loss of funding opportunities if data for their next project are routinely
248 mined by other researchers.

249

250 *Student experience and training*

251 A major contribution of long-term studies is that they often provide training to PhD and other
252 postgraduate students and postdoctoral fellows. The PIs that responded to the survey, reported that
253 from their 92 projects, 630 PhDs were awarded (Figure 3a) and 658 postgraduates and 257 postdoctoral

254 fellows participated, for a total of 1,545 trainees. This represents a substantial contribution to the
255 training and development of the ecological and evolutionary biology research community. Survey
256 respondents expressed a particular concern that PDA would negatively impact this important feature of
257 long-term studies because negotiations take place among study participants before the onset of new
258 research areas (such as MSc and PhD thesis or postdoctoral research projects) to avoid overlap. Such
259 planning is undermined if outsiders are entirely free to work with available data from long-term studies
260 without taking ongoing and planned analyses by insiders into account. The risk is especially strong for
261 PhD students as part of their training involves courses, and they need more time to complete the
262 research project and publish papers than senior researchers.

263

264 **Possible solutions**

265 The verification of results is a very important requirement by journals; however, the costs of mandatory
266 archiving of data by ongoing long-term projects could outweigh the expected benefits. Having imposed a
267 requirement for PDA, journals are asking researchers to give up rights to what many consider to be their
268 intellectual property. In fact some scientists are considering copywriting their data. Journals are rightly
269 vigilant in combating plagiarism and copyright infringement; it would be appropriate for journals to be
270 just as vigilant in respecting and protecting the scientists' data.

271 A resolution to this conflict would benefit scientific progress; high quality long-term studies have been
272 responsible for a disproportionate number of publications in journals with the highest impact factors
273 [11]. Many of the 5,378 papers from 90 studies in this survey (Fig 3b) were published in prestigious
274 journals that now require PDA. To initiate a discussion about how resolution might be achieved, we
275 suggest six potential solutions.

276

277 *Promoting collaboration*

278 Opportunities for collaboration that provide added-value to science and communication between data
279 generators and potential users should be encouraged [5] rather than compulsory archiving. Most survey
280 respondents see collaborations as the most satisfactory route to data sharing. For substantial requests,
281 the original researcher can expect and deserve co-authorship. To promote better use of data and
282 collaboration with PIs, a website could be created referencing long-term studies with information such
283 as species, duration of study, location, traits measured, protocols used, etc.

284

285 *Providing primary data on a confidential basis*

286 A solution that would satisfy most PIs would be to provide tabulated summary data initially, and if that
287 data were insufficient for editors to evaluate a submitted paper, primary data could be provided on a
288 confidential basis. After the review process, the data could be destroyed and would not be available to
289 be used for any other purpose. Once the paper is published, people who want to use the data could
290 contact the PIs of the long-term study for additional data. As the survey has shown, 93% of the
291 respondents have indicated that they have supplied data on request. For example, researchers have

292 used summary data from the 40-year study on Darwin's finches [45, 60] by Peter and Rosemary Grant
293 which was deposited in Dryad, and raw data have been supplied to four others upon request.

294 *Providing a longer embargo*

295 Some journals have indicated a willingness to embargo the data for a period of one to five years from
296 publication, allowing the original researcher time to complete any related papers. This can reduce
297 concerns in the case of smaller data sets from which only a limited number of questions can be
298 answered. However, this is unlikely to solve the problem for long-term data sets, from which many
299 questions can be addressed from different perspectives and over differing lengths of time.

300 For active long-term studies (i.e. with ongoing data collection) a minimum of 10-15 years might be
301 considered more appropriate [5]. By comparison, pharmaceutical companies have a twenty-year patent
302 to recoup their investment. A similar argument could be made for the decades of research by long-term
303 project scientists [9]. Furthermore, a longer embargo would encourage data users to contact the PIs for
304 rapid access to the most up-to-date version of the database, thereby encouraging collaboration. For
305 non-active studies where data collection has ended, the case for an earlier release is stronger.

306 *Depositing data on institutional servers*

307 Centralizing the data in a single database in one location will prevent fragmentation of data on different
308 archiving sites and ensures that the data are completely secure and up to date. Data could be archived
309 on institutional servers and the institution and its staff could control access and determine if
310 collaboration is appropriate. An example of an effective approach to the management of archived data
311 held by institutions is practiced by The Netherlands Institute of Ecology where people can request the
312 data and data extraction is done by members of the Institute, provided that the applicant will use the
313 data for a well described project, commit to not sharing the data with others, and offer co-authorship to
314 the collector if the data forms an essential part of the publication. Another example of effective use of
315 institutional servers is the Archibold Biological Station in Florida. Such institutional databases also allow
316 the preservation of data and their accessibility after the PI retires [11].

317 *Increasing notification and communication*

318 If online archiving should be preferred for the physical safety of data, two improvements to present
319 practices could be made. First, as the survey demonstrated, PIs are concerned with inappropriate use of
320 data and overlap with ongoing or future projects of their own. A clear policy should be implemented by
321 journals concerning conflicts of interest between the researchers collecting and organizing the data, and
322 those who would use the data. For example, there are currently no binding protocols or codes of
323 conduct covering the presentation of, or access to, complex data that underpin analyses in publications.
324 A process with guidelines should be established by journals to ensure that PIs are aware of potential
325 studies and are satisfied with a paper based on the data they generated prior to the review process.

326 A possibility would be to implement data tracking, allowing data collectors to obtain information on who
327 is using the data and why. For example, any request for data to the Climate Change, Agriculture and
328 Food Security Data Portal, triggers an email to be sent to the PI who deposited the data. Journals should
329 have a rule that no paper is considered where the data users have not corresponded with the data
330 owners and included appropriate acknowledgement of the source of the data within the paper. A rule
331 set by journals would have a lot of clout with data users. Data tracking would also allow the PI to be

332 systematically asked to review papers based on their data. Another option would be to send an e-mail to
333 the PI every year asking whether they wish the data to be private or open access.

334 **Concluding remarks**

335 Long-term studies currently generate science with high impact in all major fields of biology. These
336 longitudinal studies began during an era when PDA did not exist. Whilst we agree that it is essential to
337 archive data so that they are not lost to science, a key concern is that recently introduced data archiving
338 regimes combined with difficulty in obtaining continuous financial support will be a disincentive both for
339 the initiation of long-term studies, and for maintenance of ongoing studies. It would be appropriate for
340 journals and data archiving institutions to enter into a dialogue with researchers about how best to
341 meet the objectives of data archiving while allowing valuable long-term studies to thrive.

342 Specifically, we recommend the development of a formal code of conduct which respects the data
343 generated through long-term studies, and i) allow tabulated summaries to be provided in the first
344 instance backed up by the confidential submission of primary data if required by editors, ii) encourage
345 collaborative research with the data collector by people wishing to use the data, iii) extend embargoes
346 on the use of archived data [5], iv) consider allowing archiving on institutional servers rather than open
347 access servers and iv) develop enforceable procedures that enable the researcher to be contacted when
348 someone wishes to access primary data. Through these modifications, a compromise could be crafted
349 that provides an advantage to the scientific community, journals, and researchers generating long-term
350 data, as well as benefiting science.

351

352 **Acknowledgements**

353 We wish to thank all of the principal investigators who provided information on their long-term studies.
354 CT was funded by the French ANR (grant ANR-12-ADAP-0006). Many thanks to Sandra Hamel, Susan
355 Alberts and Walt Koenig for constructive comments. Deborah Mills assisted in the analysis and editing
356 the manuscript.

357

358

359

360

361

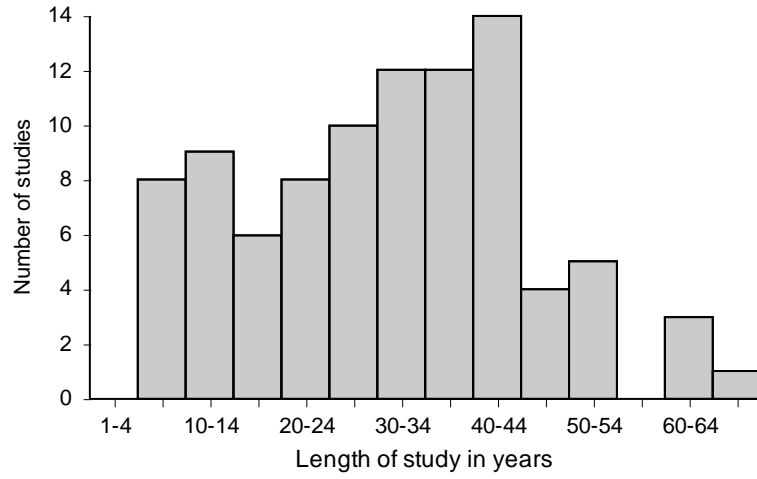
362 **References**

- 363 1 Whitlock, M.C. *et al.* (2010) Data archiving. *Amer. Nat.* 175, 145-146
364 2 Piwowar, H.A. and Vision T.J. (2013) Data reuse and the open data citation advantage. *Peer J.*
365 1:e175
366 3 Whitlock, M.C. (2011) Data archiving in ecology and evolution: best practices. *Trends Ecol. Evol.*
367 26, 61-65
368 4 Costello M.J. *et al.* (2013) Biodiversity data should be published, cited and peer reviewed.
369 *Trends Ecol. Evol.* 28, 454-461

- 370 5 Roche, D.G. *et al.* (2014) Trouble shooting public data archiving: suggestions to increase
371 participation. *PLoS Biol.* 12 (1): e10001779
- 372 6 Bloom, T., Ganley, E. and Winkler, M. (2014) Data access for the open access literature: PLoS's
373 data policy. *PLoS Biol.* 12 (2): e1001797
- 374 7 Gleditsch, N.P., Meyelits, C. and Strand, H. (2003) Posting your data: will you be scooped or will
375 you be famous? *Int. Stud Perspect* 4 (1): 89-97
- 376 8 Caetano, D.S. and Aisenberg A. (2014) Forgotten treasures: the fate of data in animal behaviour
377 studies *Anim Behav.* 98, 1-5
- 378 9 McGlynn, T. (2014) I own my data, until I don't. [SmallPondScience.com/2014/03/03 I-own-my-](http://SmallPondScience.com/2014/03/03-I-own-my-data-until-I-don-t/comm)
379 [data-until-I-don't/comm](http://SmallPondScience.com/2014/03/03-I-own-my-data-until-I-don-t/comm)
- 380 10 Kenall, A., Harold, S. and Foote, C. (2014) An open future for ecological and evolutionary data?
381 *BMC Evol. Biol.* 2014, 14, 10
- 382 11 Clutton-Brock, T. and Sheldon, B.C. (2010) Individuals and populations: the role of long-term,
383 individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol. Evol.* 25,
384 562-573
- 385 12 Coulson, J.C. (1966) The influence of the pair-bond and age on the breeding biology of the
386 kittiwake gull *Rissa tridactyla*. *J. Anim. Ecol.* 35, 269-279
- 387 13 Nisbet, I.C.T. and Dann, P. (2009) Reproductive performance of little penguins in relation to
388 year, age, pair-bond duration, breeding date and individual quality. *J. Avian Biol.* 40, 296-308
- 389 14 Ryder, J. P. (1980) The influence of age on the breeding biology of colonial nesting seabirds. In
390 *Behavior of Marine Animals, vol. 4, Marine birds* (Burger, J., Olla, B.L. and Winn, H.E. eds.), pp
391 153-168, Plenum, New York
- 392 15 Genovart, M. *et al.* (2013) Contrasting effects of climatic variability on the demography of a
393 trans-equatorial migratory seabird. *J. Anim. Ecol.* 82, 121-130
- 394 16 Becker, P.H. *et al.* (2008) Timing of initial arrival at the breeding site predicts age at first
395 reproduction in a long-lived migratory bird. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12349-12352
- 396 17 Reid, J.M. *et al.* (2010) Parentage, lifespan and offspring survival: structured variation in life
397 history in a wild population. *J. Anim. Ecol.* 79, 851-862
- 398 18 Sedinger J.S *et al.* (2008) Fidelity and breeding probability related to population density and
399 individual quality in black brant geese (*Branta bernicla nigricans*) *J. Anim. Ecol.* 77, 702-712
- 400 19 Jorgenson, J.T. *et al.* (1997) Effects of age, sex, disease and density of bighorn sheep. *Ecology*
401 78, 1019-1032
- 402 20 Frederiksen, M. *et al.* (2006) From plankton to top predators: bottom-up control of a marine
403 food web across four tropic levels. *J. Anim. Ecol.* 75, 1259-1268
- 404 21 Ratcliffe, N., Furness, R.W. and Hamer, K.C. (1998) The interactive effects of age and food
405 supply on the breeding ecology of great skuas. *J. Anim. Ecol.* 67, 853-862
- 406 22 Daunt, F. *et al.* (2014) Longitudinal biologging reveals interplay between extrinsic and intrinsic
407 carry over effects in a long-lived vertebrate. *Ecology* 95, 20177-2083
- 408 23 Mills, J.A. *et al.* (2008) The impact of climate fluctuation on food availability and reproductive
409 performance of the planktivorous red-billed gull *Larus novaehollandiae scopulinus*. *J. Anim.*
410 *Ecol.* 77, 1129-1142
- 411 24 Oro, D. and Furness, R.W. (2002) Influences of food availability and predation on survival of
412 kittiwakes. *Ecology* 83, 2516-2528
- 413 25 Korpimäki, E. (1992) Fluctuating food abundance determines the lifetime reproductive success
414 of male Tengmalm's owls. *J. Anim. Ecol.* 61, 103-111
- 415 26 Festa-Bianchet, M. (1989) Individual differences, parasites, and the costs of reproduction for
416 bighorn ewes (*Ovis canadensis*). *J. Anim. Ecol.* 58, 785-795

- 417 27 Alho, J.S. *et al.* (2012) No evidence for inbreeding avoidance through active mate choice in red-
418 billed gulls. *Behav. Ecol.* 23, 672-675
- 419 28 Ludwig, S. and Becker, P.H. (2012) Immigration prevents inbreeding in a growing colony of a
420 long-lived and philopatric seabird. *Ibis* 154, 74-84
- 421 29 Hansson, B. (2004) Marker-based relatedness predicts egg-hatching failure in great reed
422 warblers. *Conserv. Genet.* 5, 339-348
- 423 30 Hayward, A.D. *et al.* (2013) Reproductive senescence in female Soay sheep: variation across
424 traits and contributions of individual ageing and selective disappearance. *Funct. Ecol.* 27,184-
425 195
- 426 31 Rebke, M. *et al.* (2010) Reproductive improvement and senescence in a long-lived bird. *Proc.*
427 *Natl. Acad. Sci. U.S.A.* 107, 7841-7846
- 428 32 Bonneaud, C. *et al.* (2006) Complex *Mhc*-based mate choice in a wild passerine. *Proc. R. Soc.*
429 *London B. Biol. Sci.* 273, 1111-1116
- 430 33 García-Navas, V., Ortego, J. and Sanz, J.J. (2009) Heterozygosity-based assortative mating in
431 bluetits (*Cyanites caeruleus*): implications for evolution of mate choice. *Proc. R. Soc. London B*
432 *Biol. Sci.* 276, 2931-2940
- 433 34 Thomson, R. L. *et al.* (2014) Brood size manipulations in a spatially and temporally varying
434 environment: male Tengmalm's owls pass increased reproductive costs to offspring. *Oecologia*
435 176, 423-430
- 436 35 Riechert, J., Chastel, O. and Becker, P.H. (2012) Why do experienced birds reproduce better?
437 Possible endocrine mechanisms in a long-lived seabird, the common tern. *Gen. Comp.*
438 *Endocrinol.* 178, 391-399
- 439 36 Bauch, C. Becker, P.H. and Verhulst, S. (2013) Telemere length reflects phenotypic quality and
440 costs of reproduction in a long-lived seabird. *Proc. R. Soc. B* 280, 20122540
- 441 37 Bouwhuis, S. *et al.* (2010) Trans-generational effects on ageing in a wild bird population. *J. Evol.*
442 *Biol.* 23, 636-642
- 443 38 Cam, E. *et al.* (2003) Long-term fitness consequences of early conditions in the kittiwake. *J.*
444 *Anim. Ecol.* 72, 411-424
- 445 39 Newton I. (1989) *Lifetime Reproduction in Birds.* Academic Press London
- 446 40 Clutton-Brock, T.H. (1988) *Reproductive Success.* University of Chicago Press, Chicago
- 447 41 Charmantier, A. *et al.* (2014) *Quantitative Genetics in the Wild.* Oxford University Press, Oxford
- 448 42 Kruuk, L.E.B. *et al.* (2008) New answers for old questions: The evolutionary quantitative
449 genetics of wild animal population. *Ann. Rev. Ecol. Evol. Syst.* 39, 525-548
- 450 43 Charmantier, A. and Garant, D. (2005) Environmental quality and evolutionary potential:
451 lessons from wild populations. *Proc. R. Soc. London B Biol. Sci.* 272, 1415-1425
- 452 44 Merilä, J. and Hendry, A.P. (2014) Climate change, adaptation and phenotypic plasticity: the
453 problem and the evidence. *Evol. Appl.* 7, 1-14
- 454 45 Grant, P.R. and Grant, B.R. (2002) Unpredictable evolution in a 30-year study of Darwin's
455 finches. *Science* 296, 707-711
- 456 46 Townsend, A.K. *et al.* (2011) Genetic monogamy across variable demographic landscapes in
457 cooperatively breeding Florida scrub-jays. *Behav. Eco.* 22, 464-470
- 458 47 Teplitsky, C. *et al.* (2009) Heritability of fitness components in a wild bird population. *Evolution*
459 63, 716-726
- 460 48 Kruuk, L.E.B *et al.* (2000) Heritability of fitness in a wild mammal population. *Proc. Natl. Acad.*
461 *Sci. U.S.A.* 97, 698-703

- 462 49 Charmantier, A. and Gienapp, P. (2014) Climate change and timing of avian breeding and
463 migration: evolutionary versus plastic changes. *Evol. Appl.* 7, 15–28
- 464 50 Teplitsky, C. and Millien, V. (2014) Climate warming and Bergmann’s rule through time: is there
465 any evidence? *Evol. Appl.* 7, 156–168
- 466 51 Gienapp, P. *et al.* (2008) Climate change and evolution: disentangling environmental and
467 genetic responses. *Mol. Ecol.* 17, 167-178
- 468 52 Garant, D. *et al.* (2004) Evolution in a changing environment: a case study with great tit fledging
469 mass. *Amer. Nat.* 164, E115–E129
- 470 53 Cury, P.M. *et al.* (2011) Global seabird response to forage fish depletion-one third for the birds.
471 *Science* 334, 1703-1706
- 472 54 Coulson, T. and Sheldon, B. (2014) Archive your data! Animal Ecology in Focus, journal of
473 ecology.wordpress.com
- 474 55 Steen,R.G., Coasadevall, A. and Fang, F.C. (2013) Why has the number of scientific retractions
475 increased? *PLoS One* DOI:101371/journal.pone.0068397
- 476 56 Barnard, C.J. and Sibly, R.M. (1981) Producers and scroungers: a general model and its
477 application to captive flocks of house sparrows. *Anim. Behav.* 29, 543-550
- 478 57 Caraco,T. and Giraldeau, L-A. (1991) Social foraging: producing and scrounging in a stochastic
479 environment. *J. Theor. Biol.* 153, 559-583
- 480 58 Vickery, W.L. *et al.* (1991) Producers, scroungers, and group foraging. *Amer. Nat.* 137, 847-863
- 481 59 Birkhead, T. (2014) Stormy outlook for long-term ecological studies. *Nature* 514, 7523
- 482 60 Grant, P.R. and Grant, B.R. (2014) *40 years of evolution: Darwin’s finches on Daphne Major*
483 *Island. Princeton University Press*
- 484



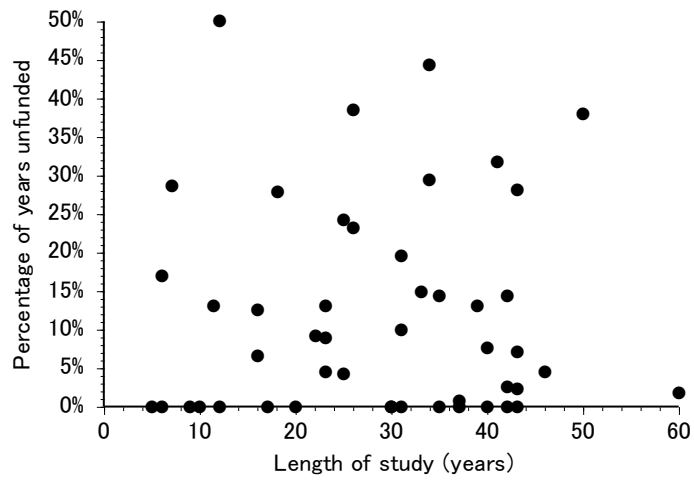
485

486 **Figure 1.** Duration of studies undertaken by the respondents in this survey

487

488

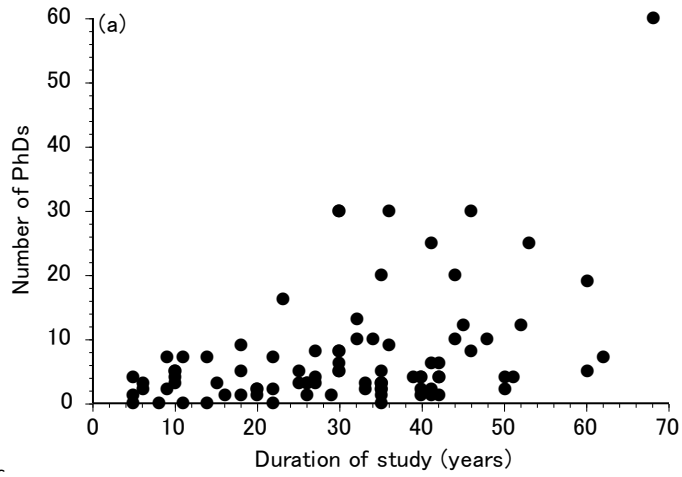
489



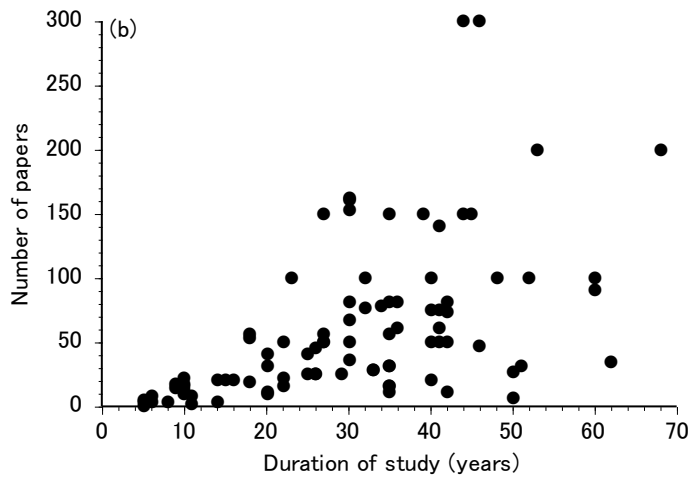
490

491 Figure 2 Duration of the study and the percentage of years unfunded

492



493 c



494

495 **Figure 3** (a) The total number of PhD students in relation to the duration of research programs (b) The number of papers produced in relation
 496 the duration of the study

497