

Aberystwyth University

Connecting the data landscape of long-term ecological studies

Culina, Antica; Adriaensen, Frank; Bailey, Liam D.; Burgess, Malcolm D.; Charmantier, Anne; Cole, Ella F.; Eeva, Tapio; Matthysen, Erik; Nater, Chloé R.; Sheldon, Ben C.; Sæther, Bernt Erik; Vriend, Stefan J.G.; Zajkova, Zuzana; Adamík, Peter; Aplin, Lucy M.; Angulo, Elena; Artemyev, Alexandr; Barba, Emilio; Barišić, Sanja; Belda, Eduardo

Published in:
Journal of Animal Ecology

DOI:
[10.32942/osf.io/6gea7](https://doi.org/10.32942/osf.io/6gea7)
[10.1111/1365-2656.13388](https://doi.org/10.1111/1365-2656.13388)
[10.5061/dryad.51c59zw6r](https://doi.org/10.5061/dryad.51c59zw6r)

Publication date:
2021

Citation for published version (APA):

Culina, A., Adriaensen, F., Bailey, L. D., Burgess, M. D., Charmantier, A., Cole, E. F., Eeva, T., Matthysen, E., Nater, C. R., Sheldon, B. C., Sæther, B. E., Vriend, S. J. G., Zajkova, Z., Adamík, P., Aplin, L. M., Angulo, E., Artemyev, A., Barba, E., Barišić, S., ... Visser, M. E. (2021). Connecting the data landscape of long-term ecological studies: The SPI-Birds data hub. *Journal of Animal Ecology*, 90(9), 2147-2160.
<https://doi.org/10.32942/osf.io/6gea7>, <https://doi.org/10.1111/1365-2656.13388>,
<https://doi.org/10.5061/dryad.51c59zw6r>

Document License
CC BY-SA

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

1 **Connected data landscape of long-term ecological studies:**

2 **the SPI-Birds data hub**

3

4

5 Antica Culina^{1#}, Frank Adriaensen², Liam D. Bailey^{1,3}, Malcolm D. Burgess^{4,5}, Anne
6 Charmantier⁶, Ella F. Cole⁷, Tapio Eeva⁸, Erik Matthysen², Chloé R. Nater⁹, Ben C. Sheldon⁷,
7 Bernt-Erik Sæther⁹, Stefan J.G. Vriend⁹, Peter Adamík¹⁰, Lucy M. Aplin¹¹, Elena Angulo¹²,
8 Alexandr Artemyev¹³, Emilio Barba¹⁴, Sanja Barišić¹⁵, Eduardo Belda¹⁶, C. Can Bilgin¹⁷, Josefa
9 Bleu¹⁸, Christiaan Both¹⁹, Sandra Bouwhuis²⁰, Claire Branston²¹, Juli Broggi¹², Terry Burke²²,
10 Andrey Bushuev²³, Carlos Camacho²⁴, Daniela Campobello^{25,26}, David Canal²⁷, Alejandro
11 Cantarero²⁴, Samuel P. Caro⁶, Maxime Cauchoix²⁸, Alexis Chaine²⁸, Mariusz Cichoń²⁹, Davor
12 Ćiković¹⁵, Camillo A. Cusimano²⁵, Caroline Deimel³⁰, André Dhondt³¹, Niels J. Dingemanse³²,
13 Blandine Doligez³³, Davide M. Dominoni²¹, Claire Doutrelant⁶, Szymon M. Drobniak^{34,35}, Anna
14 Dubiec³⁶, Marcel Eens², Kjell Einar Erikstad³⁷, Silvia Espín³⁸, Damien R. Farine^{39,40,41}, Jordi
15 Figuerola⁴², Pınar Kavak Gülbeyaz^{17,43}, Arnaud Grégoire⁶, Ian R. Hartley⁴⁴, Michaela Hau³⁰,
16 Gergely Hegyi⁴⁵, Sabine Hille⁴⁶, Camilla A. Hinde⁴⁷, Benedikt Holtmann³², Tatyana Ilyina²³,
17 Caroline Isaksson²⁴, Arne Iserbyt², Elena Ivankina⁴⁸, Wojciech Kania⁴⁹, Bart Kempnaers⁵⁰, Anvar
18 Kerimov²³, Jan Komdeur¹⁹, Peter Korsten⁵¹, Miroslav Král¹⁰, Miloš Krist¹⁰, Marcel Lambrechts⁶,
19 Carlos E. Lara⁵², Agu Leivits⁵³, András Liker⁵⁴, Jaanis Lodjak⁵⁵, Marko Mägi⁵⁵, Mark C.
20 Mainwaring⁵⁶, Raivo Mänd⁵⁵, Bruno Massa²⁵, Sylvie Massemin¹⁸, Jesús Martínez-Padilla⁵⁷,
21 Tomasz D. Mazgajski³⁶, Adele Mennerat⁵⁸, Juan Moreno⁵⁹, Alexia Mouchet³², Shinichi
22 Nakagawa³⁵, Jan-Åke Nilsson²⁴, Johan Nilsson²⁴, Ana Cláudia Norte⁶⁰, Kees van Oers^{1,61}, Markku
23 Orell⁶², Jaime Potti¹², John L. Quinn⁶³, Denis Réale⁶⁴, Tone Kristin Reiertsen³⁷, Balázs Rosivall⁴⁵,
24 Andrew F Russel⁶⁵, Seppo Rytönen⁶², Pablo Sánchez-Virosta³⁸, Eduardo S. A. Santos⁶⁶, Julia

25 Schroeder⁶⁷, Juan Carlos Senar⁶⁸, Gábor Seress⁵⁴, Tore Slagsvold⁶⁹, Marta Szulkin⁷⁰, Céline
26 Teplitsky⁶, Vallo Tilgar⁵⁵, Andrey Tolstoguzov⁷¹, János Török⁴⁵, Mihai Valcu⁵⁰, Emma Vatka⁷²,
27 Simon Verhulst¹⁹, Hannah Watson²⁴, Teru Yuta^{73,74}, José M. Zamora-Marín⁷⁵ & Marcel E. Visser¹

¹ Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

² Department of Biology, University of Antwerp, Antwerp, Belgium

³ Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

⁴ RSPB Centre for Conservation Science, The Lodge, Sandy, UK

⁵ Centre for Research in Animal Behaviour, University of Exeter, Exeter, UK

⁶ CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France

⁷ Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK

⁸ Department of Biology, University of Turku, Turku, Finland

⁹ Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

¹⁰ Department of Zoology, Faculty of Science, Palacký University, Olomouc, Czech Republic

¹¹ Max Planck Institute of Animal Behavior, Radolfzell, Germany

¹² Department of Evolutionary Ecology, Estación Biológica de Doñana-CSIC, Seville, Spain

¹³ Institute of Biology of the Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk, Russia

¹⁴ Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Paterna, Spain

¹⁵ Institute of Ornithology, Croatian Academy of Sciences and Arts, Zagreb, Croatia

¹⁶ Universitat Politècnica de València, Gandia, Valencia, Spain

¹⁷ Biodiversity and Conservation Lab, Department of Biology, METU, Ankara, Turkey

¹⁸ Université de Strasbourg, CNRS, Strasbourg, France

¹⁹ Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen

²⁰ Institute of Avian Research, Wilhelmshaven, Germany

²¹ Institute of Biodiversity, Animal Health & Comparative Medicine, Univ. of Glasgow, Glasgow, UK

²² Department of Animal & Plant Sciences, University of Sheffield, Sheffield, UK

²³ Department of Vertebrate Zoology, Faculty of Biology, Lomonosov Moscow State Univ., Moscow, Russia

²⁴ Department of Biology, Lund University, Lund, Sweden

²⁵ Stazione Ornitologica, Monreale, Italy

²⁶ Department STEBICEF, Università degli Studi di Palermo, Palermo, Italy

²⁷ Institute of Ecology and Botany, Centre for Ecological Research, Vácrátót, Hungary

²⁸ Station d'Ecologie Théorique et Expérimentale du CNRS (UMR5321), Moulis, France

²⁹ Institute of Environmental Sciences, Jagiellonian University, Kraków

³⁰ Max Planck Institute for Ornithology, Evolutionary Physiology Group, Seewiesen, Germany

³¹ Lab of Ornithology, Cornell University, Ithaca, USA

³² Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany

³³ CNRS, Department of Biometry & Evolutionary Biology, University of Lyon, Villeurbanne, France

34 Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland
35 Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences,
University of New South Wales, Sydney, Australia
36 Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
37 Norwegian Institute for Nature Research (NINA), Tromsø, Norway
38 Area of Toxicology, Department of Health Sciences, University of Murcia, Murcia, Spain
39 Department of Collective Behavior, Max Planck Institute of Animal Behavior, Konstanz, Germany
40 Department of Biology, University of Konstanz, Konstanz, Germany
41 Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany
42 Department of Wetland Ecology. Estación Biológica de Doñana (CSIC), Sevilla, Spain
43 Environmental Engineering Department, Hacettepe University, Ankara, Turkey
44 Lancaster Environment Centre, Lancaster University, Lancaster, UK
45 Department of Systematic Zoology and Ecology, Eötvös Loránd University, Budapest, Hungary
46 Institute of Wildlife Biology and Game Management, University of Natural Resources and Life
Sciences Vienna, Vienna, Austria
47 Behavioural Ecology Group, Biological Sciences, Anglia Ruskin University, Cambridge, UK
48 Zvenigorod Biological Station, Faculty of Biology, Lomonosov Moscow State University, Moscow,
Russia
49 Ornithological Station, Museum & Institute of Zoology, Polish Academy of Sciences, Warsaw,
Poland
50 Department Behavioural Ecology & Evolutionary Genetics, Max Planck Inst. for Ornithology,
Seewiesen, Germany
51 Department of Animal Behaviour, Bielefeld University, Bielefeld, Germany
52 Department of Zoology, University of Otago, Dunedin, New Zealand
53 Department of Nature Conservation, Environmental Board, Estonia
54 MTA-PE Evolutionary Ecology Research Group, University of Pannonia, Veszprém, Hungary
55 Department of Zoology, Inst. of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
56 Division of Biological Sciences, University of Montana, Missoula, USA
57 Department of Biodiversity Conservation and Ecosystem Restoration, Pyrenean Institute of Ecology
(CSIC), Jaca, Spain
58 Department of Biological Sciences, University of Bergen, Bergen, Norway
59 Museo Nacional de Ciencias Naturales (CSIC), Depart. de Ecología Evolutiva, Madrid, Spain
60 University of Coimbra, MARE - Marine and Environmental Sciences Centre, Department of Life
Sciences, Coimbra, Portugal
61 Behavioural Ecology Group, Department of Animal Sciences, Wageningen University & Research,
Wageningen, The Netherlands
62 Ecology and Genetics Research Unit, University of Oulu, Finland
63 School of Biological, Earth and Environmental Sciences, University College Cork, Ireland
64 Département des Sciences Biologiques, Université du Québec A Montréal, Montréal, Canada
65 Centre for Ecology and Conservation, Univ. of Exeter, Tremough Campus, Penryn, Cornwall, UK
66 BECO do Departamento de Zoologia, Universidade de São Paulo, São Paulo, Brazil
67 Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, UK
68 Evolutionary and Behavioural Ecology Research Unit, Museu de Ciències Naturals de Barcelona,
Barcelona, Spain
69 Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of
Oslo, Oslo, Norway

- ⁷⁰ Centre of New Technologies, University of Warsaw, Warsaw, Poland
- ⁷¹ Institute of Biology, Karelian Scientific Center, Russian Academy of Sci., Petrozavodsk, Russia
- ⁷² Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological & Environmental Sciences, University of Helsinki, Finland
- ⁷³ Yamashina Institute for Ornithology, Abiko, Japan
- ⁷⁴ Graduate School of Environment Science, Hokkaido University, Sapporo, Japan
- ⁷⁵ Department of Zoology and Physical Anthropology, University of Murcia, Murcia, Spain
- 28 # Corresponding author: a.culina@yahoo.com or a.culina@nioo.knaw.nl

29 **Abstract**

- 30 1. The integration and synthesis of the data in different areas of science is drastically slowed
31 and hindered by a lack of standards and networking programmes. Long-term studies of
32 individually marked animals are not an exception. These studies are especially important
33 as instrumental for understanding evolutionary and ecological processes in the wild.
34 Further, their number and global distribution provides a unique opportunity to assess the
35 generality of patterns and to address broad-scale global issues (e.g. climate change).
- 36 2. To solve data integration issues and enable a new scale of ecological and evolutionary
37 research based on long-terms studies of birds, we have created the SPI-Birds Network and
38 Database (www.spibirds.org) – a large-scale initiative that connects data from, and
39 researchers working on, studies of wild populations of individually recognizable (usually
40 ringed) birds. Within a year of the establishment, SPI-Birds counts 120 members working
41 on more than 80 populations, with data concerning breeding attempts of almost a million
42 individual birds over a 1700 cumulative years, and counting.
- 43 3. SPI-Birds acts as a data hub and a catalogue of studied populations. It prevents data loss,
44 secures easy data finding, use and integration, and thus facilitates collaboration and
45 synthesis. We provide community-derived data and meta-data standards and improve data
46 integrity guided by of Findable, Accessible, Interoperable, and Reusable (FAIR), and
47 aligned with the existing metadata languages (e.g. ecological meta-data language).
- 48 4. The encouraging community involvement stems from SPI-Bird's decentralized approach:
49 research groups retain full control over data use and their way of data management, while
50 SPI-Birds creates tailored pipelines to convert each unique data format into a standard
51 format. We outline the lessons learned, so that other communities (e.g. those working on
52 other taxa) can adapt our successful model. Creating community-specific hubs (such as

53 ours, COMADRE for animal demography, etc.) will aid much-needed large-scale
54 ecological data integration.

55 **Keywords:** database | birds | research network | data standards | meta-data standards | long-term
56 studies | data hub | data management | FAIR data

57

58

59

60

61 **Introduction**

62 *The importance of long-term individual-based studies*

63 Long-term individual-based studies of animals in their natural environment underpin our
64 understanding of evolutionary and ecological patterns and processes in wild populations (Clutton-
65 Brock & Sheldon 2010). These studies considerably increase our ability to establish the links
66 among genes, individual traits (including physiology and behaviour), fitness, and the environment
67 (Broggi et al. 2005, Korsten et al. 2013, Charmantier et al. 2014, Teplitsky et al. 2014, Schroeder
68 et al. 2015, Johnston et al. 2016, Laine et al 2016, Martínez-Padilla et al. 2017, Bonnet et al 2019,
69 Class et al. 2019). They further document the responses of natural populations to changing
70 environments (Clutton-Brock et al. 1992, Grant & Grant 2002, Coltman et al. 2003, Réale et al.
71 2003, Robertson et al. 2006, Barbraud et al. 2008, Burger et al. 2012, Eeva et al. 2015, Espín et al.
72 2016, Johnson et al. 2016, Dobson et al. 2017, Keogan et al. 2018, Camacho et al. 2019, Derry et
73 al. 2019, Mennerat et al. 2019, Paniw et al. 2019), and facilitate evidence-based conservation
74 (Stokes 2008, Tylianakis et al. 2008, Vatka et al. 2014, Festa-Bianchet et al. 2019).

75 The first large-scale individual-based field studies of vertebrates were conducted on birds and birds
76 remain the most commonly studied group (Clutton-Brock & Sheldon 2010, Radschuk et al. 2019).
77 Several types of birds (e.g. hole-nesting passerines, colonially breeding seabirds, or fairy-wrens)
78 have proven to be highly suitable for long-term individual-based monitoring of reproduction and
79 survival. Some of the longest-running field studies with over 65 years of non-interrupted time series
80 focus on hole-nesting birds (e.g. Lack 1954, Lack 1966, Kluijver 1951, Ahola et al. 2007) - well
81 suited to detailed study as they often breed at high densities in nest-boxes (Dhondt 2007,
82 Lambrechts et al. 2010), which allows for easy monitoring of the breeding performance (e.g. lay
83 date, clutch size, nesting success) and capture of a large number of birds. Up to now studies cover
84 species with different life-histories over a wide latitudinal and longitudinal range, and in a variety
85 of habitat types, including urban habitats (van den Steen et al. 2009, Andersson et al. 2015, Salmón
86 et al. 2016, Charmantier et al. 2017, Corsini et al. 2017, Senar et al. 2017, Seress et al. 2018).
87 Importantly, these long-term datasets covering a wide range of environmental variables make it
88 possible to answer questions that were not anticipated at the onset of data collection (e.g. influence
89 of global warming on phenology, Visser et al. 1998; effects of habitat fragmentation, Dhondt 2007).

90 The main asset of individual-based bird studies is not only the long temporal scale, but also the
91 high degree of spatial replication provided by multiple studies conducted simultaneously (Korsten
92 et al. 2010, Dingemanse et al. 2012). The amount of information available when studies are
93 combined has the potential to bring our understanding of ecological and evolutionary processes to
94 entirely new levels, and has, not surprisingly, led to a number of collaborative projects (e.g. Visser
95 et al. 2003, Both et al. 2004, Møller et al. 2006, Sæther et al. 2007, Eeva et al. 2011, Szöllösi et al.
96 2011, Laine et al. 2016, Vaugoyeau et al. 2016, Wilkins et al. 2016, Keogan et al. 2018, Samplonius
97 et al. 2018, , Loukola et al. 2020), and we provide some examples in more detail in Box 1. This
98 large-scale synthesis (including the meta-analysis context, (Culina et al. 2018a, Siepielski et al.
99 2017, Siepielski et al. 2019) is especially important for capturing the diversity of biological systems

100 and the variation in ecological conditions that are experienced by different populations. Which
101 processes may be described as being general? Which processes can be identified as being more
102 specific to certain environmental conditions? Only when we have answers to these questions, we
103 can make predictions and tackle global issues, such as habitat degradation, animal welfare, or
104 global warming, and gain insights into reproducibility of findings based on ecological time-series.

105 Over time, individual-based studies have become more complex and in addition to data on breeding
106 parameters, other types of data have been collected (e.g. morphological, behavioural,
107 physiological, genetic and genomic). Further, the number of potential relational links to other
108 sources, such as biological samples, climatic data and individual movement data has increased.
109 With the increasing extent and complexity of datasets we urgently need to address data archiving,
110 standards and integration, not only for individual based-studies but in all branches where many
111 independent research groups collect similar but differently managed, and consequently under-
112 exploited, data (the long-tail of science, Palmer et al. 2007, Wallis et al. 2013). In these branches,
113 transition to Findable, Accessible, Interoperable, and Reusable (FAIR) data (Wilkinson 2016;
114 Box2) is more urgent, but also more challenging compared to fields where data standards have
115 been set up at the very start (e.g. genomics).

116 Scientific collaborations that involve large-scale sharing of data have been shown to generate
117 significant insights, but we can only guarantee this with adequate mechanisms in place to align,
118 store, and advertise the data that are available for such endeavours. Examples of projects that
119 successfully integrate animal data across a large number of studies are Movebank database
120 (<https://www.movebank.org/>, Kranstauber et al. 2011) on animal movement data, EURING Data
121 Bank (<https://euring.org/>, du Feu et al. 2016) that stores encounter records of ringed birds, and
122 COMADRE database on animal demography (Salguero-Gómez et al. 2016).

123

124 *Barriers to collaboration*

125 Ideally, data should be openly archived in a way that supports FAIR principles (Wilkinson 2016),
126 as increasingly mandated by funders (Roche et al. 2014, Culina et al. 2018b). For this, all data
127 should be in a single, standard format, and accompanied by rich meta-data that include the
128 description of the data collection protocols, and support data finding and reuse. In practice, this is
129 difficult to achieve.

130 The core *cultural/sociological reasons* that prevent open data are the lack of incentives, the fear of
131 being scooped, and worries about losing control and overview over the way the data are interpreted
132 and used (Roche et al. 2014, Evans 2016). The latter is not without good reason: it is easy to
133 misinterpret data collected under specific ecological conditions and to misunderstand how
134 variables were derived (Nelson 2009, Mills et al. 2016). Even when researchers are willing to adopt
135 common data standards, they might lack the technical knowledge or time. Yet, inspiring examples
136 of overcoming these barriers exist. For example, all national bird ringing schemes originally used
137 their own data storing format, but ultimately agreed on one common output format, creating the
138 European Union for Bird Ringing (EURING, du Feu et al. 2016). Now, all bird ringing data can be
139 brought together at the European level.

140 The core *practical obstacles* to effective data reuse and collaboration are the lack of: i) meta-data
141 standards to describe populations, ii) data standards, and iii) a central registry of all the populations
142 (Culina et al. 2018b). To find datasets, researchers commonly search the published literature, then
143 contact data owners (who are not always readily reachable, e.g. if they change institution or retire)
144 to determine whether the data are suited for an intended project and whether their owner is willing
145 to share them. This process can take up to a year, and sometimes it fails (personal experience of
146 the authors). If the data are obtained, the user needs to understand the specific conditions of data
147 collection (e.g. specific field protocols, ecologically relevant conditions), the data structure and

148 vocabularies. Groups/researchers store data in different types of databases and formats, use
149 different vocabularies to name data elements (e.g. different languages) or different coding for the
150 same data element (e.g. some record hatching date as day 1, others as day 0). Thus, data owners
151 usually require much time to extract and compile the data and provide meta-data to meet the user's
152 needs. This process needs to be repeated for each new collaborative project. Reformatting data is
153 not only time consuming, but may also increase the risk of introducing errors.

154 Cultural and technical barriers must be solved in parallel; and until open data practices become the
155 norm and researchers recognise their benefits, it is crucial to encourage and enable proper data
156 archiving and establish meta-data and data standards. To achieve FAIR data, and to increase and
157 facilitate collaboration and data synthesis, we created Studies of Populations of Individuals – Birds
158 (SPI – Birds) Network and Database. To overcome cultural barriers, we opted for an approach
159 where data owners keep full control over the use of their data (if they do not wish otherwise), and
160 their own way of data management, with SPI-Birds converting these data into a standard format.
161 These FAIR data can later easily be made open access fully (e.g. GenBank, plant trait database
162 Try, Kattge et al. 2011), or selectively (e.g. Movebank, Kranstauber et al. 2011).

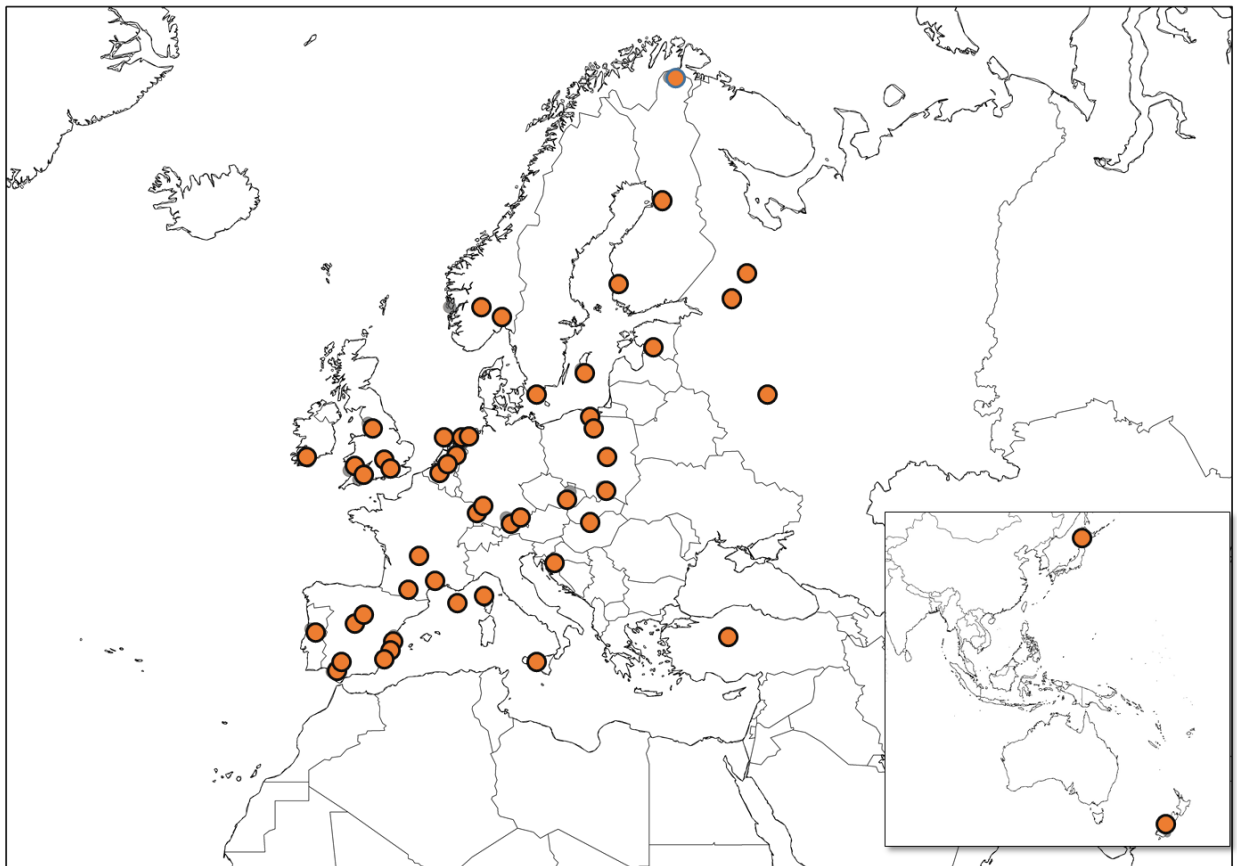
163

164 **SPI-Birds: Connecting researchers and data**

165 SPI-Birds Network and Database (www.spibirds.org) is a grassroots initiative that connects
166 researchers working on populations of birds in which individuals are uniquely marked. The main
167 goals are to: (1) increase the coordination and collaboration between research groups; (2) host the
168 registry of populations and equalise the visibility among research groups; (3) buffer against data
169 loss and provide long-term access to datasets; (4) ensure data quality and integrity; and (5) facilitate
170 data use, and give appropriate credit for data use. To achieve these goals we: (a) derive meta-data
171 attributes that describe populations (Box 2); (b) centrally archive version-controlled primary data

172 from research groups, with attached conditions of data use; (c) derive data standards with controlled
173 vocabularies and convert primary data format into a common standard; (d) conduct data quality
174 checks; (e) run a series of technical reports on the impact that protocols for data collection may
175 have on derived variables; (f) provide expert advice to researchers setting up new populations; and
176 (g) provide an online interface to find and request data, and maintain outreach activities.

177 To date, we count more than 120 members from 21 countries, monitoring over 80 populations of
178 17 mostly hole-nesting passerine species (Fig. 1). However, as a part of our long-term goal, we are
179 actively reaching out to researchers that work on other species groups across the world, and so far



180 have incorporated some of these into the database (e.g. owls, seabirds, dunnocks).

181 **Figure 1.** A map showing the location of the populations with the data hosted in the SPI-Birds
182 database as of July 2020.

183

184 *Community data standards*

185 To facilitate data compatibility and integration, SPI-Birds has already created data standards for
186 storing breeding-season data on individually monitored birds. They are described in detail on the
187 SPI-Birds GitHub repository (Culina et al. 2019). The standards are designed to cover the data
188 fields that are common across most individual-based bird studies, and are aligned with the
189 standards suggested by the Ecological Meta-Data Language (EML, Jones et al. 2019) and the
190 principles of FAIR data (Wilkinson 2016). They are dynamic and can be further extended or
191 adjusted to accommodate the breeding biology (e.g. cooperative breeders) of species yet to be
192 included into the database.

193 SPI-Birds creates tailored pipelines to convert data from each research group/contributor (i.e.
194 primary data format, Box 2) into the standard format. We also hope that this community standard
195 will be adopted by both new and existing research groups to archive their data. The existing groups
196 will be more likely to start using the standard format once their old data have been converted into
197 it by SPI-Birds. We further plan to extend this format (and create new standards) to accommodate
198 other information (e.g. genetic, hormonal, colouration, ecotoxicological, behavioural data).
199 Currently, each population's meta-data clearly indicates whether this additional information has
200 been collected, and the corresponding data tables can be stored at SPI-Birds (although not yet
201 standardized).

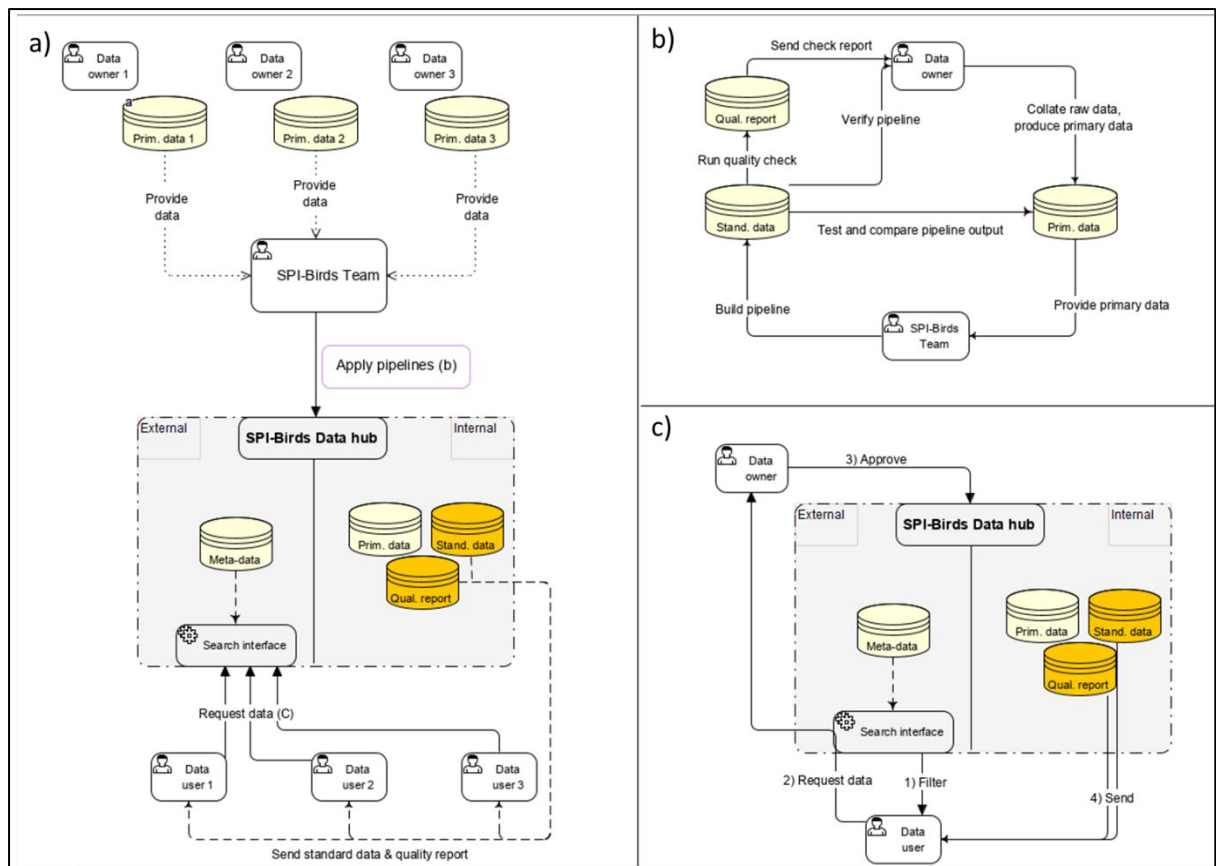
202

203 *Data processing: integration, quality checks, yearly updates*

204 Fig. 2a provides an overview of the SPI-Birds data flow (data collection, standardization, request,
205 and provisioning). First, data owners upload their primary data (Box 2) to SPI-Birds. Primary data
206 come in various storing formats of different complexity, from spreadsheet files (e.g. MS Excel)
207 and simple self-contained databases (e.g. MS Access), to dedicated database servers (e.g. MySQL).
208 Tailored pipelines are then constructed (Fig. 2b) for each dataset to convert primary data into the

209 community standard. The pipeline code is discussed with the data owners (e.g. discussion of how
 210 fields in the primary data are coerced into corresponding fields in the standard format) to ensure
 211 maximum accuracy. Pipeline construction sometimes requires several iterations before an accurate
 212 pipeline is created. At this point, the pipeline can be confidently applied. The primary data and data
 213 in the standard format are stored within the secure SPI-Birds data hub (Fig. 2a), at the file server
 214 cluster of the Netherlands Institute of Ecology (NIOO-KNAW) and backed up seven times a week.
 215 The pipelines are publicly available via GitHub.

216



217

218 **Figure 2.** Overview of SPI-Birds infrastructure. Main data workflow (a) consists of provisioning
 219 of primary data, data processing - standardization and quality checks (b), and data request and use
 220 (c). The internal part of the SPI-Birds data hub stores versioned data in the primary and the standard
 221 format, with an accompanying quality report for each dataset. Users can search meta-data and
 222 request data (c) via the external part of the SPI-Bird data hub (i.e. website). Data is sent to the user

223 (if approved by the data owner, or if data are fully open access) in the community standard format,
224 together with the data-quality report(s).

225 *Standard quality check* is applied to the standardized data. It involves (automated) checks for
226 missing data, formats of variables (e.g. date, integer), inconsistencies between variables (e.g. false
227 brood assignment), and unexpected values within variables. The output of the standard quality
228 check are two types of flags: ‘warnings’ (i.e. values that are uncommon or unusual) and ‘likely
229 errors’ (i.e. values that should be impossible). In discussion with the data owner, warnings and
230 likely errors are resolved, if at all possible, and the quality check is updated. If data owners decide
231 to address the ‘warnings’ and ‘likely errors’ and update their own primary data, these updated
232 primary data will then be stored in the SPI-Birds data hub, under the version control system.
233 Finally, any remaining unresolved flagged records appear in the quality check report that is sent to
234 the user. The ‘warnings’ and ‘likely errors’ list is part of the meta-data for each version of the
235 dataset.

236 For all populations with ongoing data collection primary data are updated to a new version every
237 year and may include not only additional data from the past 12-months but also corrections of
238 errors found in earlier data. We store all versions of the primary data following these yearly
239 updates. This way, we aid to the reproducibility of results based on the version of the data used for
240 the analysis.

241

242 *Data use: discovery, provisioning, terms of use*

243 Populations hosted at SPI-Birds can be searched via SPI-Birds website based on meta-data (e.g.
244 species studied, country, length of data collection, variables measured). Once the relevant
245 populations have been identified, data can be requested using the SPI-Birds request form. Unless
246 the data owners have made their data fully open access, data requests are sent for approval to the
247 data owner. If approval is given, standardized data from the requested population(s), accompanied
248 by the standard quality check report, are delivered to the user, and the data owner is informed about

249 the data sharing. When the user requests multiple datasets, all datasets (in a standard format) are
250 compiled and sent to the user. To give credit to those who have been collecting/managing the raw
251 data, as a minimal requirement for data use (i.e. when the data owner does not request any other
252 conditions of data use) we ask that the data owner and their funding source(s) (as clearly stated in
253 the metadata) are explicitly acknowledged upon data use (e.g. in the acknowledgment section of
254 an article). Further SPI-Birds encourage citation of the dataset source (or related publication) via
255 DOIs (digital object identifiers). The detailed Data Access Policy can be found on our website.

256

257 *Understanding data and their limitations*

258 As discussed above, datasets come with errors and limitations. While SPI-birds increases data
259 integrity and quality, the standard data may still contain errors, and data from different populations
260 might still not be entirely comparable. Thus, to enable users to understand how the primary data
261 and the final dataset (in the standard format) were derived and to highlight potential limitations in
262 the dataset, we provide several documents (as a part of the meta-data). These include the description
263 of the study site (e.g. location, size, habitat type), data collection protocols, and the list of any initial
264 quality checks on the primary data conducted by a data owner. This way, users can better
265 understand how the primary data were derived. Next, we provide a detailed description of decisions
266 and assumptions made during the conversion of data from primary to standard format (with all the
267 pipelines openly available via GitHub), details on quality checks conducted by SPI-Birds and the
268 resulting quality report. Finally, we publish a series of ‘technical reports’ on the SPI-Birds website,
269 where we discuss a range of topics related to methodological conventions (e.g. conversion from
270 one type of tarsus measurement method to another type) and potential biases induced by
271 methodological approaches to data collection (e.g. impact of the frequency at which nests are

272 checked on the estimation of laying date, the impact of nest box design on the vital rates, such as
273 survival of young, Lambrechts et al. 2010).

274

275 **Lessons learned – creating a community data hub**

276 The need to adopt global meta-data and data standards in ecology and evolution is growing (Poisot
277 et al. 2019, Schneider et al. 2019). We strongly believe, and our example supports, that the best
278 way to achieve the addition of global standards is to first create standards for well-defined
279 communities (Poisot et al. 2019). When research communities that work on a similar type of data
280 have established their own standards, it becomes easier to scale up to even larger, more global
281 standards (e.g. EML, Jones et al. 2019). Lessons learned from the SPI-Birds example can be useful
282 to research communities where many researchers (groups) collect data of a comparable type (or
283 purpose), but where research protocols and data management are not uniform (i.e. the so-called
284 long-tail of science, Palmer et al. 2007, Wallis et al. 2013).

285 We suggest four key points in establishing a common database and community data standards in
286 the long-tail of science: (1) How to start: define what your community is and what its needs (and
287 fears) are. Aim to identify researchers/groups that belong to your community. This is largely a
288 snowballing process – once you locate several members, ask them to identify others. Ideally,
289 organize a kick-off meeting to discuss the aims, distribution of tasks, and further steps. From our
290 experience, it is important to have at least several research groups keen on the project at the start.
291 Each community has its needs and fears. It is important to consider these when deciding on the best
292 working model. For example, our success in mobilizing members largely comes from a
293 decentralized approach: data owners keep full control (i.e. ownership) over their data, and over
294 their data management practices. (2) Keep it rolling: involvement and engagement: We found it
295 essential to enable all the members to have the opportunity to contribute to the decisions made. For

296 example, once we created data standards with a smaller subset of researchers, this standard was
297 then opened for comments from all the members. Second, it is important, especially at the start, to
298 show that the project is active. We suggest publishing a newsletter every month or two, and creating
299 a social media account. For example, we tweet about each data set we receive, keeping the
300 community informed of our continuous growth. Third, organise workshops/meetings where the
301 community physically (or virtually) comes together. (3) Funding: Plan to allow for different
302 funding scenarios. We find that it is best to plan finances in steps (if no long-term large funding is
303 available at the very start). Make sure that the first step – what you want to achieve at the minimum
304 – is financially covered from the start. This must include securing a permanent, long-term platform
305 to archive the datasets. After that, plan in five-year (or similar) steps. Here, make sure that in the
306 worst-case scenario (no further funding secured) each step is maintainable with a minimal financial
307 and personal commitment. For example, our first step was to integrate data on hole-nesting
308 passerines in Eurasia, and this period was financially covered by a grants held by participating
309 individuals, and volunteer contributions from several members. After this initial phase the database
310 can be kept functional with a minimal investment (e.g. storage capacity). In the next step, we plan
311 to increase our scope. At this stage, our project has already proven successful, which makes it more
312 attractive for longer-term support (e.g. European open science funds). Finally, we ask (but do not
313 mandate) that those whose research plans rely on the collective power of datasets hosted at SPI-
314 Birds to allocate some of their resources to the SPI-Birds initiative.

315

316 **Vision for an integrated future**

317 SPI-Birds is a large-scale initiative to integrate data on individual-based studies of breeding birds
318 and to connect all researchers working on populations with individually marked birds. With this
319 paper we also call for additional members to join our fast growing community. Further

320 developments of the database include: (1) covering other species and a wider geographical areas;
321 (2) integrating and standardizing other data types (e.g. hormonal, behavioural); (3) connecting with
322 ongoing centralized efforts to map the full spectrum of different types of data on birds that can
323 complement each other. Here, the main collaborators are scientific groups that centralize the
324 collection of complementary types of data (e.g. genetic data from the great tit HapMap project,
325 Spurgin et al. 2019; Movebank, Kranstauber et al. 2011, Fiedler and Davidson 2012; EURING, du
326 Feu et al. 2016). Within this scope, we can connect individual-level data hosted at SPI-Birds to
327 other types of data on the same individuals based on their unique ID and provide even more
328 comprehensive information on individuals across their full life-cycle. A second target group are
329 citizen science projects such as NestKast (de Jong et al. 2018), or the Woodland Trust phenology
330 network (<https://naturescalendar.woodlandtrust.org.uk/>). SPI-Birds can also serve as a platform to
331 enable better resource allocation between research groups. For example, while a data owner might
332 have the data, they might lack funds to analyse them. On the other hand, a data user might have
333 funds or even apply for funds based on these data. SPI-Birds can thus help pull the resources (data
334 and funds) together, thereby enabling scientific projects, and progress, where it may otherwise be
335 unlikely to occur. Finally, during the unforeseen international crisis, such as caused by a novel
336 corona virus during writing of this contribution, SPI-Birds provided an excellent platform to update
337 and document field situations and to mitigate the unbalanced effects of the crisis on research
338 groups. We hope that initiatives such as SPI-Birds can truly help a transition to a new level of
339 ecological synthesis.

340

341

342

Box 1 –Examples of using multiple wild populations

(a) Assessing the ability to substitute space-for-time

Within the scope of understanding and predicting ecological and evolutionary responses to climate change, sampling and studying multiple populations of the same species across latitudinal or altitudinal gradient may provide insights into adaptation to climate variation, if we assume that time can be substituted by space in the processes involved (Blois et al. 2013, Phillimore et al. 2010). For example, Bay et al. (2018) sampled yellow warblers (*Setophaga petechia*) across their breeding range to analyse genomic variation across space and environments (climate, vegetation type, and elevation). Assuming that the current spatial variation in traits of this species may provide information on temporal variation in the future, this study suggested that those yellow warbler populations that have already experienced the largest population declines, require the greatest shifts in allele frequencies to keep pace with future climate change (i.e. are most genetically vulnerable). Similarly, urban-driven evolutionary adaptation is a fascinating process that not only can be followed in time, but also across space, and fostering long-term ecological and evolutionary monitoring in urban areas is key (Szulkin et al. 2020). In urban evolutionary biology, the spatial dimension is particularly valuable from an empirical perspective as it allows researchers to take advantage of replicated urbanisation gradients, where each city or urban area acts as independent urbanisation replicate (Vaugoyeau et al. 2016, Szulkin et al. 2020, Santangelo et al. 2020).

(b) Using spatial replication to infer causal relationships

Spatial variation in local temperature trends across long-term population studies allows researchers to separate effects of climate change from confounding correlates which may also be changing over time. As we, unfortunately, have no replicate world without climate change, it is often difficult to attribute changes in local phenotypic distributions to temperature change, rather than to the multitude of other environmental changes that may happen simultaneously. For example, based on local trends of spring temperatures and laying dates in 25 long-term populations of *Ficedula* flycatchers across Europe, Both et al. (2004) showed that many populations did not exhibit a significant trend towards earlier breeding, but altogether, there was a clear negative population-level correlation between the trend in laying date and the trend in temperature. In a similar analysis on great tits *Parus major* and blue tits *Cyanistes caeruleus*, such an effect of local temperature was not found: rather, populations originally having a low frequency of second broods did advance, whereas populations in which second broods used to be common did not advance their laying dates (Visser et al., 2003). These examples nicely illustrate how both within- and between-species comparisons of long-term studies deepen our understanding of how organisms may adapt to climate change.

(c) Comparisons of evolutionary potential

Evolutionary potential depends on the genetic architecture of traits. From a quantitative genetics perspective, this architecture is summarized in G , the additive genetic (co)variance matrix. Comparisons of evolutionary potential across populations or species enable us to evaluate the generality of evolutionary constraints (Agrawal & Stinchcombe, 2009) and to gain insight into the evolution of the underlying genetic architecture (McGlothlin et al., 2018; Stepan, Phillips, & Houle, 2002). For example, using long-term datasets with pedigree information, Teplitsky et al. (2014) assessed the expected constraints on evolutionary responses of morphological traits in ten populations of seven wild bird species. Based on estimated G matrices and selection gradients for four morphological traits, their results suggest that genetic correlations may reduce the expected rate of evolution by 28% on average, even for traits such as morphological traits, that are generally thought to have a high evolutionary potential.

In terms of the evolution of genetic architecture, Delahaie et al. (2017) showed that the genetic architecture of life history and morphological traits is relatively conserved across populations of blue tits inhabiting contrasting habitats. Additionally, Martínez-Padilla et al. (2017) compiled all published estimates of additive genetic variation of morphological traits quantified from 20 long-term and individually-monitored populations of 12 wild European bird species. They found that the evolutionary potential of morphological traits decreases as environmental conditions approaches the extremes, either being favourable or unfavourable. Stronger selection pressures that erode additive genetic variation when environmental conditions were unfavourable or high intraspecific competition in favourable environmental conditions may explain the pattern. These examples illustrate the need of larger-scale studies, both in terms of geography and phylogeny, to fully address the question of the evolution of genetic architecture in wild populations.

(d) Resolving methodological issues

Long-term individual level studies often vary in protocols, applied methodologies and approaches to data collection. Using many long-term datasets may help identify such variation, and point towards those variables that can have potentially significant impacts on how results are interpreted, especially at the between-study level. Møller et al. (2014) targeted one important, strongly varying, component of long-term hole-nesting bird studies: nest-box design. Their study included reproductive data of four bird species: blue and great tits, and pied and collared flycatchers. They have found significant positive relationship between nest-box floor area and clutch size, as well as extensive between-species variation in the strength of this relationship, indicating that variation in study design at the between-population level should always be included as it may prove an important predictor of some of the observed inter-population variation

Box 2 – Glossary

Individual-based studies of birds – Individual birds are marked with rings engraved with a unique identifying number. Birds are captured (or observed), often over subsequent years, and data on individual characteristics and/or breeding parameters (e.g. laying date, clutch size, number of hatchlings and fledglings, partner) are collected. This information directly links to fitness and thus can be used to study different ecological and evolutionary processes, such as selection on individual traits or population-dependent processes (e.g. density-dependent selection). Other types of data are also increasingly collected, e.g. behavioural, hormonal, genetic or genomic, fine-scale environmental data (including e.g. pollutant data).

Open Data – Data that anyone is free to use, reuse and redistribute — subject, at most, to the requirement to attribute or share-alike, <https://creativecommons.org/licenses/by-sa/2.5/>

FAIR data – FAIR data are equivalent to open data. FAIR data are structured and described in a way that supports their Findability, Accessibility, Interoperability, and Reusability, for both machines and humans.

Metadata – Data that describe datasets. Metadata comprise information explaining the purpose and origin of data, methods used to acquire them, the structure of the data, time references, geographical location, brief description of the study site(s), creator, access conditions and terms of use.

Data owner – A person or institution that has collected the raw data and/or is hosting the primary data.

Data user – A person interested in using the data owned by the data owner. Data owners can be data users of someone else's data.

Raw data – Data as collected in the field.

Primary data – Data stored locally by each research group. Primary data might differ from raw data because of (1) errors made during transcribing raw data into primary data or (2) correction of obvious errors in raw data during transcribing them into primary data (3) primary data contain some derivate of primary data (e.g. average value for a repeated measurement of an individual)

Primary data format – A format in which primary data are stored. This includes the way that data are divided among different tables, the variables recorded, names of these variables, and how values of these variables are expressed.

Standard (community) data format – A format agreed upon within the research community to record and archive data. The standard format defines the way data are organized among different tables, the

vocabularies used to describe the data elements (names of the variables), and conventions used to express the values of the variables.

Standard quality check – A range of checks to test the quality and integrity of the primary data converted into the standard format. Each check differentiates between two main types of flags: ‘warnings’ (values that are uncommon or unusual) and ‘likely errors’ (values that seem impossible).

Data hub – A central location to physically store (archive) all data for a certain domain.

Pipeline – A set of code functions and commands used to convert data provided in the primary format into the standard format. A pipeline usually has a hierarchical structure (outputs of one component of the processing sequence are fed to the next step) and often is modular (non-necessary components can be removed or changed to modify the final structure of output data).

343

344

345 **Funding**

346 The SPI-Birds have been supported by an NWO personal grant (grant number 016.Veni.181.054)
347 to AC, and a Research Council of Norway grant: 223257 (SFF-III) and 267511 (EVOCLIM).

348 **Acknowledgments**

349 We would like to thank Alastair Wilson and Lisette Senerpont Domis for their valuable comments
350 on the manuscript. We would also like to acknowledge the enormous efforts of all current and past
351 fieldworkers that have contributed to data collection in the populations we list.

352 **Data availability**

353 Data have not been used in this manuscript.

354 **Authors contributions**

355 A. Culina: Conceptualization, Methodology, Investigation, Resources, Data Curation, Writing -
356 Original Draft, Visualization, Supervision, Project administration, Funding acquisition; F.
357 Adriaensen, L.D. Bailey, M.D. Burgess, A. Charmantier, E.F. Cole, T. Eeva, E. Matthysen, C. R.
358 Nater, B.C. Sheldon, B-E. Sæther, S.J.G. Vriend, M.E. Visser: Conceptualisation, Methodology,
359 Resources, Data curation, Writing - review & editing. In addition, B-E. Sæther also: Funding. Other
360 co-authors: Resources, Data Curation, Writing - review & editing

361

362

363 References

- 364 Agrawal, A. F. & Stinchcombe J.R. How much do genetic covariances alter the rate of adaptation?
365 *Proceedings of the Royal Society B-Biological Sciences* **276**(1659), 1183-1191 (2009).
366
- 367 Ahola, M.P., Laaksonen, T., Eeva, T. & Lehikoinen, E. Climate change can alter competitive
368 relationships between resident and migratory birds. *Journal of Animal Ecology*, **76**, 1045-1052
369 (2007).
370
- 371 Andersson, M. N., et al. Composition of physiologically important fatty acids in great tits differs
372 between urban and rural populations on a seasonal basis. *Frontiers in Ecology and Evolution* **3**, 93
373 (2015).
374
- 375 Barbraud, C., Marteau, C., Ridoux, V., Delord, K. & Weimerskirch, H. Demographic response of
376 a population of white-chinned petrels *Procellaria aequinoctialis* to climate and longline fishery
377 bycatch. *Journal of Applied Ecology* **45**, 1460-1467 (2008).
378
- 379 Bay R.A., Harrigan, R.J., Underwood, V.L, Gibbs, H.L., Smith, T.B. & Ruegg, K. Genomic signals
380 of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83-86
381 (2018).
382
- 383 Bonnet, T., et al. The role of selection and evolution in changing parturition date in a red deer
384 population. *PLOS Biol.* **17**, e3000493 (2019).
385
- 386 Both, C., et al. Large-scale geographical variation confirms that climate change causes birds to lay
387 earlier. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**(1549),
388 1657–1662 (2004).
389
- 390 Broggi, J., Hohtola, E., Orell, M. & Nilsson, J-Å. Local adaptation to winter conditions in a
391 passerine spreading north: a common-garden approach. *Evolution* **59**, 1600-1603 (2005).
392
- 393 Burger, C., et al. Climate change, breeding date and nestling diet: how temperature differentially
394 affects seasonal changes in Pied flycatcher diet depending on habitat variation. *Journal of Animal*
395 *Ecology* **81**(4), 926–936 (2012).
396
- 397 Camacho, C., Martínez-Padilla, J., Canal, D. & Potti, J. Long-term dynamics of phenotype-
398 dependent dispersal within a wild bird population. *Behavioral Ecology* **30**, 548-556 (2019).
399
- 400 Charmantier, A., Garant, D. and & Kruuk L. E. B. *Quantitative genetics in the wild.* (Oxford
401 University Press, 2014).
402
- 403 Charmantier, A., Demeyrier, V., Lambrechts, M., Perret, S. & Grégoire, A. Urbanization Is
404 associated with divergence in pace-of-life in great tits. *Front. Ecol. Evol.* **5**, 53 (2017).
405
- 406 Class, B., Brommer, J. & van Oers, K. Exploratory behavior undergoes genotype–age interactions
407 in a wild bird. *Ecology and Evolution* **9**, 8987–8994 (2019).
408
- 409 Clutton-Brock, T., Price, O., Albon, S. & Jewell, P. Early development and population fluctuations
410 in Soay sheep. *Journal of Animal Ecology* **61**(2), 381-396 (1992).

411
412 Clutton-Brock, T. & Sheldon, B.C. Individuals and populations: the role of long-term, individual-
413 based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*
414 **25**(10), 562–73 (2010).
415
416 Coltman, D. W., et al. Undesirable evolutionary consequences of trophy hunting. *Nature*,
417 **426**(6967), 655-658 (2003).
418
419 Corsini, M., Dubiec, A., Marrot, P. & Szulkin, M. Humans and tits in the city: quantifying the
420 effects of human presence on great tit and blue tit reproductive trait variation. *Frontiers in Ecology*
421 *and Evolution* **5**, 82 (2017).
422
423 Culina, A., Crowther, T. W., Ramakers, J. J. C., Gienapp, P., & Visser, M. E. How to do meta-
424 analysis of open datasets: comment. *Nature Ecology and Evolution* **2**, 1053-1056 (2018a).
425
426 Culina, A., et al. Navigating the unfolding open data landscape in ecology and evolution. *Nature*
427 *Ecology & Evolution* **2**(3), 420–26 (2018b).
428
429 Culina, A., Bailey, L.D, Vriend, S.J.G, & Visser, M.E. *Standard protocol for the collection of*
430 *individual level data.*
431 [https://github.com/SPI_Birds/documentation/blob/master/standard_protocol/SPI_Birds_ProtoPro](https://github.com/SPI_Birds/documentation/blob/master/standard_protocol/SPI_Birds_ProtoPro_v1.0.0.pdf)
432 [_v1.0.0.pdf](https://github.com/SPI_Birds/documentation/blob/master/standard_protocol/SPI_Birds_ProtoPro_v1.0.0.pdf) (2019).
433
434 Delahaie, B., et al. "Conserved G-matrices of morphological and life-history traits among
435 continental and island blue tit populations." *Heredity* **119** (2), 76-87 (2017).
436
437 Derry, A. M., et al. Conservation through the lens of (mal) adaptation: Concepts and meta-analysis.
438 *Evolutionary Applications* **12**(7), 1287-1304 (2019).
439
440 Dhondt, A. A. in *Ecology and behavior of chickadees and titmice: an integrated approach* (KA
441 Otter, Editor), What drives differences between North American and Eurasian tit studies (Oxford
442 University Press, 2007).
443
444 Dingemanse, N. J., et al. Variation in personality and behavioural plasticity across four populations
445 of the great tit *Parus major*. *Journal of Animal Ecology* **81**, 116– 126 (2012).
446
447 Dobson, F.S., Becker, P.H., Arnaud, C.M., Bouwhuis, S., & Charmantier, A. Plasticity results in
448 delayed breeding in a long-distant migrant seabird. *Ecology and Evolution* **7**, 3100-3109 (2017).
449
450 du Feu, C. R., Clark, J. A., Schaub, M., Fiedler, W., & Baillie, S. R. The EURING Data Bank – A
451 critical tool for continental-scale studies of marked birds. *Ringing & Migration* **31**, 1– 18 (2016).
452
453 Eeva, T., et al. Geographical trends in the yolk carotenoid composition of the pied flycatcher
454 (*Ficedula hypoleuca*). *Oecologia* **165**, 277-287 (2011).
455
456 Eeva, T. & Lehikoinen, E. Long-term recovery of clutch size and egg shell quality of the pied
457 flycatcher (*Ficedula hypoleuca*) in a metal polluted area. *Environmental Pollution* **201**, 26-33
458 (2015).
459

460 Espín et al. Tracking pan-continental trends in environmental contamination using sentinel raptors-
461 what types of samples should we use? *Ecotoxicology* **25**(4), 777-801 (2016).
462

463 Evans, S.R. Gauging the purported costs of public data archiving for long-term population studies.
464 *PLoS Biol* **14**(4), e1002432 (2016).
465

466 Festa-Bianchet, M., Cote, S., Hamel, S. & Pelletier, F. Long-term studies of bighorn sheep and
467 mountain goats reveal fitness costs of reproduction. *J Animal Ecology* **88**, 1118-
468 1133 (2019).
469

470 Fiedler W. & Davidson, S. Movebank: eine offene Internetplattform für Tierwanderungsdaten
471 [Movebank: an open internet platform for animal movement data]. *Vogelwarte* **50**(1), 15–20
472 (2012).
473

474 Grant, P. R. & Grant, B. R. Unpredictable evolution in a 30-year study of Darwin’s finches. *Science*
475 **296**, 707–117 (2002).
476

477 de Jong, M., et al. Timing of avian breeding in an urbanized world. *Ardea* **106**, 31 – 38 (2018).
478

479 Johnston, S. E., Béréños, C., Slate, J., & Pemberton, J.M. Conserved genetic architecture
480 underlying individual recombination rate variation in a wild population of Soay sheep (*Ovis aries*).
481 *Genetics* **203**, 583-598 (2016).
482

483 Kattge J., et al. TRY – a global database of plant traits. *Global Change Biology* **17**, 2905–2935
484 (2011).
485

486 Keogan, K., et al. Global phenological insensitivity to shifting ocean temperatures among seabirds.
487 *Nature Climate Change* **8**, 313-318 (2018).
488

489 Kluijver, H.N..The population ecology of the great tit, *Parus m. major* L. *Ardea* **39**(1-3), 1-135
490 (1951).
491

492 Korsten, P., et al. Association between DRD4 gene polymorphism and personality variation in
493 great tits: a test across four wild populations. *Molecular Ecology* **19**, 832-843 (2010).
494

495 Korsten, P., van Overveld, T., Adriaensen, F., & Matthysen, E. Genetic integration of local
496 dispersal and exploratory behaviour in a wild bird. *Nature Communications* **4**, 2362 (2013).
497

498 Kranstauber, B., et al. The Movebank data model for animal tracking. *Environmental Modelling &*
499 *Software* **26**(6), 834–835 (2011).
500

501 Jones, M.B., O’Brien, M., Mecum, B., Boettiger, C. Schildhauer, M. Et al. Ecological Metadata
502 Language version 2.2.0. KNB Data Repository (2019). doi:10.5063/F11834T2
503

504 Lack, D. *The Natural Regulation of Animal Numbers* (Clarendon Press, 1954).
505

506 Lack, D. *Population Studies of Birds* (Oxford University Press, 1966).
507

508 Laine, V., et al. Evolutionary signals of selection on cognition from the great tit genome and
509 methylome. *Nat. Commun.* **7**, 10474 (2016).

510
511 Lambrechts, M.M., et al. The design of artificial nestboxes for the study of secondary hole-nesting
512 birds: a review of methodological inconsistencies and potential biases. *Acta Ornithol.* **45**, 1–26
513 (2010).
514
515 Loukola, O. J., et al. The roles of temperature, nest predators and information parasites for
516 geographical variation in egg covering behaviour of tits (*Paridae*). *Journal of Biogeography*
517 (2020).
518
519 Martínez-Padilla, J., et al. Evolvability meets biogeography: evolutionary potential decreases at
520 high and low environmental favourability. *Proceedings of the Royal Society B: Biological Sciences*
521 **284**(1856), 20170516 (2017).
522
523 McGlothlin, J. W., et al. Adaptive radiation along a deeply conserved genetic line of least resistance
524 in *Anolis* lizards. *Evolution Letters* **2**(4), 310-322 (2018).
525
526 Mennerat, A., Charmantier, A., Perret, P., Hurtrez-Boussès, S. & Lambrechts, M.M. Parasite
527 intensity is driven by temperature in a wild bird. Preprint at:
528 <https://www.biorxiv.org/content/10.1101/323311v1> (2019).
529
530 Mills, J.A., et al. Archiving primary data: solutions for long-term studies. *Trends in Ecology &*
531 *Evolution* **30**(10), 581–89 (2015).
532
533 Møller, A.P., et al. An analysis of continent-wide patterns of sexual selection in a passerine bird.
534 *Evolution* **60**, 856-868 (2006).
535
536 Møller, A. P., et al. Variation in nest size in relation to clutch size in birds. *Ecol. Evol.* **4**, 3583—
537 3595 (2014).
538
539 Nelson, B. Empty archives. *Nature* **461**, 160-163 (2009).
540
541 Paniw, M.,*Maag, N., Cozzi, G., Clutton-Brock, T., & Ozgul, A. Life history responses of meerkats
542 to seasonal changes in extreme environments. *Science* **363**, 631-635 (2019).
543
544 Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. Ecological data should not be
545 so hard to find and reuse. *Trends in Ecology & Evolution* **34**(6), 494–496 (2019).
546
547 Radchuk, V., et al. Adaptive responses of animals to climate change are most likely insufficient.
548 *Nature Communications* **10**(1), 1–14 (2019).
549
550 Réale, D., McAdam, A., Boutin, S., & Berteaux, D. Genetic and plastic responses of a northern
551 mammal to climate change. *Proceedings of the Royal Society of London. Series B: Biological*
552 *Sciences* **270**(1515), 591-596 (2003).
553
554 Robertson, B. C., Elliott, G. P., Eason, D. K., Clout, M. N., & Gemmill, N. J. Sex allocation theory
555 aid species conservation. *Biol. Lett.* **2**, 229-231 (2006).
556
557 Roche, D. G., et al. Troubleshooting public data archiving: suggestions to increase participation.
558 *PLoS Biol* **12**(1), e1001779 (2014).
559

560 Sæther, B-E., et al. The extended moran effect and large-scale synchronous fluctuations in the size
561 of great tit and blue tit populations. *Journal of Animal Ecology* **76**, 315-325 (2007).
562

563 Salguero-Gómez, R., Jones, O.R., Archer, C.R., Bein, C., de Buhr, H et al. COMADRE: a global
564 data base of animal demography. *Journal of Animal Ecology* **85**, 371-384 (2016).
565

566 Salmón, P., Nilsson, J. F. , Nord, A., Bensch, S., & Isaksson, C. Urban environment shortens
567 telomere length in nestling great tits, *Parus Major*. *Biology Letters* **12**(6), 20160155 (2016).
568

569 Samplonius, J. M., et al. Phenological sensitivity to climate change is higher in resident than in
570 migrant bird populations among European cavity breeders. *Glob. Change Biol.* **24**, 3780– 3790
571 (2018).
572

573 Santangelo, J. S., et al. In: *Urban Evolutionary Biology* (ed. Szulkin M., Munshi-South J., &
574 Charmantier, A.) Urban environments as a framework to study parallel evolution. (Oxford
575 University Press, 2020).
576

577 Schneider et al. Towards an ecological trait-data standard. *Methods Ecol. Evol.* **10**, 2006–2019
578 (2019).
579

580 Schroeder, J., Rees, M., Nakagawa, S., & Burke, T. Reduced fitness in progeny from old parents
581 in a wild population. *PNAS* **112**, 4021-4025 (2015).
582

583 Senar, J.C., et al. Urban great tits (*Parus major*) show higher distress calling and pecking rates than
584 rural birds across Europe. *Front. Ecol. Evol.* **5**, 163 (2017).
585

586 Seress, G., et al. Impact of urbanization on abundance and phenology of caterpillars and
587 consequences for breeding in an insectivorous bird. *Ecological Applications* **28**, 1143-1156 (2018).
588

589 Siepielski, A. M., et al. Precipitation drives global variation in natural selection. *Science* **355**(6328),
590 959-962 (2017).
591

592 Siepielski, A. M., et al. No evidence that warmer temperatures are associated with selection for
593 smaller body sizes. *Proceedings of the Royal Society B: Biological Sciences* **286**(1907), 20191332
594 (2019).
595

596 Spurgin et al. The great tit HapMap project: a continental-scale analysis of genomic variation in a
597 songbird. Preprint at: <https://www.biorxiv.org/content/10.1101/561399v1> (2019).
598

599 Van den Steen, E., et al. Brominated flame retardants and organochlorines in the European
600 environment using great tit eggs as a biomonitoring tool. *Environment International*, **35**(2), 310-
601 317 (2009).
602

603 Stepan, S. J., Phillips, P. C., & Houle, D. Comparative quantitative genetics: evolution of the G-
604 matrix. *Trends in Ecology & Evolution*, **17**(7), 320-327 (2002).
605

606 Szulkin, M., Munshi-South, J., & Charmantier, A. In: *Urban Evolutionary Biology* (ed. Szulkin,
607 M, Munshi-South J, & Charmantier A.) Introduction. (Oxford University Press, 2020a).
608

609 Szulkin, M., et al. In: *Urban Evolutionary Biology* (ed. Szulkin, M, Munshi-South J, & Charmantier
610 A.) How to quantify urbanisation when testing for urban evolution? (Oxford University Press,
611 2020b).
612

613 Teplitsky, C., et al. Assessing multivariate constraints to evolution across ten long-term avian
614 studies. *PLoS ONE* **9**(3), e90444 (2014).
615

616 Tylianakis, J.M., Raphael K. Didham, Bascompte J., Wardle, D.A. Global change and species
617 interactions in terrestrial ecosystems. *Ecology Letters*, **11**, 1351–1363 (2008).

618 Votka, E., et al. Nest site selection of a primary hole-nesting passerine reveals means to developing
619 sustainable forestry. *Journal of Avian Biology* **45**(2), 187-196 (2014).
620

621 Vaugoyeau, E., et al. Interspecific variation in the relationship between clutch size, laying date and
622 intensity of urbanisation in four species of hole-nesting birds. *Ecology & Evolution* **6**, 5907–5920
623 (2016).
624

625 Visser M. E., van Noordwijk, A.J., Tinbergen J. M., & Lessells C. M. Warmer springs lead to mis-
626 timed reproduction in great tits (*Parus major*). *Proc. R. Soc. Lond. B.* **265**, 1867-1870 (1998).
627

628 Visser, M. E., et al. Variable responses to large-scale climate change in European *Parus*
629 populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**(1513),
630 367–72 (2003).
631

632 Wallis, J. C., Rolando E., & Borgman C. L. If we share data, will anyone use them? Data sharing
633 and reuse in the long tail of science and technology. *PLoS ONE* **8**(7), e67332 (2013).
634

635 Wilkins, M. R., et al. Phenotypic differentiation is associated with divergent sexual selection
636 among closely related barn swallow populations. *J. Evol. Biol.* **29**, 2410-2421 (2016).
637

638 Wilkinson, M. D. The FAIR Guiding Principles for scientific data management and stewardship.
639 *Sci. Data* **2**, 1–9 (2016).
640
641