

1

Manuscript number: NATECOLEVOL-19036421D

Length of text and legends: 5,980

Number of references: 118

Number of display items: 3 Figures & 3 Boxes

2 **Open Science principles for accelerating trait-based science across the Tree of Life**

3 Gallagher, Rachael V.¹, Falster, Daniel S.², Maitner, Brian S.³, Salguero-Gómez ,
4 Roberto^{4,5,6}, Vandvik, Vigdis^{7,8}, Pearse, William D.⁹, Schneider, Florian D.¹⁰, Kattge,
5 Jens^{11,12}, Poelen, Jorrit H.¹³, Madin, Joshua S.¹⁴, Ankenbrand, Markus J.^{15,16,17}, Penone,
6 Caterina¹⁸, Feng, Xiao³, Adams, Vanessa M.¹⁹, Alroy, John¹, Andrew, Samuel C.²⁰,
7 Balk, Meghan A.²¹, Bland, Lucie M.²², Boyle, Brad L.³, Bravo-Avila, Catherine H.^{23,24},
8 Brennan, Ian²⁵, Carthey, Alexandra J. R.¹, Catullo, Renee²⁵, Cavazos, Brittany R.²⁶,
9 Conde, Dalia A.^{27,28,29}, Chown, Steven L.³⁰, Fadrique, Belen²³, Gibb, Heloise³¹,
10 Halbritter, Aud H.^{7,8}, Hammock, Jennifer³², Hogan, J. Aaron³³, Holewa, Hamish²⁰,
11 Hope, Michael²⁰, Iversen, Colleen M.³⁴, Jochum, Malte^{18,12,35}, Kearney, Michael³⁶,
12 Keller, Alexander^{15,15}, Mabee, Paula³⁷, Manning, Peter³⁸, McCormack, Luke³⁹,
13 Michaletz, Sean T.⁴⁰, Park, Daniel S.⁴¹, Perez, Timothy M.^{23,24}, Pineda-Munoz, Silvia⁴²,
14 Ray, Courtenay A.⁴³, Rossetto, Maurizio^{44,45}, Sauquet, Hervé^{44,46,2}, Sparrow, Benjamin⁴⁷,
15 Spasojevic, Marko J.⁴⁸, Telford, Richard J.^{7,8}, Tobias, Joseph A.⁴⁹, Violle, Cyrille⁵⁰,
16 Walls, Ramona⁵¹, Weiss, Katherine C. B.⁴³, Westoby, Mark¹, Wright, Ian J.¹, Enquist,
17 Brian J.^{3,52}

18

19 ¹Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia , ²Evolution and
20 Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of
21 New South Wales, Sydney, Australia, ³Department of Ecology and Evolutionary Biology, University
22 of Arizona, Tucson, Arizona, USA, ⁴Department of Zoology, Oxford University, Oxford, UK, ⁵Centre
23 for Biodiversity and Conservation Science, University of Queensland, Queensland,
24 Australia, ⁶Evolutionary Demography Laboratory, Max Plank Institute for Demographic Research,
25 Rostock, Germany, ⁷Department of Biological Sciences, University of Bergen, Bergen,
26 Norway, ⁸Bjerknes Centre for Climate Research, University of Bergen, Norway, ⁹Ecology Center and
27 Department of Biology, Utah State University, Utah, USA, ¹⁰Private residence, ¹¹Max Planck Institute
28 for Biogeochemistry, Jena, Germany, ¹²German Centre for Integrative Biodiversity Research (iDiv)
29 Halle-Jena-Leipzig, Leipzig, Germany, ¹³400 Perkins Street Apt 104, Oakland, California, United

30 States, ¹⁴Hawai'i Institute of Marine Biology, University of Hawai'i at Manoa, Hawai'i,
31 USA, ¹⁵Department of Bioinformatics, Biocenter, University of Wuerzburg, Wuerzburg,
32 Germany, ¹⁶Center for Computational and Theoretical Biology, Biocenter, University of Wuerzburg,
33 Wuerzburg, Germany, ¹⁷Comprehensive Heart Failure Center, University Hospital Wuerzburg,
34 Wuerzburg, Germany, ¹⁸Institute of Plant Sciences, University of Bern, Bern, Switzerland, ¹⁹Discipline
35 of Geography and Spatial Sciences, University of Tasmania, Hobart, Australia, ²⁰Commonwealth
36 Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia, ²¹Bio5 Institute,
37 University of Arizona, Tucson, Arizona, USA, ²²School of Life and Environmental Sciences, Centre
38 for Integrative Ecology, Deakin University, Victoria, Australia, ²³Department of Biology, University of
39 Miami, Florida, USA, ²⁴Fairchild Tropical Botanic Garden, Florida, USA, ²⁵Research School of
40 Biology, Australian National University, Canberra, Australia, ²⁶Department of Ecology, Evolution,
41 and Organismal Biology, Iowa State University, Ames, Iowa, USA, ²⁷Species360 Conservation Science
42 Alliance, Bloomington, Minnesota, USA, ²⁸Interdisciplinary Center on Population Dynamics,
43 University of Southern Denmark, Odense, Denmark, ²⁹Department of Biology, University of Southern
44 Denmark, Odense, Denmark, ³⁰School of Biological Sciences, Monash University, Victoria,
45 Australia, ³¹Department of Ecology, Environment and Evolution and Centre for Future Landscapes, La
46 Trobe University, Victoria, ³²National Museum of Natural History, Smithsonian Institution,
47 Washington DC, USA, ³³International Center for Tropical Botany, Department of Biological Sciences,
48 Florida International University, Florida, USA, ³⁴Climate Change Science Institute and Environmental
49 Sciences Division, Oak Ridge National Laboratory, Tennessee, USA, ³⁵Leipzig University, Institute of
50 Biology, Leipzig, Germany, ³⁶School of BioSciences, The University of Melbourne, Victoria,
51 Australia, ³⁷Department of Biology, University of South Dakota, South Dakota, USA, ³⁸Senckenberg
52 Biodiversity and Climate Research Centre (SBIK-F), Frankfurt,, ³⁹Center for Tree Science, The Morton
53 Arboretum, Lisle, Illinois, USA, ⁴⁰Department of Botany and Biodiversity Research Centre, University
54 of British Columbia, British Columbia, Canada, ⁴¹Department of Organismic and Evolutionary Biology
55 and Harvard University Herbaria, Harvard University, Cambridge, Massachusetts, USA, ⁴²School of
56 Biological Sciences, School of Earth & Atmospheric Sciences, Georgia Institute of Technology,
57 Georgia, USA, ⁴³School of Life Sciences, Arizona State University, Tempe, Arizona, USA, ⁴⁴National
58 Herbarium of New South Wales, Royal Botanic Gardens and Domain Trust, Sydney, NSW,

59 Australia,⁴⁵Queensland Alliance of Agriculture and Food Innovation, University of Queensland,
60 Brisbane, Qld, Australia,⁴⁶Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech,
61 Université Paris-Saclay, Orsay, France, ⁴⁷TERN / School of Biological Sciences, Faculty of Science,
62 The University of Adelaide, Adelaide, Australia, ⁴⁸Department of Evolution, Ecology, and Organismal
63 Biology, University of California Riverside, California, USA, ⁴⁹Department of Life Sciences, Imperial
64 College London, Silwood Park, UK,⁵⁰CEFE, CNRS, Univ Montpellier, Université Paul Valéry
65 Montpellier, Montpellier, France, ⁵¹CyVerse, University of Arizona, Tucson, AZ, USA, ⁵²Santa Fe
66 Institute, Santa Fe, New Mexico, USA

67 **SUMMARY**

68 Synthesising trait observations and knowledge across the Tree of Life remains a grand
69 challenge for biodiversity science. Species traits are widely used in ecological and
70 evolutionary science, and new data and methods have proliferated rapidly. Yet accessing
71 and integrating disparate data sources remains a considerable challenge, slowing
72 progress toward a global synthesis to integrate trait data across organisms. Trait science
73 needs a vision for achieving global integration across all organisms. Here, we outline
74 how the adoption of key Open Science Principles – open data, open source, and open
75 methods – is transforming trait science, increasing transparency, democratising access,
76 and accelerating global synthesis. To enhance widespread adoption of these principles,
77 we introduce the Open Traits Network (OTN), a global, decentralised community
78 welcoming all researchers and institutions pursuing the collaborative goal of
79 standardising and integrating trait data across organisms. We demonstrate how
80 adherence to Open Science principles is key to the OTN community and outline five
81 activities that can accelerate the synthesis of trait data across the Tree of Life, thereby
82 facilitating rapid advances to address scientific inquiries and environmental issues.
83 Lessons learned along the path to a global synthesis of trait data will provide a
84 framework for addressing similarly complex data science and informatics challenges.

85 INTRODUCTION

86 Traits, broadly speaking, are measurable attributes or characteristics of organisms. Traits
87 related to function (e.g. leaf size, body mass, tooth size, or growth form) are often used
88 to understand how organisms interact with their environment and other species via key
89 vital rates such as survival, development, and reproduction¹⁻⁵.

90 Trait-based approaches have long been used in systematics and macroevolution to
91 delineate taxa and reconstruct ancestral morphology and function⁶⁻⁸ and to link
92 candidate genes to phenotypes⁹⁻¹¹. The broad appeal of the trait concept is its ability to
93 facilitate quantitative comparisons of biological form and function. Traits also allow us
94 to mechanistically link organismal responses to abiotic and biotic factors with
95 measurements that are, in principle, relatively easy to capture across large numbers of
96 individuals. For example, appropriately chosen and defined traits can help identify
97 lineages that share similar life-history strategies for a given environmental regime^{12,13}.
98 Documenting and understanding the diversity and composition of traits in ecosystems
99 directly contributes to our understanding of organismal and ecosystem processes,
100 functionality, productivity, and resilience in the face of environmental change¹⁴⁻¹⁹.

101 In light of the multiple applications of trait data to address challenges of global
102 significance (Box 1), a central question remains: How can we most effectively advance
103 the synthesis of trait data within and across disciplines? In recent decades, the collection,
104 compilation, and availability of trait data for a variety of organisms has accelerated
105 rapidly. Substantial trait databases now exist for plants²⁰⁻²³, reptiles^{24,25}, invertebrates^{23,26-}
106 ²⁹, fish^{30,31}, corals³², birds^{23,33,34}, amphibians³⁵, mammals^{23,36-38}, and fungi^{23,39} and parallel
107 efforts are no doubt underway for other taxa. Though considerable effort has been made
108 to quantify traits for some groups (e.g. Fig. 1), substantial work remains. To develop and
109 test theory in biodiversity science, much greater effort is needed to fill in trait data across

110 the Tree of Life by combining and integrating data and trait collection efforts.

111 [INSERT FIGURE 1]

112

113 **Current barriers to global trait-based science**

114 Despite the recognised importance of traits, several common research practices limit our
115 capacity for meaningful synthesis across the Tree of Life. These practices include failure
116 to publish usable datasets alongside new findings⁴⁴, missing or inadequate metadata⁴⁵,
117 minimal descriptions of methods used to collate, clean, and analyse trait datasets in
118 published works⁴⁶, and inadequate coordination between researchers and institutions
119 with common goals, such as filling strategic spatial or taxonomic gaps in trait
120 knowledge^{47,48}. Our limited ability to access and redistribute trait data contributes to the
121 widespread reproducibility crisis within science⁴⁹. Any study relying on data that cannot
122 easily be re-used introduces barriers to verifying the claims made by those studies and
123 thereby questions the reproducibility of the science⁵⁰, which is becoming of prime
124 importance to many scientific journals. Such limitations have been common within trait-
125 based science.

126 Access to data is not the only impediment to a global synthesis of trait
127 knowledge. Barriers to synthesis exist because researchers and institutions are
128 apprehensive that the time and resources they spend to create new observations or share
129 legacy data (e.g. observations from field guides, specimens, or publications without data
130 supplements) will not be recognised. Identifying who should receive credit for
131 contributing trait observations (whether via co-authorship or other formal recognition) is
132 a complex issue, particularly where data involve a chain of expertise (e.g. when trait data
133 are extracted from taxonomic treatments involving specimen collectors, digitisers,
134 taxonomists, and curators). Funding bodies are often reluctant to support data

135 management, limiting recognition of the sizeable effort expended on creating bespoke
136 solutions to curating and harmonising trait data from different sources⁵⁰.

137 Opportunities exist for expanding the spatial and taxonomic coverage of trait
138 observations, particularly by strengthening interdisciplinary connections across single
139 organismic groups. Despite certain plant traits (e.g. growth form, height, and leaf size)
140 being carefully catalogued in taxonomic species descriptions⁵¹, these data have only
141 recently been exchanged with large-scale databases such as TRY²¹ or BIEN⁵². Although
142 several informatics challenges in biodiversity science have now been overcome (e.g.
143 synthesising global species occurrence information⁵³ and sharing genetic data on
144 individuals⁵⁴), trait science lacks a vision for achieving global integration across all
145 organisms. We argue that this is not simply a failure of the traits community to learn
146 from existing successful networks. Instead, cataloguing traits is a more complex task that
147 is highly context-dependent and therefore needs a more refined network model than that
148 offered by a centralised repository.

149 We propose that widespread adoption of key Open Science principles (Box 2)
150 could be transformative for trait science in achieving a global synthesis. These principles
151 would lay a strong foundation for transparency, reproducibility, and recognition and
152 encourage a culture of data sharing and collaboration beyond established networks.
153 Openness reinforces the scientific process by allowing increased scrutiny of methods and
154 results, resulting in the deeper exploration of findings and their significance^{46,55-58}. The
155 scope of trait science would increase if researchers and institutions: 1) made datasets
156 available in machine-accessible formats under clear licensing arrangements; 2) created
157 and adopted standardised protocols, handbooks or metadata formats for data collection,
158 documentation, and management (see⁵⁴⁻⁵⁶); and 3) created human-centred networks to
159 reduce the complexity of integrating existing data from disparate sources (e.g. specimens,

160 published literature, citizen science initiatives^{57,58}, and large-scale digitisation efforts).
161 These different sources exhibit systematic differences in error rates, validation, context,
162 reproducibility, and objectivity relative to field-collected trait observations. Without a
163 model of recognition that embraces transparency and fairness, much trait data will
164 remain hidden from science.

165

166 **Introducing the Open Traits Network: a collaborative initiative for accelerating trait** 167 **data synthesis**

168 The Open Traits Network (OTN) is a global, decentralised community welcoming all
169 researchers and institutions pursuing the collaborative goal of standardising and
170 integrating trait data across all organisms. We promote five main objectives built upon
171 Open Science ideals that could transform trait science:

- 172 1. Openly sharing data, methods, protocols, codes, and workflows;
- 173 2. Citing original data collectors and providing scholarly credit;
- 174 3. Providing appropriate metadata together with trait observations;
- 175 4. Collecting trait data following reproducible, standardised methods and protocols
176 (when available), or committing to their development;
- 177 5. Providing training resources in trait collection and database construction using Open
178 Science principles.

179 We envision a future for trait research where protocols for data exchange and re-
180 use are transparent, research findings are reproducible, and all trait data (either newly
181 collected or from legacy sources) are openly available to the research community and
182 broader public. While several network models exist in trait research (Fig. 2), the OTN
183 adopts a decentralised but connected structure with an emphasis on bringing people
184 together through data and expertise.

185 [INSERT FIGURE 2]

186 Often, groups building smaller-scale databases do so in isolation, using their own
187 tools and workflows tailored to their research question; they are decentralised and
188 disconnected (Fig. 2A). Decentralisation has certain advantages, including retaining the
189 power to determine which traits are most useful in a study system and how they should
190 be compiled. There is little formal support or interaction across this style of network, so
191 researchers often collect redundant data and develop similar tools for data collection,
192 cleaning, and integration, which can lead to duplication of effort. There are many small,
193 isolated, and heterogeneous data sources of this sort, increasing the disconnect between
194 pools of trait data⁵⁹.

195 For some organisms, centralised hubs exist to aggregate and standardise trait data
196 across disparate sources (see ^{21,32,60-64}) (Fig. 2B). These trait repositories have become the
197 main access point for trait data on well-studied taxa such as plants and corals, but they
198 remain mostly isolated, limiting the sharing of expertise and information across taxa. As
199 these repositories continue to grow, difficulties with data integration and synthesis will
200 also increase due to the momentum of entrenched workflows and exchange protocols
201 that may not be interoperable.

202 Some successful large-scale initiatives have followed the centralised and
203 connected network model (e.g. the Global Biodiversity Information Facility⁵³ and
204 GenBank⁵⁴). These platforms mandate strict data exchange protocols to facilitate
205 synthesis using standardised vocabularies (e.g. the Darwin Core⁶⁵ and Humboldt Core⁶⁶).
206 These protocols have been central to the explosive growth of biodiversity data as they
207 facilitate the exchange of information using common data formats⁶⁵⁻⁶⁷. Ontologies that
208 provide unified terms and concepts necessary to represent traits have been developed
209 (e.g. Uberon, the multispecies anatomy ontology for animals⁶⁸, and TOP, the Thesaurus

210 of Plant characteristics⁶⁹). These provide integration with other data types (e.g. genetic
211 and environmental) and their corresponding ontologies (e.g. Gene Ontology⁷⁰ and
212 Environmental Ontology⁷¹).

213 Despite these successes, we argue that a centralised and connected network
214 structure will not facilitate trait data synthesis. Trait observations are highly nuanced and
215 hierarchical. Describing multiple aspects of a phenotype for any organism with traits is
216 not amenable to a simplified set of exchange fields that apply across the Tree of Life.
217 While the centralised and connected model (Fig. 2B) does have benefits, it lacks the
218 necessary flexibility to connect trait data where ontologies and exchange formats do not
219 exist. The likely result is that established trait networks will remain isolated and
220 disconnected.

221 The decentralised but connected model (orange connections in Fig. 2C) adopted
222 by the OTN maintains the key advantages of a decentralised network (e.g. taxon- or
223 discipline-specific decision making) while enhancing the level of connectivity among
224 groups, allowing for easier sharing of expertise, tools, and data. These network
225 characteristics also buffer against node loss (e.g. due to lack of funding). Decentralised
226 and connected networks are characterised by socially mediated improvements in
227 learning⁷² as they capitalise on the aggregated judgement of many experts rather than
228 singular opinions⁷³. The OTN model capitalises on existing connections within
229 disciplines and links domains across the Tree of Life to disseminate knowledge about
230 traits. By recognising the importance of specialist taxon groups (light green nodes in Fig.
231 2C) and accommodating their needs into the development of cross-domain tools for
232 synthesis (dark green nodes in Fig. 2C), the OTN model will be particularly beneficial
233 for low-profile taxa that may not be accommodated by a centralised effort to synthesise
234 data. The OTN's open, decentralised network structure will allow researchers to retain

235 agency and independence while also creating a collaborative effort to minimise the
236 duplication of effort.

237

238 **How (and why) to participate in the OTN**

239 The OTN seeks to broaden its membership by lowering barriers to inclusion and
240 advocating for approaches to trait science that benefit data custodians. New members
241 can join the OTN via our website (www.opentraits.org) through two mechanisms: (i)
242 adding a member profile (e.g. name, location, expertise, and collaboration statement)
243 and/or (ii) registering their open-source (or embargoed) trait datasets in the OTN Trait
244 Dataset Registry (see *Activity 1* below). The registry contains metadata for trait datasets
245 and links users to the open dataset. New entries to the Registry will be reviewed by OTN
246 members before being added. This step will facilitate interaction between new and
247 established OTN members and encourage deeper collaboration. Once registered,
248 members will receive regular updates about the OTN, including newly registered trait
249 datasets, notifications about upcoming chances for face-to-face meetings, and funding
250 opportunities. Members will also benefit from the OTN through the sharing of resources,
251 funding calls and workshops where appropriate.

252 OTN membership spans scientists (and institutions) with high-level expertise in
253 trait data science and synthesis activities through to those with strong motivations to
254 work with traits but little expertise. The OTN has already conducted an international
255 workshop facilitated by an open call for participants, with more workshops planned.
256 Following this initial communication process, we are currently sharing ideas and act
257 upon them within subgroups. Being a decentralised network, the OTN does not need to
258 rely on funding and dedicated personnel to complete tasks, though larger goals will
259 benefit from financial support. Instead, we will communicate the joint aims and gaps

260 between network nodes (Fig. 2) and arrange workshops and activities where necessary.

261 We recognise that altruism is unlikely to offer enough motivation to ensure
262 widespread participation in the OTN. The sharing of trait datasets is not merely a
263 technical problem to be solved; it relies on custodians having the skills, incentives, and
264 motivation to contribute. The key incentives for individuals to join the OTN include
265 increasing the findability of their data and expertise and having access to a ready-made
266 network of trait scientists and institutions engaging in relevant initiatives. Data are a
267 powerful asset for researchers, and release under open-license schemes accompanied by
268 well-defined metadata offers great potential for new collaborations and increased
269 visibility. A persistent concern is that scientists will lose control of their hard-earned data
270 under open licensing, though this underestimates the potential for new collaborations
271 and may unnecessarily increase distrust within the scientific community⁷⁴. Access to
272 scientific networks can provide valuable exposure and connection⁵⁶, particularly for
273 early-career researchers and those in developing nations, although it is important to
274 understand the risks involved. By emphasising the importance of community
275 engagement and support, the OTN seeks to make trait-data sharing and synthesis an
276 opportunity for all involved rather than simply a technical challenge to be solved.

277

278 **Milestones toward an open approach to trait-based science**

279 We highlight five OTN activities (several of which are already operational) that
280 demonstrate the power of a decentralised and connected network to increase knowledge
281 transfer in trait science. Trait scientists have made significant achievements in key areas,
282 such as the synthesis of large numbers of observations within taxonomic groups²⁰⁻³⁹ and
283 the development of theory and frameworks to use these data when testing ideas and
284 large-scale empirical studies¹⁻¹⁹. However, basic foundations are still lacking to quantify

285 how and why traits vary across organisms.

286

287 ***Activity 1: Maintaining a global registry of trait-based initiatives***

288 Several data gaps impede synthetic analyses across taxa, geographical locations, and
289 ontogeny. The heterogeneous ways in which trait data have been collected to date have
290 resulted in a patchy and unrepresentative data landscape across trait types, taxa, regions,
291 and times of the year^{75,76}. The OTN bridges these gaps by maintaining a Trait Dataset
292 Registry that can be accessed at <http://opentraits.org/datasets.html>.

293 The OTN Registry contains information on existing open (or embargoed) datasets
294 so that gaps can be identified and ultimately filled through collective effort. Core
295 information for the registry includes Digital Object Identifier (DOI), taxonomic
296 coverage, curator, and format. The OTN Registry also provides the opportunity for
297 contributors to identify if and where code to process and manipulate raw data is located
298 (see *Activity 2*). As it develops, the OTN Registry will relate trait concepts to ontologies
299 provided through the Open Biomedical Ontologies Foundry⁷⁷. The OTN Registry maps
300 to several Open Science principles (e.g. Open Source, Open Data, and Open Access; Box
301 2) and is designed to support data retrieval and integration.

302 The OTN does not place restrictions on what members may consider traits of
303 importance to a taxonomic group. Most traits can be measured from individuals and fit
304 into existing definitions, though this may not be appropriate for organisms where
305 individual or taxonomic boundaries are unclear (e.g. microbes⁷⁸ and fungi⁷⁹). It can be
306 argued that traits encompass emergent properties of populations (e.g. abundance and
307 geographic range size) or represent interactions among species (e.g. diet type). Within
308 the OTN, we believe that more important than imposing strict definitions around traits is
309 engaging the community in discussion about the utility of available data for answering

310 novel ecological and evolutionary questions.

311

312 ***Activity 2: Sharing reproducible workflows and tools for aggregating trait data***

313 The OTN leverages collaborative software development via platforms like GitHub

314 (<https://github.com/>) to create modular, open-source software to access, harmonise,

315 and re-use data with seamless piping of data from one software tool to the next. OTN

316 contributors have already developed several open-source tools such as the *traitdataform*

317 package, which assists R users to format their data and harmonise units

318 (<http://ecologicaltraitdata.github.io/traitdataform>). The code for the Coral Traits

319 database³² (<https://github.com/jmadin/traits>) could be modified to guide the creation of

320 databases on other organisms. The FENNEC project provides a tool for accessing and

321 viewing community trait data as a self-hosted website service⁸⁰

322 (<https://github.com/molbiodiv/fennec>). The OTN can act as a connector between

323 developers and the broader community seeking to synthesise trait data, facilitating the

324 training of scientists in all aspects of reproducible data management.

325

326 ***Activity 3: Advocating for a free flow of data and appropriate credit***

327 One goal of the OTN is to increase the use of open datasets and to ensure due credit is

328 given to researchers who collect or synthesise primary data. Without effective reward or

329 motivation for collecting new trait observations or sharing legacy data, a trait synthesis

330 across the Tree of Life will remain unattainable. Currently, motivation for collecting and

331 sharing new primary data is not strong and direct funding for trait data management is

332 scarce.

333 The OTN can strengthen the attribution of credit to data providers and promote

334 new data collection via two paths. Firstly, the OTN will encourage citation back to

335 primary source via a permissive license model that secures authorship attribution (e.g.
336 Creative Commons Attribution 4.0 Int; CC BY 4.0) and the use of DOIs and ORCID
337 identifiers. Open-access datasets with a DOI can be tracked to understand patterns of re-
338 use and to assess the impact of the author's decision to share.

339 There is an important distinction between sharing data within a network and
340 making data publicly available under an open license. Clear license arrangements
341 increase visibility and promote fair attribution and citation (e.g. using Creative
342 Commons licenses such as CC-BY or CC0). CC-BY requires attribution (i.e. citation) to
343 the original creator whereas CC0 doesn't legally require users of the data to cite the
344 source, though this does not affect ethical norms for attribution in research communities
345 (<https://creativecommons.org/share-your-work/public-domain/cc0/>). Identifying who
346 should be credited for prior work on legacy data is complicated by the involvement of
347 many individuals. This issue could be solved, in part, by inviting organisations to be
348 named as contributors or co-authors on outputs using their data or (looking forward)
349 implementing new ways of documenting who should be credited for making specimens
350 or datasets usable in trait-science.

351 Incentives to collect new trait data can be linked to the Open Science practice of
352 pre-registration. In pre-registration, authors archive a public proposal for research
353 activities (e.g. via the Centre for Open Science; <https://cos.io/prereg/>) which, if
354 approved, may receive in-principle acceptance from participating journals. As of March
355 2019, 168 journals are willing to give in-principle acceptance following pre-review of the
356 study design prior to conducting field or experimental work. Ten of these participating
357 journals regularly feature papers on trait-based science (e.g. *BMC Ecology* and *Ecology and*
358 *Evolution*). We envision a situation where the OTN Trait Registry (*Activity 1*) could be
359 used to identify spatial or taxonomic gaps in trait data that could be coupled to pre-

360 registered hypotheses. Together, pre-registration and in-principle acceptance of findings
361 could incentivise the collection of new data, circumventing the growing reliance on
362 available data with known gaps.

363

364 *Activity 4: Creating a trait core to facilitate synthesis and standardisation*

365 Trait science requires its own ‘core’ terminology or data standard that is flexible enough
366 to capture the complexity of trait data. Building on efforts to standardise occurrence data
367 (i.e. Darwin Core⁶⁵) and biological inventories (i.e. Humboldt Core^{47,66}), the OTN
368 envisions a trait core offering a set of cross-domain metadata standards and controlled
369 vocabularies that are (ideally) connected to trait ontologies via unambiguous identifiers.
370 This standard terminology would be implemented across trait-data publications, unifying
371 data in decentralised repositories as well as centralised data portals.

372 A trait core would allow trait data to be: (i) interpreted accurately within the
373 context of their collection (i.e. including information on associated data on factors such
374 as environmental conditions at collection sites, taxa covered, data custodians, or
375 collection methods); and (ii) known by compatible terms so that observations of similar
376 phenomena can be grouped and compared (i.e. what is meant by ‘generation time’ or
377 ‘establishment’ across taxonomic groups^{81,82}). Existing initiatives may provide logical
378 cornerstones for referencing terms and concepts, including Ecological Metadata
379 Language⁴⁵. Several initiatives implement the Ecological Metadata Language (e.g. The
380 Knowledge Network for Biocomplexity⁸³, Darwin Core⁶⁵, and Humboldt Core⁶⁶) and the
381 use of referencing terms from anatomy or phenotype ontologies (e.g. the Plant
382 Ontology⁷³ and the Vertebrate Trait Ontology⁷⁴) to relate traits to publicly defined terms,
383 allowing annotated data to be processed computationally⁷⁷.

384 Progress towards a trait core is already being made through the development of a

385 prototypal Ecological Trait Standard⁸⁴ (Box 3). However, the development and adoption
386 of a trait core requires consultation and coordination within the broader scientific
387 community, a goal which the OTN is ideally placed to advance. The OTN can mobilise
388 expertise for cross-domain workshops and advocate for funding, which allows not only
389 meetings of experts but also the creation of cyber-infrastructure for synthesis nodes (dark
390 green nodes in Fig. 2C). Links to emerging initiatives for biodiversity data
391 standardisation (e.g. Species Index of Knowledge⁶⁴) will also be vital for success, as will
392 ratification of the core through the Biodiversity Information Standards (TDWG,
393 www.tdwg.org).

394

395 *Activity 5: Facilitating consistent approaches to measuring traits within major groups*

396 The OTN will share new developments towards protocols and handbooks for major
397 clades that standardise approaches to capture trait observations. Protocols are necessary
398 because downstream activities such as developing metadata standards (*Activity 4*) will be
399 impossible if trait measurement protocols do not exist. Some research communities have
400 adopted standardised terms^{63,69} and data collection protocols (e.g. plants^{20,85-88},
401 invertebrates^{29,89-91}, mammals³⁶, and aquatic life^{30,32,92}), though these may not always fit
402 the requirements of some studies (e.g. where trait variability rather than the average trait
403 of species is targeted⁹³). Protocols and handbooks may not emerge rapidly and should
404 have the flexibility to be open to innovation through a commitment to version control
405 and updates as techniques evolve. Two versions of the plant trait measurement
406 handbook have been published^{85,94} and several online resources exist that can be updated
407 regularly (see⁹⁵).

408 Standardising approaches to trait measurement across research communities will
409 reduce ambiguity when aggregating data and improve the quality of resulting datasets.

410 Integrating trait standardisation and databasing into taxonomic workflows constitutes a
411 challenge and an opportunity⁷ that holds the promise of bridging the long disconnect
412 between structural and functional traits. The presence of a range of biodiversity
413 collections personnel in the OTN and an open invitation for more to join is expected to
414 catalyse the adoption of trait-based thinking into taxonomic practices.

415

416 **Concluding remarks**

417 This is the opportune time to push towards a new approach to sharing and synthesising
418 trait data across all organisms. Trait science has great potential to increase its taxonomic,
419 phylogenetic, and spatial scopes by leveraging data-science tools, embracing Open
420 Science principles, and creating stronger connections between researchers, institutions,
421 publishers, and funding bodies. We hope that trait enthusiasts, regardless of field and
422 research stage, will engage with the OTN via our website (www.opentraits.org) and help
423 build new connections between disciplines, institutions, and taxonomic domains. By
424 adding metadata profiles for datasets to the OTN Trait Dataset Registry, trait collection
425 efforts become more findable, as do the researchers who have compiled them. We
426 envision that by connecting people with common goals, we can work collectively
427 towards a synthesis of global trait data to preserve the nuances of taxon-specific
428 expertise while also facilitating collaboration across domains. We urge scientists and
429 institutions keen to commit to Open Science principles to make use of existing resources,
430 including those offered by the Centre for Open Science (<https://cos.io/>), the Open
431 Science Training Handbook (<https://open-science-training-handbook.gitbook.io/book/>),
432 the Open Science Training Initiative (<http://www.opensciencetraining.com/index.php>),
433 and FOSTER (<https://www.fosteropenscience.eu/toolkit>).

434 To support and expand the activities of the OTN, we will grow membership and

435 develop communities around synthesis nodes to undertake key activities and secure
436 funding support, in particular for the development of a trait core. Funding for
437 international workshops, technical support, and implementation meetings could drive a
438 new era of trait-based synthesis that mirrors the achievement of similar initiatives such as
439 GBIF, which now houses >1 billion occurrence records.

440 By supporting a reciprocal exchange of expertise and outputs using Open Science
441 principles between researchers and institutions, we can mobilise data for a cross-taxa,
442 worldwide, trait-based data resource to examine, understand, and predict nature's
443 responses to global change. As a better-connected OTN emerges, data streams and
444 coordination will improve, allowing us to deliver information to support globally
445 important research agendas (Box 1) as well as specific data and knowledge to the public
446 through integration with third-party portals. Lessons learned along the path to a global
447 synthesis of trait data across all organisms will provide a framework for addressing
448 similarly complex, context-dependent challenges in biodiversity informatics and beyond.

449

450 **CORRESPONDENCE**

451 Rachael V. Gallagher, Department of Biological Sciences, Macquarie University, North
452 Ryde, NSW 2109 Australia

453 Email: rachael.gallagher@mq.edu.au

454 Phone: +61 2 9850 8404

455

456 **ACKNOWLEDGEMENTS**

457 Ideas presented stem from initial discussions at three international meetings – the
458 Australian National Climate Change Adaptation Research Facility Roundtable on
459 Species Traits, the iDigBio ALA Traits workshop, and the preliminary Open Traits

460 workshop held at the Ecological Society of America. RVG is supported by an Australian
461 Research Council DECRA Fellowship (DE170100208). DSF is supported by an
462 Australian Research Council Future Fellowship (FT160100113). RS-G is supported by
463 NERC R/142195-11-1. WDP is supported by NSF ABI-1759965, NSF EF-1802605, and
464 USDA Forest Service agreement 18-CS-11046000-041. AK received financial support for
465 MJA by the German Research Foundation (DFG KE1743/7-1). CMI was supported by
466 the Biological and Environmental Research program in the United States Department of
467 Energy's Office of Science. CP is supported by the DFG Priority Program 1374. MJ was
468 supported by the German Research Foundation within the framework of the Jena
469 Experiment (FOR 1451) and by the Swiss National Science Foundation. SPM was
470 supported by the Benson Fund from the Department of Paleobiology, National Museum
471 of Natural History. STM is supported by SERDP project RC18-1346. BJE was supported
472 by NSF Grants DEB0133974, HDR1934790, and EF1065844, a Leverhulme Trust
473 Visiting Professorship Grant, and an Oxford Martin School Fellowship. Peer-review
474 from Nicolas Gross and two anonymous colleagues was greatly appreciated.

475

476 **AUTHOR CONTRIBUTIONS**

477 R.V.G wrote the manuscript with contributions from D.S.F., B.S.M., R.S-G.,V.V.,
478 W.D.P., F.D.S., J.K., J.H.P., J.S.M., M.J.A., C.P., X.F., V.M.A., J.A., S.C.A., M.A.B.,
479 L.M.B., B.L.B., C.H.B-A., I.B., A.J.R.C., R.C., B.R.C., D.A.C., S.L.C., B.F., H.G.,
480 A.H.H., J.H., J.A.H., H.H., M.H., C.M.I., M.J., M.K., A.K., P.M., P.Manning, L.M.,
481 S.T.M., D.S.P., T.M.P., S.P-M., C.A.R., M.R., H.S., B.S., M.A., R.J.T., J.A.T., C.V.,
482 R.W., K.C.B.W., M.W., I.J.W., and B.J.E.

483

484 **COMPETING INTERESTS**

485

486 The authors declare no competing interests.

487

488 **BOXES**

489 **Box 1: Research programs dependent on comprehensive trait data across the Tree of**

490 **Life**

491 Access to open data on the traits of all organisms will allow the pursuit of long-standing
492 questions in ecological science, including:

493

494 **Defining major axes of strategy variation across the Tree of Life**

495 Measurements of organismal traits can be used to identify the position of species along
496 trade-off spectrums that shape fitness. For instance, traits (e.g. leaf nitrogen content, leaf
497 mass per area, seed mass, and maximum height) have been used to capture trade-offs
498 between plant species at global scales⁹⁶. With access to open data on the traits of a wide
499 breadth of organisms, we will be able to identify major axes of functional specialisation
500 across the Tree of Life. Traits such as adult body mass, offspring mass at independence,
501 mass-specific metabolic rate, and body temperature capture core differences in the
502 ecological strategies within and across groups⁹⁷⁻⁹⁹. Though there is tremendous potential
503 in the comparison of traits across the Tree of Life via the OTN, consideration of scale
504 and biology are needed to identify where it is appropriate to make such comparisons¹⁰⁰.

505

506 **Conservation of functional diversity in protected areas**

507 Reserve-selection procedures¹⁰¹ seek to identify representative networks of potential
508 protected areas at the lowest financial cost. Species are often the biological unit being
509 targeted regardless of the functional diversity they collectively represent. However, there

510 is an increased focus on ecosystem services and protecting ecosystem function by better
511 representing phylogenetic or functional diversity¹⁰². With trait data from the Tree of Life,
512 systematic conservation planning¹⁰³ could: (i) optimise reserve designs to maximise the
513 conservation of ecosystem function predicted from species traits or (ii) assess the
514 adequacy of current protected areas in representing ecosystem function.

515 Reserve design typically targets species or populations, not ecosystem functions,
516 and is based on three fundamental principles: comprehensiveness, adequacy, and
517 representativeness¹⁰³. Using theoretical predictions about the relationship between traits
518 and ecosystem functions we could predict which traits, or range of trait values¹⁰⁴, should
519 be preserved when designing protected area networks. In this context, *comprehensiveness*
520 would allow for the inclusion of the full range of ecosystem functions (as captured by
521 traits of species present) recognised at an appropriate scale within and across relevant
522 ecosystem units (e.g. bioregions and biomes); *adequacy* would seek to maintain the
523 viability and integrity of ecosystem function and model how functional redundancy (i.e.
524 different species performing the same functional role¹⁰⁵) may scale across landscapes to
525 identify the minimum viable reserve size for maintaining ecosystem functions; and
526 *representativeness* would seek to capture the diversity of functions and the gradients across
527 which they occur, including the level of intraspecific variation and plasticity inherent to
528 the trait being examined.

529

530 **Strengthening predictions of the effects of global change on biodiversity**

531 Traits already provide valuable information for models predicting global change impacts
532 on the biosphere. Access to large-scale data on plant traits allows the distribution of traits
533 to be captured in Earth System Models rather than only modelling competition among
534 broad plant functional types¹⁰⁶⁻¹⁰⁸. Species-based predictions from correlative niche

535 models may also benefit from the integration of data on species biology to capture
536 mechanistic links between function and the environment¹⁰⁹⁻¹¹².

537 The next generation of Earth system models may also benefit from integration of
538 open trait data for (at least) two types of organisms: (i) terrestrial vertebrates and (ii)
539 microbes. Terrestrial vertebrates provide key ecological functions (e.g. dispersal and
540 disturbance¹¹³) and their loss may result in significant changes to ecosystem function¹¹⁴.
541 However, capturing the influence of terrestrial megafauna on forest structure, function,
542 and biogeochemical cycles will be improved by access to data on traits (e.g. size and diet)
543 to parameterise process-based models. Access to knowledge about the traits of soil-borne
544 microbes, whose activities increase atmospheric warming via the decomposition of soil
545 carbon, may help predict planetary responses to climate change¹¹⁵⁻¹¹⁷. Molecular-level
546 traits in microbes (e.g. enzyme activity) allow the estimation of decomposition rates but
547 are difficult to measure, amplifying the need to share these data openly.

548 **Box 2: Using Open Science principles in trait research**

549 Open Science principles outline a movement towards making all aspects of the scientific
550 process transparent and accessible to a wide audience⁵⁸. Open Science principles are
551 rapidly being adopted across the sciences.

552 [INSERT FIGURE BOX1]

553 In this context, knowledge is considered open if anyone can freely access, use, modify,
554 and share it, subject at most to measures that preserve provenance and openness
555 (<http://opendefinition.org/>). Several pronouncements about Open Science principles
556 have already been made, including the Berlin Declaration
557 (<https://openaccess.mpg.de/Berlin-Declaration>), the Bouchout Declaration
558 (<http://www.bouchoutdeclaration.org/declaration/>), and the Denton Declaration
559 (<https://openaccess.unt.edu/denton-declaration>) on open access to science data. Other
560 initiatives champion open practices such as the Bari Manifesto on interoperability¹¹⁸ and
561 the FORCE 11 network, which developed the 'Joint Declaration of Data Citation
562 Principles' (<https://www.force11.org/datacitationprinciples>) and the 'FAIR' principles
563 (<https://www.force11.org/group/fairgroup/fairprinciples>). The FAIR principles address
564 several of the challenges facing trait-based research, namely making data findable,
565 accessible, interoperable, and reusable.

566 **Box 3: A pipeline for harmonising trait data from disparate sources**

567 Preparation and re-use of open trait data requires the following action steps⁸⁴:

568 **1. Data collection and data handling within the project context.** Applying project-

569 specific methodology and storing data in tables suitable for data analysis, applying

570 project-specific terms for taxa, traits, and column labels. Adopting standard

571 methodologies and terminologies from the start will simplify steps 2 and 3 and facilitate

572 data publication. OTN *Activity 5* aims to build consensus for common trait definitions

573 and measurement methodologies for major organismal groups.

574 **2. Harmonisation of taxa and traits into standard terms using ontologies.** Prior to

575 publication, all taxa should be harmonised into accepted names linked to ontologies

576 using uniform resource identifiers. Those can be provided in metadata or secondary data

577 tables. Ontologies for traits are scarce, but if available should also be referred to via URIs

578 to deliver unambiguous trait definitions. OTN *Activity 2* will foster the development of

579 ontologies linking trait data to publicly available resources.

580 **3. Standardisation of table descriptors and metadata using a standard vocabulary.**

581 Data should be published in tables using standard terms for column names, such as those

582 provided by the Ecological Trait-data Standard vocabulary (ETS; doi:

583 10.5281/zenodo.2605377). This ETS implements a minimal terminology that can be

584 adapted to include traits from a variety of organisms and uses of uniform resource

585 identifiers for taxa, traits, methods, and units, thereby following the standards for a

586 semantic web of scientific data. Metadata should point to the applied standard to

587 facilitate interpretation by humans and machines. OTN *Activity 4* will engage in the

588 cross-domain, community-based development for standard vocabularies for trait data.

589 **4. Publication of data and upload to a public repository.** Open Access file hosting

590 services offer permanent hosting and findability of data by assigning a DOI to data

591 publications and stating authorship and conditions for re-use under Creative Commons
592 licenses. OTN *Activity 3* will support recognition for data publications and thereby
593 mitigate the investment into data standardisation for smaller research projects.

594 **5. Synthesis of trait data and re-use in downstream products.** Open-access data
595 publications with high-quality metadata are a valuable complement for meta-analysis or
596 the functional analysis of abundance data. By keeping original author terms and values,
597 the quality of the derived datasets can be assured and controlled for a better integration
598 of multiple datasets. Availability of such high-quality data will facilitate reproducibility
599 and enable computer-aided analysis of large databases. OTN *Activity 1* supports the
600 findability of trait data by creating a public registry, and OTN *Activity 2* will develop
601 tools to aid the compilation of databases.

602

603 REFERENCES

- 604 1 Adler, P. B. *et al.* Functional traits explain variation in plant life history strategies.
605 *Proceedings of the National Academy of Sciences* **111**, 740-745 (2014).
- 606 2 Chapin III, F. S., Autumn, K. & Pugnaire, F. Evolution of suites of traits in response
607 to environmental stress. *The American Naturalist* **142**, S78-S92 (1993).
- 608 3 Chown, S. L. & Gaston, K. J. Macrophysiology—progress and prospects. *Functional*
609 *Ecology* **30**, 330-344 (2016).
- 610 4 Kooijman, S. A. L. M. & Kooijman, S. A. L. M. *Dynamic energy and mass budgets in*
611 *biological systems*. (Cambridge University Press, 2000).
- 612 5 Diaz, S., Cabido, M. & Casanoves, F. Plant functional traits and environmental filters
613 at a regional scale. *Journal of Vegetation Science* **9**, 113-122 (1998).
- 614 6 Harmon, L. J. *et al.* Early bursts of body size and shape evolution are rare in
615 comparative data. *Evolution: International Journal of Organic Evolution* **64**, 2385-
616 2396 (2010).
- 617 7 Sauquet, H. & Magallón, S. Key questions and challenges in angiosperm
618 macroevolution. *New Phytologist* (2018).
- 619 8 Sneath, P. H. & Sokal, R. R. *Numerical taxonomy. The principles and practice of*
620 *numerical classification*. (1973).
- 621 9 Edmunds, R. C. *et al.* Phenoscope: identifying candidate genes for evolutionary
622 phenotypes. *Molecular Biology and Evolution* **33**, 13-24 (2015).
- 623 10 Mungall, C. J. *et al.* The Monarch Initiative: an integrative data and analytic platform
624 connecting phenotypes to genotypes across species. *Nucleic Acids Research* **45**,
625 D712-D722 (2016).

- 626 11 Gkoutos, G. V., Schofield, P. N. & Hoehndorf, R. The anatomy of phenotype
627 ontologies: principles, properties and applications. *Briefings in Bioinformatics* **19**,
628 1008-1021 (2017).
- 629 12 Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. & Wright, I. J. Plant ecological
630 strategies: some leading dimensions of variation between species. *Annual Review of*
631 *Ecology and Systematics* **33**, 125-159 (2002).
- 632 13 Kiørboe, T., Visser, A. & Andersen, K. H. A trait-based approach to ocean ecology.
633 *ICES Journal of Marine Science* (2018).
- 634 14 Kunstler, G. *et al.* Plant functional traits have globally consistent effects on
635 competition. *Nature* **529**, 204 (2016).
- 636 15 Laughlin, D. C. Nitrification is linked to dominant leaf traits rather than functional
637 diversity. *Journal of Ecology* **99**, 1091-1099 (2011).
- 638 16 Finegan, B. *et al.* Does functional trait diversity predict above-ground biomass and
639 productivity of tropical forests? Testing three alternative hypotheses. *Journal of*
640 *Ecology* **103**, 191-201, doi:10.1111/1365-2745.12346 (2015).
- 641 17 Laigle, I. *et al.* Species traits as drivers of food web structure. *Oikos* **127**, 316-326
642 (2018).
- 643 18 Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a
644 metabolic theory of ecology. *Ecology* **85**, 1771-1789 (2004).
- 645 19 West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of
646 allometric scaling laws in biology. *Science* **276**, 122-126 (1997).
- 647 20 Iversen, C. M. *et al.* A global Fine-Root Ecology Database to address below-ground
648 challenges in plant ecology. *New Phytologist* **215**, 15-26 (2017).
- 649 21 Kattge, J. *et al.* TRY—a global database of plant traits. *Global Change Biology* **17**,
650 2905-2935 (2011).
- 651 22 Bernhardt-Römermann, M., Poschlod, P. & Hentschel, J. BryForTrait—A life-history
652 trait database of forest bryophytes. *Journal of Vegetation Science* **29**, 798-800 (2018).
- 653 23 Bennett, J. M. *et al.* GlobTherm, a global database on thermal tolerances for aquatic
654 and terrestrial organisms. *Scientific Data* **5**, 180022 (2018).
- 655 24 Meiri, S. Traits of lizards of the world: Variation around a successful evolutionary
656 design. *Global Ecology and Biogeography* **27**, 1168-1172 (2018).
- 657 25 Myhrvold, N. P. *et al.* An amniote life-history database to perform comparative
658 analyses with birds, mammals, and reptiles. *Ecology* **96**, 3109-3109 (2015).
- 659 26 Schäfer, R. B. *et al.* A trait database of stream invertebrates for the ecological risk
660 assessment of single and combined effects of salinity and pesticides in South-East
661 Australia. *Science of the Total Environment* **409**, 2055-2063 (2011).
- 662 27 Bland, L. Global correlates of extinction risk in freshwater crayfish. *Animal*
663 *Conservation* **20**, 532-542 (2017).
- 664 28 Brun, P., Payne, M. R. & Kiørboe, T. A trait database for marine copepods. *Earth*
665 *System Science Data* **9**, 99-113, doi:10.5194/essd-9-99-2017 (2017).
- 666 29 Parr, C. L. *et al.* GlobalAnts: a new database on the geography of ant traits
667 (Hymenoptera: Formicidae). *Insect Conservation and Diversity* **10**, 5-20 (2017).
- 668 30 Froese, R. & Pauly, D. (Fisheries Centre, University of British Columbia, 2010).
- 669 31 Frimpong, E. A. & Angermeier, P. L. Fish traits: a database of ecological and life-
670 history traits of freshwater fishes of the United States. *Fisheries* **34**, 487-495 (2009).
- 671 32 Madin, J. S. *et al.* The Coral Trait Database, a curated database of trait information for
672 coral species from the global oceans. *Scientific Data* **3**, 160017 (2016).
- 673 33 Garnett, S. T. *et al.* Biological, ecological, conservation and legal information for all
674 species and subspecies of Australian bird. *Scientific Data* **2**, 150061 (2015).

675 34 Wilman, H. *et al.* EltonTraits 1.0: Species-level foraging attributes of the world's
676 birds and mammals: Ecological Archives E095-178. *Ecology* **95**, 2027-2027 (2014).

677 35 Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C. & Costa, G. C.
678 AmphiBIO, a global database for amphibian ecological traits. *Scientific Data* **4**,
679 170123 (2017).

680 36 Jones, K. E. *et al.* PanTHERIA: a species-level database of life history, ecology, and
681 geography of extant and recently extinct mammals. *Ecology* **90**, 2648-2648 (2009).

682 37 Faurby, S. *et al.* PHYLACINE 1.2: The phylogenetic atlas of mammal macroecology.
683 *Ecology* (2018).

684 38 Galán-Acedo, C., Arroyo-Rodríguez, V., Andresen, E. & Arasa-Gisbert, R.
685 Ecological traits of the world's primates. *Scientific Data* **6**, 55 (2019).
686 39 <https://github.com/traitecoevo/fungaltraits>

687 40 Maitner, B. S. *et al.* The bien r package: A tool to access the Botanical Information
688 and Ecology Network (BIEN) database. *Methods in Ecology and Evolution* **9**, 373-
689 379 (2018).

690 41 Jetz, W., Thomas, G., Joy, J., Hartmann, K. & Mooers, A. The global diversity of
691 birds in space and time. *Nature* **491**, 444 (2012).

692 42 Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny.
693 *American Journal of Botany* **105**, 302-314 (2018).

694 43 Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other
695 things). *Methods in Ecology and Evolution* **3**, 217-223 (2012).

696 44 Sholler, D., Ram, K., Boettiger, C. & Katz, D. S. Enforcing public data archiving
697 policies in academic publishing: A study of ecology journals. *Big Data & Society* **6**,
698 2053951719836258 (2019).

699 45 Fegraus, E. H., Andelman, S., Jones, M. B. & Schildhauer, M. Maximizing the value
700 of ecological data with structured metadata: an introduction to Ecological Metadata
701 Language (EML) and principles for metadata creation. *The Bulletin of the Ecological*
702 *Society of America* **86**, 158-168 (2005).

703 46 Parker, T. H. *et al.* Transparency in ecology and evolution: real problems, real
704 solutions. *Trends in Ecology & Evolution* **31**, 711-719 (2016).

705 47 Hortal, J. *et al.* Seven shortfalls that beset large-scale knowledge of biodiversity.
706 *Annual Review of Ecology, Evolution, and Systematics* **46**, 523-549 (2015).

707 48 Cornwell, W. K., Pearse, W. D., Dalrymple, R. L. & Zanne, A. E. What we (don't)
708 know about global plant diversity. *bioRxiv*, doi:10.1101/404376 (2018).

709 49 Stodden, V., Seiler, J. & Ma, Z. An empirical analysis of journal policy effectiveness
710 for computational reproducibility. *Proceedings of the National Academy of Sciences*
711 **115**, 2584-2589 (2018).

712 50 Lowndes, J. S. S. *et al.* Our path to better science in less time using open data science
713 tools. *Nature Ecology & Evolution* **1**, 0160 (2017).

714 51 Weigelt, P., König, C. & Kreft, H. GIFT—A Global Inventory of Floras and Traits for
715 macroecology and biogeography. *Journal of Biogeography* (2019).
716 52 <http://bien.nceas.ucsb.edu/bien/>

717 53 <https://www.gbif.org/>

718 54 <https://www.ncbi.nlm.nih.gov/genbank/>

719 55 Parker, T., Nakagawa, S., Gurevitch, J. & participants, I. w. Promoting transparency
720 in evolutionary biology and ecology. *Ecology Letters* **19**, 726-728 (2016).

721 56 McKiernan, E. C. *et al.* Point of view: How open science helps researchers succeed.
722 *Elife* **5**, e16800 (2016).

723 57 Munafò, M. R. *et al.* A manifesto for reproducible science. *Nature Human Behaviour*
724 **1**, 0021, doi:10.1038/s41562-016-0021 (2017).

725 58 Nosek, B. A. *et al.* Promoting an open research culture. *Science* **348**, 1422-1425
726 (2015).

727 59 Farley, S. S., Dawson, A., Goring, S. J. & Williams, J. W. Situating ecology as a big-
728 data science: Current advances, challenges, and solutions. *BioScience* **68**, 563-576
729 (2018).

730 60 Falster, D. S. *et al.* BAAD: a Biomass And Allometry Database for woody plants.
731 *Ecology* **96**, 1445-1445 (2015).

732 61 Salguero-Gómez, R. *et al.* COMADRE: a global data base of animal demography.
733 *Journal of Animal Ecology* **85**, 371-384 (2016).

734 62 Salguero-Gómez, R. *et al.* The COMPADRE Plant Matrix Database: an open online
735 repository for plant demography. *Journal of Ecology* **103**, 202-218 (2015).

736 63 Marques, G. M. *et al.* The AmP project: comparing species on the basis of dynamic
737 energy budget parameters. *PLoS Computational Biology* **14**, e1006100 (2018).

738 64 Conde, D. A. *et al.* Data gaps and opportunities for comparative and conservation
739 biology. *Proceedings of the National Academy of Sciences* **116**, 9658-9664 (2019).

740 65 Wieczorek, J. *et al.* Darwin Core: an evolving community-developed biodiversity data
741 standard. *PloS One* **7**, e29715 (2012).

742 66 Guralnick, R., Walls, R. & Jetz, W. Humboldt Core—toward a standardized capture of
743 biological inventories for biodiversity monitoring, modeling and assessment.
744 *Ecography* **41**, 713-725 (2018).

745 67 Deans, A. R. *et al.* Finding our way through phenotypes. *PLoS Biology* **13**, e1002033
746 (2015).

747 68 Haendel, M. A. *et al.* Unification of multi-species vertebrate anatomy ontologies for
748 comparative biology in Uberon. *Journal of Biomedical Semantics* **5**, 21 (2014).

749 69 Garnier, E. *et al.* Towards a thesaurus of plant characteristics: an ecological
750 contribution. *Journal of Ecology* **105**, 298-309 (2017).

751 70 Consortium, G. O. The Gene Ontology resource: 20 years and still GOing strong.
752 *Nucleic Acids Research* **47**, D330-D338 (2018).

753 71 Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J. & Lewis, S. E. The
754 environment ontology: contextualising biological and biomedical entities. *Journal of*
755 *Biomedical Semantics* **4**, 43 (2013).

756 72 Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the
757 wisdom of crowds. *Proceedings of the National Academy of Sciences*, 201615978
758 (2017).

759 73 Page, S. E. *The Difference: How the Power of Diversity Creates Better Groups,*
760 *Firms, Schools, and Societies-New Edition.* (Princeton University Press, 2008).

761 74 Tenopir, C. *et al.* Data sharing by scientists: practices and perceptions. *PloS One* **6**,
762 e21101 (2011).

763 75 Tyler, E. H. *et al.* Extensive gaps and biases in our knowledge of a well-known fauna:
764 implications for integrating biological traits into macroecology. *Global Ecology and*
765 *Biogeography* **21**, 922-934 (2012).

766 76 Kissling, W. D. *et al.* Towards global data products of Essential Biodiversity
767 Variables on species traits. *Nature Ecology & Evolution*, 1 (2018).

768 77 <http://www.obofoundry.org>

769 78 Lajoie, G. & Kembel, S. W. Making the Most of Trait-Based Approaches for
770 Microbial Ecology. *Trends in Microbiology* (2019).

771 79 Dawson, S. K. *et al.* Handbook for the measurement of macrofungal functional traits:
772 A start with basidiomycete wood fungi. *Functional Ecology* **33**, 372-387 (2019).

773 80 Ankenbrand, M. J., Hohlfield, S. C., Weber, L., Förster, F. & Keller, A. Functional
774 exploration of natural networks and ecological communities. *Methods in Ecology and*
775 *Evolution* **9**, 2028-2033 (2018).

776 81 Gaillard, J.-M. *et al.* Generation time: a reliable metric to measure life-history
777 variation among mammalian populations. *The American Naturalist* **166**, 119-123
778 (2005).

779 82 Steiner, U. K., Tuljapurkar, S. & Coulson, T. Generation time, net reproductive rate,
780 and growth in stage-age-structured populations. *The American Naturalist* **183**, 771-
781 783 (2014).

782 83 Andelman, S. J., Bowles, C. M., Willig, M. R. & Waide, R. B. Understanding
783 environmental complexity through a distributed knowledge network. *BioScience* **54**,
784 240-246 (2004).

785 84 Schneider, F. D. *et al.* Towards an Ecological Trait-data Standard. *Methods in*
786 *Ecology & Evolution*, <https://doi.org/10.1111/2041-210X.13288> (2019).

787 85 Perez-Harguindeguy, N. *et al.* A new handbook for standardised measurement of
788 plant functional traits worldwide. *Australian Journal of Botany* **64**, 715-716 (2013).

789 86 Fang, J. *et al.* Methods and protocols for plant community inventory. *Biodiversity*
790 *Science* **17**, 533-548 (2009).

791 87 Sack, L. *et al.* A unique web resource for physiology, ecology and the environmental
792 sciences: PrometheusWiki. *Functional Plant Biology* **37**, 687-693 (2010).

793 88 Bjorkman, A. D. & Myers-Smith, I. H. Tundra Trait Team: A database of plant traits
794 spanning the tundra biome. *Global Ecology and Biogeography* (2018).

795 89 Moretti, M. *et al.* Handbook of protocols for standardized measurement of terrestrial
796 invertebrate functional traits. *Functional Ecology* **31**, 558-567 (2017).

797 90 Ferris, H. NEMAPLEX: The Nematode-Plant Expert Information System. *Online at:*
798 *http://plpnemweb.ucdavis.edu/nemaplex/Nemaplex.htm* [Accessed 25 April 2005]
799 (2001).

800 91 Tennessen, J. M., Barry, W. E., Cox, J. & Thummel, C. S. Methods for studying
801 metabolism in *Drosophila*. *Methods* **68**, 105-115 (2014).

802 92 Palomares, M. & Pauly, D. SeaLifeBase. *World Wide Web Electronic Publication*, <
803 *www.sealifebase.org, version*>(12/2010) (2010).

804 93 Le Bagousse-Pinguet, Y. *et al.* Traits of neighbouring plants and space limitation
805 determine intraspecific trait variability in semi-arid shrublands. *Journal of Ecology*
806 **103**, 1647-1657 (2015).

807 94 Cornelissen, J. *et al.* A handbook of protocols for standardised and easy measurement
808 of plant functional traits worldwide. *Australian Journal of Botany* **51**, 335-380 (2003).
809 95 http://prometheuswiki.org/tiki-custom_home.php.

810 96 Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167
811 (2016).

812 97 Andersen, K. H. *et al.* Characteristic sizes of life in the oceans, from bacteria to
813 whales. *Annual Review of Marine Science* **8**, 217-241 (2016).

814 98 Neuheimer, A. B. *et al.* Adult and offspring size in the ocean over 17 orders of
815 magnitude follows two life history strategies. *Ecology* **96**, 3303-3311 (2015).

816 99 Ernest, S. M. *et al.* Thermodynamic and metabolic effects on the scaling of
817 production and population energy use. *Ecology Letters* **6**, 990-995 (2003).

818 100 Weiss, K. C. & Ray, C. A. Unifying functional trait approaches to understand the
819 assemblage of ecological communities: synthesizing taxonomic divides. *Ecography*
820 (2019).

821 101 Ball, I. R., Possingham, H. P. & Watts, M. Marxan and relatives: software for spatial
822 conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods*
823 *and computational tools*, 185-195 (2009).

824 102 Pollock, L. J., Thuiller, W. & Jetz, W. Large conservation gains possible for global
825 biodiversity facets. *Nature* **546**, 141 (2017).

826 103 Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405**, 243
827 (2000).

828 104 Gross, N. et al. Functional trait diversity maximizes ecosystem multifunctionality.
829 *Nature Ecology & Evolution* **1**, 0132 (2017).

830 105 Loreau, M. Does functional redundancy exist? *Oikos* **104**, 606-611 (2004).

831 106 van Bodegom, P. M., Douma, J. C. & Verheijen, L. M. A fully traits-based approach
832 to modeling global vegetation distribution. *Proceedings of the National Academy of*
833 *Sciences* **111**, 13733-13738 (2014).

834 107 Sakschewski, B. et al. Leaf and stem economics spectra drive diversity of functional
835 plant traits in a dynamic global vegetation model. *Global Change Biology* **21**, 2711-
836 2725 (2015).

837 108 Butler, E. E. et al. Mapping local and global variability in plant trait distributions.
838 *Proceedings of the National Academy of Sciences* **114**, E10937-E10946 (2017).

839 109 Kearney, M. & Porter, W. Mechanistic niche modelling: combining physiological and
840 spatial data to predict species' ranges. *Ecology Letters* **12**, 334-350 (2009).

841 110 Fordham, D. A. et al. How complex should models be? Comparing correlative and
842 mechanistic range dynamics models. *Global Change Biology* **24**, 1357-1370 (2018).

843 111 Enriquez-Urzelai, U., Kearney, M. R., Niecieza, A. G. & Tingley, R. Integrating
844 mechanistic and correlative niche models to unravel range-limiting processes in a
845 temperate amphibian. *Global Change Biology* (2019).

846 112 Benito Garzón, M., Robson, T. M. & Hampe, A. Δ Trait SDMs: species distribution
847 models that account for local adaptation and phenotypic plasticity. *New Phytologist*
848 (2019).

849 113 Berzaghi, F. et al. Assessing the role of megafauna in tropical forest ecosystems and
850 biogeochemical cycles—the potential of vegetation models. *Ecography* **41**, 1934-1954
851 (2018).

852 114 Galetti, M. & Dirzo, R. Ecological and evolutionary consequences of living in a
853 defaunated world. *Biological Conservation* **163**, 1-6 (2013).

854 115 Huang, Y. et al. Orchimic (v1. 0), a microbe-mediated model for soil organic matter
855 decomposition. *Geoscientific Model Development* **11**, 2111-2138 (2018).

856 116 McGuire, K. L. & Treseder, K. K. Microbial communities and their relevance for
857 ecosystem models: decomposition as a case study. *Soil Biology and Biochemistry* **42**,
858 529-535 (2010).

859 117 Todd-Brown, K. E., Hopkins, F. M., Kivlin, S. N., Talbot, J. M. & Allison, S. D. A
860 framework for representing microbial decomposition in coupled climate models.
861 *Biogeochemistry* **109**, 19-33 (2012).

862 118 Hardisty, A. R. et al. The Bari Manifesto: An interoperability framework for essential
863 biodiversity variables. *Ecological Informatics* **49**, 22-31 (2019).

864

865

866

867

868

869 **FIGURE CAPTIONS**

870 **Figure 1.** Mammal, bird, and plant phylogenies coloured according to the number of
871 traits for which we have data for each species and lineage. The plant phylogeny is
872 sparsely populated for traits but contains more taxa ($n = 10,596$) than the mammal and
873 bird phylogenies ($n = 5,747$ and $9,993$, respectively). Trait data were downloaded from
874 the following references^{25,34,40}. We counted the number of traits present across these
875 datasets for each species and mapped those onto phylogenies using posteriors^{37,41} and a
876 random subset of plant species within a single phylogeny⁴². Terminal branches
877 (representing species) and ancestral lineages (using ancestral state reconstruction⁴³) were
878 coloured according to the number of reconstructed traits. Note that this is an exploratory
879 analysis conducted purely to show variation in the availability of trait data across
880 taxonomic groups.

881

882 **Figure 2.** Architectures of three alternative networks in which research groups (nodes)
883 interact in collecting and organising trait data. Black nodes are individuals, groups, or
884 institutions conducting projects. Light green nodes are those harmonising data and
885 developing protocols, where node size is proportional to available resources. Dark green
886 nodes are synthesis nodes that collect standardised trait data and knowledge. (A) Groups
887 are disconnected and decentralised, risking duplication of effort (often the status quo);
888 (B) Groups are linked to a centralised repository, potentially limiting innovation; (C) The
889 Open Traits Network, represented by orange lines. Nodes are linked within biological
890 domains (e.g. plants or marine) and include expertise from diverse disciplines (e.g.
891 systematics, palaeobiology, ecology, and biomechanics) allowing for more efficient and
892 specialised decisions about trait collection. Data synthesis across domains or disciplines

893 is facilitated by joining nodes based on common workflows, theoretical frameworks, and
894 data-sharing protocols that adhere to the principles of the Open Traits Network.

895

896 **Figure BOX1.** The six core principles of Open Science and their potential benefits to
897 trait science. Three Open Science Principles are particularly relevant to the Open Traits
898 Network and trait-based research more broadly: Open Data, Open Source, and Open
899 Methods.

900