This is the Accepted Manuscript version of the following article:


which has been published in final form at: https://doi.org/10.1038/s41559-020-1109-6

Publisher: Springer

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


1Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia, 2Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia, 3Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA, 4Department of Zoology, Oxford University, Oxford, UK, 5Centre for Biodiversity and Conservation Science, University of Queensland, Queensland, Australia, 6Evolutionary Demography Laboratory, Max Plank Institute for Demographic Research, Rostock, Germany, 7Department of Biological Sciences, University of Bergen, Bergen, Norway, 8Bjerknes Centre for Climate Research, University of Bergen, Norway, 9Ecology Center and Department of Biology, Utah State University, Utah, USA, 10Private residence, 11Max Planck Institute for Biogeochemistry, Jena, Germany, 12German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, 13400 Perkins Street Apt 104, Oakland, California,
United States, 14Hawai‘i Institute of Marine Biology, University of Hawai‘i at Manoa, Hawai‘i, USA, 15Department of Bioinformatics, Biocenter, University of Wuerzburg, Wuerzburg, Germany, 16Center for Computational and Theoretical Biology, Biocenter, University of Wuerzburg, Wuerzburg, Germany, 17Comprehensive Heart Failure Center, University Hospital Wuerzburg, Germany, 18Institute of Plant Sciences, University of Bern, Bern, Switzerland, 19School of Geography, University of Tasmania, Hobart, Australia, 20Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia, 21Bio5 Institute, University of Arizona, Tucson, Arizona, USA, 22School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Victoria, Australia, 23Department of Biology, University of Miami, Coral Gables, Florida, USA, 24Fairchild Tropical Botanic Garden, Coral Gables, Florida, USA, 25Research School of Biology, Australian National University, Canberra, Australia, 26Department of BioSciences, Program in Ecology and Evolutionary Biology, Rice University, Houston, Texas, USA, 27Species360 Conservation Science Alliance, Bloomington, Minnesota, USA, 28Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark, 29Department of Biology, University of Southern Denmark, Odense, Denmark, 30School of Biological Sciences, Monash University, Victoria, Australia, 31Department of Biology, University of Miami, Florida, USA, 32Department of Ecology, Environment and Evolution and Centre for Future Landscapes, La Trobe University, Victoria, 33National Museum of Natural History, Smithsonian Institution, Washington DC, USA, 34International Center for Tropical Botany, Department of Biological Sciences, Florida International University, Florida, USA, 35Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, USA, 36Leipzig University, Institute of Biology, Leipzig, Germany, 37School of BioSciences, The University of Melbourne,
Victoria, Australia, 38Department of Biology, University of South Dakota, South Dakota, USA, 39Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, 40Center for Tree Science, The Morton Arboretum, Illinois, USA, 41Department of Botany and Biodiversity Research Centre, University of British Columbia, British Columbia, Canada, 42Department of Organismic and Evolutionary Biology and Harvard University Herbaria, Harvard University, Cambridge, Massachusetts, USA, 43Fairchild Tropical Botanic Garden, Florida, USA, 44School of Biological Sciences, School of Earth & Atmospheric Sciences, Georgia Institute of Technology, Georgia, USA, 45School of Life Sciences, Arizona State University, Tempe, Arizona, USA, 46National Herbarium of New South Wales, Royal Botanic Gardens and Domain Trust, Sydney, NSW, Australia, 47Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France, 48 School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia, 49TERN / School of Biological Sciences, Faculty of Science, The University of Adelaide, Adelaide, Australia, 50Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, California, USA, 51Department of Life Sciences, Imperial College London, Silwood Park, UK, 52CEFE, CNRS, Univ Montpellier, Université Paul Valéry Montpellier, Montpellier, France, 53CyVerse, University of Arizona, Tucson, AZ, USA, 54Santa Fe Institute, Santa Fe, New Mexico, USA

**Corresponding author**

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

Email: rachael.gallagher@mq.edu.au

Phone: +61 2 9850 8404
SUMMARY

Synthesising trait observations and knowledge across the Tree of Life remains a grand challenge for biodiversity science. Species traits are now widely used in ecological and evolutionary science, and new data and methods have proliferated rapidly. Accessing and integrating disparate data sources, however, remains a considerable challenge, slowing progress toward a global synthesis that integrates trait data across different organisms. Here, we outline how the adoption of key Open Science Principles – open data, open source, and open methods – are transforming trait science, increasing transparency, democratising access, and accelerating global synthesis. To enhance widespread adoption of these principles, we introduce the Open Traits Network (OTN) – a global, decentralised community welcoming all researchers and institutions pursuing the collaborative goal of standardising and integrating trait data across all organisms. We demonstrate how adherence to Open Science principles is key to the OTN community and outline five activities which can accelerate the synthesis of trait data across the Tree of Life, and thereby facilitate rapid advances in our ability to address pressing scientific questions and environmental issues. Lessons learned along the path to a global synthesis of trait data across all organisms will provide a framework for addressing similarly complex, context-dependent informatics challenges.
INTRODUCTION

Traits, broadly speaking, are measurable attributes or characteristics of organisms. Traits related to function – such as leaf size, body mass, tooth size or growth form – are often used to understand how organisms interact with their environment and with other species via key vital rates such as survival, development, and reproduction\textsuperscript{1-5}. Trait-based approaches have long been used in systematics and macroevolution to delineate taxa and reconstruct ancestral morphology and function\textsuperscript{6-8} and to link candidate genes to phenotypes\textsuperscript{9-11}. The broad appeal of the trait concept is its ability to facilitate quantitative comparisons of biological form and function. Traits also allow us to mechanistically link organismal responses to abiotic and biotic factors using measurements that are, in principle, relatively easy to capture across large numbers of individuals. For example, appropriately chosen and defined traits can help identify different lineages that share similar life-history strategies for a given environmental regime\textsuperscript{12,13}. Thus, documenting and understanding the diversity and composition of traits in ecosystems directly contributes to our understanding of organismal and ecosystem processes, functionality, productivity, and resilience in the face of environmental change\textsuperscript{14-19}.

In recent decades there has been an acceleration in the collection, compilation, and availability of trait data for a variety of organisms. Substantial trait databases now exist for plants\textsuperscript{20-23}, reptiles\textsuperscript{24,25}, invertebrates\textsuperscript{23,26-29}, fish\textsuperscript{30,31}, corals\textsuperscript{32}, birds\textsuperscript{23,33,34}, amphibians\textsuperscript{35}, mammals\textsuperscript{23,34,36-38}, and fungi\textsuperscript{23,39} and parallel efforts are no doubt underway for other taxa. Though considerable effort has been made to quantify traits for some key groups (e.g., Fig. 1), substantial work remains. In order to advance knowledge by developing and testing theory in biodiversity science much greater effort is needed to fill in trait data across the Tree of Life via combining and integrating, or synthesising, data and trait collection efforts.
There are two central questions that together define the scope and nature of the challenge to trait science. First (Box 1), how can we most effectively advance the synthesis of trait data within and across disciplines to address questions of global significance? Second, how can we best deliver accurate and understandable biodiversity knowledge to non-academic audiences who should have equitable access to quality data on the traits of Earths’ species?

**Figure 1.** Mammal, bird, and plant phylogenies coloured according to the number of individual traits for which we have data for each species and lineage. The plant phylogeny is sparsely populated for traits but has many more taxa than either the mammal or bird phylogenies. Trait data were downloaded from\textsuperscript{25,34,40}, the number of traits present across these datasets for each species counted, and then mapped onto a single phylogenies from the posteriors of\textsuperscript{37,41}, and a random subset of plant species within a single phylogeny from\textsuperscript{42}. Terminal branches (representing species) and ancestral lineages (using ancestral state reconstruction\textsuperscript{43}) were then coloured according to the number of (reconstructed) traits. Note that this is an exploratory analysis, conducted purely to show the variation across taxonomic groups in the amount of available trait data.

**Barriers to a global trait-based science**

Despite the recognised importance of traits, several common research practices currently limit our capacity for meaningful synthesis across the Tree of Life. These practices include failure to publish usable datasets alongside new findings\textsuperscript{44}, missing or inadequate metadata\textsuperscript{45} which impedes the findability and interpretation of data, minimal descriptions of methods used to collate, clean and analyse trait datasets in published
works\textsuperscript{46}, and inadequate coordination between researchers and institutions with common goals, such as filling strategic spatial or taxonomic gaps in trait knowledge\textsuperscript{47,48}. Our limited ability to access and redistribute trait data also directly contributes to the widespread reproducibility crisis within science\textsuperscript{49}. Any study relying on data that is not able to be easily re-used introduces barriers to verifying the claims within the paper and thereby the reproducibility of the science\textsuperscript{50}. Such limitations have been common within trait-based science.

Access to data is not the only impediment to a global synthesis of trait knowledge. Barriers to synthesis also exist because researchers and institutions are apprehensive that the time and resources they spend to create new observations or share legacy data (\textit{e.g.}, observations from field guides, specimens, publications without data supplements) will not be appropriately recognised. Identifying who should receive credit for contributing trait observations – whether via co-authorship, or other formal recognition – can be complicated, particularly where data involve a chain of expertise (\textit{e.g.}, when trait data are extracted from taxonomic treatments which involve specimen collectors, digitisers, taxonomists, and curators). Also, research funding bodies are often reluctant to support data management as a significant cost, limiting recognition of the sizeable current effort expended on creating bespoke solutions to curating and harmonising trait data from different sources\textsuperscript{50}.

Major opportunities exist for expanding the spatial and taxonomic coverage of trait observations, particularly by strengthening interdisciplinary connections across single organismic groups. For example, despite certain traits of plants – such as growth form, height, and leaf size – being carefully catalogued in taxonomic species descriptions worldwide\textsuperscript{51}, these data have only recently been systemically exchanged with large-scale trait databases, such as TRY\textsuperscript{21} or BIEN\textsuperscript{32}. Although several difficult informatics
challenges in biodiversity science have now been overcome (e.g., synthesising global species occurrence information\textsuperscript{53}; sharing genetic data on individuals\textsuperscript{54}) trait science lacks a vision for achieving global integration across all organisms. We argue that this is not simply a failure of the traits community to learn from existing successful networks, but instead that cataloguing traits is a more complex task which is highly context-dependent and therefore needs a more nuanced network model than that offered by a centralised repository.

Here, we propose that widespread adoption of key Open Science Principles (Box 2) could be transformative for trait science in achieving a global synthesis. These principles would lay a strong foundation for transparency, reproducibility and recognition and encourage a culture of data sharing and collaboration beyond established networks. Openness reinforces the scientific process by allowing increased scrutiny of methods and results, resulting in the deeper exploration of findings and their significance\textsuperscript{46,55-58}. The scope of trait science would rapidly increase if researchers and institutions: 1) made datasets available under clear licensing arrangements in machine-accessible formats; 2) created and adopted standardised protocols, handbooks or metadata formats for data collection, documentation, and management (e.g., see\textsuperscript{54-56}); and 3) create human-centred networks which can reduce the complexity of integrating existing legacy data from disparate sources (e.g., specimens, published literature, citizen science initiatives\textsuperscript{57,58}, and large scale digitisation efforts (e.g., Biodiversity Heritage Library)). These different sources have systematic differences in error rates, validation, context, reproducibility, and objectivity relative to field-collected trait observations. Similarly, without a model of recognition that embraces transparency and fairness, much trait data will remain effectively hidden from science.
Introducing the Open Traits Network (OTN): a collaborative initiative for accelerating trait data synthesis

The OTN is a global, decentralised community welcoming all researchers and institutions pursuing the collaborative goal of standardising and integrating trait data across all organisms. We promote five main objectives built upon Open Science ideals which could be transformative for trait science:

1. Openly sharing data, methods, protocols, code, and workflows;
2. Appropriately citing original data collectors, and providing scholarly credit;
3. Providing appropriate metadata together with trait observations;
4. Collecting of trait data following reproducible, standardised methods and protocols (when available) or committing to their development;
5. Providing training resources in trait collection and database construction using Open Science principles.

We envision a future for trait research where protocols for data exchange and reuse are transparent, research findings are reproducible, and all trait data – either newly collected or from legacy sources – are openly available to the research community and broader public. While several network models exist in trait research (Fig. 2), the Open Traits Network (OTN) adopts a decentralised but connected structure with an emphasis on bringing people together via their data and expertise.
Figure 2. Architectures of three alternative networks in which research groups (nodes) interact in collecting and organising trait data. Black nodes are individuals, groups, or institutions pursuing trait projects. Light green nodes are those harmonising data and developing protocols, where node size is proportional to available resources. Dark green nodes are synthesis nodes, which collect the flow of standardised trait data and knowledge. (A) Groups are disconnected and decentralised, risking duplication of effort (often the status quo); (B) Groups are inflexibly linked to a centralised repository, potentially limiting innovation; (C) The Open Traits Network, which is represented by orange connecting lines. Nodes are linked together within biological domains (e.g., plants, marine) and include expertise from many disciplines (e.g., systematics, palaeobiology, ecology, biomechanics) allowing for more efficient and specialised decisions about trait collection. However, data synthesis across domains or disciplines is then facilitated by joining nodes based on common workflows, theoretical frameworks, and data sharing protocols which adhere to the guiding principles of the Open Traits Network. Central synthesising nodes may be, for example, the Smithsonian Encyclopedia of Life https://eol.org/.

In many instances, groups building smaller-scale databases do so in isolation, using their own tools and workflows tailored to their research question: they are decentralised and disconnected (Fig. 2A). Decentralisation has certain advantages, including retaining the power of the relevant researchers to determine which traits are most useful in their study system and how they should be compiled. However, there is little formal support or interaction across this style of network, so researchers often collect redundant data, use different data structures, and may develop similar tools for data collection, cleaning, and integration in isolation, which can lead to duplication of effort. There are many small, isolated, and heterogeneous data sources of this sort, increasing the disconnect between pools of trait data. For some organisms, centralised hubs already exist which aggregate and
standarise trait data from across disparate sources (e.g.,\textsuperscript{21,32,60-64}) (Fig. 2B). These trait repositories have increasingly become the main access point for trait data on well-studied taxa such as plants and corals, but they remain mostly isolated from each other, limiting the sharing of expertise and information. As these repositories continue to grow, difficulties with integration and synthesis of the data they contain will also increase due to the momentum of entrenched workflows and exchange protocols which may not be interoperable.

Some highly successful large-scale initiatives have followed the centralised and connected network model (e.g., the Global Biodiversity Information Facility\textsuperscript{53}, GenBank\textsuperscript{54}). These platforms mandate strict data exchange protocols to facilitate synthesis using standardized, controlled vocabularies – such as Darwin Core\textsuperscript{65} and Humboldt Core\textsuperscript{66}. These protocols have been central to the explosive growth of biodiversity data because they facilitate free exchange of information using common data formats\textsuperscript{65-67}. Further, ontologies which provide the unified terms and concepts necessary to represent traits have been already been developed (e.g., Uberon – the multispecies anatomy ontology for animals\textsuperscript{68}, TOP – the Thesaurus of Plant characteristics\textsuperscript{69}). These provide integration and with other data types, such as genetic and environmental data, and their corresponding ontologies (e.g., Gene Ontology\textsuperscript{70}, Environmental Ontology\textsuperscript{71}).

Despite these previous successes, we argue that a centralised and connected network structure will not rapidly facilitate trait data synthesis. Trait observations are highly-nuanced and hierarchical. Further, describing multiple aspects of a phenotype for any organism via traits is not amenable to an overly simplified set of exchange fields which apply uniformly across the Tree of Life. While the centralised and connected model (Fig. 2B) does have benefits, it lacks the necessary flexibility to dynamically connect trait data where ontologies and exchange formats do not exist. The result is that
established trait networks will remain isolated and disconnected.

The decentralised but connected model (represented by the orange connections in Fig. 2C) adopted by the OTN maintains the key advantages of a decentralised network (e.g., taxon/discipline specific decision-making) while enhancing the level of connectivity among groups, allowing for easier aggregation and sharing of expertise, tools, and data. These network characteristics may also buffer against node loss (e.g., due to lack of funding). Decentralised and connected networks are known to be characterised by socially-mediated improvements in learning\textsuperscript{72} as they capitalise on the aggregated judgement of many, rather than the singular opinion of an expert\textsuperscript{73}. The OTN model capitalises on existing robust connections within disciplines, but links between domains of the Tree of Life to more efficiently disseminate granular knowledge about traits. By maintaining recognition of the importance of specialist taxon groups (Fig. 2C; light green nodes) for devising relevant protocols and standards and accommodating their needs into the development of cross-domain tools for synthesis (Fig. 2C; dark green nodes), the OTN model may be particularly beneficial for ‘low-profile’ taxa (c.f. plants, mammals, corals) which may not be accommodated by a centralized effort to synthesise trait data. The OTNs open, decentralised network structure is well-placed to leverage the power of existing centralised hubs, allowing individual researchers to retain their own agency and independence while also joining them together to minimise redundancy of effort.

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

The OTN actively seeks to broaden its membership by lowering barriers to inclusion and advocating for approaches to trait science which benefit data custodians. For instance, new members can easily join the OTN via our website (www.opentracts.org) through two mechanisms: (i) adding a member profile (name, location, expertise, collaboration...
statement), and/or (ii) registering their open source (or embargoed) trait datasets in the OTN Trait Dataset Registry (see Activity 1 below). The registry contains metadata for trait datasets and links users to the open dataset. New entries to the Registry will be reviewed by OTN members before being added. This step is intended to facilitate interactions between new and established OTN members and encourage deeper collaborative engagement. Once registered, members will receive intermittent updates about the OTN, including newly registered trait datasets, notifications about upcoming chances for face-to-face meetings, and funding opportunities. Members are asked to actively support the OTN by identifying where their projects can make use of the network through, for instance, promoting new workshops, resources, or funding calls that adhere to Open Science principles.

OTN membership spans scientists (and institutions) with high-level expertise in trait data science and synthesis activities, through to those with strong motivations to work with traits but little formal expertise. The OTN has already conducted an international workshop facilitated by an open call for participants, with more planned. Following this initial workshop and communication process, we now share ideas and an awareness for common problems and goals and act upon them within subgroups of members or in bilateral exchange. An advantage of the OTN is its flexibility – being a decentralised network, we do not need to rely on funding and dedicated personnel to complete tasks, though larger goals will benefit from financial support. Instead, we provide a means to communicate the joint aims and gaps between network nodes (Fig. 2), and to arrange workshops and complete activities.

We recognise that altruism is unlikely to offer strong enough motivation to ensure widespread participation in the OTN. The sharing of trait datasets is not merely a technical problem to be solved; it relies on custodians having the skills, incentives and
motivation to want to contribute. Motivation to share data will need to be particularly strong for those whose data are newly collected. The key incentives for individuals to join the OTN include increasing the findability of their data and expertise and having access to a ready-made network of trait scientists and institutions engaging in trait-based initiatives. Data are a powerful asset for researchers, and release under permissive open license schemes accompanied by well-defined metadata offers great potential for new collaborations and increased visibility of the researcher's effort. A persistent concern is that scientists will lose control of their hard-won trait data by making it open, though this underestimates the potential for new collaborations to arise and may unnecessarily increase distrust among the scientific community. For many, particularly early-career researchers and those in developing nations, access to networks of like-minded scientists can provide valuable exposure and connection. By emphasising the importance of community engagement and support, the OTN seeks to make trait data sharing and synthesis an opportunity for all involved, rather than simply a technical challenge to be solved.

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

Here, we highlight five OTN activities, several of which are already operational, that demonstrate the power of a decentralised and connected network to increase knowledge transfer in trait science. Trait scientists have already made significant achievements in key areas, such as the synthesis of large numbers of observations within particular taxonomic groups, the development of theory and frameworks to utilise this data in the testing of ideas, and large-scale empirical studies. Lacking, however, are some basic foundations which will allow us to quantify how and why traits vary across all organisms. Primary amongst these missing pieces are our inability to find trait
datasets and, once found, access them freely; and a lack of tools for standardising the collection and synthesis of data, such as common metadata standards and reproducible workflows.

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

Several trait data gaps impede synthetic analyses across taxa, space, and ontogeny. Specifically, the heterogeneous ways in which trait data have been collected to date have resulted in a patchy and unrepresentative trait landscape across trait types, taxa, regions, and times of the year. The OTN is bridging these gaps by now maintaining a Trait Dataset Registry that can be accessed at [www.opentraits.org/registry](http://www.opentraits.org/registry) and is designed to help: (i) increase findability of open trait datasets, including those which are newly-collected and under embargo (ii) identify taxonomic and geographic gaps in trait knowledge (iii) prioritise trait collection, and (iv) allow researchers to avoid inadvertent duplication of efforts when collating trait data.

The OTN Registry contains information on existing open (or embargoed) datasets so that gaps can be easily identified, and ultimately filled, through collective effort. Core information for the registry includes Digital Object Identifier (DOI), taxonomic coverage, curator, and format. Importantly, the OTN Registry also provides the opportunity for contributors to identify if and where code to process and manipulate their raw data is located (see Activity 2 below). As it develops, the OTN Registry will relate trait concepts to ontologies provided through the Open Biomedical Ontologies Foundry. Thus, the OTN registry maps to several Open Science principles (Box 2; e.g., Open Source, Open Data, Open Access) and is designed, from the ground up, to support resolving the issue of data findability and integration.

The OTN does not place restrictions on what members may consider the traits of
importance to a taxonomic group. Most traits can be measured on individuals and easily fit into existing broad definitions, though this may not be appropriate for organisms where individual or taxonomic boundaries are unclear (e.g., microbes\textsuperscript{78}, fungi\textsuperscript{79}).

Equally, it can be argued that traits can also encompass emergent properties of populations (e.g., abundance, geographic range size) or represent interactions between species (e.g., diet type, such as carnivore). More important than imposing strict definitions around traits, is engaging the community in discussion about the utility of available data for answering novel ecological and evolutionary questions.

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

The OTN leverages collaborative software development via platforms like GitHub (https://github.com/) to create modular open source software for access and harmonisation and re-use of trait data, with seamless piping of data from one software tool to the next. OTN contributors have already developed several open source tools. For instance, the `traitdataform` package assists R users to format their data and harmonise units (http://ecologicaltraitdata.github.io/traitdataform); the code for the Coral Traits database\textsuperscript{32} (https://github.com/jmadin/trait) could be easily modified to guide the creation of databases on other organisms; and the FENNEC project provides a tool for accessing and viewing community trait data as a self-hosted website service\textsuperscript{80} (https://github.com/molbiodiv/fennec). The OTN can act as a connector for interactions between developers and the broader community seeking to synthesise trait data, facilitating the training of scientists in all aspects of reproducible data management.

**Activity 3: Advocating for a free flow of data and appropriate recognition of efforts**

A goal of the OTN is to increase the use of open datasets and to make sure credit is given
to researchers who collect or synthesise primary data on species traits. Without effective reward or motivation for collecting new trait observations or liberating legacy data, a broad trait synthesis across the Tree of Life will remain unattainable. Currently, however, motivation for collecting and sharing new primary data is not strong and direct funding for trait data management is scarce. Open access to readily findable data is central to improving our capacity to aggregate data across the Tree of Life.

The OTN can strengthen the attribution of credit to data providers and promote new data collection via two paths. Firstly, by encouraging citation back to primary source via a permissive license model that secures authorship attribution, such as Creative Commons Attribution 4.0 Int (CC BY 4.0) and the use of DOI’s and ORCID identifiers to attribute the use of data records. Open access datasets which have been assigned a DOI can be readily tracked to understand patterns of re-use and to assess the impact of the author’s decision to share. DOI’s create a direct path to the source of trait observations and are routinely issued when data are shared on repositories such as Zenodo, figshare, and Dryad.

There is an important distinction between sharing data within a network and making data publicly available under an open license. Clear license arrangements increase visibility and promote fair attribution/citation (e.g., using creative commons licenses such as CC-BY or CC0). CC-BY requires attribution (i.e., citation) to the original creator, whereas CC0 doesn’t legally require users of the data to cite the source, though does not affect the ethical norms for attribution in scientific and research communities (https://creativecommons.org/share-your-work/public-domain/cc0/). However, identifying who should be credited for prior work on legacy data is complicated by the involvement of many individuals (e.g., when traits are collated from taxonomic treatments which involve many experts such as specimen collectors, digitisers,
taxonomists, and curators). This issue could be solved, in part, by inviting entire organisations to be named as contributors or co-authors on outputs using their data, or – looking forward – implementing new ways of documenting who should be credited for making individual specimens or taxonomic groups usable in trait-science.

Secondly, incentive to collect new trait data can be linked to the Open Science practice of pre-registration. In pre-registration, authors publicly archive a proposal for research activities (e.g., via the Centre for Open Science; https://cos.io/prereg/) which, if approved, may then receive in-principle acceptance from participating journals. As of March 2019, 168 journals are already willing to give in-principle acceptance of research activities following pre-review of the study design and prior to conducting field or experimental work. Approximately ten of these participating journals regularly feature papers on trait-based science, including *BMC Ecology* and *Ecology and Evolution*. We envision a situation where the OTN Trait Registry (*Activity 1*) could be used to identify important spatial or taxonomic gaps in trait data which can be coupled to pre-registered hypotheses. Together, pre-registration and in-principle acceptance of findings have the potential to incentivise the collection of new data, circumventing a growing reliance on available data with known gaps.

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

Trait science urgently requires its own ‘core’ terminology or data standard that is flexible enough to capture the inherent complexity of trait data. Building upon efforts to standardise occurrence data (*i.e.*, Darwin Core65) and biological inventories (*i.e.*, Humboldt Core66)47 which have been instrumental for mobilising biodiversity data, the OTN envisions a trait core which is flexible enough to offer a set of cross-domain metadata standards and controlled vocabularies which are (ideally) connected to trait
ontologies via unambiguous identifiers. Such a standard terminology would be implemented across trait-data publications, unifying data in decentralised repositories as well as centralised data portals.

A trait core would allow trait data to be: (i) accurately interpreted within the context of their collection (i.e., including information on associated data on factors like the environmental conditions at collection sites, taxa covered (and their trophic level), data custodians, or collection methods); and (ii) known by compatible terms so that observations of similar phenomena across domains can be grouped and compared (i.e., what is meant by ‘generation time’ or ‘establishment’ in different taxonomic groups). Existing initiatives may provide logical cornerstones for referencing terms and concepts, including Ecological Metadata Language (EML). Several current initiatives implement EML (e.g., The Knowledge Network for Biocomplexity; Darwin Core; Humboldt Core) and the use of referencing terms from anatomy or phenotype ontologies (e.g., the Plant Ontology; the Vertebrate Trait Ontology) to relate traits semantically to publicly-defined terms, allowing this annotated data to be processed computationally.

Progress towards a trait core is already being made through the development of a prototype Ecological Trait Standard (ETS) (Box 3). However, the further development and adoption of a trait core requires consultation and coordination in a broader community of researchers – a goal which the OTN is ideally-placed to advance. The OTN can mobilise expertise for a series of cross-domain workshops and advocate for funding which allows not only meetings of experts, but also the creation of appropriate cyber-infrastructure for synthesis nodes (Fig. 2C; dark green nodes). Links to emerging initiatives for biodiversity data standardisation – such as the Species Index of Knowledge – will also be vital for success, as will ratification of the core through the Biodiversity Information Standards (TDWG, www.tdwg.org).
Activity 5: Facilitating consistent approaches to measuring traits within major groups

The OTN will share new developments towards protocols and handbooks for major clades that standardise approaches to capture trait observations. Protocols are necessary because downstream activities such as developing metadata standards (Activity 4) will be impossible to create if trait measurement protocols do not exist. Some trait-research communities have already adopted standardised lists of terms\textsuperscript{63,69} and trait data collection protocols (\textit{e.g.}, plants\textsuperscript{20,85-88}, invertebrates\textsuperscript{29,89-91}, mammals\textsuperscript{36}, aquatic life\textsuperscript{30,32,92}), though these may not always fit to the requirements of some studies, particularly those where variability in a trait is targeted as opposed to capturing the average trait of species\textsuperscript{93}. New protocols and handbooks may not emerge rapidly and should have the flexibility to be open to innovation through a commitment to creating versions and updates as techniques evolve. For instance, two versions of the plant trait measurement handbook have been published\textsuperscript{85,94} and several online resources exist which can be dynamically updated through time (\textit{e.g.},\textsuperscript{95}).

Standardising approaches to trait measurement \textit{a priori} across communities of researchers will reduce ambiguity when aggregating data and improve the quality of resulting datasets. Also, integrating trait standardization and databasing into taxonomic workflows represents both a considerable challenge and opportunity\textsuperscript{7}, but also holds the promise of bridging a long disconnect between structural and functional traits. The presence of a range of biodiversity collections personnel in the OTN, and an open invitation for more to join, has the potential to catalyse adoption of trait-based thinking into taxonomic practices.

Concluding remarks
This is the opportune time to push towards a new approach in the sharing and synthesis of trait data across all organisms. Trait science has clear potential to rapidly increase its taxonomic, phylogenetic, and spatial scope by leveraging the expansion of data science tools, embracing Open Science principles, and creating stronger connections between researchers, institutions, publishers, and funding bodies.

We hope that trait enthusiasts, regardless of field and research stage, will engage with the OTN via our website (www.opentraits.org) and help to build new connections between disciplines, institutions, and taxonomic domains. By adding metadata profiles for datasets to the OTN Trait Dataset Registry, trait collection efforts become more findable, as do the researchers who have spent hours compiling them. We envision that by connecting people with common goals we can work more collectively towards a synthesis of global trait data which preserves the nuance of taxon specific expertise, while also facilitating collaboration across domains. We also urge scientists and institutions keen to commit to Open Science principles in their trait research to make use of existing resources, including those offered by the Centre for Open Science (https://cos.io/), the Open Science Training Handbook (https://open-science-training-handbook.gitbook.io/book/), the Open Science Training Initiative (http://www.openscienctraining.com/index.php), and FOSTER (https://www.fosteropenscience.eu/toolkit).

Ultimately, to support and expand the activities of the OTN, we need to grow membership and develop communities around synthesis nodes which can push forward key activities and secure funding support, in particular for the development of a trait core. Funding would allow for the necessary international workshops, technical support, and implementation meetings to drive a new era of trait-based synthesis that mirrors the achievement of initiatives like GBIF which now houses over 1 billion occurrence
records.

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

ACKNOWLEDGEMENTS

Ideas presented stem from initial discussions at three separate international meetings – the Australian National Climate Change Adaptation Research Facility Roundtable on Species Traits, the iDigBio ALA Traits workshop, and a preliminary Open Traits workshop held at the Ecological Society of America. RVG is supported by an Australian Research Council DECRA Fellowship (DE170100208). DSF is supported by an Australian Research Council Future Fellowship (FT160100113). RS-G is supported by NERC R/142195-11-1. WDP is supported by NSF ABI-1759965, NSF EF-1802605, and USDA Forest Service agreement 18-CS-11046000-041. AK received financial support for MJA by the German Research Foundation (DFG KE1743/7-1). CMI was supported by the Biological and Environmental Research program in the United States Department of Energy’s Office of Science. CP is supported by the DFG Priority Program 1374. MJ was supported by the German Research Foundation within the framework of the Jena
Experiment (FOR 1451) and by the Swiss National Science Foundation. SPM was supported by the Benson Fund from the Department of Paleobiology, National Museum of Natural History. STM is supported by SERDP project RC18-1346. BJE was supported by NSF Grants DEB0133974, HDR1934790, and EF1065844, a Leverhulme Trust Visiting Professorship Grant, and an Oxford Martin School Fellowship. Peer-review from Nicholas Gross and two anonymous colleagues helped to improve the manuscript.

**BOXES**

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

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

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

Measurements of the traits of organisms can be used to index the position of species along trade-off spectrums which shape their fitness. For instance, traits – such as leaf nitrogen content, leaf mass per area, seed mass, and maximum height – have been used to capture key trade-offs between plant species at global scales⁹⁶. Moving beyond plants, and with access to open data on the traits of a wide breadth of organisms, we will be able to identify major axes of functional specialisation across the Tree of Life. Traits such as adult body mass, offspring mass at independence, mass-specific metabolic rate, and body temperature capture core differences in the ecological strategies within and across diverse groups⁹⁷-⁹⁹ and are therefore promising candidates for comparison across the Tree of Life. Though there is tremendous potential in the comparison of traits across the tree of life.
via the OTN, consideration of scale and biology are needed to identify when these comparisons are meaningful.\textsuperscript{100}

\textbf{Conservation of functional diversity in global protected areas}

Reserve selection procedures\textsuperscript{101} seek to identify representative networks of potential protected areas at the lowest financial cost. Species are typically the biological unit being targeted, regardless of the functional diversity they collectively represent; however, there is an increased focus on ecosystem services and protecting ecosystem function by better representing phylogenetic or functional diversity\textsuperscript{102}. With trait data from across the Tree of Life systematic conservation planning\textsuperscript{103} could: (i) optimize reserve designs to maximise conservation of ecosystem function predicted from species traits or (ii) comprehensively assess the adequacy of current protected areas in representing ecosystem function.

Reserve design typically targets species or populations, not ecosystem functions, and is based on three fundamental principles: comprehensiveness, adequacy, and representativeness\textsuperscript{102}. Using theoretical predictions about the relationship between traits and ecosystem functions we may be able to explicitly predict which traits should be actively preserved when designing and refining protected area networks. In this context, \textit{comprehensiveness} would allow for the inclusion of the full range of ecosystem functions – as captured by traits of species present – recognized at an appropriate scale within and across relevant ecosystem units (e.g. bioregions, biomes); \textit{adequacy} would seek to maintain the viability and integrity of ecosystem function and model how functional redundancy (i.e., different species performing the same functional role\textsuperscript{104}) may scale across landscapes in order to assess the minimum viable reserve size for maintaining a particular ecosystem function; and \textit{representativeness} would seek to capture the diversity of functions and the
gradients across which they occur, including the level of intraspecific variation and plasticity inherent to the trait being examined. Open trait data across all organisms – including information on their abundance – is integral to parameterising model selection procedures for maximising ecosystem function.

**Strengthening predictions of the effects of global change on biodiversity.** Traits are already proving valuable for improving models that predict global change impacts on the biosphere. For instance, access to large-scale data on the traits of plants allows the distribution of traits to be captured in Earth System Models (ESMs) rather than competition among broad plant functional types\(^{105-107}\). Species-based predictions from correlative niche models may similarly benefit from the integration of data on species' physiology or biology to capture mechanistic links between function and environment\(^{108-111}\).

The next generation of Earth system models may also specifically benefit from integration of open trait data for (at least) two types of other organisms: (i) terrestrial vertebrates and (ii) microbes. Terrestrial vertebrates are involved in several ecological roles in ecosystems (e.g. dispersal, disturbance)\(^{112}\) and their loss may result in significant changes to ecosystem function\(^{113}\). However, capturing the influence of terrestrial megafauna on forest structure, function, and biogeochemical cycles will be improved by access to data on key traits, such as size and diet, for parameterising process-based models. Similarly, access to knowledge about the traits of soil-borne microbes – whose activity may increase atmospheric warming via acceleration of processes such as decomposition of soil carbon – may assist our capacity to predict planetary responses to climate change\(^{114-116}\). For instance, molecular-level traits in microbes – such as enzyme activity – allow estimation of decomposition rate but are difficult to measure, amplifying the need to openly share this type of primary data.
Box 2: Using Open Science principles in trait research

Open Science principles outline a movement towards making all aspects of the scientific process transparent and accessible to a wide audience\(^{51,52}\) and are rapidly being adopted across the sciences.

**Figure B1.** The six core principles of Open Science and their potential benefits to trait science. Three Open Science Principles are particularly relevant to the Open Traits Network and trait-based research more broadly: Open Data, Open Source, and Open Methods.

In this context, knowledge is considered open if anyone can freely access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness ([http://opendefinition.org/](http://opendefinition.org/)). Several pronouncements about Open Science principles have already been made, including the Berlin ([https://openaccess.mpg.de/Berlin-Declaration](https://openaccess.mpg.de/Berlin-Declaration)), Bouchout ([http://www.bouchoutdeclaration.org/declaration/](http://www.bouchoutdeclaration.org/declaration/)), and Denton Declarations ([https://openaccess.unt.edu/denton-declaration](https://openaccess.unt.edu/denton-declaration)) on open access to science data. Other initiatives champion some open practices such as the Bari Manifesto on interoperability\(^{53}\) and the FORCE 11 network, which developed the 'Joint
Declaration of Data Citation Principles’ ([https://www.force11.org/datacitationprinciples](https://www.force11.org/datacitationprinciples)) and ‘FAIR’ principles ([https://www.force11.org/group/fairgroup/fairprinciples](https://www.force11.org/group/fairgroup/fairprinciples)). The FAIR principles address several of the major challenges facing trait-based research, namely making data Findable, Accessible, Interoperable, and Reusable.

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

Preparation and re-use of open trait data requires action in the following steps:

1. **Data collection and data handling within the project context.** This usually applies project-specific methodology and stores data in data tables suitable for the target type of analysis, applying own terms for taxa, traits, and column labels. Adopting standard methodologies and terminologies from the start will greatly simplify steps 2 and 3 and facilitate data publication. Note that OTN Activity 5 aims at consensus building for common trait definitions and methodologies for measurement within the major groups of organisms.

2. **Harmonization of taxa and traits into standard terms using ontologies.** Before publication, all taxa should be harmonized into accepted names which are linked to ontologies using uniform resource identifiers (URIs). Those can be provided in metadata or secondary data tables. Ontologies for traits are scarce, but if available should also be referred to via URIs to deliver unambiguous trait definitions. OTN Activity 2 will foster the development of ontologies linking trait data to publicly available resources.

3. **Standardization of table descriptors and metadata using a standard vocabulary.** Data should be published in tables using standard terms for column names, such as provided by the Ecological Trait-data Standard vocabulary (ETS; doi: 10.5281/zenodo.2605377). This ETS implements a minimal trait-data terminology
which can be adapted to include traits from a variety of organisms and makes use of Uniform Resource Identifiers (URI) for taxa and trait definitions, methods and units, thereby following the standards for a semantic web of scientific data. Metadata should point to the applied standard to make interpretation for humans and computational tools feasible. OTN Activity 4 will engage in a cross-domain community-based development for standard vocabularies for trait data.

4. **Publication of data and upload to a public repository.** Open Access file hosting services offer permanent hosting and findability of data by assigning a DOI to the data publication and stating authorship and conditions for re-use under Creative Commons licenses. OTN Activity 3 will support the appropriate recognition for data publications and thereby mitigate the investment into data standardisation for smaller research projects.

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

REFERENCES


Kunstler, G. *et al.* Plant functional traits have globally consistent effects on competition. *Nature* 529, 204 (2016).


30 Froese, R. & Pauly, D. Fishbase. (Fisheries Centre, University of British Columbia, 2010).
39 https://github.com/traitecoevo/fungaltraits
52 http://bien.nceas.ucsb.edu/bien/
https://www.gbif.org/
http://www.obofoundry.org
95 http://prometheuswiki.org/tiki-custom_home.php


