Towards an Ecological Trait-data Standard

Florian D. Schneider[[1]](#footnote-1),[[2]](#footnote-2), Malte Jochum[[3]](#footnote-3), Gaetane Le Provost2,   
Caterina Penone3, Andreas Ostrowski[[4]](#footnote-4), Nadja K. Simons[[5]](#footnote-5)

# Abstract

1. The assessment and analysis of functional traits has been established in all domains of ecological research, promising great advancements for biodiversity and climate research. A wealth of data is available from multiple sources and for many organism groups but is lacking standardisation due to heterogeneous data formats and data types.
2. To harmonize and standardise datasets, we present a general vocabulary and structure for an ecological trait-data standard for storing and sharing functional trait data.
3. We further provide an R-package that assists in transferring own data structures into the data standard and to harmonize and merge trait datasets from heterogeneous sources.
4. With these decentralised tools for traitdata management, we intend to facilitate the exchange and analysis of trait data for ecological research.

# Key-words:

functional ecology, traits, semantic web, ontologies, data standard

# Introduction

Functional traits are phenotypic characteristics that are related to the fitness and performance of an organism (McGill et al. 2006; Violle et al. 2007). The scope of trait-based research spans from the evolutionary basis of individual-level properties to global patterns of biodiversity and ecosystem functioning.

Comparative studies of traits inform researchers about physiological and evolutionary trade-offs, and help identifying economic strategies of organisms beyond taxonomy (Wright et al. 2004; Gross et al. 2009; Menezes et al. 2010; Forister et al. 2015; Siefert et al. 2015). Inferring an unknown functional role or ecological function of individuals from their apparent features (Duarte et al. 2011; Schrodt et al. 2015) is a promising venue to bypass taxonomic impediment, i.e. the fact that a majority of species are yet undescribed and little is known of their interactions with the environment and other organisms. Mutual 'trait-matching' of co-evolved organisms enables the inference of ecological interactions from observed features of live specimens (Kalinkat et al. 2013; Dehling et al. 2014; Le Provost et al. 2017) which may be used to inform ecological networks (Digel et al. 2011; Miranda et al. 2013; Weiner et al. 2014; Schleuning et al. 2015) and parametrise ecological population-dynamic models (Schneider et al. 2012; Kalinkat et al. 2013). These analyses are also a promising venue for a multitrophic community ecology aiming to unravel dependencies and synchrony in trait spaces between interacting functional groups (Soliveres et al. 2016). Since traits are so closely related to ecological mechanisms, they have proven useful for scaling individual level physiology and behaviour to functional diversity of communities (Diaz et al. 1998; Lavorel and Garnier 2002; Villéger et al. 2008; Petchey et al. 2009; Laliberté and Legendre 2010; HilleRisLambers et al. 2012), processes at the ecosystem scale (Fontaine et al. 2005; Jennings and Pocock 2009; Moretti et al. 2013; Birkhofer et al. 2014; Gámez-Virués et al. 2015; Kühsel and Blüthgen 2015; Lajoie and Vellend 2015; Gossner et al. 2016; Gravel et al. 2016; Simons et al. 2016), as well as macro-patterns of biodiversity at the global level (Reich et al. 1997; Reich et al. 1999; Díaz et al. 2013; Díaz et al. 2016). Traits can be used to predict extinction risk (Fritz and Purvis 2010; Pearson et al. 2014) and susceptibility to environmental toxins (Rubach et al. 2011). Most importantly, the trait framework relates key ecosystem services for human well-being back to changes in biodiversity in consequence to global change or local anthropogenic land use (Díaz et al. 2007; Jennings and Pocock 2009; de Bello et al. 2010; Vandewalle et al. 2010; Díaz et al. 2011; Lavorel and Grigulis 2012; Börschig et al. 2013; Mouillot et al. 2013; Hopfenmüller et al. 2014; Allan et al. 2015; Birkhofer et al. 2015; Perović et al. 2015; Wood et al. 2015; Gossner et al. 2016; Simons et al. 2016; Birkhofer et al. 2017; Mangels et al. 2017).

In recent years, functional ecology has thus grown into a major field of ecological research, and that continues to integrate more and more fundamental and applied research questions. Given the expected future challenges in dealing with climate change and managing land use, global trait synthesis is an interdisciplinary, cross-topic field with great potential. However, it relies heavily on the broad availability and coverage of factual information and measurement data on all biotic taxa. To mobilize data for this enterprise, information must be made available digitally, harmonized according to data standards and hosted in interlinked online resources. The development of semantic-web standards (Page 2008; Wieczorek et al. 2012) and ontologies (Walls et al. 2012; Garnier et al. 2017), as well as web-services (Kattge et al. 2011; Parr et al. 2016) for trait-data management is under way. Nonetheless, individual researchers often neglect the benefit of centralised databases and, to cope with journals' and research funding agencies increasing demand for the publication of data without access restrictions (Swan 2012; Allison and Gurney 2015), instead publish primary data on file hosting services, where no data standards are forced upon the uploaded material (e.g. Figshare, Dryad). Given this prospect, there will be no lack of data, but a lack of data standardisation.

In this paper, we (1.) discuss the importance of data standards for trait-based research and identify current deficits in standardisation from an pragmatic view of data providers and data users. We further (2.) review existing trait databases and web services, as well as initiatives for standardisation. We (3.) describe the common practice for structuring trait data for ecological research. Based on these considerations, we (4.) propose a low-threshold structure and vocabulary for describing trait-datasets compatible with existing ontologies. (5.) We provide a tool for an easy adoption of this structure and for the harmonisation of data from different sources.

With an easy-to-use terminology and toolset, we hope to inform trait-data providers and trait-data users about the general importance of data standardisation and sketch a roadmap towards simple and low-threshold trait data standardisation.

# 1. Need for data standardisation

Despite the obvious benefit and grand prospects of functional approaches, an integration of several studies across taxa, trophic levels, or organisational scales remains difficult because of different organismal, methodological, and functional focus, which renders data on traits extremely heterogeneous.

Trait data have therefore been standardised into centralised databases for specific organism groups and research questions. These compiled data are either being published as static data tables or in the form of sophisticated web-services (see below for a review of current initiatives). Services attract data submissions from a focal community of researchers and map heterogeneous data into a common scheme, but more importantly, provide platforms for consensus building in methods. Furthermore, database services offer access control and data usage policies, to claim the rights of the original data providers, while simplifying data queries for synthesis researchers.

Nonetheless, open-access publications in low-threshold data repositories such as Figshare.com, Dryad (datadryad.org), Researchgate.net or Zenodo.org are gaining importance and foster a decentralised data hosting with low expectations regarding data standardisation and documentation. Although open for manual download, data on low-threshold data platforms might be stored in proprietary (e.g. .xlsx) or binary (e.g. .pdf) data formats which make a programmatical extraction tedious or impossible.

On the other hand, not all centralised databases for functional traits comply with the criteria of open access, which leaves them at a disadvantage compared to low-threshold repositories. Open access to science data from publicly funded research is a declared goal of international and national research organisations (Alliance of German Science Organisations 2010; The Royal Society Science Policy Centre 2012; Emerson et al. 2015) and increasingly required by journals and funding agencies to warrant reproducibility of science and foster public benefit. Also, factual data such as trait measurements, being not a product of a creative act, do not qualify as intellectual property and copyright does not apply (Carroll 2015; Creative Commons n.d.). However, there is still no broad consensus achieved in the scientific community whether data from public funding must be generally dedicated to the public domain, or whether the investment in production must be acknowledged by attribution or co-authorship. Low-threshold data repositories assume public domain and Creative Commons Attribution licenses compliant with the demands of journals and research funding.

Given this prospect, there will be no lack of data, but a lack of data standardisation hampering efficient future use of published datasets. To cope with this trend towards distributed data sources, data brokering services (e.g. the German Federation on Biological Data; gfbio.org) or database management systems for scientific projects (e.g. BExIS, XX ) are likely to gain importance. These services simplify and direct the standardized upload of research data and descriptive metadata into reliable and interlinked data infrastructures. One goal of such initiatives is to facilitate data publications and standardization for researchers, for instance by providing terminologies and ontologies for biodiversity data, and by consulting on publication licenses.

After all, to mobilise the broad wealth of data into structured databases, it requires a certain minimum of information from the data providers. Getting accurate and well-defined input requires improving literacy of ecological researchers on data standards and ontologies.

# 2. Overview of trait-data initiatives

In this section, we list two kinds of initiatives: the first aims at the harmonization of trait data into trait databases; the second develops trait ontologies and data standards. Often ontologies are being developed to serve a particular database project and thus are being developed in concert, as in the case of the TRY plant database (Kattge et al. 2011) and the TOP Thesaurus of plant traits (Garnier et al. 2017).

## Databases and Datasets

Plant flower, leaf and stem traits have been measured for decades, and were compiled into comprehensive databases and published as static datasets, e.g. the LEDA traitbase for the Northwest-European flora (Kleyer et al. 2008). Regarding webservices, most prominently, plant traits from many different sources have been collated into the extensive TRY database (Kattge et al. 2011). Other more specialised databases for plant traits exist, for instance flower traits (BiolFlor for German wild flowers, Klotz et al. 2002) and root measurements (Fine-root Ecology Database; Iversen et al. 2017). Traits on crop plants have been collected in agronomy databases like Planteome and BetyDB (<https://www.betydb.org/>; LeBauer et al. 2017). Trait-data for lower plants are less completely assessed.

Microbial ecology also makes great use of trait-based approaches to assess genomic function and describe functional diversity at the community level (Fierer et al. 2012; Fierer et al. 2014; Krause et al. 2014) derived from metagenomic analysis (Torsvik and Øvreås 2002; Langille et al. 2013). Databases are at hand to interpret gene sequences, so called 'operational taxonomic units' (OTUs), in terms of their functional role (e.g. the KEGG orthology, Kanehisa et al. 2012).

In the animal kingdom, n unknown plethora of single datasets has been published along with article publications (Sokol et al. 2011; Price et al. 2013; e.g., Gossner et al. 2015a, Gossner et al. (2015b), Renner et al. (2014); Birkhofer et al. 2017; van der Plas et al. 2017). In the past two decades, many datasets have been aggregated and harmonized into greater collections with particular organismal focus (Jones et al. 2009; Kissling et al. 2014; Myhrvold et al. 2015; Iversen et al. 2017; Oliveira et al. 2017; e.g. Ricklefs 2017). Specialised trait-databases have been created that cover different organism groups, interaction types, or ecosystems. For instance, the BETSI database collects traits on soil invertebrates (<http://betsi.cesab.org/>; Pey et al. 2014), and the Carabids.org web portal collects traits of carabid beetles (<http://www.carabids.org/>). Unlike for plants, a single unified webservice and harmonizing scheme for animal trait data is lacking. It has been pointed out that data on animals, covering multiple trophic levels and having evolved very different body plans will not be a simple task and go far beyond the efforts produced for plant traits (Moretti et al. 2017).

A general database for major physiological and life-history traits of all taxa is the Encyclopedia of Life Trait Bank project (Parr et al. 2016, see below for more detail).

A global overview of existing trait data for all taxa and trait types is difficult to obtain. As an attempt to track databases and unstandardised datasets from independent publications, we initiated a living spreadsheet (<https://goo.gl/QxzfHy>) which lists published trait datasets with their their taxonomic focus and extend of traits covered, their location on the internet and the terms of use. We invite data owners and users to register further datasets qualifying as trait data.

## Thesauri and Ontologies

A glossary or thesaurus of terms would be primarily a listing of definitions. Expanding on this, ontologies are linking individual terms with each other by formally defining the relationship between terms, to the ends of enabling a computational interpretation of data.

Lists of trait definitions have been developed for plants, for instance in the TOP Thesaurus of plant traits, which is employed in the TRY database (Garnier et al. 2017), the Planteome ontology for crop plants [Walls et al. (2012); Jaiswal et al. (2005); www.plantontology.org], or the Flora Phenotype Ontology (<http://obofoundry.org/ontology/flopo.html>, Hoehndorf et al. 2016). On the side of animal traits, a couple of initiatives published comprehensive trait ontologies, like for instance a Thesaurus for Soil Invertebrate Trait-based Approaches (T-SITA, <http://t-sita.cesab.org/>, Pey et al. 2014), the Hymenoptera Anatomy Ontology (Yoder et al. 2010, <http://hymao.org/>), or the vertebrate trait ontology (<http://obofoundry.org/ontology/vt.html>, Park et al. 2013). Hosting services for biological ontologies and terminologies are the OBO foundry (<http://obofoundry.org/>) or ontobee (<http://www.ontobee.org/>).

Relating the trait definition to methodological definitions on trait measurements are key for harmonizing trait data already at the physical process of measuring. For instance, handbooks for measurement protocols provide such a methodological standardization for plants (Cornelissen et al. 2003; Perez-Harguindeguy et al. 2013) or invertebrates (Moretti et al. 2017) and should be referenced when defining traits in Thesauri.

While trait thesauri define traits for focal groups of organisms, they are not specifying the formats or structures in which trait data should be stored. The Encyclopedia of Life (EOL) project has advanced on the topic by developing the database framework TraitBank (Parr et al. 2016), which is to date the most general approach to providing a semantic structure for trait data tables. The framework employs established terms provided by the Darwin Core Standard (DWC, Wieczorek et al. 2012), relates trait definitions to specific trait ontologies and phenotypic and anatomical ontologies, and maps taxa to global identifiers in taxonomic hierarchies of multiple name service providers (e.g. GBIF Taxonomy) to capture synonyms, misspellings and controversies (Parr et al. 2016). Additional layers of information capture bibliographic reference, multimedia archives and ecological interactions. TraitBank invites data submissions to the EOL database in a structured Darwin Core Archive. Another initiative aimed for a mobilization of measurements from museal collection data collated in the VertNet database into harmonized traitdata (Guralnick et al. 2016).

However, the Terms of the Darwin Core Standard and available Extensions, TraitBank and VertNet, are responses to a demand in Biodiversity Data Management for structured input from the research communities. The practical importance of these frameworks for data providers and data users is less obvious and often not appropriately considered when storing or uploading data to low-threshold file-publishing services. None of the aggregated and harmonized databases for birds, amphibians, or mammals employs such ontologies or semantic web standards. One reason certainly is compatibility: to date, the existing trait data standards do not cover all of the aspects required in the various use-cases of trait-based research. The other reason is competence: trait-data providers are not trained in the demands of biodiversity data-management and complying with the apparently non-intuitive data structures is an investment without immediate pay-off.

# 3. The structure of trait datasets

As a minimum consensus, trait datasets may be defined as follows: A trait dataset contains quantitative measurements or qualitative facts (i.e. trait values) about physical phenotypic characteristics of fitness, behaviour or performance (i.e. traits) observed on an individual organism or parts of an individual (i.e. an entity) that can be assigned to a biological taxon (i.e. a species or higher-level taxon) .

Quantitative measurements are values obtained either by direct morphological or physiological measurement on single specimens (Fig. 1*a*), by aggregation of replicated measurements on multiple entities (Fig. 1*b*) or by estimation of means or ranges for the species as reported in literature or other published sources (e.g. databases, Fig. 1*c*). Qualitative facts are assignments of an entity to a categorical level, e.g. of a behavioural or life-history trait (Fig. 1.*d*). The entity or observation (i.e. the occurrence) to which the reported measurement or fact applies may differ in resolution -- depending on the scientific question -- and could be a sub-sample or bodypart, an individual specimen, an entire species or higher-level taxon.  
This definition refines the entity-quality model (EQ) as described by Garnier et al. (2017), where a trait observation is phrased as 'an entity having a quality'.

Trait-datasets usually take different formats. For instance, if trait data have been collated at the species level from different literature sources or from expert knowledge, or as aggregated measurements collated from raw data, they usually are reported in a speciestrait matrix (a.k.a. wide-table format, Fig. 2*a*), with a column of trait values for each trait recorded and a row for each species (or taxon) for which data were available. This format is usually reporting missing data as 'NA'. It may store additional information (e.g. on variation of means or literature source) in secondary columns or accompanying datasheets. The matrix format is widely used for the production of lookup-tables at the species level, which may either report qualitative facts (from literature or expert knowledge) or aggregated quantitative traits (averages or ranges for this taxon). Quantitative traits at this resolution for instance may be used for the calculation of community weighted means or functional diversity metrics at the community level (Laliberté and Legendre 2010). Also at this level of information, missing entries, e.g. about the trait values or behaviour and functional role of a rarely described species, may be inferred from these tables combined with phylogenies ('gap filling', Swenson 2014) and using Bayesian approaches (Schrodt et al. 2015). However, the tables do not allow to estimate within-species variation or assess the quality of the reported measurement or fact (Guralnick et al. 2016).

Phenotypic variation of traits within species would usually be recorded per observation, i.e. each individual specimen would be recorded in a row, with the respective trait values stored in columns. In these occurrence data tables (Fig. 2*b*), each row is centered around a single physical occurrence of a species. Those data are common in investigations of evolutionary trade-offs and trait correlation, or of intra-specific variation along environmental gradients. Occurrence data would also capture phenotypic variation arising from morphotypes and sexual dimorphism. This format is the most intuitive for recording own empirical measurements and therefore is common for measured quantitative data and rarely found for reported qualitative facts obtained from the literature.

Both the matrix format and the occurrence table format have the different trait values which have been collected or measured on one entity stored in multiple columns. Each trait column represents one measured value. The width of the table varies with the number of traits included. Merging data from different sources requires manual harmonization of these structures.

Computationally more effective and allowing for highest flexibility is the storage of traitdata in 'tidied' long-table formats (Wickham 2014; Parr et al. 2016), where each row is reserved for a single measurement or fact of a specific trait, referenced to a single specimen which is assigned to a specific taxon (i.e., one occurrence). This allows for storing repeated measurements on a single individual by linking the data of separate rows via a unique specimen ID. Also, multivariate trait measurements can be recorded in this format by linking multiple rows via a unique measurement identifier. Long-table dataset purport multiple advantages for the data manipulation (e.g. filtering, sub-setting and aggregating data), visualization (e.g. plot measured values by factor variable or taxon) and statistical modelling (e.g. ANOVA for testing difference of trait value by sex) (Wickham 2014). They further draw from a defined set of columns, which simplifies merging of datasets.

The latter format thus provides the highest resolution and is ideal for storing raw data of trait measurements in huge databases and for linking them via the respective unique identifiers to additional information, such as occurrenceID or locationID. Each row of the dataset can be interpreted as a statement of an 'entity x having a qualitative/quantitative feature y' (Fig. 2*c*).

# 4. Proposing an ecological traitdata standard

With this paper, we propose a versatile, minimal data structure, as well as a defined vocabulary that captures the different degrees of resolution and measurement detail for multiple use cases of trait-based research. We were drawing on the combined expertise of empirical biodiversity researchers (data providers), biodiversity synthesis researchers (data users), and biodiversity informatics researchers (data managers).  
We were particularly considering the work of Kattge et al. (2011), and Garnier et al. (2017), as well as Parr et al. (2016) to ensure compatibility of datasets with major web-services and existing standards for biodiversity data management.

The online reference glossary of terms and description of the proposed structure is found at <https://ecologicaltraitdata.github.io/TraitDataStandard/>.

The vocabulary proposed will not capture the entire complexity of trait-based research in ecology and evolution. We invite researchers to comment on and contribute to the next iterations of this standard and develop own applications and ontologies.

## structure for trait data

As described above, tidy long-table formats are the most predictable and compatible format for the core content of trait datasets and should be used for standalone uploads on low-threshold repositories.

To enable a better standardisation of nomenclatures and values, the entries of the core dataset should be related to semantic web references using Uniform Resource Identifiers (URIs) keep a duplicate record of accepted taxon names and standard trait names. In these duplicate columns, values should be reported as 'harmonized' entries, which are linked to semantic ontologies and globally valid identifiers, consistent with semantic web standards (Berners-Lee et al. 2001; Page 2008; Wieczorek et al. 2012). If data contain multi-layered information on the measurement or occurrence (including geolocation and date information), further columns may be added to the core set (see description for Extensions below). For reasons of reproducibility and openness, data should not be uploaded in proprietary spreadsheet formats (like '.xlsx') but rather in comma-separated text files ('.csv' or '.txt').

The metadata of any dataset that employs this data structure should refer to the online version as "Schneider et al. 2017 Ecological Traitdata Standard v1.0, DOI: XXXX.xxxx, URL: <https://ecologicaltraitdata.github.io/TraitDataStandard/v1.0/>". In addition to the versioned online reference, the dataset should also cite this paper for an explanation of the rationale.

In very complex datasets repetition of verbose entries and long character strings in a single data table should be avoided (Wickham 2014). Therefore we recommend use of Darwin Core Archives (DwC-A, Robertson et al. 2009), the structure proposed for GBIF (GBIF 2017, <http://tools.gbif.org/dwca-assistant/>) or EOL TraitBank (Parr et al. 2016, <http://eol.org/info/cp_archives>). This also integrates a defined framework for metadata in the Ecological Metadata Language (KNB 2011), which should be used for describing and linking the data tables.

In some contexts, it may be safer to add such information directly as own columns of the data table to facilitate an analysis of co-factors and correlations further down the road. This way, if datasets of different source are merged, the information is readily available without the risk of breaking the reference to an external datasheet. The trade-off is data-consistency and readability *vs.* avoidance of content duplication.

Other cases, where key data columns would be placed in the same table as the core data are traits assessed on a higher level of organisation, e.g. microbial functional traits assessed at the community level taken from a soil sample. Here, location or measurement information are in the primary focus of the investigation (see vocabulary extensions below). A general definition, whether a column describes asset data or is part of the central dataset is ill advised. Therefore, our glossary of terms and its extensions should be used to describe the scientific data according to the study context.

## vocabulary for trait data

Column names and, where possible, field entries should follow the defined vocabulary listed at <https://ecologicaltraitdata.github.io/TraitDataStandard/>. Individual terms should be referenced via their global identifier, which is hosted by the GFBio Terminology Service and also can be accessed programmatically (i.e. via the Application Programming Interface, API). Wherever our glossary refines or duplicates existing terms from other ontologies for biological data, like the Glossary of EOL (<http://eol.org/info/516>) and Darwin Core (<http://rs.tdwg.org/dwc/terms/>), we indicate this in the fields 'refines' or 'identical', respectively.

### Core columns

To describe the minimum core content, where each row is the reported measurement or fact for a single observation, the following columns are required: a value (column traitValue) and -- for numeric values -- a standard unit (traitUnit); a descriptive trait name (traitName); the scientific taxon for which the measurement or fact was obtained (scientificName). For these core values, we highly recommend to use unambiguous and self-explanatory vocabularies for trait names and taxa. However, to have the data compatible with own databases or analytical code, it might be necessary to use abbreviations or identifiers instead.

### Relate to global taxon and trait definitions

It is indispensable to relate the user-defined vocabulary to a consensus standard of taxon names or or look-up tables of traits. This is achieved by adding globally valid identifiers for taxa (taxonID) and trait definitions (traitID), complemented by the more readable accepted names (ScientificNameStd and traitNameStd, respectively), according to a public ontology (e.g. GBIF Backbone Terminology; for *Bellis perennis*, the taxonID would be '<https://www.gbif.org/species/3117424>'; in TOP Thesaurus, the traitID for 'fruit mass' would be '<http://top-thesaurus.org/annotationInfo?viz=1>&&trait=Fruit\_mass').  
By allowing for a double record of both user-specific and standardised entries, we acknowledge the fact that most authors have their own schemes for standardisation and may refer to different scientific community standards (as practised in TRY; Kattge et al. 2011). This redundancy of data enables continuity on the author and data owner side and quality checking and comparability on the data user side. Keeping an additional reference to a kingdom and class (using columns kingdom, class as described by DWC and suggested for EOL Traitbank) will avoid ambiguities with taxonomic homonyms. It is best practice to keep a record of the reported taxon rank (in column taxonRank) for filtering purposes for traits recorded at the family or genus level. This information can be used for trait inference of higher taxonomic resolution using hierarchical probabilistic matrix factorisation (Schrodt et al. 2015).

The existing thesauri and ontologies for traits differ widely in terms of hierarchical depths and detail, as well as curation efforts and measures for peer-reviewed quality control. The accessibility by software differs widely. For biodiversity data, the OBO Foundry (<http://www.obofoundry.org/>) or the GFBio Terminology service (<https://terminologies.gfbio.org/browse/>), provide hosting for advanced trait ontologies and offer API services for computational access. More compatibility and availability of trait thesauri could greatly improve the possibilities for a comparison of trait data from different sources.

If no published trait definition is available that can be referenced, trait-datasets should be accompanied by a dataset-specific glossary of traits, or thesaurus. A thesaurus provides a "controlled vocabulary designed to clarify the definition and structuring of key terms and associated concepts in a specific discipline" (Laporte et al. 2013; Garnier et al. 2017). Thus, in its simplest form, a trait thesaurus would assign trait names with A) an unambiguous definition of the trait and B) an expected format (e.g. units or legit factor levels) of measured values or reported facts. It might additionally provide semantic relationships between terms for deriving a hierarchical or tree-based classification of traits. By providing a minimal vocabulary for this case (<https://ecologicaltraitdata.github.io/TraitDataStandard/#terms-for-traitlists-a-trait-thesaurus>), we hope to facilitate the publication of trait thesauri developed for the own project context which always should be referenced in the core trait dataset. The thesaurus might accompany the core data in a Darwin Core Archive, or be published on any other stable webservice.

See our online reference for recommendations how to publish a minimal Thesaurus compliant with semantic web standards [<http://ecologicaltraitdata.github.io/TraitDataStandard/#terms-for-traitlists-a-trait-thesaurus>]. We also provide an example thesaurus, that has been developed to capture traits assessed within the German Biodiversity Exploratories project on Github (<http://ecologicaltraitdata.github.io/TraitDataList/>).

### Metadata

Information about the authorship and ownership of the data and the terms of use should be considered when sharing and working with trait datasets. We define a set of terms (<https://ecologicaltraitdata.github.io/TraitDataStandard/#metadata-vocabulary>) that allows to relate trait data to authorship and define a bibliographic reference and license model.

In the case of raw measurement data, these information apply to the entire trait dataset, which classifies them as metadata stored in the information structure where the data are published. In cases, where individual measurements and facts from different sources are compiled into a trait database, these metadata information must be linked to the measurement level, via an unambiguous datasetID and a descriptive datasetName.

### Extensions for additional data layers

Beyond measurement units or higher taxon information, further information might be available that may not be core data, but are related to the individual or specimen, or to the reported fact, measurement or sampling event.

The data standard provides two extensions of the vocabulary that should be used to describe these information:

* the occurrence extension (<https://ecologicaltraitdata.github.io/TraitDataStandard/#extension-occurrence>) contains vocabulary to describe information on the level of individual specimens, such as physiological specifications of the phenotype, such as sex, life stage or age. This also includes the date and geographical location, as well as the method of sampling and preservation, which provides an important resource to analyse trait variation due to differences in space and time.
* the measurementOrFact extension (<https://ecologicaltraitdata.github.io/TraitDataStandard/#extension-measurement-or-fact>) provides terms to describe information at the level of single measurements or reported facts, such as the original literature from where the value is cited, the method of measurement or statistical method of aggregation. It aims at important information to track potential sources of noise or bias in measured data (e.g. variation in measurement method) or aggregated values (e.g. statistical method applied), as well as the source of reported facts (e.g. literature source or expert reference).

Many terms of these extensions refine terms of the Darwin Core Standard and their own [MeasurementOrFact](http://rs.tdwg.org/dwc/terms/index.htm#measureindex) and [Occurrence](http://rs.tdwg.org/dwc/terms/index.htm#Occurrence) extensions and EOL TraitBank's use of those terms (<http://eol.org/info/structured_data_archives>).

Same as for taxon or trait information, these additional layers of information can either be added as extra columns to the core dataset or, to avoid redundancy and duplication of content, be kept in separate data sheets of a Darwin Core Archive (see above). A unique identifier would link to these other datasheets, encoding each individual occurrence of a species (occurrenceID), single measurements or reported facts (measurementID), locations of sampling (locationID) and sampling campaigns (eventID). Some data-types may directly refer to existing global identifiers for occurrence IDs, e.g. a GBIF URI or a museum collection code referencing the precise specimen at which the measurement was taken and that is publicly available.

# 5. Computational tools for producing compliant data

To access data from public data-bases, the R-package 'traits' (Chamberlain et al. 2017) contains functions to extract trait data via several open API interfaces including Birdlife, EOL TraitBank or BetyDB. The package 'TR8' provides similar access to plant traits from a list of databases (including LEDA, BiolFlor and Ellenberg values; Bocci 2015) and aggregates them into a species-trait matrix. However, none of these packages provides harmonization of traits into a unified scheme to facilitate further analysis.

We provide the R package 'traitdataform' to assist in producing data compliant with the trait data standard proposed here. There are two major use cases for the package:

* preparing of own trait datasets for publication on low-threshold file-servers and project databases, and
* harmonising trait datasets from different sources by moulding them into a unified format.

A comprehensive documentation of the package is found in the package vignette, and on its Github repository (<https://github.com/EcologicalTraitData/traitdataform>) and the documentation website (<http://EcologicalTraitData.github.io/traitdataform/>).

The package is under continuous open source development and invites participation in development, comments or bug reports via the Github Issue page (<https://github.com/fdschneider/traitdataform/issues>).

The key function of the package is as.traitdata() which moulds a species-trait-matrix or occurrence table into a measurement longtable format. It also maps column names into terms provided in the trait data standard and adds metadata as attributes to the output object. This example converts an own file 'data.csv' into a dataset of long-table structure that employs the standard vocabulary for core data:

library("traitdataform")  
dataset <- as.traitdata(read.csv("path/to/data.csv"),   
 traits = c("body\_length", "antenna\_length",   
 "metafemur\_length"),  
 units = "mm",  
 taxa = "name\_correct",   
 keep = c(locationID = "location")  
 )

The column containing taxon names is given in parameter 'taxa'. Note that the parameter 'keep' specifies and renames any data that should be maintained in the output. The parameter 'units' is used to specify the units of measurement.

Scientific taxon names are matched automatically to the GBIF Backbone Taxonomy and added to the core data by calling the function standardize.taxonomy(). Similarly, the standardize.traits() function matches user-provided trait names onto a given trait vocabulary and links it to an existing trait ontology or a user-provided table of trait definitions, i.e. a thesaurus of traits. The same function harmonises trait values into target units and legit factor levels. The following example harmonizes traits based on a minimal thesaurus, referencing trait names with globally valid URIs defined by the BETSI thesaurus of soil invertebrate traits:

traitlist <- as.thesaurus(  
 body\_length = as.trait("body\_length", traitUnit = "mm", traitUnitStd = "mm",   
 traitID = "http://t-sita.cesab.org/BETSI\_vizInfo.jsp?trait=Body\_length"),   
 antenna\_length = as.trait("antenna\_length", traitUnitStd = "mm",   
 traitID = "http://t-sita.cesab.org/BETSI\_vizInfo.jsp?trait=Antenna\_length"),  
 metafemur\_length = as.trait("metafemur\_length", traitUnit = "mm",  
 traitID = "http://t-sita.cesab.org/BETSI\_vizInfo.jsp?trait=Femur\_length")  
)  
  
datasetStd <- standardize.traits(dataset, thesaurus = traitlist)

The function as.thesaurus() provides a structured object that is required by the function standardize.traits(). Other ways of defining a thesaurus object are documented in the package vignette and function documentation (?as.thesaurus). Future iterations of the R package will aim at automatising the generation of trait thesaurus objects from globally available Ontologies.

The package functions form a toolchain where each function can be piped as an input into the next. A wrapper function standardize() applies all functions sequentially, making transferring and harmonising traitdata as simple as:

datasetStd <- standardize(read.csv("path/to/data.csv"),  
 thesaurus = traitlist,  
 taxa = "name\_correct",   
 units = "mm"  
 )

Datasets that have been produced by these functions can easily be appended using the rbind() function of R base, while maintaining any available metadata information as separate column entries.

To merge datasets with additional information on the occurrence or measurement level, secondary data tables can be added as columns of the core dataset according to a unique identifier using the function merge(). This enables an easy handling of data sources that have been stored in a Darwin Core Archive format.

Since the intention of the package is to facilitate upload of own data to hosting services but also simplify the access to published trait data, the package offers direct access to trait datasets that have been published in the public domain or under creative commons licenses (Gossner et al. 2015a; Gossner et al. 2015b; Myhrvold et al. 2015; Ricklefs 2017; van der Plas et al. 2017). See a full list of linked datasets on the package website or by calling data(package = "traitdataform"). We invite users and authors of datasets to add further data to the package and thereby contribute to this registry for distributed trait datasets.

# Conclusion

In this paper, we propose a general, low-threshold structure and vocabulary for ecological trait data. We also provide a toolbox in R to transfer and harmonize data into this scheme.

To date, existing initiatives and centralised infrastructures for standardisation and data hosting are focused on a constrained organism group, ecosystem type or region. To make these heterogeneous sources of data more accessible and interlinkable, we propose a traitdata standard terminology and formatting scheme that can be applied in multiple contexts, for instance for building an own taxon-specific database or in user-side compilations of trait data from different sources.

The tools proposed here help to standardise trait datasets before upload to central as well as distributed data repositories. In times of increasing demand for open research data and international platforms for biodiversity data management, the development of meaningful terminologies for the standardisation of biodiversity data is more than essential: defined ontologies enable researchers to relate published datasets to each other to achieve a greater synthesis, thereby opening the doors for new insights on global biodiversity patterns. In terms of data science, semantic web ontologies are a prerequisite for advanced big data mining using machine learning techniques.

We therefore encourage the development of trait thesauri in semantic ontologies. The biggest challenge in community efforts of standardisation of traits may be the investment in consensus building which leads to an acceptance and establishment of the methodological and conceptual definitions of traits. This effort is time consuming and lengthy, but it returns synergies and great scientific benefit by enabling a global synthesis, as the example of the global plant trait database TRY shows intricately (Díaz et al. 2016).

Furthermore, the tools provided here can be applied to harmonise and combine data from heterogeneous sources. Such tools to handle the extensive raw data resources on traits are a prerequisite for research on trait variation along environmental gradients or intra-specific variation. It also opens the ground for a development of best practice principles for the filtering of raw measurements when creating species-level mean values. We hope this aids research on the response of traits to environmental and anthropogenic drivers, synthesis on trait patterns on all ecological scales (Messier et al. 2010), as well as multi-trophic trait analysis (Allan et al. 2015; Soliveres et al. 2016).

The trait data standard as well as the R package are published for open source community development. Contributions are invited and vital for the future development of the data standard and an extension and improvement of the R package.

# Acknowledgements

Thanks to all respondents to an internal online survey on trait data for the Biodiversity Exploratories project and to Matthias Biber, Diana Bowler, Klaus Birkhofer, Runa Boeddinghaus, Martin Gossner, Catrin Westphal for comments on the manuscript, the trait data standard and the R-package.

We thank the managers of the three Exploratories, Kirsten Reichel-Jung, Juliane Vogt, Miriam Teuscher, and all former managers for their work in maintaining the plot and project infrastructure; Christiane Fischer and Jule Mangels for giving support through the central office, Michael Owonibi for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project.

The work has been partly funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories" (DFG-Refno.). Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG).

M.J. was supported by the German Research Foundation within the framework of the Jena Experiment (FOR 1451) and by the Swiss National Science Foundation.

# Authors' contributions

FDS, MJ, GP, AO, CP, and NKS conceived the ideas for the manuscript and critically revised the glossary of terms of the trait data standard; FDS led the writing of the manuscript; NKS authored the list of traits; FDS developed the R package. All authors contributed critically to the drafts and gave final approval for publication.

# References

Allan, E., P. Manning, F. Alt, J. Binkenstein, S. Blaser, N. Blüthgen, S. Böhm, et al. 2015. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. Ecology Letters 18:834–843.

Alliance of German Science Organisations. 2010. Principles for the Handling of Research Data. Schwerpunktinitiative “Digitale Information”. <http://www.allianzinitiative.de/en/core-activities/research-data/principles/#c2363>.

Allison, L., and R. Gurney, eds. 2015. A Place to Stand: E-Infrastructures and Data Management for Global Change Research (COMMUNITY EDITION).

Berners-Lee, T., J. Hendler, and O. Lassila. 2001. The semantic web. Scientific american 284:28–37.

Birkhofer, K., M. M. Gossner, T. Diekötter, C. Drees, O. Ferlian, M. Maraun, S. Scheu, et al. 2017. Land-use type and intensity differentially filter traits in above-and below-ground arthropod communities. Journal of Animal Ecology 86:511–520.

Birkhofer, K., H. G. Smith, W. W. Weisser, V. Wolters, and M. M. Gossner. 2015. Land-use effects on the functional distinctness of arthropod communities. Ecography 38:889–900.

Birkhofer, K., V. Wolters, and T. Diekötter. 2014. Grassy margins along organically managed cereal fields foster trait diversity and taxonomic distinctness of arthropod communities. Insect conservation and diversity 7:274–287.

Bocci, G. 2015. TR8: An R package for easily retrieving plant species traits. Methods in Ecology and Evolution 6:347–350.

Börschig, C., A.-M. Klein, H. von Wehrden, and J. Krauss. 2013. Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. Basic and Applied Ecology 14:547–554.

Carroll, M. W. 2015. Sharing Research Data and Intellectual Property Law: A Primer. PLOS Biology 13:e1002235.

Chamberlain, S., Z. Foster, I. Bartomeus, D. LeBauer, and D. Harris. 2017. Traits: Species Trait Data from Around the Web.

Cornelissen, J. H. C., S. Lavorel, E. Garnier, S. Diaz, N. Buchmann, D. E. Gurvich, P. B. Reich, et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian journal of Botany 51:335–380.

Creative Commons. n.d. Data. <https://wiki.creativecommons.org/wiki/Data>.

de Bello, F., S. Lavorel, S. Díaz, R. Harrington, J. H. C. Cornelissen, R. D. Bardgett, M. P. Berg, et al. 2010. Towards an assessment of multiple ecosystem processes and services via functional traits. Biodiversity and Conservation 19:2873–2893.

Dehling, D. M., T. Töpfer, H. M. Schaefer, P. Jordano, K. Böhning-Gaese, and M. Schleuning. 2014. Functional relationships beyond species richness patterns: Trait matching in plantbird mutualisms across scales. Global ecology and biogeography 23:1085–1093.

Diaz, S., M. Cabido, and F. Casanoves. 1998. Plant functional traits and environmental filters at a regional scale. Journal of vegetation science 9:113–122.

Digel, C., J. O. Riede, and U. Brose. 2011. Body sizes, cumulative and allometric degree distributions across natural food webs. Oikos 120:503–509.

Díaz, S., J. Kattge, J. H. C. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, et al. 2016. The global spectrum of plant form and function. Nature 529:167–171.

Díaz, S., S. Lavorel, S. McIntyre, V. Falczuk, F. Casanoves, D. G. Milchunas, C. Skarpe, et al. 2007. Plant trait responses to grazing a global synthesis. Global Change Biology 13:313–341.

Díaz, S., A. Purvis, J. H. C. Cornelissen, G. M. Mace, M. J. Donoghue, R. M. Ewers, P. Jordano, et al. 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. Ecology and Evolution 3:2958–2975.

Díaz, S., F. Quétier, D. M. Cáceres, S. F. Trainor, N. Pérez-Harguindeguy, M. S. Bret-Harte, B. Finegan, et al. 2011. Linking functional diversity and social actor strategies in a framework for interdisciplinary analysis of nature’s benefits to society. Proceedings of the National Academy of Sciences 108:895–902.

Duarte, L., M. Carlucci, C. Fontana, S. Hartz, and V. Pillar. 2011. Plant diaspore traits as indicators of mutualistic interactions in woody vegetation patches developing into a grassland-forest mosaic. Community Ecology 12:126–134.

Emerson, C., E. M. Faustman, M. Mokrane, and S. Harrison. 2015. *World Data System (WDS) Data Sharing Principles*. Zenodo.

Fierer, N., A. Barberán, and D. C. Laughlin. 2014. Seeing the forest for the genes: Using metagenomics to infer the aggregated traits of microbial communities. Frontiers in Microbiology 5.

Fierer, N., J. W. Leff, B. J. Adams, U. N. Nielsen, S. T. Bates, C. L. Lauber, S. Owens, et al. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences of the United States of America 109:21390–21395.

Fontaine, C., I. Dajoz, J. Meriguet, and M. Loreau. 2005. Functional Diversity of PlantPollinator Interaction Webs Enhances the Persistence of Plant Communities. PLOS Biology 4:e1.

Forister, M. L., V. Novotny, A. K. Panorska, L. Baje, Y. Basset, P. T. Butterill, L. Cizek, et al. 2015. The global distribution of diet breadth in insect herbivores. Proceedings of the National Academy of Sciences 112:442–447.

Fritz, S. A., and A. Purvis. 2010. Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. Conservation Biology 24:1042–1051.

Garnier, E., U. Stahl, M.-A. Laporte, J. Kattge, I. Mougenot, I. Kühn, B. Laporte, et al. 2017. Towards a thesaurus of plant characteristics: An ecological contribution. Journal of Ecology 105:298–309.

Gámez-Virués, S., D. J. Perović, M. M. Gossner, C. Börschig, N. Blüthgen, H. de Jong, N. K. Simons, et al. 2015. Landscape simplification filters species traits and drives biotic homogenization. Nature Communications 6:8568.

GBIF. 2017. Darwin Core Archives - How-to Guide. <http://eol.org/info/structured_data_archives>.

Gossner, M. M., T. M. Lewinsohn, T. Kahl, F. Grassein, S. Boch, D. Prati, K. Birkhofer, et al. 2016. Land-use intensification causes multitrophic homogenization of grassland communities. Nature 540:266–269.

Gossner, M. M., N. K. Simons, R. Achtziger, T. Blick, W. H. Dorow, F. Dziock, F. Köhler, et al. 2015a. A summary of eight traits of Coleoptera, Hemiptera, Orthoptera and Araneae, occurring in grasslands in Germany. Scientific Data 2:150013.

Gossner, M. M., N. K. Simons, L. Höck, and W. W. Weisser. 2015b. Morphometric measures of Heteroptera sampled in grasslands across three regions of Germany. Ecology 96:1154–1154.

Gravel, D., C. Albouy, and W. Thuiller. 2016. The meaning of functional trait composition of food webs for ecosystem functioning. Phil. Trans. R. Soc. B 371:20150268.

Gross, N., G. Kunstler, P. Liancourt, F. De Bello, K. N. Suding, and S. Lavorel. 2009. Linking individual response to biotic interactions with community structure: A trait-based framework. Functional Ecology 23:1167–1178.

Guralnick, R. P., P. F. Zermoglio, J. Wieczorek, R. LaFrance, D. Bloom, and L. Russell. 2016. The importance of digitized biocollections as a source of trait data and a new VertNet resource. Database 2016.

HilleRisLambers, J., P. Adler, W. Harpole, J. Levine, and M. Mayfield. 2012. Rethinking Community Assembly through the Lens of Coexistence Theory. Annual Review of Ecology, Evolution, and Systematics 43:227–248.

Hoehndorf, R., M. Alshahrani, G. V. Gkoutos, G. Gosline, Q. Groom, T. Hamann, J. Kattge, et al. 2016. The flora phenotype ontology (FLOPO): Tool for integrating morphological traits and phenotypes of vascular plants. Journal of Biomedical Semantics 7:65.

Hopfenmüller, S., I. Steffan-Dewenter, and A. Holzschuh. 2014. Trait-Specific Responses of Wild Bee Communities to Landscape Composition, Configuration and Local Factors. PLOS ONE 9:e104439.

Iversen, C. M., M. L. McCormack, A. S. Powell, C. B. Blackwood, G. T. Freschet, J. Kattge, C. Roumet, et al. 2017. A global Fine-Root Ecology Database to address below-ground challenges in plant ecology. New Phytologist 215:15–26.

Jaiswal, P., S. Avraham, K. Ilic, E. A. Kellogg, S. McCouch, A. Pujar, L. Reiser, et al. 2005. Plant Ontology (PO): A Controlled Vocabulary of Plant Structures and Growth Stages. International Journal of Genomics. Research article, <https://www.hindawi.com/journals/ijg/2005/373740/abs/>.

Jennings, N., and M. J. O. Pocock. 2009. Relationships between Sensitivity to Agricultural Intensification and Ecological Traits of Insectivorous Mammals and Arthropods. Conservation Biology 23:1195–1203.

Jones, K. E., J. Bielby, M. Cardillo, S. A. Fritz, J. O’Dell, C. D. L. Orme, K. Safi, et al. 2009. PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90:2648–2648.

Kalinkat, G., F. D. Schneider, C. Digel, C. Guill, B. C. Rall, and U. Brose. 2013. Body masses, functional responses and predator–prey stability. Ecology Letters 16:1126–1134.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research 40:D109–D114.

Kattge, J., S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, G. Bönisch, E. Garnier, et al. 2011. TRY a global database of plant traits. Global Change Biology 17:2905–2935.

Kissling, W. D., L. Dalby, C. Fløjgaard, J. Lenoir, B. Sandel, C. Sandom, K. Trøjelsgaard, et al. 2014. Establishing macroecological trait datasets: Digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. Ecology and Evolution 4:2913–2930.

Kleyer, M., R. Bekker, I. Knevel, J. Bakker, K. Thompson, M. Sonnenschein, P. Poschlod, et al. 2008. The LEDA Traitbase: A database of life-history traits of the Northwest European flora. Journal of Ecology 96:1266–1274.

Klotz, S., I. Kühn, W. Durka, and G. Briemle. 2002. BIOLFLOR: Eine Datenbank mit biologisch-ökologischen Merkmalen zur Flora von Deutschland (Vol. 38). Bundesamt für naturschutz Bonn.

KNB. 2011. Ecological Metadata Language (EML) Specification. The Knowledge Network for Biocomplexity. <https://knb.ecoinformatics.org/#external//emlparser/docs/eml-2.1.1/index.html>.

Krause, S., X. Le Roux, P. A. Niklaus, P. M. Van Bodegom, J. T. Lennon, S. Bertilsson, H.-P. Grossart, et al. 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Frontiers in Microbiology 5.

Kühsel, S., and N. Blüthgen. 2015. High diversity stabilizes the thermal resilience of pollinator communities in intensively managed grasslands. Nature Communications 6:7989.

Lajoie, G., and M. Vellend. 2015. Understanding context dependence in the contribution of intraspecific variation to community traitenvironment matching. Ecology 96:2912–2922.

Laliberté, E., and P. Legendre. 2010. A distance-based framework for measuring functional diversity from multiple traits. Ecology 91:299–305.

Langille, M. G. I., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31:814–821.

Laporte, M.-A., E. Garnier, and I. Mougenot. 2013. A faceted search system for facilitating discovery-driven scientific activities: A use case from functional ecology. Semantics for Biodiversity (S4BioDiv 2013) 25.

Lavorel, S., and É. Garnier. 2002. Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail. Functional ecology 16:545–556.

Lavorel, S., and K. Grigulis. 2012. How fundamental plant functional trait relationships scale-up to trade-offs and synergies in ecosystem services. Journal of Ecology 100:128–140.

Le Provost, G., N. Gross, L. Börger, H. Deraison, M. Roncoroni, and I. Badenhausser. 2017. Trait-matching and mass effect determine the functional response of herbivore communities to land-use intensification. Functional Ecology.

LeBauer, D., R. Kooper, P. Mulrooney, S. Rohde, D. Wang, S. P. Long, and M. C. Dietze. 2017. BETYdb: A yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. GCB Bioenergy n/a–n/a.

Mangels, J., K. Fiedler, F. D. Schneider, and N. Blüthgen. 2017. Diversity and trait composition of moths respond to land-use intensification in grasslands: Generalists replace specialists. Biodiversity and Conservation 1–21.

McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby. 2006. Rebuilding community ecology from functional traits. Trends in Ecology & Evolution 21:178–185.

Menezes, S., D. J. Baird, and A. M. V. M. Soares. 2010. Beyond taxonomy: A review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. Journal of Applied Ecology 47:711–719.

Messier, J., B. J. McGill, and M. J. Lechowicz. 2010. How do traits vary across ecological scales? A case for trait-based ecology. Ecology Letters 13:838–848.

Miranda, M., F. Parrini, and F. Dalerum. 2013. A categorization of recent network approaches to analyse trophic interactions. Methods in Ecology and Evolution 4:897–905.

Moretti, M., F. de Bello, S. Ibanez, S. Fontana, G. B. Pezzatti, F. Dziock, C. Rixen, et al. 2013. Linking traits between plants and invertebrate herbivores to track functional effects of land-use changes. Journal of Vegetation Science 24:949–962.

Moretti, M., A. T. Dias, F. Bello, F. Altermatt, S. L. Chown, F. M. Azcárate, J. R. Bell, et al. 2017. Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. Functional Ecology 31:558–567.

Mouillot, D., N. A. J. Graham, S. Villéger, N. W. H. Mason, and D. R. Bellwood. 2013. A functional approach reveals community responses to disturbances. Trends in Ecology & Evolution 28:167–177.

Myhrvold, N. P., E. Baldridge, B. Chan, D. Sivam, D. L. Freeman, and S. K. M. Ernest. 2015. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. Ecology 96:3109–3109.

Oliveira, B. F., V. A. São-Pedro, G. Santos-Barrera, C. Penone, and G. C. Costa. 2017. AmphiBIO, a global database for amphibian ecological traits. Scientific Data 4:sdata2017123.

Page, R. D. M. 2008. Biodiversity informatics: The challenge of linking data and the role of shared identifiers. Briefings in Bioinformatics 9:345–354.

Park, C. A., S. M. Bello, C. L. Smith, Z.-L. Hu, D. H. Munzenmaier, R. Nigam, J. R. Smith, et al. 2013. The Vertebrate Trait Ontology: A controlled vocabulary for the annotation of trait data across species. Journal of Biomedical Semantics 4:13.

Parr, C. S., K. S. Schulz, J. Hammock, N. Wilson, P. Leary, J. Rice, C. Jr, et al. 2016. TraitBank: Practical semantics for organism attribute data. Semantic Web 7:577–588.

Pearson, R. G., J. C. Stanton, K. T. Shoemaker, M. E. Aiello-Lammens, P. J. Ersts, N. Horning, D. A. Fordham, et al. 2014. Life history and spatial traits predict extinction risk due to climate change. Nature Climate Change 4:217–221.

Perez-Harguindeguy, N., S. Diaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. Bret-Harte, et al. 2013. New handbook for standardised measurement of plant functional traits worldwide. Australian Journal of botany 61:167–234.

Perović, D., S. Gámez-Virués, C. Börschig, A.-M. Klein, J. Krauss, J. Steckel, C. Rothenwöhrer, et al. 2015. Configurational landscape heterogeneity shapes functional community composition of grassland butterflies. Journal of applied ecology 52:505–513.

Petchey, O. L., E. O’Gorman, and D. F. Flynn. 2009. A functional guide to functional diversity measures. In S. Naeem, D. E. Bunker, A. Hector, M. Loreau, & C. Perrings, eds., Biodiversity, ecosystem functioning, and human wellbeing: An ecological and economic perspective (pp. 49–59). Oxford University Press, New York, USA.

Pey, B., M.-A. Laporte, J. Nahmani, A. Auclerc, Y. Capowiez, G. Caro, D. Cluzeau, et al. 2014. A Thesaurus for Soil Invertebrate Trait-Based Approaches. PLOS ONE 9:e108985.

Price, S. A., J. J. Tavera, T. J. Near, and P. Wainwright. 2013. Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. Evolution 67:417–428.

Reich, P. B., D. S. Ellsworth, M. B. Walters, J. M. Vose, C. Gresham, J. C. Volin, and W. D. Bowman. 1999. Generality of leaf trait relationships: A test across six biomes. Ecology 80:1955–1969.

Reich, P. B., M. B. Walters, and D. S. Ellsworth. 1997. From tropics to tundra: Global convergence in plant functioning. Proceedings of the National Academy of Sciences 94:13730–13734.

Renner, S. C., M. M. Gossner, T. Kahl, E. K. Kalko, W. W. Weisser, M. Fischer, and E. Allan. 2014. Temporal changes in randomness of bird communities across Central Europe. PloS one 9:e112347.

Ricklefs, R. E. 2017. Passerine morphology: External measurements of approximately one-quarter of passerine bird species. Ecology 98:1472–1472.

Robertson, T., M. Döring, J. Wieczorek, R. De Giovanni, and D. Vieglais. 2009. Darwin Core Text Guide. <http://rs.tdwg.org/dwc/terms/guides/text/index.htm>.

Rubach, M. N., R. Ashauer, D. B. Buchwalter, H. J. De Lange, M. Hamer, T. G. Preuss, K. Töpke, et al. 2011. Framework for traits-based assessment in ecotoxicology. Integrated environmental assessment and management 7:172–186.

Schleuning, M., J. Fründ, and D. García. 2015. Predicting ecosystem functions from biodiversity and mutualistic networks: An extension of trait-based concepts to plantanimal interactions. Ecography 38:380–392.

Schneider, F. D., S. Scheu, and U. Brose. 2012. Body mass constraints on feeding rates determine the consequences of predator loss. Ecology Letters 15:436–443.

Schrodt, F., J. Kattge, H. Shan, F. Fazayeli, J. Joswig, A. Banerjee, M. Reichstein, et al. 2015. BHPMF a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. Global Ecology and Biogeography 24:1510–1521.

Siefert, A., C. Violle, L. Chalmandrier, C. H. Albert, A. Taudiere, A. Fajardo, L. W. Aarssen, et al. 2015. A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. Ecology Letters 18:1406–1419.

Simons, N. K., W. W. Weisser, and M. M. Gossner. 2016. Multi-taxa approach shows consistent shifts in arthropod functional traits along grassland land-use intensity gradient. Ecology 97:754–764.

Sokol, E. R., E. F. Benfield, L. K. Belden, and H. Maurice Valett. 2011. The assembly of ecological communities inferred from taxonomic and functional composition. The American Naturalist 177:630–644.

Soliveres, S., F. van der Plas, P. Manning, D. Prati, M. M. Gossner, S. C. Renner, F. Alt, et al. 2016. Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. Nature 536:456–459.

Swan, A. 2012. Policy guidelines for the development and promotion of open access. Open guidelines series. UNESCO, Paris, France.

Swenson, N. G. 2014. Phylogenetic imputation of plant functional trait databases. Ecography 37:105–110.

The Royal Society Science Policy Centre. 2012. *Science as an open enterprise*. The Royal Society.

Torsvik, V., and L. Øvreås. 2002. Microbial diversity and function in soil: From genes to ecosystems. Current Opinion in Microbiology 5:240–245.

van der Plas, F., R. van Klink, P. Manning, H. Olff, and M. Fischer. 2017. Sensitivity of functional diversity metrics to sampling intensity. Methods in Ecology and Evolution 8:1072–1080.

Vandewalle, M., F. De Bello, M. P. Berg, T. Bolger, S. Dolédec, F. Dubs, C. K. Feld, et al. 2010. Functional traits as indicators of biodiversity response to land use changes across ecosystems and organisms. Biodiversity and Conservation 19:2921–2947.

Villéger, S., N. W. H. Mason, and D. Mouillot. 2008. New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology. Ecology 89:2290–2301.

Violle, C., M.-L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, and E. Garnier. 2007. Let the concept of trait be functional! Oikos 116:882–892.

Walls, R. L., B. Athreya, L. Cooper, J. Elser, M. A. Gandolfo, P. Jaiswal, C. J. Mungall, et al. 2012. Ontologies as integrative tools for plant science. American Journal of Botany 99:1263–1275.

Weiner, C. N., M. Werner, K. E. Linsenmair, and N. Blüthgen. 2014. Land-use impacts on plantpollinator networks: Interaction strength and specialization predict pollinator declines. Ecology 95:466–474.

Wickham, H. 2014. Tidy data. Journal of Statistical Software 59:1–23.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, et al. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PloS one 7:e29715.

Wood, S. A., D. S. Karp, F. DeClerck, C. Kremen, S. Naeem, and C. A. Palm. 2015. Functional traits in agriculture: Agrobiodiversity and ecosystem services. Trends in Ecology & Evolution 30:531–539.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-Bares, et al. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

Yoder, M. J., I. Mikó, K. C. Seltmann, M. A. Bertone, and A. R. Deans. 2010. A Gross Anatomy Ontology for Hymenoptera. PLOS ONE 5:e15991.

1. Corresponding author: florian.dirk.schneider@gmail.com [↑](#footnote-ref-1)
2. Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany [↑](#footnote-ref-2)
3. Institute of Plant Sciences, University of Bern, Bern, Switzerland [↑](#footnote-ref-3)
4. Department of Mathematics and Computer Science, Friedrich-Schiller-Universität Jena, Jena, Germany [↑](#footnote-ref-4)
5. Terrestrial Ecology Research Group, Department for Ecology and Ecosystem Management, Center for Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany [↑](#footnote-ref-5)