

A generic structure for plant trait databases

Jens Kattge^{1*}, Kiona Ogle², Gerhard Bönisch¹, Sandra Díaz³, Sandra Lavorel⁴, Joshua Madin⁵, Karin Nadrowski⁶, Stephanie Nöllert¹, Karla Sartor⁷ and Christian Wirth^{1,6}

¹Max-Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany; ²Department of Botany, University of Wyoming, Dept. 3165, 1000 E. University Ave, Laramie, WY 82071, USA; ³Instituto Multidisciplinario de Biología Vegetal (CONICET–UNC) and FCEfYN, Universidad Nacional de Córdoba, C. Correo 495, 5000 Córdoba, Argentina; ⁴Laboratoire d'Ecologie Alpine, UMR 5553 CNRS – Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France; ⁵Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; ⁶Department of Special Botany and Functional Biodiversity Research, University of Leipzig, Johannisallee 21-23, 04103 Leipzig, Germany; and ⁷Biological Laboratories, Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave, Cambridge, MA 02138, USA

Summary

1. Plant traits are fundamental for understanding and predicting vegetation responses to global changes, and they provide a promising basis towards a more quantitative and predictive approach to ecology. As a consequence, information on plant traits is rapidly accumulating, and there is a growing need for efficient database tools that enable the assembly and synthesis of trait data.

2. Plant traits are highly heterogeneous, exhibit a low degree of standardization and are linked and interdependent at various levels of biological organization: tissue, organ, plant and population. Therefore, they often require ancillary data for interpretation, including descriptors of the biotic and abiotic environment, methods and taxonomic relationships.

3. We introduce a generic database structure that is tailored to accommodate plant trait complexity and is consistent with current theoretical approaches to characterize the structure of observational data. The over-arching utility of the proposed database structure is illustrated based on two independent plant trait database projects.

4. The generic database structure proposed here is meant to serve as a flexible blueprint for future plant trait databases, improving data discovery, and ensuring compatibility among them.

Key-words: ancillary data, bio-informatics, covariates, dimensional data model, eco-informatics, functional biodiversity, hierarchical data structure, relational database, star-scheme

Introduction

There is a critical need for integrated analyses in ecology to better understand and manage Earth's biological resources (Clark *et al.* 2001). This raises significant challenges in accessing relevant data, including the development of global data information systems (Scholes *et al.* 2008), of which integrated plant trait databases must be a keystone. Plant traits – morphological, anatomical, physiological or phenological features measurable at the individual level (Violle *et al.* 2007) – reflect the outcome of evolutionary processes in the context of abiotic and biotic environmental constraints (Grime *et al.* 1997; Westoby *et al.* 2002; Díaz *et al.* 2004; Valladares, Gianoli & Gomez 2007). Information on a set of traits may therefore be a more objective predictor of ecosystem dynamics and function-

ing than, for example, species identity or functional group classification (McGill *et al.* 2006).

Plant trait data have been used in studies covering a diversity of topics, including plant functional ecology (Wright *et al.* 2004; Reich, Wright & Lusk 2007; Sperry, 2008), community ecology (Lavorel & Garnier 2002; Ackerly & Cornwell 2007; Messier, McGill & Lechowicz 2010), plant evolution (Moles *et al.* 2005; Cavender-Bares *et al.* 2009), macroecological theory (Enquist *et al.* 2007), palaeobiology (Barboni *et al.* 2004; Royer *et al.* 2007), disturbance ecology (Wirth 2005; Diaz *et al.* 2007), plant migration and invasion (Schurr *et al.* 2005; Tackenberg & Stocklin 2008), conservation biology (Kahmen, Poschlod & Schreiber 2002) and – more recently – plant geography (Swenson & Enquist 2007; Swenson & Weiser in press). Plant trait data are also critical for parameterizing vegetation characteristics in models of ecosystem dynamics (White *et al.* 2000; Kattge *et al.* 2009) and individual-based models of plant growth and mortality (Ogle & Pacala 2009; Wirth & Lichstein

*Correspondence author. E-mail: jkattge@bgc-jena.mpg.de
Correspondence site: <http://www.respond2articles.com/MEE/>

Table 1. Examples of current major compilations of plant trait data

Database	Reference	Url
Trait–species matrix		
LEDA	Kleyer <i>et al.</i> (2008)	leda-traitbase.org
BIOPOP	Poschlod <i>et al.</i> (2003)	uni-oldenburg.de/landeco/Projects/biopop/biopop_en.htm
BiolFlor	Klotz, Kühn, & Durka (2002)	ufz.de/biolflor
ECOFLORE	Fitter & Peat (1994)	ecoflora.co.uk
SID		data.kew.org/sid
InsideWood		insidewood.lib.ncsu.edu
PLANTSdata	Green (2009)	bricol.net/downloads/data/PLANTSdatabase
Wood Density	Chave <i>et al.</i> (2009)	
BROT	Paula <i>et al.</i> (2009)	uv.es/jgpausas/brot.htm
Traits in their environmental or experimental context		
ECOCRAFT	Medlyn & Jarvis (1999)	
MARIWENN	Ollivier, Baraloto, & Marcon (2007)	ecofog.cirad.fr/Mariwenn
ALTA	Lavorel <i>et al.</i> (2009)	
VISTA	Garnier <i>et al.</i> (2007)	
ArtDeco	Cornwell <i>et al.</i> (2008)	
Intraspecific variability and trait–trait correlation in an environmental or experimental context		
GlopNet	Wright <i>et al.</i> (2004)	bio.mq.edu.au/~iwright//glopian.htm
CORDOBASE	Díaz <i>et al.</i> (2004)	
Meta-Phenomics	Poorter <i>et al.</i> (2010)	
FET		bgc-jena.mpg.de/bgc-organisms/pmwiki.php/Research/FET
TRY		try-db.org

2009). Furthermore, plant traits provide complementary information to earth observations like remote sensing (Ollinger *et al.* 2008), measurements of atmospheric CO₂ fluxes and concentrations (e.g. FluxNet¹ and GlobalView²), forest inventories (e.g. the US Forest Inventory³), and global biodiversity assessments (Scholes *et al.* 2008).

The scientific focus determines the database requirements: the structure of a database should be as complex as necessary but as simple as possible, because the data compilation itself tends to introduce its own level of complexity. Major compilations of plant trait data have been, and are still being developed for different purposes: to characterize the species–trait matrix for specific geographical regions or selected traits, to compile trait measurements in the context of their environmental or experimental conditions, or to analyse correlations between a limited number of different traits accounting for intraspecific variability in an environmental context (for references see Table 1). Data compilations to characterize species–trait matrices compile one to several trait values per species and trait, but do not need to characterize the intraspecific variability in an environmental context in detail, as they are focussed on the species' mean state. Compilations focussed on trait measurements in the context of environmental or experimental information account for intra- and interspecific variability, but they do not necessarily support the correlation analysis between different traits, as they only allow indirect deducing if different measurements had been conducted on the same object

from information about the environmental/experimental context (location, plot, experimental treatment, individual, etc.). This indirect identification easily becomes ambiguous and may complicate analyses of trait correlations (see Appendix S1). Data compilations that account for the fact that different traits have been measured on the same object are often two-dimensional spreadsheets (tables) with traits and additional information (ancillary data) in separate columns and observations in rows. These two-dimensional spreadsheets are convenient for the compilation of a limited number of traits and ancillary data, but they become increasingly inconvenient as the number of traits and/or ancillary variables increase (Madin *et al.* 2007). This limitation leads to researchers often only including a limited amount of ancillary information and important details of the data collection reside in written notes or the researcher's memory (Michener *et al.* 1997).

In the past few years, the focus of plant trait research included the characterization of ecological strategies and biodiversity in a quantitative, functional context as a function of the environment and phylogenetic constraints, which provide conditions for a more quantitative and predictive community ecology (McGill *et al.* 2006) and for a more realistic representation of vegetation in earth system models (Lavorel *et al.* 2007). This requires the compilation of several traits in combination with a detailed characterization of their intra- and interspecific variability, trait–trait correlations, abiotic and biotic environment, and phylogeny. Statistical tools and modelling concepts to accomplish these demands are being developed with strong momentum, e.g. the adaptation of hierarchical Bayesian concepts to ecological analyses (Ogle & Barber 2008), or the implementation of adaptive strategies to vegetation modelling (Scheiter & Higgins 2009), while dedicated structures for plant trait databases have not been published so far, and ecological

¹FluxNet: <http://daac.ornl.gov/FLUXNET/index.cfm>.

²Globalview: <http://www.esrl.noaa.gov/gmd/ccgg/globalview>.

³US Forest Inventory and Analysis National Program: <http://www.fia.fs.fed.us>.

researchers had to develop them for each individual application.

The need for integrated analyses poses another requirement for the development of database structures: interoperability – the seamless merging of disparate information from different databases into an integrated form for analysis. Here, ontologies, formal models that use mathematical logics to define concepts and their relationships (Madin *et al.* 2008), provide a mechanism for creating a common computer-interpretable basis for interoperability. Ontologies have been successfully used in genetics, microbiology, and medicine, especially when used for interoperability among different databases and organizations. In ecology, where data structures are more complex, framework ontologies have only recently been laid out, which show considerable overlap (Madin *et al.* 2008). Consistency with these ontology schemata should ensure compatibility between different datasets and therefore be another requirement for future trait databases. However, simultaneously accounting for the various aspects of plant traits and the need for database interoperability poses a new challenge to the development of plant trait databases.

In the following, we analyse the structure of plant trait data to derive the key requirements for the development of plant trait databases ('entity-relationship analysis', Chen 1976). Based on this we propose a generic structure, which is supposed to serve as a blueprint to help ecologists develop plant trait databases according to their individual needs. Finally, we illustrate consistency with current framework ontologies and improvements in access and interoperability of disparate plant trait data by two recently developed trait databases using the generic database structure presented here.

Entity-relationship analysis of plant traits

PLANT TRAITS ARE A HETEROGENEOUS GROUP OF DATA WITH LOW DEGREE OF STANDARDIZATION

The term 'plant trait' is used in a wide range of contexts and there have been numerous attempts to define, conceptualize and categorize plant traits. Violle *et al.* (2007) propose a general definition of plant traits: 'Any morphological, physiological or phenological feature measurable at the individual level, from the cell to the whole-organism level,...' According to this definition, plant traits may characterize anything from a leaf dark respiration rate to a maximum plant height, referring to an organ vs. the whole plant, measured with different instruments (e.g. gas exchange instrument vs. hypsometer), reported in different units (e.g. $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ vs. m), and with different modes of standardization (e.g. at 25 °C in the dark vs. mean of tallest 5% of individuals ever measured), and so on. There have been substantial efforts to develop standardized trait definitions and measurement protocols, e.g. Cornelissen *et al.* (2003) and in the context of the LEDA project (Knevel *et al.* 2003; Kleyer *et al.* 2008), but these standards are only available for a limited set of traits, are often ignored, and do not help when it comes to the vast literature dating prior to 2003. Some traits cannot be measured but need to be estimated

by fitting models to raw data (parameter-based traits, e.g. $V_{\text{cmax},25}$, the maximum carboxylation rate at 25 °C or parameters defining seed dispersal kernels). In summary, plant traits can be seen as a heterogeneous group of information with relatively low degree of standardization (Table 2).

VARIABILITY AS SIGNAL AND NOISE IN COMPARATIVE TRAIT ANALYSIS: THE NEED FOR ANCILLARY DATA

Understanding and quantifying the variability of plant traits at the species or functional group level in response to changes in environmental drivers are of utmost importance in global change research and evolutionary ecology. Traits that are treated as qualitative, like leaf habit (e.g. needle-leaved, broad-leaved) or wood porosity type (e.g. ring-porous, diffuse-porous) are often nearly invariant within species, even though in some cases they are much more variable than studies suggest (e.g. flower colour, dispersal mode). Traits that are treated as quantitative, like leaf mass per area or leaf dry matter content vary substantially between and within individuals of a species (Messier, McGill & Lechowicz 2010). This intraspecific trait variability can optimize plant performance and fitness in response to abiotic and biotic constraints and is the consequence of genetic variation and phenotypic plasticity, the latter being the environmentally contingent trait expression of a given genotype (Fig. 1). The capacity for expressing trait variability may differ between species and developmental stages and may be constrained by trade-offs between different traits (Valladares, Gianoli & Gomez 2007). This variability may be the signal of interest, but may also represent unwanted 'noise' in a comparative analysis (Fig. 1).

For qualitative traits, which are almost invariant at the species-level (e.g. wood porosity type), we can assign each species i its unambiguous trait value θ_i . For quantitative traits with a low degree of variability (e.g. density of cell walls), it is often sufficient to calculate the mean trait value $\bar{\theta}_i$ if several measurements per species are available. In this case $\bar{\theta}_i$ is a stochastic (or uncertain) quantity, and this uncertainty should be propagated into a comparative analysis. However, the majority of quantitative traits, including many of the most relevant traits for ecosystem functioning, are strongly modulated by the environment. In this case, the absence of quantifying the variability in the trait substantially reduces the information content in the data and can lead to inference problems (e.g. ecological fallacy, Robinson 1950). Using the mean value will yield an unrealistic point estimate contingent to the particular individuals and localities that happen to be represented in the dataset (Albert *et al.* in press). Here, we need to model the variable trait in response to 'driving variables' (covariates). For example, estimates of maximum photosynthesis rates should account for the influence of irradiance, temperature, water availability, and air humidity during the measurement period as well as plant and leaf developmental stage (Kattge & Knorr 2007). The result is a standardized rate θ_i^* , i.e. a predicted rate at some reference state of the covariates (e.g. maximum photosynthesis at light saturation and 25 °C), coefficients that quantify the influence of the respective covariates (variability

Table 2. Common types of data in plant trait databases, their sources and application in ecological research

	Definition	Example	Typical sources	Application
Categorical – phylogenetic	Taxonomical descriptor	Species Genus	Floras Species monographs	Grouping variable Random or fixed factor in mixed effects model
Categorical – nominal	Qualitative feature that cannot be expressed numerically	Photosynthetic pathway Composite leaves Deciduous Weed Animal dispersed	Floras Species monographs Physiological literature Trait databases	Grouping Definition of plant functional types
Categorical – ordinal	Qualitative feature that describes a performance intensity relative to other plants	Shade tolerance Frost tolerance Resprouting capacity	Literature on plant indicator values Specialized databases Applied sciences (forestry, agriculture)	Grouping Classification Covariate Definition of plant functional types
Categorical – environment	Qualitative feature describing a plant's growth environment	Slope exposure Growth chamber Soil classification	Vegetation descriptions Original publications	Filtering Classification Random factor in mixed effects model
Quantitative – environment	Quantitative measure describing a plant's growth environment	Annual precipitation Fertilizer application Clay content LAI of community	Original publications Climate/Soil databases Experimental setup	Covariate Driver in model inversion
Quantitative plant state	Quantity characterizing the (often transient) state of a plant	Height Growth rate Specific leaf area	Original publications Inventories	Covariate Model validation
Quantitative plant trait	Quantity characterizing a typical (non-transient) feature of a taxonomic unit (e.g. family, species)	Maximum height Specific leaf area of sun leaves Stem wood density	Original publication Trait databases	Dependent variable Covariate Model parameter
Parameter-based trait	Statistical parameter of a function relating a plant state to an environmental variable	V_{cmax} Rate constant of leaf litter decomposition Q_{10} of leaf respiration	Original publications Meta-analyses	Parameter in process model

coefficients), and measures of their uncertainty and covariance. Comparing between species requires a standardization of traits, which in turn requires ancillary information about the environmental context, the type and amount of which depends on the trait and its inherent variability.

Moreover, ancillary data are often used for filtering, classification, and for accounting for data heterogeneity (Table 2). For example, filter variables may be used to tailor the dataset to the research questions. An analysis only interested in the traits of mature trees under field conditions might filter the data by 'tree age' and 'growth environment'. Ecological data are inherently stochastic and often structured by processes (e.g. 'random effects') that do not necessarily reflect the impacts of treatments or covariates of interest. For example, in a study on specific leaf area, the data may come from 15 different laboratories and these laboratories may have employed slightly different methods. In this case, it is not desirable to estimate separate parameters for each laboratory, and yet we must account for this extra variability to obtain accurate parameter estimates for the quantities of interest. Thus, providing information on laboratories or research groups allows one to treat

them as random effects. It should be noted that traits might themselves be used for filtering, classification and grouping of another trait. For example, the variable 'photosynthetic pathway' may be used as a grouping variable in a study on leaf nitrogen concentrations or as a response variable in a study on the distribution of photosynthetic pathways.

TRAIT CORRELATIONS AND THE IDENTITY PRINCIPLE

Correlations between traits are important as they often reflect optimality principles (Wright *et al.* 2004; Reich *et al.* 2006, 2008) and allow us to draw conclusions about the underlying evolutionary forces (Reich *et al.* 2003). Modellers employ them to exclude non-reasonable parameter combinations (Moorcroft, Hurtt & Pacala 2001) and for improved parameterization (Kattge *et al.* 2009). Most traits involved in these correlations and trade-offs vary from organ to organ, from individual to individual, and they covary over time in response to ontogenetic processes and environmental fluctuations. Correlations will appear progressively weaker if the different traits have been measured at different points in time or on

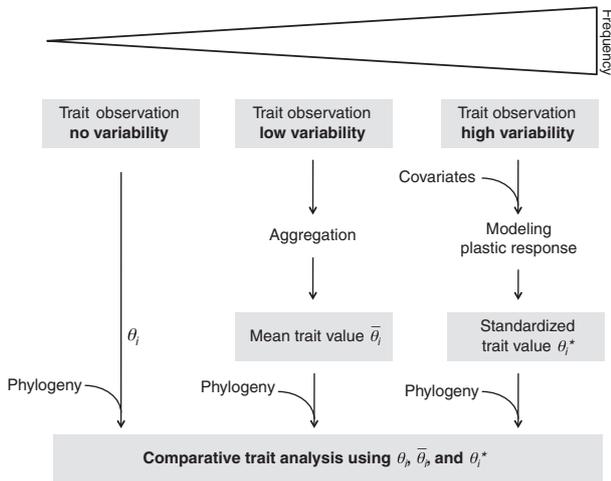


Fig. 1. Different levels of variability of plant traits and statistical treatment. No variability: qualitative traits invariable at the respective level of phylogeny (e.g. leaf habit at species-level); each species i is assigned its trait value θ_i . Low variability – quantitative traits with a low degree of variability (e.g. lignin content of bark): calculate the mean trait value $\bar{\theta}_i$ from several measurements for species i . High variability – quantitative traits with a high degree of variability (e.g. photosynthetic capacity): model the plastic response in relation to factors that affect it (cf. ‘covariates’); the result, θ_i^* , is the standardized (predicted) trait value at a reference state of the covariates. Finally, a comparative analysis requires phylogenetic (or taxonomic) information as a predictor or grouping variable. Frequency indicates the relative occurrence of the different types of traits: most traits are characterized by a high variability and only a few show no variability.

different organs of the same individual, different individuals of the same population, etc. (Fig. 2). Keeping track of the identities of the objects and the temporal context associated with a

trait measurement (identity principle) within a database is therefore important if one wishes to control for the ‘degree of relatedness’ when analysing correlations.

HIERARCHICAL STRUCTURE

The objects from which traits are measured are often nested in an observational hierarchy (Madin *et al.* 2007; Messier, McGill, & Lechowicz 2010). Each level of the hierarchy can be characterized by a unique set of traits and ancillary data that provide important context for other levels in the hierarchy (Fig. 2). An example is the hierarchy of ecological organization: the leaf is part of a branch, characterized by its length and particular position in the tree canopy; the branch is part of a tree, which may be healthy or damaged, young or old, and so on; the tree is part of a stand, which could be, for example, an uneven-aged mixed or mono-specific even-aged stand. The stand is part of a larger landscape unit, which may be characterized by shallow or deep soils. Since all the levels of this ecological hierarchy are connected, the influence of landscape (or site) conditions propagates along the hierarchy from the level of the stand to the individual leaf and cells.

TRAIT INFORMATION IS ONLY AS GOOD AS THE RESPECTIVE TAXONOMIC INFORMATION

Finally, the trait information is only as good as the taxonomic information. This poses an additional challenge to the development of a database structure. One problem that is very difficult to resolve is the mis-identification of species. Here, the collection of specimen of each studied individual may help and the plant trait database should facilitate the link from species name

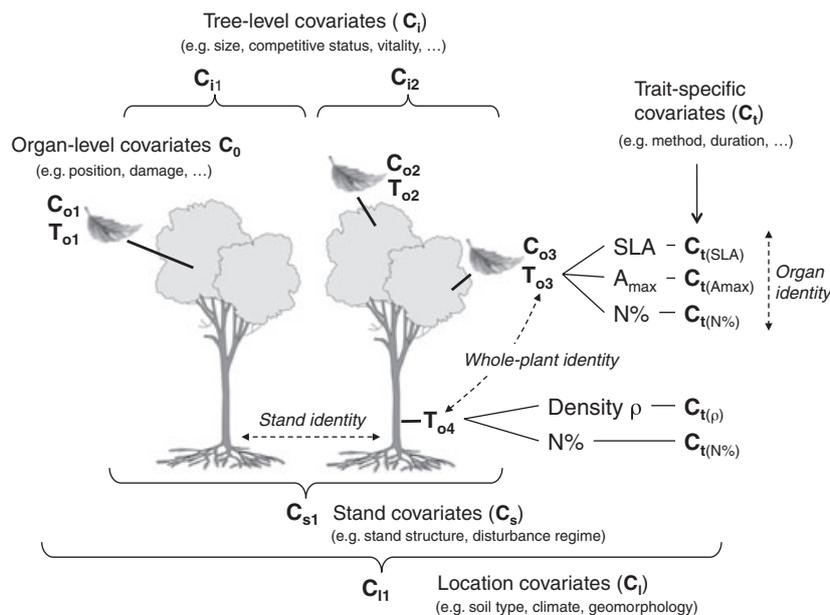


Fig. 2. Representation of the identity principle and hierarchical structure. T, vector of traits. The example traits are specific leaf area (SLA), nitrogen concentration (N%), maximum photosynthesis rate (A_{max}), and wood density (ρ). C, vectors of covariates (or ancillary data). Covariates either characterize trait measurements directly (trait covariates, C_t) or the hierarchical context of trait measurements, i.e. organ covariates (C_o), individual tree (or plant) covariates (C_i), stand covariates (C_s), location covariates (C_l) and time covariates (not shown). This scheme allows the measured objects to be unambiguously characterized and various types of identities specified (examples in italics).

to the respective specimen. Additional problems are different species concepts used by different floras, the synonymy of plant names, and the ongoing development and updating of species names and the deep taxonomy (Berendsohn & Geoffroy 2007). Assuming a good representation according to the current taxonomic concepts, what happens to the database 6 months, 2 years, a decade from now, when many of those species have been lumped, split, renamed, synonymized, etc.? Names are not static. The generic database structure cannot solve these problems, but it has to provide the respective concepts to enable the ecologist to treat these problems appropriately, e.g. by introducing a versioning system and facilitating links to specimen compilations.

A generic structure for plant trait databases

As a consequence of the above arguments, a plant trait database needs to provide the appropriate structure to (i) characterize each trait entry in detail, which is necessary due to the heterogeneity and relatively little standardization of plant traits, and (ii) place it in its specific biotic and abiotic context, accommodating ancillary data, the degree of relatedness of different measurements, inherent hierarchical structures and taxonomic specifications.

CHARACTERIZING TRAIT AND ANCILLARY DATA AS MEASUREMENTS ON SPECIFIC OBJECTS

Despite their heterogeneity, all plant trait data can be characterized as being measurable characteristics of specific objects: e.g. the length of a leaf or the height of a plant (cf. Madin *et al.* 2007). This is also true for ancillary data, like latitude and longitude of a location or the name of the person that has conducted the measurements. In this context, even the taxonomic classification can be addressed as measurable characteristics of specific objects: e.g. '*Quercus robur* L.' is the binomial expression of the characteristic 'species', like 'tree' is an expression of the categorical trait 'growth form'. In terms of data structure, there is hence no principle difference between trait data and ancillary data, including the taxonomic specification, and we

propose to treat them identically as measurements of specific objects (cf. Madin *et al.* 2007).

MEASUREMENTS ARE AGGREGATED TO OBSERVATIONS

All measurements that have been taken on the same object for the same time are directly related to each other. We consider this aspect of different measurements being 'related to the same object and time' as the most important relationship among traits and between traits and ancillary data (the identity principle). We therefore propose to directly keep track of this relationship in the database and link all individual measurements taken at the same time on the same object to a unique 'observation' identifier.

In accordance with the hierarchical structure of traits and ancillary data we propose observations to be hierarchically nested, and influences on a higher level of the hierarchy, like stand, are propagated along the hierarchy to the lower levels, like individual leaf and cells (Fig. 2). Due to this hierarchically nested structure, different observations provide context for each other, and thus facilitate the comprehensive description of abiotic and biotic environmental conditions.

THE DATABASE STRUCTURE

The enfolding database structure is characterized by two key aspects: 'measurement' and 'observation' (Fig. 3). Measurement integrates all information directly related to a specific measurement, like name of trait or ancillary data, measurement standard, value, unit and precision. Relating all of this information to a measurement facilitates the detailed characterization of each database entry of a trait or ancillary data. The aggregation of different measurements to observations facilitates the realization of the most important relationship between traits and ancillary data: 'being related to the same object in time'. Finally, observations are hierarchically nested, which facilitates the comprehensive characterization of the abiotic and biotic context of each measurement, accounting for the degree of relatedness.

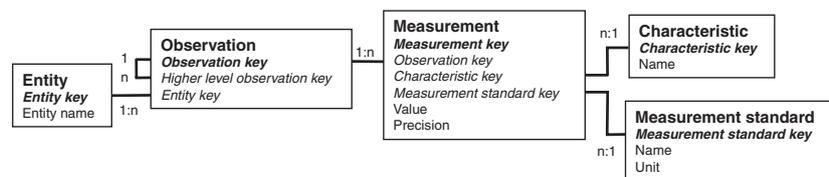


Fig. 3. Core tables (boxes) and relationships (connectors) of the proposed generic structure for plant trait databases: a dimensional data model realized in a relational database. 'Observation' is the central table of this conceptual framework, indicated by the 1:n (one to many) relationship between Observation and Measurement table: each observation can be characterized by measurements in n dimensions (traits and ancillary data). Each measurement is characterized by a value, precision (if appropriate), characteristic, and a measurement standard. Traits and ancillary data are defined in the Characteristic table. Measurements on the same object in time are aggregated to an observation. Observations are embedded in a hierarchy of observations on different levels, which is realized via the link within the observation table. Each Observation is related to a real world object, here called entity. No specification into traits, ancillary data or taxonomy here. In this generic representation of the database structure, all of these data are identically treated as measurable characteristics of specific objects. Bold italics: primary key (uniquely identifies each row in a table); italics: foreign keys (a key stored in one table which refers to a primary key in another table, used to establish a relationship between two tables); plain text: data entries.

Due to the aggregation of different measurements to observations and their hierarchical nesting, observation becomes the central element of the database structure. Such a centralized structure is called a 'dimensional' database, because the central element is characterized by attributes in different dimensions (Kimball & Ross 2002). In our case, the central element 'observation' is characterized by measurements of traits and ancillary data, where each trait and each kind of ancillary data represent a dimension, such that this results in a database structure with n dimensions. Dimensional databases with one central element are called 'star-schema' (Kimball & Ross 2002).

Realized as a relational database, the dimensional structure is appropriate to compile diverse data, like plant traits and their ancillary data, because a relational database is highly flexible and can be easily adapted to the required level of complexity (e.g. adding new dimensions or subcentres to characterize the measurements). Finally, the proposed generic structure for plant trait databases can be characterized as a hierarchically nested relational-dimensional database structure.

Two example databases using the generic database structure

The proposed generic database structure, designed to be a blueprint for the development of plant trait databases for individual purposes, has so far been used in the development of three plant trait databases. Here, we will present two contrasting databases – the Functional Ecology of Trees (FET) database and the TRY database – and illustrate how the generic database structure ensures data interoperability.

The FET⁴ database (Fig. 4) was collaboratively developed to address questions in the field of comparative trait research, functional biodiversity, and biogeochemical and demographic modelling. It was designed to compile data for a wide range of tree physiological, morphological, anatomical, and life-history traits accompanied by a wide range of ancillary data related to several hierarchical levels of organization (organ, individual, forest stand, site), with the goal to disentangle phylogenetic, environmental, and disturbance-related influences on plant trait values (Kattge *et al.* 2009; Kutsch *et al.* 2009). Data are primarily derived from the literature (data summaries) and direct experiments (raw data). The database supports the standardization of numerical and textual data during manual data entry.

The TRY⁵ database (Fig. 5) is a communal effort to merge information about ecological traits of plant species from differ-

ent existing datasets at a global scale (Lavorel *et al.* 2007). The specific traits, ancillary data and formats of data to be compiled within the TRY database were not known during the development of the database. The objective was therefore to design a database structure that is flexible enough to compile data for any plant trait and ancillary data contributed in any format without compromising the integrity of the contributed data.

SPECIFIC ASPECTS OF THE TWO DATABASES

The FET database provides a *modular organization* of templates (groups of tables with preset formats) to compile trait and context-specific data. Each dimension of the traits and their context is characterized by a separate template. The modular organization conserves the transparency of the database and allows for exchanging or adding of templates. For example, it is easy to add a new trait with its own set of trait-specific ancillary data. The users can therefore easily adapt the database according to their specific needs and it still is compatible with other users' databases. The modular organization facilitates highly standardized data input via specific entry forms. These entry forms support the consistent design of experiments and extraction of data from literature sources and facilitate the standardized manual input of data into the FET database. An *inbuilt hierarchical structure* promotes the quantification of plant traits or ancillary data at the prescribed levels of organs, individuals, stands and sites (Figs 1 and 5 and Appendix S2). There are sets of ancillary data at each of these levels, which may be continuous variables or binary, categorical, or ordinal variables, in which case, the possible entries are predefined.

The major requirement of the TRY database was to respect the integrity of the contributed data, such that they can be reproduced from the database in their originally contributed form. This requirement was addressed through the use of an automated algorithm for importation of the original data values with minimal manual interference. By assembling numerous databases, the TRY project revealed a vast heterogeneity of the contributed trait information. For example, up to six different units were provided per trait; 138 different categories were provided for the categorical trait plant growth form; leaf toughness was provided as three different measures and as a categorical trait; and, identical traits come with different names (e.g. specific leaf area and leaf mass area). The database structure proved to be appropriate to deal with this heterogeneity and to support its standardization.

TREATMENT OF TAXONOMIC SPECIFICATION

A separate 'Species' table is used in FET and TRY to organize the taxonomy. This table contains fields for the 'original species name' and 'accepted species' identifier. An entry in the field 'original species name' consists of the binomial species name, which should be accompanied by the authority and, if possible, a reference to the regional flora. The original species name is on the one hand linked via the observation table to the data source, which is either a literature source or a specimen, and on

⁴The FET database is being collaboratively developed by the Organismic Biogeochemistry group at the Max-Planck-Institute for Biogeochemistry in Jena (Germany), the Department of Special Botany and Functional Biodiversity Research at the University of Leipzig (Germany) and the Ogle-Labat the University of Wyoming, Laramie (USA).

⁵The TRY initiative (try-db.org) has been developed under the framework of the International Geosphere-Biosphere Programme (IGBP) in the context of the Fast-Track Initiative 'Refining Plant Functional Classifications for Earth System Modelling' (<http://www.igbp.kva.se/page.php?pid=369>).

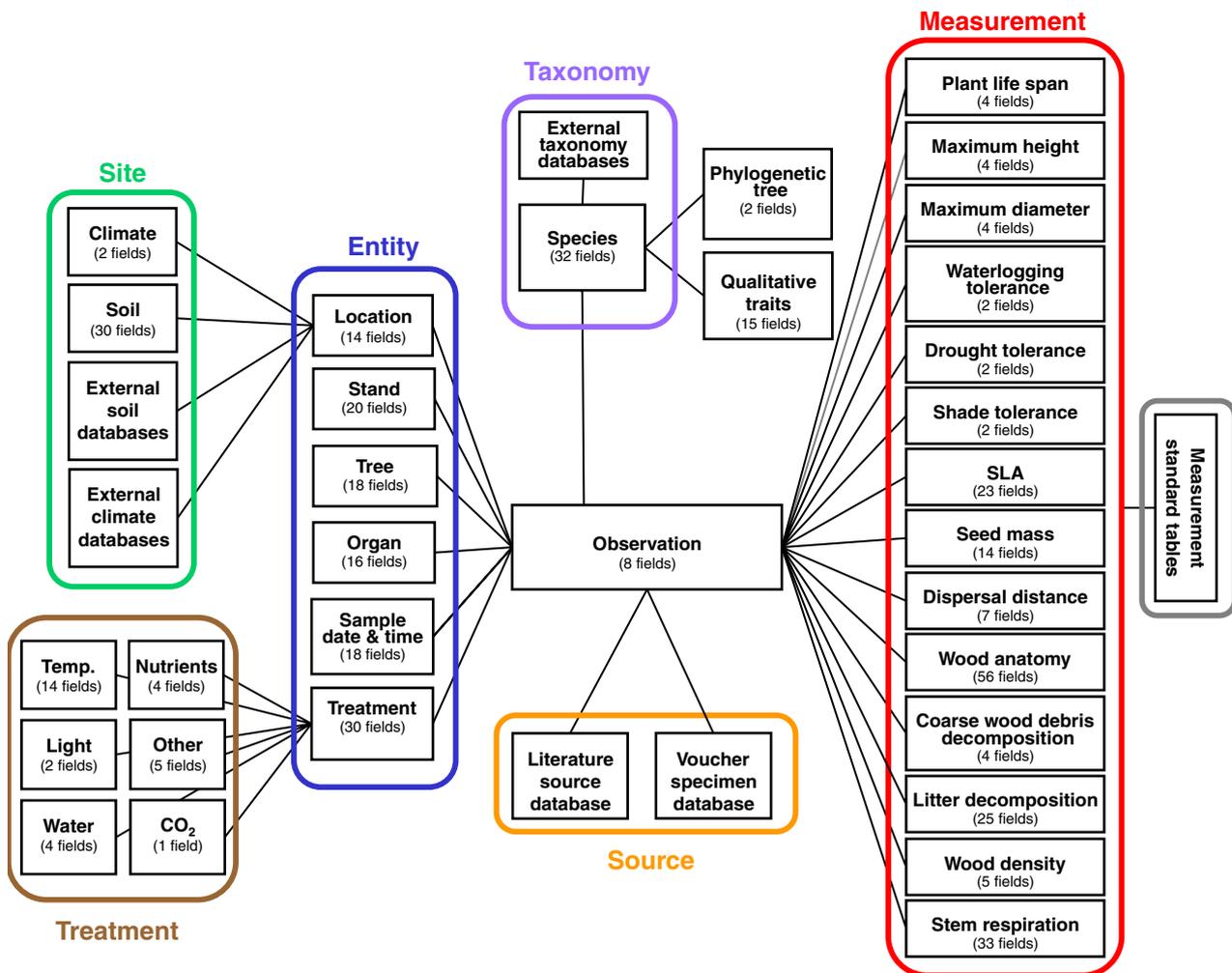


Fig. 4. Simplified relationships in the FET database. All information is linked through the central Observation table. Measurements are separated into traits (Trait data and Measurement standard tables) and tables providing information about the context of trait measurements with an inbuilt hierarchical structure (Entity, Site, Treatment, Taxonomy/phylogeny, and Source information). For each trait value entered in the trait measurement tables, information is entered, when available, on study site details, manipulative treatments and other study/sampling information. The trait data are linked to this information and a species, which is linked to external taxonomy datasets, to a position in the phylogenetic relationships, and to categorical traits that are invariable within species. Each box represents a core table in the database, and the current numbers of fields in each table are indicated.

the one hand to an accepted species identifier, which is linked to an official plant name index, e.g. IPNI.⁶ Conserving the originally contributed species name and linking it to accepted species lists ensures that the species classification can be adapted to changes of the official taxonomy without losing the original information.

INTEROPERABILITY OF DIFFERENT DATASETS

Interoperability ensures that information – data and their relationships – from different datasets can seamlessly be merged for subsequent integrated analysis. Ontologies provide a formal mechanism for defining terms and their relationships using mathematical logic, to facilitate the interoperability of different

datasets. One ontology scheme that is gathering support in ecology and environmental science is the Extensible Observation Ontology (OBOE, see Appendix S3), developed by the SEEK project.⁷ Like in the generic database structure presented here, in OBOE the information content of observational data is structured along ‘Measurement’ and ‘Observation’: measurements are aggregated to observations, which in turn are hierarchically nested. The generic database structure proposed here and the two databases presented are therefore consistent with this general structure of the OBOE ontology scheme, which facilitates interoperability, the seamless merging of disparate information from different databases into an integrated form for analysis (details see Appendix S4).

⁶IPNI: International Plant Names Index, <http://www.ipni.org>.

⁷Science Environment for Ecological Knowledge (SEEK): <http://seek.ecoinformatics.org>.

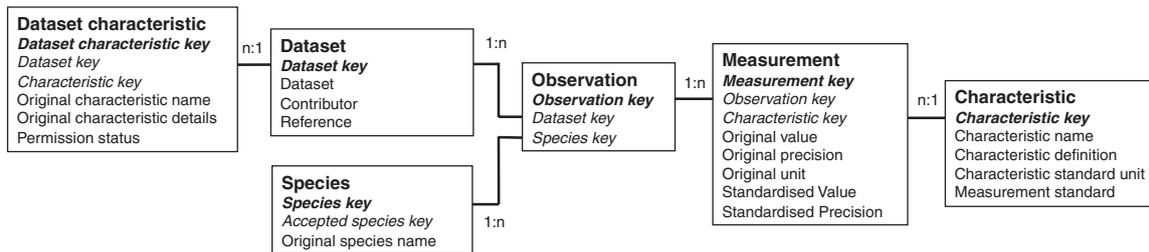


Fig. 5. Core tables, relationships and data entry-types of the TRY database (each box represents a table). Each observation is characterized by one to several measurements, the respective dataset and the name of a species. Each measurement is linked to a characteristic (either trait or ancillary data). The measurement standard is specified in the Characteristic table. In the DataSet table each contributed dataset is at least characterized by its name and the names of contributors. The additional DataSet Characteristic table facilitates the import of the original name of each trait and ancillary data, and a specific characterization of measurement details for each dataset. The automated import of contributed data as original entries realizes data integrity (original value, original precision, original unit, original species name and original characteristic name). Standardized values and species names are added to the original entries.

Discussion

In this article, we analysed the inherent structure of plant trait data to determine an appropriate database structure for capturing existing and future plant trait data. This entity-relationship analysis uncovered several key aspects that must be considered to develop successfully an appropriate and interoperable database structure: plant traits are a heterogeneous group of data with a relatively low degree of standardization; a wide range of equally heterogeneous ancillary data are required to characterize the biotic and abiotic environment and the measurement methods; it is important to keep track of object identities in order to specify the ‘degree of relatedness’ relevant for the analysis of trait correlation, of the hierarchical structure of biological systems and of the taxonomy, which is constantly under revision. The resulting generic database structure for plant trait data followed three principles: (i) traits and ancillary data are treated identically as measurements of objects, (ii) measurements are aggregated to observations, and (iii) observations are hierarchically nested along the hierarchy of biological integration. These principles resulted in a hierarchically nested relational-dimensional database structure. Finally, we illustrate the realization of this structure with two plant trait databases developed for different purposes, and we show that the new structure is consistent with the existing ontology schemes in ecology, which facilitate interoperability among plant traits and other ecological and environmental databases. The realized databases may be very detailed (e.g. FET) or comparatively simple (e.g. TRY). However, the underlying database structure is straightforward and is based on a dimensional data model that facilitates easy and unambiguous data entry, queries, and data quality assurance (Kimball & Ross 2002).

The database software needs not be specified, because the structure can be implemented in any software that facilitates the construction of a relational database. The appropriate software depends on user preference and application, and will probably depend on whether the database shall facilitate Internet-based user access. The expected size of the database will in most cases not be an issue, as all common database programmes provide sufficient storage capacity for plant trait

databases. Compared to other ecological data streams, the quantity of plant trait and ancillary data to be stored is relatively small (e.g. 750 megabytes in the case of the TRY data base, realized in Microsoft Office Access 2007, version 12.0.6535.5005; Microsoft Corporation, Redmond, Washington, USA), because the individual trait measurements are often manual, and therefore relatively time-consuming and expensive. This is in contrast to other ecological databases, e.g. remote sensing or eddy covariance data, which contain data that are sampled automatically and may be weighed in the order of giga- to terabytes.

We are not aware of any other trait database that simultaneously accounts for (i) different measurements being taken on the same object and (ii) the hierarchical nesting of information. Two-dimensional spreadsheet datasets often account for different measurements being taken on the same object (observation), which obviously has been identified by ecologists as being a pivotal characteristic of plant trait data. Yet, spreadsheet datasets are limited with respect to the number of different traits and ancillary data, and the representation of the hierarchical structure. During the transition from two-dimensional spreadsheets to a relational database the characteristic ‘being measured on the same object’ gets lost if the relational database does not provide the respective structure.

IDENTICAL TREATMENT OF TRAITS, TAXONOMY AND ANCILLARY DATA

The proposed generic structure for plant trait databases is not defined explicitly in terms of plant traits, taxonomy and ancillary data, but it is based on measurements of objects. Traits and covariates are only defined *a posteriori* as specific characteristics of the measurements. This is a major advantage because only in some cases are traits well defined (Cornelissen *et al.* 2003), while in most cases their definition is either vague or deliberately modified to address a specific scientific question. For example, in the literature, leaf nitrogen content could be reported as the average leaf nitrogen content of all leaves of an individual plant, the nitrogen content of a single leaf, the nitrogen content of a single leaf per dry mass or leaf area, the nitrogen content per leaf area of a sun-exposed leaf, or the nitrogen

content per leaf area of a sun-exposed leaf during the peak growth period. The compilation of the individual measurements together with key ancillary data exactly describing the measurement provides the opportunity to make this decision and select the relevant data *a posteriori*, as dictated by the ecological question and subsequent statistical analysis (see also Appendix S5: The representation of parameter-based traits). The transparent aggregation of measurements to observations and the hierarchical structure of observations avoids duplicating information and allows tracking post-sampling data processing (Ellison *et al.* 2006).

Even the treatment of taxonomy fits into this concept of 'being a measurement of a specific object', although the classification systems are not always unambiguous (e.g. synonymy), different classification systems exist in parallel and are overlapping (regional floras, global name indices), and these classification systems are constantly changing (Berendsohn & Geoffroy 2007). Here, the compilation of the original species name related to an accepted species name has proven convenient and flexible. The original names are unambiguously linked via the Observation table to the data source (literature source, specimen), while the accepted names are linked to 'official' lists of species names, and thus make use of the treatment of synonyms in these lists and are able to follow changes. Dealing with species information in a few specific tables is a major advantage over using species names independently in several separate spreadsheets. Thus, the database structure supports the adaptation of taxonomy within the database to keep track with changes in external taxonomy sources.

DEFINITION OF OBSERVATION: WEAKNESS OR STRENGTH?

The central element of the proposed database structure is the observation. An observation is defined by measurements on the same object in space and time. Deciding which measured values belong to the same observation is flexible and can be subjective. The decision to be made is: what is to be considered an object in space and time? Two measurements on one leaf may be considered to belong to the same or different observations, depending on the perspective of the researcher (they will still be related on a higher level of the hierarchy). This subjectivity does not present a weakness but a strength of this approach, because the decision of what is to be considered a group of information that belongs to the same observation is most vivid at the time of data acquisition. Thus, the aggregation of measurements to observations already constitutes a knowledge component (Baumeister *et al.* 2007) stored in the database and ready to be reused by later projects.

Due to the separation of measurements, the aggregation of measurements to observations and the hierarchical arrangement of observations, the generic database structure realized in an relational database is extremely flexible, and facilitates the consistent compilation of data on higher levels of the biological organization, e.g. community-level data, in combination with plant trait data. This flexibility in consistence with the major ontology schemes (e.g. OBOE) makes the generic structure not

only appropriate for plant trait databases, but also applicable in other contexts (e.g. databases to compile data for scientific projects in general), where different kinds of data are to be compiled in combination with several ancillary data. First applications of the generic database structure for such project databases are currently being tested.

Conclusions and perspectives

Based on a comprehensive examination of plant traits with respect to data compilation, we have developed a generic dimensional database structure that follows three key principles: (i) traits and ancillary data are identically treated as measurements of specific objects; (ii) measurements related to the same object and time are aggregated to observations; and (iii) observations are hierarchically nested from organ to ecosystem. This database structure is consistent with main ontology frameworks (e.g. OBOE) that are currently being developed in ecology for improving data interoperability among research efforts. We illustrate the over-arching utility of the proposed database structure using two independent plant trait database projects. The generic database structure will serve as a flexible blueprint for future plant trait databases, improving data discovery, and ensuring compatibility among them.

Acknowledgements

The authors wish to thank Jens Nieschulze and Brian McGill for essential input in the context of eco-informatics, the anonymous referees and the handling editor for valuable comments that helped to substantially improve the manuscript. K.O. and K.S. were supported by National Science Foundation (NSF) grants awarded to K.O. in 2003 and 2006 (#0630119) and by NSF grant #EPS-0447681. The FET database was supported by the German Science Foundation (DFG) through the BEAM project of C.W. within the Biodiversity Exploratories. The development of the TRY database was supported by IGBP, GLP, DIVERSITAS, QUEST and the French GIS Climate Environment and Society consortium.

References

- Ackerly, D.D. & Cornwell, W.K. (2007) A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecology Letters*, **10**, 135–145.
- Albert, C., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P. & Lavorel, S. (2010) Intraspecific functional variability: extent, structure and sources of variation within a French alpine catchment. *Journal of Ecology*, **98**, 604–613.
- Barboni, D., Harrison, S.P., Bartlein, P.J., Jalut, G., New, M., Prentice, I.C., Sanchez-Goni, M.F., Spessa, A., Davis, B. & Stevenson, A.C. (2004) Relationships between plant traits and climate in the Mediterranean region: a pollen data analysis. *Journal of Vegetation Science*, **15**, 635–646.
- Baumeister, J., Reutelschöfer, J., Nadrowski, K. & Misok, A. (2007) *Using Knowledge Wikis to Support Scientific Communities*. SCOOP'07: Proceedings of 1st Workshop on Scientific Communities of Practice, Bremen, Germany.
- Berendsohn, W.G. & Geoffroy, M. (2007) Networking taxonomic concepts – uniting without "unitary-ism". *Biodiversity Databases – Techniques, Politics, and Applications* (eds G. Curry & C. Humphries), pp. 13–22. CRC Taylor & Francis, Baton Rouge, LA.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.
- Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G. & Zanne, A.E. (2009) Towards a world wide wood economics spectrum. *Ecology Letters*, **12**, 351–366.

- Chen, P.P.-S. (1976) The entity-relationship model – toward a unified view of data. *ACM Transactions of Database Systems*, **1**, 9–36.
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, J.R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H. & Wear, D. (2001) Ecological forecasts: an emerging imperative. *Science*, **293**, 657–660.
- Cornelissen, J.H.C., Lavorel, S., Garnier, E., Diaz, S., Buchmann, N., Gurvich, D.E., Reich, P.B., ter Steege, H., Morgan, H.D., van der Heijden, M.G.A., Pausas, J.G. & Poorter, H. (2003) A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, **51**, 335–380.
- Cornwell, W.K., Cornelissen, J.H.C., Amatangelo, K., Dorrepaal, E., Eviner, V.T., Godoy, O., Hobbie, S.E., Hoorens, B., Kurokawa, H., Pérez-Harguindeguy, N., Queded, H.M., Santiago, L.S., Wardle, D.A., Wright, I.J., Aerts, R., Allison, S.D., van Bodegom, P., Brovkin, V., Chatain, A., Callaghan, T.V., Diaz, S., Garnier, E., Gurvich, D.E., Kazakou, E., Klein, J.A., Read, J., Reich, P.B., Soudzilovskaia, N.A., Vaieretti, M.V. & Westoby, M. (2008) Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. *Ecology Letters*, **11**, 1065–1071.
- Diaz, S., Lavorel, S., McIntyre, S., Falczuk, V., Casanoves, F., Milchunas, D.G., Skarpe, C., Rusch, G., Sternberg, M., Noy-Meir, I., Landsberg, J., Zhang, W., Clark, H. & Campbell, B.D. (2007) Plant trait responses to grazing – a global synthesis. *Global Change Biology*, **13**, 313–341.
- Diaz, S., Hodgson, J.G., Thompson, K., Cabido, M., Cornelissen, J.H.C., Jalili, A., Montserrat-Martí, G., Grime, J.P., Zarrinkamar, F., Asri, Y., Band, S.R., Basconcelo, S., Castro-Diez, P., Funes, G., Hamzehee, B., Khoshnevi, M., Pérez-Harguindeguy, N., Pérez-Rontomé, M.C., Shirvany, F.A., Vendramini, F., Yazdani, S., Abbas-Azimi, R., Bogaard, A., Boustani, S., Charles, M., Dehghan, M., de Torres-Espuny, L., Falczuk, V., Guerrero-Campo, J., Hynd, A., Jones, G., Kowsary, E., Kazemi-Saeed, F., Maestro-Martínez, M., Romo-Díez, A., Shaw, S., Siavash, B., Villar-Salvador, P. & Zak, M.R. (2004) The plant traits that drive ecosystems: evidence from three continents. *Journal of Vegetation Science*, **15**, 295–304.
- Ellison, A.M., Osterweil, L.J., Clarke, L., Hadley, J.L., Wise, A., Boose, E., Foster, D.R., Hanson, A., Jensen, D., Kuzeja, P., Riseman, E. & Schultz, H. (2006) Analytic webs support the synthesis of ecological datasets. *Ecology*, **87**, 1345–1358.
- Enquist, B.J., Kerkhoff, A.J., Stark, S.C., Swenson, N.G., McCarthy, M.C. & Price, C.A. (2007) A general integrative model for scaling plant growth, carbon flux, and functional trait spectra. *Nature*, **449**, 218–222.
- Fitter, A.H. & Peat, H.J. (1994) The ecological flora database. *Journal of Ecology*, **82**, 415–425.
- Garnier, E., Lavorel, S., Ansquer, P., Castro, H., Cruz, P., Dolezal, J., Eriksson, O., Fortunel, C., Freitas, H., Golodets, C., Grigulis, K., Jouany, C., Kazakou, E., Kigel, J., Kleyer, M., Lehsten, V., Lepš, J., Meier, T., Pakeman, R., Papadimitriou, M., Papanastasis, V.P., Queded, H., Quéfier, F., Robson, M., Roumet, C., Rusch, G., Skarpe, C., Sternberg, M., Theau, J.-P., Thébault, A., Vile, D. & Zarovani, M.P. (2007) Assessing the effects of land-use change on plant traits, communities and ecosystem functioning in grasslands: a standardized methodology and lessons from an application to 11 European sites. *Annals of Botany*, **99**, 967–985.
- Green, W. (2009) *USDA PLANTS Compilation, Version 1, 09-02-02*. Available at: <http://bricol.net/downloads/data/PLANTSdatabase/NRCS:ThePLANTSDatabase>. Available at: <http://plants.usda.gov> (accessed on 1 February 2009). National Plant Data Center, Baton Rouge, LA, USA.
- Grime, J.P., Thompson, K., Hunt, R., Hodgson, J.G., Cornelissen, J.H.C., Rorison, I.H., Hendry, G.A.F., Ashenden, T.W., Askew, A.P., Band, S.R., Booth, R.E., Bossard, C.C., Campbell, B.D., Cooper, J.E.L., Davison, A.W., Gupta, P.L., Hall, W., Hand, D.W., Hannah, M.A., Hillier, S.H., Hodgkinson, D.J., Jalili, A., Liu, Z., Mackey, J.M.L., Matthews, N., Mowforth, M.A., Neal, A.M., Reader, R.J., Reiling, K., Ross-Frazer, W., Spencer, R.E., Sutton, F., Tasker, D.E., Thorpe, P.C. & Whitehouse, J. (1997) Integrated screening validates primary axes of specialisation in plants. *Oikos*, **79**, 259–281.
- Kahmen, S., Poschold, P. & Schreiber, K.F. (2002) Conservation management of calcareous grasslands. Changes in plant species composition and response of functional traits during 25 years. *Biological Conservation*, **104**, 319–328.
- Kattge, J. & Knorr, W. (2007) Temperature acclimation in a biochemical model of photosynthesis: a reanalysis of data from 36 species. *Plant, Cell and Environment*, **30**, 1176–1190.
- Kattge, J., Knorr, W., Raddatz, T. & Wirth, C. (2009) Quantifying photosynthetic capacity and nitrogen use efficiency for earth system models. *Global Change Biology*, **15**, 976–991.
- Kimball, R. & Ross, M. (2002) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2nd edn. John Wiley and Sons, Inc., New York.
- Kleyer, M., Bekker, R.M., Knevel, I.C., Bakker, J.P., Thompson, K., Sonnenschein, M., Poschold, P., van Groenendael, J.M., Klimes, L., Klimesova, J., Klotz, S., Rusch, G.M., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Dannemann, A., Endels, P., Götzenberger, L., Hodgson, J.G., Jackel, A.-K., Kühn, I., Kunzmann, D., Ozinga, W.A., Römermann, C., Stadler, M., Schlegelmilch, J., Steendam, H.J., Tackenberg, O., Wilmann, B., Cornelissen, J.H.C., Eriksson, O., Garnier, E. & Peco, B. (2008) The LEDA Trait-base: a database of life-history traits of the Northwest European flora. *Journal of Ecology*, **96**, 1266–1274.
- Klotz, S., Kühn, I. & Durka, W. (2002) *BIOLFLOR – Search and Information System on Vascular Plants in Germany*. Umweltforschungszentrum Leipzig-Halle GmbH, Leipzig.
- Knevel, I.C., Bekker, R.M., Bakker, J.P. & Kleyer, M. (2003) Life-history traits of the Northwest European flora: the LEDA database. *Journal of Vegetation Science*, **14**, 611–614.
- Kutsch, W.L., Wirth, C., Kattge, J., Noellert, S., Herbst, M. & Kappen, L. (2009) Ecophysiological characteristics of large, mature trees and consequences for old-growth functioning. *Old-growth Forests: Function, Fate and Value. Ecological Studies*, Vol. 207 (eds C. Wirth, G. Gleixner & M. Heimann), pp. 57–79. Springer, New York/Berlin/Heidelberg.
- Lavorel, S. & Garnier, E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16**, 545–556.
- Lavorel, S., Diaz, S., Prentice, I.C. & Leadley, P. (2007) Refining plant functional classifications for earth system modeling. *Global Land Project (GLP) Newsletter*, **2**, 14.
- Lavorel, S., Gachet, S., Sahl, A., Colace, M.-P., Gaucherand, S., Burylo, M. & Bonet, R. (2009) A plant functional traits database for the Alps – understanding functional effects of changed grassland management. *Data Mining for Global Trends in Mountain Biodiversity* (eds C. Körner & E. Spehn), pp. 106–123. CRC Press/Taylor and Francis, Boca Raton, FL.
- Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D. & Villa, F. (2007) An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, **2**, 279–296.
- Madin, J.S., Bowers, S., Schildhauer, M.P. & Jones, M.B. (2008) Advancing ecological research with ontologies. *Trends in Ecology & Evolution*, **23**, 159–168.
- McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, **21**, 178–185.
- Medlyn, B.E. & Jarvis, P.G. (1999) Design and use of a database of model parameters from elevated CO₂ experiments. *Ecological Modelling*, **124**, 69–83.
- Messier, J., McGill, B.J. & Lechowicz, M.J. (2010) How do traits vary across ecological scales? A case for trait-based ecology. *Ecology Letters*, **13**, 838–848.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. & Stafford, S.G. (1997) Non-geospatial metadata for the ecological sciences. *Ecological Applications*, **7**, 330–342.
- Moles, A.T., Ackerly, D.D., Webb, C.O., Tweddle, J.C., Dickie, J.B., Pitman, A.J. & Westoby, M. (2005) Factors that shape seed mass evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10540–10544.
- Moorcroft, P.R., Hurtt, G.C. & Pacala, S.W. (2001) A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*, **71**, 557–585.
- Ogle, K. & Barber, J.J. (2008) Bayesian data-model integration in plant physiological and ecosystem ecology. *Progress in Botany*, **69**, 281–311.
- Ogle, K. & Pacala, S.W. (2009) A modeling framework for inferring tree growth and allocation from physiological, morphological, and allometric traits. *Tree Physiology*, **29**, 587–605.
- Ollinger, S.V., Richardson, A.D., Martin, M.E., Hollinger, D.Y., Frolking, S.E., Reich, P.B., Plourde, L.C., Katul, G.G., Munger, J.W., Oren, R., Smith, M.-L., Paw, K.T., Bolstad, P.V., Cook, B.D., Day, M.C., Martin, T.A., Monson, R.K. & Schmid, H.P. (2008) Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19336–19341.
- Ollivier, M., Baraloto, C. & Marcon, E. (2007) A trait database for Guianan rain forest trees permits intra- and inter-specific contrasts. *Annals of Forest Science*, **64**, 781–786.

- Paula, S., Arianoutsou, M., Kazanis, D., Tavsanoğlu, Ç., Lloret, F., Buhk, C., Ojeda, F., Luna, B., Moreno, J.M., Rodrigo, A., Espelta, J.M., Palacio, S., Fernández-Santos, B., Fernandes, P.M. & Pausas, J.G. (2009) Fire-related traits for plant species of the Mediterranean Basin. *Ecology*, **90**, 1420.
- Poorter, H., Niinemets, Ü., Walter, A., Fiorani, F. & Schurr, U. (2010) A method to construct dose–response curves for a wide range of environmental factors and plant traits by means of a meta-analysis of phenotypic data. *Journal of Experimental Botany*, **61**, 2043–2055.
- Poschlod, P., Kleyer, M., Jackel, A.K., Dannemann, A. & Tackenberg, O. (2003) BIOPOP – a database of plant traits and Internet application for nature conservation. *Folia Geobotanica*, **38**, 263–271.
- Reich, P.B., Wright, I.J. & Lusk, C.H. (2007) Predicting leaf physiology from simple plant and climate attributes: a global GLOPNET analysis. *Ecological Applications*, **17**, 1982–1988.
- Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M. & Walters, M.B. (2003) The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences*, **164**, 143–164.
- Reich, P.B., Tjoelker, M.G., Machado, J.-L. & Oleksyn, J. (2006) Universal scaling of respiratory metabolism, size and nitrogen in plants. *Nature*, **439**, 457–461.
- Reich, P.B., Tjoelker, M.G., Pregitzer, K.S., Wright, I.J., Oleksyn, J. & Machado, J.L. (2008) Scaling of respiration to nitrogen in leaves, stems and roots of higher land plants. *Ecology Letters*, **11**, 793–801.
- Robinson, W.S. (1950) Ecological correlations and the behavior of individuals. *American Sociological Review*, **15**, 351–357.
- Royer, D.L., Sack, L., Wilf, P., Lusk, C.H., Jordan, G.J., Niinemets, U., Wright, I.J., Westoby, M., Cariglino, B., Coley, P.D., Cutter, A.D., Johnson, K.R., Labandeira, C.C., Moles, A.T., Palmer, M.B. & Valladares, F. (2007) Fossil leaf economics quantified: calibration, Eocene case study, and implications. *Paleobiology*, **33**, 574–589.
- Scheiter, S. & Higgins, S.I. (2009) Impacts of climate change on the vegetation of Africa: an adaptive dynamic vegetation modelling approach. *Global Change Biology*, **15**, 2224–2246.
- Scholes, R.J., Mace, G.M., Turner, W., Geller, G.N., Jurgens, N., Larigauderie, A., Muchoney, A., Walther, B.A. & Mooney, H.A. (2008) Toward a global biodiversity observing system. *Science*, **22**, 1044–1045.
- Schurr, F.M., Bond, W.J., Midgley, G.F. & Higgins, S.I. (2005) A mechanistic model for secondary seed dispersal by wind and its experimental validation. *Journal of Ecology*, **93**, 1017–1028.
- Sperry, J.S., Mainzer, F.C. & McCulloh, K.A. (2008) Safety and efficiency conflicts in hydraulic tecture: scaling from tissues to trees. *Plant, Cell and Environment*, **31**, 632–654.
- Swenson, N.G. & Enquist, B.J. (2007) Ecological and evolutionary determinants of a key plant functional trait: wood density and its community-wide variation across latitude and elevation. *American Journal of Botany*, **94**, 451–459.
- Swenson, N. & Weiser, M. (2010) Plant geography upon the basis of functional traits: an example from eastern North American trees. *Ecology*, **91**, 2234–2241.
- Tackenberg, O. & Stocklin, J. (2008) Wind dispersal of alpine plant species: a comparison with lowland species. *Journal of Vegetation Science*, **19**, 109–118.
- Valladares, F., Gianoli, E. & Gomez, J.M. (2007) Ecological limits to plant phenotypic plasticity. *New Phytologist*, **176**, 749–763.
- Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E. (2007) Let the concept of trait be functional! *Oikos*, **116**, 882–892.
- Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A. & Wright, I.J. (2002) Plant ecological strategies: some leading dimensions of variation between species. *Annual Review of Ecology and Systematics*, **33**, 125–159.
- White, M.A., Thornton, P.E., Running, S.W. & Nemani, R.R. (2000) Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls. *Earth Interactions*, **4**, 1–85.
- Wirth, C. (2005) Fire regime and tree diversity in boreal forests: implications for the carbon cycle. *Forest Diversity and Function: Temperate and Boreal Systems*, **176**, 309–344.
- Wirth, C. & Lichstein, J.W. (2009) The imprint of succession on old-growth forest carbon balances insights from a trait-based model of forest dynamics. *Old-growth Forests: Function, Fate and Value* (eds C. Wirth, G. Gleixner & M. Heimann), pp. 81–113. Springer, New York/Berlin/Heidelberg.
- Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M.L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J. & Villar, R. (2004) The worldwide leaf economics spectrum. *Nature*, **428**, 821–827.

Received 23 April 2010; accepted 17 August 2010

Handling Editor: Robert P. Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Data base systems commonly used by ecologists to compile plant trait data.

Appendix S2. The inbuilt hierarchical structure of the FET database.

Appendix S3. OBOE ontology scheme.

Appendix S4. Interoperability – the seamless merging of disparate information from different databases into an integrated form for analysis.

Appendix S5. The representation of parameter-based traits.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.