Journal of Ecology

Journal of Ecology 2017, **105**, 298–309 doi: 10.1111/1365-2745.12698

Towards a thesaurus of plant characteristics: an ecological contribution

Eric Garnier*[,1,2], Ulrike Stahl[3,4,5], Marie-Angélique Laporte[1,4,6], Jens Kattge[3,4], Isabelle Mougenot[7], Ingolf Kühn[4,5,8], Baptiste Laporte[2], Bernard Amiaud[9,10], Farshid S. Ahrestani[11,12], Gerhard Bönisch[3], Daniel E. Bunker[13], J. Hans C. Cornelissen[14], Sandra Díaz[15], Brian J. Enquist[16], Sophie Gachet[17], Pedro Jaureguiberry[15], Michael Kleyer[18], Sandra Lavorel[19], Lutz Maicher[20,21], Natalia Pérez-Harguindeguy[15], Hendrik Poorter[22], Mark Schildhauer[23], Bill Shipley[24], Cyrille Violle[1], Evan Weiher[25], Christian Wirth[4,26], Ian J. Wright[27] and Stefan Klotz[5]

[1]Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175), CNRS – Université de Montpellier – Université Paul-Valéry Montpellier – EPHE, 34293 Montpellier Cedex 5, France; [2]CEntre for the Synthesis and Analysis of Biodiversity (CESAB-FRB), 13100 Aix-en-Provence, France; [3]Max-Planck-Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07743 Jena, Germany; [4]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany; [5]Department of Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Str. 4, 06120 Halle, Germany; [6]Bioversity International, via dei Tre Denari, 174/a Maccarese, Rome, Italy; [7]UMR 228 ESPACE-DEV, Maison de la Télédétection, 34093 Montpellier, France; [8]Martin-Luther-University Halle-Wittenberg (MLU), Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany; [9]UMR 1137, Ecologie et Ecophysiologie Forestière, Université de Lorraine, 54506 Vandoeuvre-les-Nancy, France; [10]UMR 1137, Ecologie et Ecophysiologie Forestière, INRA, 54280 Champenoux, France; [11]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA; [12]Frontier Wildlife Conservation, Mumbai 400007, India; [13]Department of Biological Sciences, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA; [14]Department of Ecological Science, Vrije Universteit, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; [15]IMBIV (CONICET-UNC) & FCEFyN, Universidad Nacional de Córdoba, Av. Vélez Sarsfield 1611, X5016GCA Córdoba, Argentina; [16]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA; [17]Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), Aix Marseille Université, CNRS, IRD, Avignon Université, Campus St-Jérôme Case 421, 13397 Marseille, France; [18]Institute of Biology and Environmental Sciences, University of Oldenburg, 26111 Oldenburg, Germany; [19]Laboratoire d'Ecologie Alpine (UMR 5553), CNRS – Université Joseph Fourier, 38041 Grenoble, France; [20]Computer Science Department, University of Jena, 07743 Jena, Germany; [21]Competitive Intelligence Group, Fraunhofer MOEZ, Neumarkt 9-19, 04109 Leipzig, Germany; [22]Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany; [23]National Center for Ecological Analysis and Synthesis (NCEAS), UC Santa Barbara, Santa Barbara, CA 93101-5504, USA; [24]Département de biologie, Université de Sherbrooke Sherbrooke QC J1K 2R1, Canada; [25]Department of Biology, University of Wisconsin, Eau Claire, WI 54702-4004, USA; [26]Institute for Systematic Botany and Functional Biodiversity, University of Leipzig, Johannisallee 21, 04103 Leipzig, Germany; and [27]Department of Biological Sciences, Macquarie University, Sydney NSW 2109, Australia

Summary

1. Ecological research produces a tremendous amount of data, but the diversity in scales and topics covered and the ways in which studies are carried out result in large numbers of small, idiosyncratic data sets using heterogeneous terminologies. Such heterogeneity can be attributed, in part, to a lack of standards for acquiring, organizing and describing data. Here, we propose a terminological resource, a Thesaurus Of Plant characteristics (TOP), whose aim is to harmonize and formalize concepts for plant characteristics widely used in ecology.

2. TOP concentrates on two types of plant characteristics: traits and environmental associations. It builds on previous initiatives for several aspects: (i) characteristics are designed following the entity-quality (EQ) model (a characteristic is modelled as the 'Quality' <Q> of an 'Entity' <E>) used in

*Correspondence author. E-mail: eric.garnier@cefe.cnrs.fr

the context of Open Biological Ontologies; (ii) whenever possible, the Entities and Qualities are taken from existing terminology standards, mainly the Plant Ontology (PO) and Phenotypic Quality Ontology (PATO) ontologies; and (iii) whenever a characteristic already has a definition, if appropriate, it is reused and referenced. The development of TOP, which complies with semantic web principles, was carried out through the involvement of experts from both the ecology and the semantics research communities. Regular updates of TOP are planned, based on community feedback and involvement.

3. TOP provides names, definitions, units, synonyms and related terms for about 850 plant characteristics. TOP is available online (www.top-thesaurus.org), and can be browsed using an alphabetical list of characteristics, a hierarchical tree of characteristics, a faceted and a free-text search, and through an Application Programming Interface.

4. *Synthesis*. Harmonizing definitions of concepts, as proposed by TOP, forms the basis for better integration of data across heterogeneous data sets and terminologies, thereby increasing the potential for data reuse. It also allows enhanced scientific synthesis. TOP therefore has the potential to improve research and communication not only within the field of ecology, but also in related fields with interest in plant functioning and distribution.

Key-words: biodiversity, controlled vocabulary, ecology, informatics, ontology, plant characteristics, plant environmental association, plant trait, semantics, thesaurus

Introduction

Among the impediments that currently slow the progress of ecology towards a '*data-intensive**' or 'big data' science (Kelling *et al.* 2009; Michener & Jones 2012; Hampton *et al.* 2013; terms in italics followed by* are defined in Table 1), two appear especially salient. The first one pertains to the integrative nature of ecology, which requires combining information from multiple spatial and temporal scales, levels of organization and disciplines. The second impediment relates to the practice of ecology: while some coordinated studies carried out at large scales exist, the vast majority of ecological data are collected by researchers working independently and with little coordination among them. Together, these limitations result in the generation of numerous small data sets whose forms, contents and semantics can be highly specific to a particular research question or even researcher (Heidorn 2008; Michener & Jones 2012; Hampton *et al.* 2013).

Semantic heterogeneity is often overlooked but is a potential source of high confusion. Different data sets often have variables and *concepts** that have different meanings across disciplines, scales, levels of organization or even worse, among researchers of the same field. Also, a lack of either coordination and/or recognized terminological or *ontological** standards within ecology can limit our ability to integrate and compare data across studies. This issue stems from a failure to describe and define concepts, and results in various forms of terminological uncertainty (Herrando-Pérez, Brook & Bradshaw 2014 and references therein). Overall, semantic heterogeneity seriously impedes data integration, sharing and reuse (Kattge *et al.* 2011b; Reichman, Jones & Schildhauer 2011; Parr *et al.* 2016), and ultimately, impedes discovery and advancement of knowledge towards a unified foundation for ecological science (Madin *et al.* 2008 and references therein). The harmonization of definitions and concepts is a fundamental contribution to the emerging discipline of ecoinformatics (Jones *et al.* 2006; Michener & Jones 2012), whose long-term objective is to allow both scientists and computers to communicate more effectively with one another (Michener & Brunt 2000; Walls *et al.* 2012, 2014; Parr *et al.* 2016). The work presented here provides a common semantic resource to better integrate plant characteristics for ecology.

There is now a growing consensus that a functional approach to biodiversity has a strong potential to address many pending questions in ecology and evolution (McGill *et al.* 2006; Lavorel *et al.* 2007; Garnier & Navas 2012; Enquist *et al.* 2015; Garnier, Navas & Grigulis 2016 for detailed reviews). As in other fields of ecology, however, primary data are mostly collected by research groups working independently, while at the same time many concepts remain poorly, inconsistently, or only implicitly defined. Examples of semantic confusion in the field of functional ecology and how these can induce misunderstanding and/or mistakes are given in the next section, which demonstrate that although vocabularies and standards for particular aspects of biodiversity data do exist, these actually lack many of the necessary terms to describe the different dimensions of biodiversity, including its functional facet (see e.g. Walls *et al.* 2014 for a synthesis).

The aim of this work is to report on the development of a terminological resource for major plant characteristics used in functional ecology, entitled TOP: a *Thesaurus** *Of* *Plant characteristics**. TOP is Web-accessible, and built according to the *SKOS** data model, a recommendation of the World Wide Web Consortium (http://www.w3.org/TR/skos-reference/). In TOP, a plant characteristic is defined as 'a feature of an individual plant, plant population or plant species, describing either a plant trait or an environmental association' (Fig. 1). A *plant trait** is defined as 'any morphological, anatomical, biochemical, physiological or phenological heritable feature measurable at the individual level, from the cell to

Table 1. Glossary of selected terms and expressions used in this article

Term	Definition	Reference
Application Programming Interface (API)	A set of protocols used by programmers to create applications for a specific operating system or to interface between the different modules of an application	http://dictionary.reference.com
Common name	A name which is generally preferred and used by the community	This paper
Concept	Ideas, notions or objects and events; the units of thought; here made explicit by name, definition, URI and reference	SKOS recommandation, http://www.w3.org/TR/skos-reference/
Data-intensive science	An emerging way of conducting science as a result of the accumulation of large quantities of data, and from the need for new analysis techniques	Kelling *et al.* (2009)
Entity	Something that has a real existence (used in the entity-quality formalism)	
Environmental association	A non-random association of individual plants, plant populations or plant species with particular characteristics of the environment	This paper
Facet	A common feature shared by a set of objects	http://www.mumia-network.eu/index.php/working-groups/wg4
Faceted search	A technique for accessing a collection of information, allowing users to explore by filtering available information	http://www.mumia-network.eu/index.php/working-groups/wg4
Formal name	A unique name that is still understandable to people and which reflects the EQ model	This paper
Local identifier	A fragment identifier, part of the URI, corresponding to the unique identifier for a concept. In the context of TOP a 6 character string without any meaning beginning with 'TOP'	http://www.w3.org/TR/rdf11-concepts/#section-IRIs
Metadata	Data documentation representing the higher level information or instructions that describe the content, context, quality, structure, provenance and accessibility of a data object	Michener *et al.* (1997)
Ontology	An explicit specification of a conceptualization. Formal model of a domain of interest, i.e. of its objects and their relationships	Gruber (1995)
Plant characteristic	A feature of an individual plant, plant population or plant species, describing either a plant trait or an environmental association	This paper
Plant trait	Any morphological, anatomical, biochemical, physiological or phenological heritable feature measurable on an individual plant	Violle *et al.* (2007), as modified by Garnier, Navas & Grigulis (2016)
Quality	A specific feature of an entity (in the entity-quality formalism)	Mungall *et al.* (2010)
Related concepts	Two concepts that are 'connected' by an associative link	SKOS recommendation, http://www.w3.org/TR/skos-reference/
Semantic annotation	The process of attaching names, attributes, comments, descriptions, etc. to a document or to a selected part in a text	http://www.ontotext.com
Semantic Web	Refers to the World Wide Web Consortium's (W3C) vision of a global Web of linked data. Semantic Web technologies provide standard ways to describe and access resources on the Web. Linked data are empowered by technologies such as RDF, SPARQL, OWL and SKOS.	http://www.w3.org/standards/semanticweb/
Semantic Web standards	Specifications of Semantics Web technologies	http://www.w3.org/standards/semanticweb/
SKOS (Simple Knowledge Organization System)	SKOS provides a standard way to represent knowledge organization systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way	SKOS web site home page: www.w3.org/2004/02/skos/intro
Standard	A published reference whose diffusion and utilization are widespread and recognized by a large proportion of those working in the domain	http://www.iso.org/iso/home/standards.htm
Term	A word or compound word that is a name or label for some concept	
Thesaurus	A controlled vocabulary designed to clarify the definition and structuring of key terms and associated concepts in a specific discipline	Laporte, Mougenot & Garnier (2012)
URI (Uniform Resource Identifier)	A string of characters used to identify uniquely the name of a resource on the web	

the whole-organism level' (Violle *et al.* 2007 as modified by Garnier, Navas & Grigulis 2016); an *environmental association** is defined as 'a non-random association of individual plants, plant populations or plant species with particular characteristics of the environment' (based on Underwood, Chapman & Crowe 2004). We present a first version of the thesaurus covering about 850 plant characteristics building, whenever possible, on existing *standards** defined in the context of terminological [e.g. Plant Ontology (PO): Cooper *et al.* 2013; Phenotypic Quality Ontology (PATO): Mungall *et al.* 2010; Plant Trait Ontology (TO): Jaiswal *et al.* 2002] and methodological (Knevel *et al.* 2005; Pérez-Harguindeguy *et al.* 2013) initiatives (see the 'General principles' section below). Each characteristic is considered as a concept for which TOP provides at least a name, definition, reference and a unique identifier on the web.

Three overarching principles have guided the development of TOP. First, TOP results from a collective effort of about 20

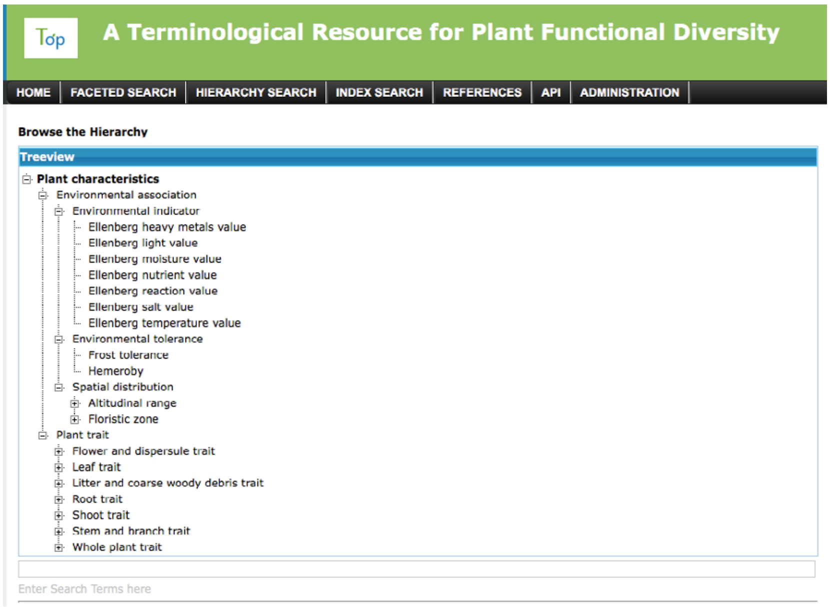

Fig. 1. Screenshot of the 'Hierarchy Search' page of the TOP web site showing the upper levels of the hierarchy used to organize the plant characteristics. [Colour figure can be viewed at wileyonlinelibrary.com]

experts in the field of plant functional diversity, ensuring a consensus basis for its semantic content (discussed in e.g. Deans *et al.* 2015; Parr *et al.* 2016). Second, computer scientists were involved at all stages of the process, so as to guarantee the use of relevant and up to date standards proposed in the context of the *Semantic Web** (see Laporte, Mougenot & Garnier 2012). And third, the concepts in TOP were selected on the basis of data availability in major ecological data bases such as the Ecological Flora of the British Isles (Fitter & Peat 1994: http://www.ecoflora.co.uk), BiolFlor (Klotz, Kühn & Durka 2002: http://www.biolflor.de) and TRY (Kattge *et al.* 2011a: https://www.try-db.org), assumed to give a proper reflexion of concepts widely used in the scientific field concerned.

After giving some examples of sematic confusion in the field of functional ecology, we present the general principles applied to design TOP, the type of information provided for each concept, and the web based tool with which TOP can be browsed and annotated. Finally an overview of the current TOP content and perspectives for its curation and further enrichment are presented.

The semantic bazaar in functional ecology: some examples

Table 2 illustrates the semantic confusion induced by the lack of precise terminology for selected characteristics widely used in functional ecology. We discuss below issues related to plant height, leaf size and related leaf traits, and seed size.

PLANT HEIGHT

In a widely used handbook of methods to measure plant traits (Cornelissen *et al.* 2003; Pérez-Harguindeguy *et al.* 2013), plant height is defined as: 'the shortest distance between the upper boundary of the main photosynthetic tissue on a plant and the ground level'. However, the expression 'plant height' actually applies to a number of related terms such as 'vegetative plant height', 'generative plant height', 'reproductive plant height', 'releasing height', 'canopy height', 'plant height at maturity' or 'maximum plant height' (Table 2). Moreover, plant height can be easily confused with the total length of the stem regardless of whether the plant has a prostrate, ascending, erect or liana-like growth form. In the absence of clear definitions, it is not possible to know whether these different expressions are synonymous or whether they cover different, albeit related meanings. And yet, the ecological significance may at times be quite different if by 'plant height', one is referring to the vegetative or generative height for a given species, to the length of the stem regardless whether the plant is prostrate, or to the height of an individual plant, or to the height of vegetation canopy in which the respective plant has been observed.

LEAF SIZE AND RELATED LEAF TRAITS

In the case of physical objects such as leaves, the term 'size' actually relates to any of its dimensions. Interestingly, the first classification of leaf size that was devised was based on leaf area (Raunkiaer 1934) but subsequently, 'leaf size' has been variously used to mean e.g. 'leaf area', 'leaf mass', 'leaf length' or 'leaf width' (Table 2). These different dimensions actually play very different roles in leaf function: light interception directly relates to leaf area; energy exchange between the leaf and the atmosphere is most strongly determined by leaf width (which, for a given wind speed, strongly influences the thickness of the boundary layer); and the mechanical support of leaves is more related to leaf length and mass than area. Hence, using a more precise term than 'leaf size' clarifies the issues at stake straight away.

Table 2. Examples of sources of semantic confusion encountered in the literature for selected plant characteristics. Issues concerning the first three characteristics are discussed in more details in the text. For the other ones, two short examples of major issues are commented upon

Characteristics	Related concepts used with same or unclarified meaning	Example of unit frequently used	Examples of issues
Plant height	Vegetative plant height Generative plant height Reproductive plant height Releasing height Canopy height Plant height at maturity Maximum plant height	m m m m m m m	1. Canopy height might refer to the height of an individual plant or to the vegetation canopy in which the individual plant is observed 2. Dispersal distance relates to reproductive plant height, which might be substantially different from vegetative plant height
Leaf size	Leaf area Leaf mass Leaf length Leaf width	cm^2 mg cm cm	1. Light interception, mechanical support and resistance of the boundary layer depend on different dimensions of the leaf 2. Inclusion of petiole and rachis in compound leaves (see text for further details)
Seed size	Seed volume Seed mass (dry or fresh) Largest seed dimension Mass of 'true seed' or dispersule	cm^3 mg cm mg	1. Seeds are often confounded with 'dispersules' (i.e. the seed plus various appendages such as wings or pappus), the relevant unit for dispersal 2. Seedling establishment and survival depends on the mass of the 'true seed' (i.e. embryo, endosperm and testa)
Specific leaf area (SLA)	SLA of individual leaves or of all leaves on the shoot SLA for one or two sides of the leaf SLA with or without petiole SLA of leaf or leaflet (for compound leaves)	$m^2\ kg^{-1}$ $m^2\ kg^{-1}$ $m^2\ kg^{-1}$ $m^2\ kg^{-1}$ $m^2\ kg^{-1}$	1. If assessed on whole shoots, integrates effects of shading within the canopy and of leaf ageing 2. Same as 2. under 'leaf size'
Leaf nitrogen concentration (LNC)	Leaf nitrogen content LNC expressed on a mass (dry or fresh) or area basis	mg $mg\ g^{-1}$ or $g\ m^{-2}$	1. Content scales with e.g. leaf mass, while concentration scales with activities per e.g. unit mass 2. Same as 2. under 'leaf size'
Photosynthetic rate (PS)	PS of whole shoot or individual leaves PS expressed on an area or a mass basis	$\mu mol\ m^2\ s^{-1}$ $\mu mol\ m^2\ s^{-1}$ or $nmol\ g^{-1}\ s^{-1}$	1. Same as 1. under 'SLA' 2. Expression of trade-offs among leaf traits (e.g. the 'leaf economics spectrum') much stronger when PS is expressed on a mass than on an area basis
Relative growth rate (RGR)	RGR expressed on a mass (fresh or dry), area or (e.g. height) length basis RGR of whole plant or plant parts (e.g. leaf, root)	$g\ g^{-1}\ day^{-1}$ or $cm^2\ cm^{-2}\ day^{-1}$ or $cm\ cm^{-1}\ day^{-1}$ $g\ g^{-1}\ day^{-1}$	1. RGR models are well-developed for RGR on mass basis but not for RGR on an area basis. Transfer of theoretical concepts might lead to wrong conclusions 2. Organs do not necessarily grow at the same rate, which e.g. induces shifts in biomass allocation among organs during growth
Specific root length (SRL)	SRL of the whole root system SRL of fine roots SRL of roots of a specific order SRL of roots or rhizomes	$m\ g^{-1}$ $m\ g^{-1}$ $m\ g^{-1}$ $m\ g^{-1}$	1. Root traits, including SRL, vary tremendously with root order and root diameter at given order 2. Often, roots and rhizomes are lumped, although both organs are functionally different
Flowering time	Onset of flowering Time of peak flowering Flowering duration Flowering frequency	Julian day Julian day Number of days Number per year	1. Onset of flowering might occur substantially earlier than peak of flowering 2. Some annual species flower several times per year (e.g. *Poa annua*)
Frost tolerance	Frost hardiness Frost resistance	Dependent on how it is assessed	1. Resistance reduces frost damage while tolerance reduces the negative fitness impacts of damage 2. Can be defined in terms of mortality or in terms of loss of mass/reduced growth/damage

Another important issue when dealing with leaf traits relates to the specific portion of the leaf that is actually under consideration: is it the whole leaf, including the petiole, and rachis in compound leaves? Or is it only the leaf blade and leaflets? Since the petiole and rachis can make up a substantial part of total leaf mass, especially in compound leaves, including or excluding one or the other may lead to important differences in e.g. Specific leaf area (SLA) values (the ratio of leaf area to leaf mass). Considering the whole leaf or only the leaf blade and/or leaflets may also have important implications for a number of other traits, including photosynthetic rate and chemical composition (Table 2), the nutrient concentrations and physiological activities of petiole/rachis material being generally lower than those of the leaf lamina.

SEED SIZE

As for leaves, the relevant 'seed size' characteristics will depend on the function of interest: reproduction, dispersion and colonization of new areas, persistence during periods of unfavourable conditions, or overall cost to the parent plant. A first issue concerns the morphological confusion between dispersule (or propagule = the unit of seed, fruit or spore as it is dispersed: Pérez-Harguindeguy *et al.* 2013), and the seed. The dispersule may correspond with the seed. However, in many species, it is composed of the seed plus surrounding structures, i.e. various appendages such as wings or pappi. The size (and shape) of these appendages should obviously be taken into account if the focus is on dispersal. And, if cost to parent plant is under consideration, then the constructions costs of these appendages are also relevant. By contrast, it is the volume rather than the mass of the entire reproductive dispersule which will determine whether it can penetrate the litter layer or will remain on the surface. However, if the focus is on seedling establishment and survival, the amount of reserves stored in the seed will be of prime importance, and the mass of the 'true seed' (i.e. the embryo, endosperm and testa) will be a more relevant trait to assess.

A number of less detailed examples are given for other concepts in Table 2. Beyond the fact that improving semantics will help make sure that we are indeed speaking a common language, it can also encourage creativity by identifying interesting or important new or largely overlooked characteristics. For example the ratio of reproductive to vegetative height (sometimes called 'relative prominence of inflorescence') is thought to confer ability to escape domestic herbivores in grasses (McIntyre *et al.* 1999). Similarly, the ratio of whole leaf SLA to blade/leaflet SLA might be useful to assess differentiation of functions in leaves, e.g. transport vs. carbon uptake. By pointing out the variety of ways seemingly singular characteristics have been defined, we can then also promote innovation by thinking more about how to make good use of the variety of concepts which have been loosely defined so far.

Whatever the case, this section demonstrates that improving the semantics of concepts pertaining to the functional facet of biodiversity can only be beneficial to the field.

General principles of terminology development

TOP is designed to serve as a terminological resource for the characterization of concepts pertaining to the two types of plant characteristics introduced above: plant traits and environmental associations, and provides simple semantic relationships among these concepts. As the first aim of TOP is to reduce semantic heterogeneity for plant characteristics, it does not address methodological issues: TOP defines *what* a plant characteristic is, but not *how* this characteristic can be measured. Information on measurement protocols are clearly needed to interpret data (cf. Discussion section and Fig. 5), but the development of methods is a different issue that has been addressed separately by the community (see e.g. Hendry & Grime 1993; Knevel *et al.* 2005; Sack *et al.* 2010; Pérez-Harguindeguy *et al.* 2013). The next paragraphs describe the major features of the methodology followed up for the development of TOP.

A COLLECTIVE INITIATIVE

The TOP initiative developed from early efforts to define methodological standards for the measurement of 28 plant traits (Cornelissen *et al.* 2003), which came with a series of accurate definitions of terms. The TRY initiative, set up in 2007, made it obvious that these initial efforts had to be expanded to a much wider array of terms and expressions (see above). This led to a series of workshops between 2009 and 2015, involving experts in the fields of plant functional ecology and informatics. During the first phase of this work (2009–2011), 15 experts from the community working with plant functional traits world-wide contributed to the construction of a preliminary version of TOP, in close interaction with experts in Web semantics. Based on the broad overview of traits compiled in the TRY data base and their original names (Kattge *et al.* 2011b), concepts and associated definitions were taken from reference publications in the field (Hendry & Grime 1993; Cornelissen *et al.* 2003; Knevel *et al.* 2005) for a set of approximately 130 traits.

This initial list of concepts, associated definitions and a first hierarchy among traits were made available to the experts *via* a web-based interface, ThesauForm (Laporte, Mougenot & Garnier 2012). ThesauForm, which is based on SKOS *Semantic Web standards** (Miles & Bechhofer 2009: http://www.w3.org/TR/skos-reference), facilitated an efficient involvement of the scientific community during the definition of individual concepts, and promoted consensus building that will help ensure community acceptance of the thesaurus. The preliminary version that resulted from this work was used during the second phase (2012–2015) for the further development of TOP and the definitions of rules for constructing new terms.

BUILDING ON EXISTING MODELS AND VOCABULARIES

To be consistent with both previous and ongoing developments of standards in related fields, TOP builds – whenever possible – on existing principles and vocabularies. Most relevant in this context are: (i) the EQ formalism to model plant characteristics and (ii) definitions for entities and qualities provided by external sources and the reuse of concepts.

The entity-quality model

Plant characteristics are modelled on the basis of the EQ model, which is also used for the description of phenotypes in the field of genetics (see e.g. Mungall *et al.* 2010). These descriptions consist of the *entity** that is observed (for example: a leaf, a set of intraspecific populations or a species), and the specific *quality** of that entity (for example: area, mass, colour, geographic distribution). A characteristic is therefore composed of a combination of at least one entity (noted <E> hereafter) and one quality (noted <Q> hereafter), and is defined as ‘an entity having a quality’ (for instance ‘leaf area’ (leaf [<E>] area [<Q>]), see Table 3). In the case of plant traits, the entity refers to the individual plant or parts thereof. The case is less straightforward for environmental associations, which can be defined for plant individuals – from the level of cells to the whole organism -, populations, a set of populations or a species. For example, frost tolerance can be assessed in many different ways, either e.g. from laboratory experiments conducted at the level of organs or whole individuals, or derived from population or species distribution ranges (see Table 2 and e.g. Bannister 2007 for a review). When the entity is not precisely known for a particular environmental association, we therefore use the term ‘plant’ as a generic entity in the formal name, but the entity or entities that can potentially be associated with the quality are specified in the definition.

Definitions from external sources and reuse of concepts

Whenever possible, the definition of the entity is based on the PO (http://www.plantontology.org/: Jaiswal *et al.* 2002; Cooper *et al.* 2013), while the qualities are based on definitions in the PATO (http://www.obofoundry.org/ontology/pato.html: Mungall *et al.* 2010).

Whenever a characteristic in TOP had already been defined in the context of a well-established vocabulary, e.g. the TO (http://www.ontobee.org/ontology/TO) or the Flora Phenotype Ontology (http://bioportal.bioontology.org/ontologies/FLOPO), the handbook for trait measurements (Pérez-Harguindeguy *et al.* 2013 and its on-line updatable version www.nucleodiversus.org/index.php?mod=page&id=79), the PrometheusWiki (Sack *et al.* 2010: http://prometheuswiki.publish.csiro.au/tiki-custom_home.php), or a well curated data base, e.g. BiolFlor (Klotz, Kühn & Durka 2002) or LEDA (Kleyer *et al.* 2008), the definition is considered and, if appropriate, re-used and referenced. If a concept exists but is not considered appropriate (e.g. it is used with different meanings, like ‘canopy height’, which may refer to the vegetative height of an individual plant or to the height of a vegetation canopy), it is mentioned in TOP as a related term to guarantee consistency with existing vocabularies and legacy data, from which this term originates.

Concepts of characteristics are reused within TOP whenever exactly the same definitions apply; for instance, the concepts ‘leaf area’ and ‘leaf dry mass’ are reused in the concept ‘leaf area per leaf dry mass’. Consistent reuse of concepts allows for cross-referencing and for the modelling of complex characteristics (e.g. ratios) based on simple structured concepts.

Table 3. Modelling plant characteristics using the entity-quality model (‘EQ’ model). Examples are for the two types of plant characteristics covered by TOP

Characteristics	Preferred name	Entity (<E>)	Quality (<Q>)
Plant trait	Leaf area	Leaf	Area
Environmental association	Ellenberg temperature value	Plant population*	Temperature indicator value according to Ellenberg

*Although the entity for Ellenberg indicator values might be thought to be ‘species’, these values have actually been defined for populations of species within Central Europe (Ellenberg *et al.* 1992).

PERSISTENCE AND CURATION

Once a concept has been approved, it will be persistent: the concept and its *Uniform Resource Identifier** (URI, a unique and persistent identifier in the World Wide Web) will continue to be available and its change history will be tracked (http://www.w3.org/Consortium/Persistence.html). Finally, regular updates of TOP are planned, based on community feedback and involvement.

Concepts – the core units of the thesaurus

The core units of TOP are the conceptualizations of plant characteristics and categories thereof (see Appendix S1 for details, Supporting Information). Each concept is characterized by a number of components, which are displayed on its individual page (Fig. 2) of the TOP web site (www.top-thesaurus.org). These are: (i) a common and a formal name; (ii) a definition with an associated reference acknowledging the source of this definition and (iii) a URI. Additional information like synonyms, abbreviations, related terms, formal measurement unit, comments and semantic relations are also given, so as to help users finding and understanding the concepts. If available, the concept-page provides links to external sources of measurement protocols and measured data.

COMMON AND FORMAL NAMES

For each concept, both a common name and a formal name are provided. The *common name** is a preferred term typically used and well-known in the scientific domain for describing that concept (e.g. leaf area, specific leaf area, or frost tolerance). The *formal name** is a unique name that is still understandable to people and which reflects the EQ model. For example, in the case of the common name 'leaf area', the formal name is the same (leaf [<E>] area [<Q>]), but in the case of the common name 'Ellenberg nutrient value' the formal name would be 'plant population (<E>) nutrient indicator value according to Ellenberg (<Q>)'. When a characteristic is complex (e.g. a ratio, flux, rate, etc.) it may consist of a combination of several entities and qualities, which translates into the corresponding combination of several EQ associations, e.g. the trait 'leaf area ratio', which has the formal name 'whole-plant leaf area per whole-plant dry mass'. 'Whole-plant leaf area' is defined in TOP as: the sum of the area (PATO:area) of all leaves (PO: leaf, TOP:leaf area) on the shoot (PO:shoot system). 'Whole-plant dry mass' is defined in TOP as: the mass (PATO:mass) of a whole plant (PO:whole plant) assessed after drying.

DEFINITION

The definition of a characteristic follows the formal name providing the entities, qualities and their relationships. As previously indicated, the definitions are based whenever possible on concepts of entities and qualities from existing vocabularies or concepts defined within TOP. The definition given for a concept is free of any information pertaining to measurement protocol or methodological information. For example, the trait 'seed dry mass' consists of the entity 'seed' and the quality 'dry mass', and the definition for this characteristic is: 'the mass of a seed being dried', and not 'the mass of a seed being dried at 95 °C for 1 h in the oven', which would then include measurement standards and protocol information.

UNIFORM RESOURCE IDENTIFIER AND AUTHORSHIP

Each concept is assigned a *local identifier** (e.g. 'TOP 25' for leaf area), which is unique within TOP. In combination with the URL of the TOP Web site, the local identifier provides a URI for each concept (e.g. www.top-thesaurus.org/trait/TOP25 for leaf area).

Authorship identifies 'who' has given the formal definition for a concept retained in TOP, and will employ ORCID identifiers (http://orcid.org/) to unambiguously refer to the person contributing the concept definition.

ADDITIONAL INFORMATION PROVIDED ON THE CONCEPT WEBPAGES

Synonyms, related terms and abbreviations (Fig. 2) are essential to enable the concept to be found under different names, which may have the same or slightly different meanings. Semantic relations (i.e. relations between concepts), which can be hierarchical or associative, provide information about the general and specific context of the concept: (i) a fragment of the hierarchical tree which organizes the TOP concepts is shown, with more general and more specific concepts surrounding the concept defined; (ii) a comment field provides relevant details and links to *related concepts** defined elsewhere (e.g. in other controlled vocabularies or ontologies, such as PO, PATO, TO or within TOP). These comments offer the opportunity to provide additional information related to the concept in an unstructured format. As additional information to users, and when available, the concept page provides links to websites with relevant method descriptions for the characteristic or/and to data bases with measurement records.

Current content, visualization and curation of TOP

The current version of TOP provides concepts for 858 plant characteristics and their categories (Fig. 3). The publication of this initial list of concepts allows for additional feedback from the community as to accuracy and completeness. A revised version of TOP is expected about 1 year after publication of the current version of the thesaurus. Any deprecated concepts will be either mapped to more accurate concepts, or otherwise re-directed to the closest similar term, with a clear annotation as to their current and former status.

TOP is freely available on the web under the URL http://www.top-thesaurus.org. To assist users in their search for

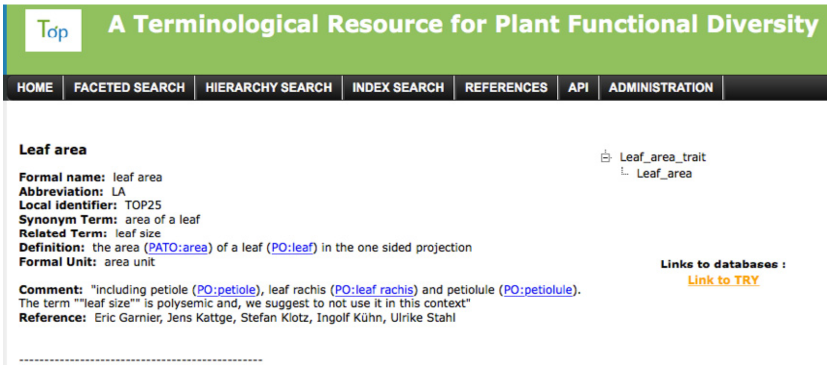

Fig. 2. Screenshot of the page from the TOP-Thesaurus web site showing the information given for plant characteristics, using the example of leaf area. [Colour figure can be viewed at wileyonlinelibrary.com]

pertinent information within TOP, the TOP website is organized in three tabs, with each tab offering users a different search option (Fig. 4): (i) an 'Index Search', available through the Index tab; (ii) a 'Hierarchy Search', which allows for browsing the thesaurus through a hierarchical tree; a free text search is also available on this tab; (iii) a '*Faceted Search**' (cf. Laporte *et al.* 2014). Additionally, an *Application Programming Interface** (API) is implemented. As all search modes are concept-based, synonyms, abbreviations and related terms can also be used to retrieve information. Further details on how to search TOP are given in Appendix S2.

Similar to other terminological (e.g. PO: Cooper *et al.* 2013) or methodological initiatives (e.g. PrometheusWiki: Sack *et al.* 2010; Plant Traits Handbook: Pérez-Harguindeguy *et al.* 2013), TOP is designed for growth and continued updating. We thus expect people with interests in any aspect of plant structure and functioning within and beyond the field of ecology to actively contribute to the development of TOP. The procedure to be implemented in this context involves the five steps described in Appendix S3.

We expect to deliver new releases of TOP at least once a year, or more frequently if many additions or changes are suggested over a short period of time. Support for multilingual versions of TOP is planned for the future.

Discussion

The thesaurus presented here provides recommendations pertaining to the characterization of concepts widely used in plant functional ecology. It aims at reducing the ambiguity of terms used to describe plant characteristics – traits and environmental associations – by formalizing the construction of the terms themselves, their definitions and how terms are inter-related. Besides its role as a terminological resource, TOP may contribute to resource discovery and interpretation in the context of data publication, sharing and access. By clarifying the semantic content of concepts, it can also encourage creativity by identifying interesting characteristics which have been overlooked so far.

Previous terminologies in the field of plant functional ecology were developed as a side activity to the design of methodological standards (Cornelissen *et al.* 2003; Pérez-Harguindeguy *et al.* 2013) and data bases (Klotz, Kühn & Durka 2002; Knevel *et al.* 2005). To our knowledge, the TOP initiative is the first coordinated action to provide a comprehensive thesaurus of concepts frequently used in this field. Building on advances from related disciplines such as plant anatomy and morphology (e.g. Cooper *et al.* 2013), it fills a gap by allowing accurate descriptions of key aspects of plant functional diversity, which are currently poorly described in existing terminological standards (Walls *et al.* 2014; but see Pey *et al.* 2014 for a comparable initiative as applied to soil invertebrates). TOP should be considered as a contribution by ecology towards the realization of 'computable phenotypes', identified as a major required breakthrough for achieving an integrative understanding across many fields in biology including genomics, evolution, ecology, breeding and systematics (Deans *et al.* 2015). As such, TOP will be made visible

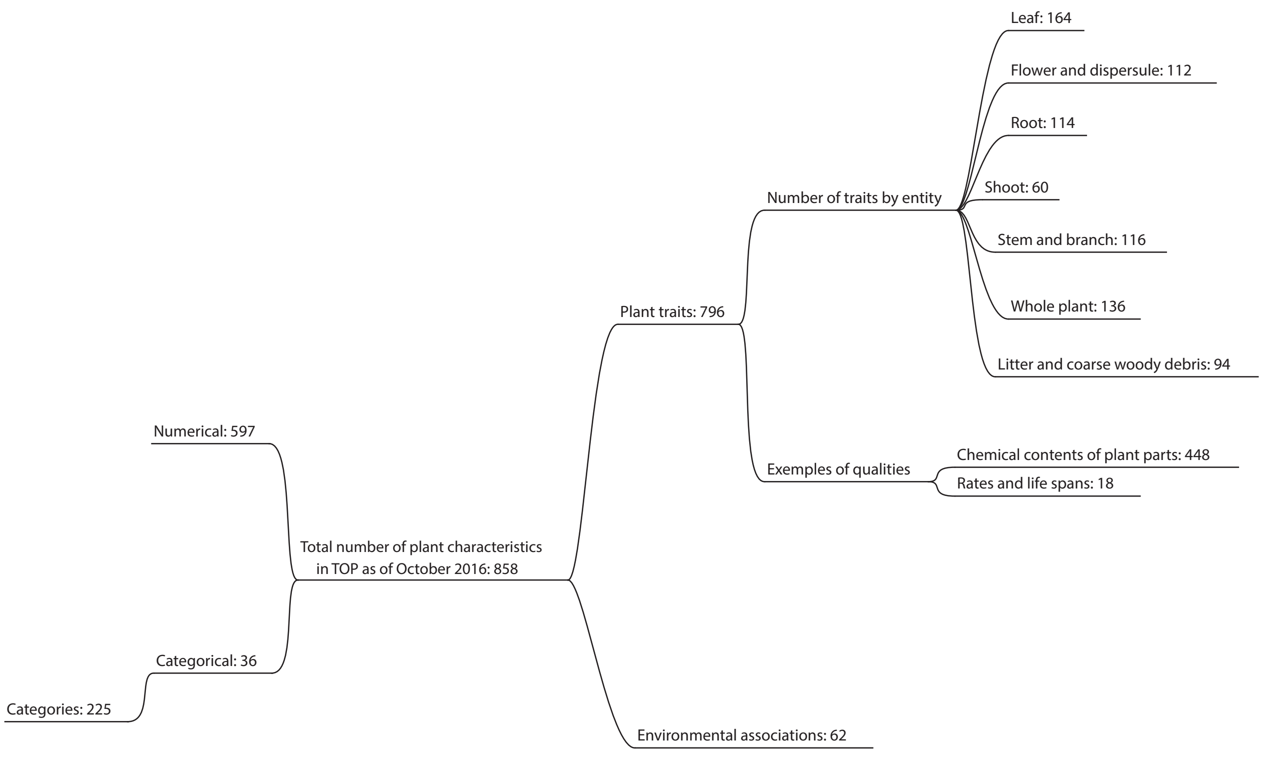

Fig. 3. Summary of TOP content as of October 2016. For qualities, only a selected number of examples are shown. See Appendix S1 for the distinction between numerical and categorical characteristics and the main text for further details.

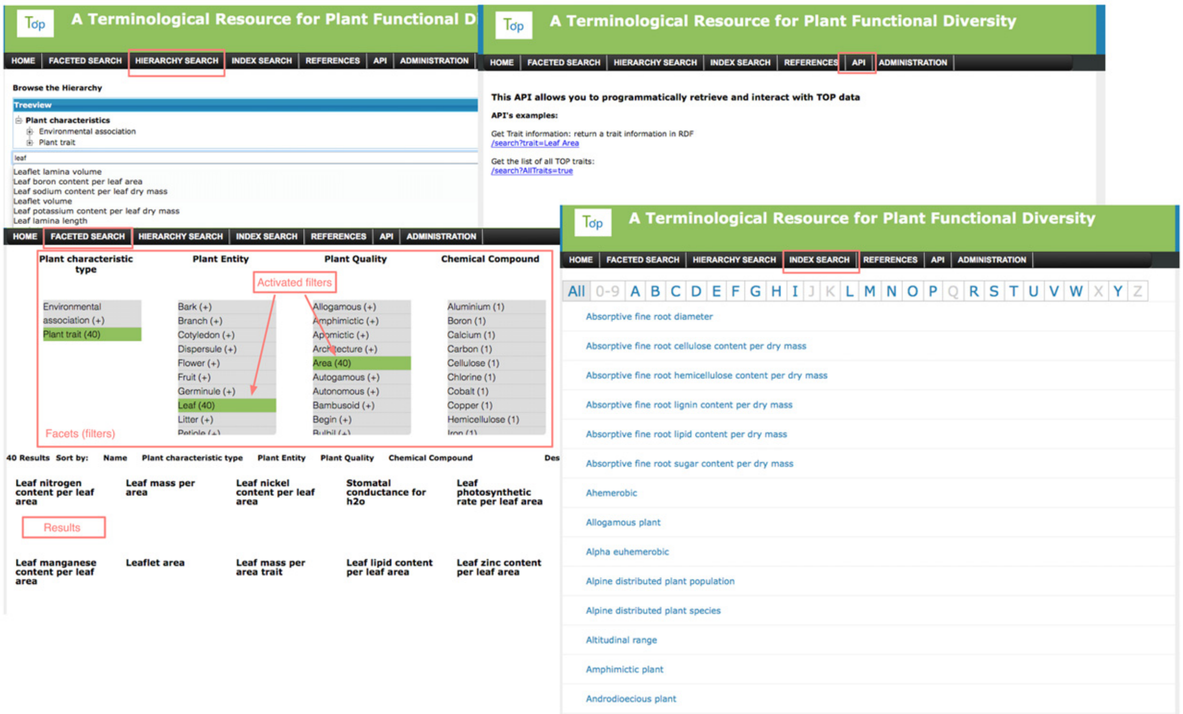

Fig. 4. A combination of screenshots illustrating the different modes that can be used to search TOP. [Colour figure can be viewed at wileyonlinelibrary.com]

and available on relevant terminological portals and registries of ontologies (e.g. BioPortal: www.bioportal.bioontology.org; ontobee: http://www.ontobee.org/; GFBio: http://www.gfbio.org). TOP developers will also engage in closer collaborations with scientists from aligned fields in biology and the environmental sciences, to improve the harmonization of terminological standards across disciplines.

TOP is deliberately focused on the semantic aspects of plant characteristics only. In particular, care was taken not to embed any ‘hidden’ method in the definitions of characteristics. In order to improve the quality and interoperability of data, the terminological resource proposed here has to be complemented by, and preferably referenced to, at least, methodological standards (e.g. Sack *et al.* 2010; Pérez-Harguindeguy *et al.* 2013) and units of expression (Fig. 5), which all constitute *metadata** information, describing the ‘who, what, when, where, and how’ about every aspect of the data (Michener 2006). The TOP concepts can actually be considered as metadata descriptors specified in the Ecological Metadata Language (EML: Michener *et al.* 1997): concept names can be mapped to ‘Variable identity’ (Class IV.B.1. of EML), concept definition to ‘Variable definition’ (Class IV.B.2.) and formal unit to ‘Units of measurement’ (Class IV.B.3.) (cf. Fig. 5).

Scientists/curators will increasingly associate particular measurements with specific, well-defined concepts. For example, with respect to the ‘seed dry mass’ mentioned above, details about the ‘dry’ in ‘dry mass’ may have to be defined in the metadata. In addition, for selecting and analysing the data, scientists might need to know ‘how dry’ a dry mass is. They may decide to exclude data (e.g. sun-dried seeds) based on this information, or to use the actual oven temperature as a covariate. In summary, a substantial part of the scientific workflow dealing with details of the plant growth history or measurement methodologies, might only be documented in the metadata (cf. Fig. 5). This means that the concepts defined in TOP gain further clarification when linked to additional meta-information in data bases. This bidirectional relationship between a concept and its application to specific measurement instances also provides opportunities for semantic sharpening and concept evolution: if a linkage (or ‘*semantic annotation**’, as it is also called) exists between a concept and some methodological details, a curator or scientist can later explore the realized methodological variations of the concept, and empirically derive limits as to what practitioners label as ‘dry mass’ (e.g. based on a histogram of drying temperatures).

TOP is expressed as a SKOS language (Miles & Bechhofer 2009: http://www.w3.org/TR/skos-reference), and complies with Semantic Web standards, providing a standard set of structures that will enable computers to operate in ways that more precisely assist data users in locating (data discovery) and processing the data of interest. Additional benefits of TOP’s adhering to Web standards is enhanced interoperability and effectiveness of automated data exchange among different sources. Simultaneous queries on different data bases will thus become possible when different data bases use TOP concepts for the semantic annotation of their data. The added value coming from curated and annotated data will thus be to ensure the quality of the data – e.g. unambiguous names and

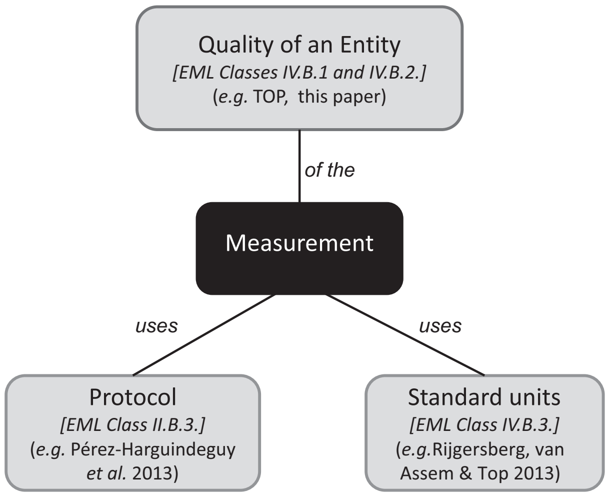

Fig. 5. sKey standards necessary to disambiguate the values assigned to a trait include adequate descriptions of (1) what is measured: these are the semantic aspects developed in this paper (the Quality of an Entity), (2) how it is measured: this corresponds to the use of standardized protocols when available, or the accurate report of the methods followed, and (3) how it is expressed: this corresponds to the units used when coding the trait value. Examples of references for each of these three standards are given between parentheses. These standards correspond to four classes of descriptors used in EML (Ecological Metadata Language: Michener *et al.* 1997), given between brackets: class IV.B. refers to 'Data structural descriptors – Variable information', with subclasses 1, 2 and 3 corresponding to 'Variable identity', 'Variable definition' and 'Units of measurement', respectively; class II.B.3. refers to 'Research origin descriptors – Specific subproject description – Research methods'. See text for further details.

definitions for the characteristics – coming from different data sources and produced in different contexts, as well as to facilitate the integration of data, the reproducibility of science, and the synthesis of data.

The EQ model as applied here to plant characteristics forms the basis of several ontologies used to describe phenotypes in e.g. genomics and biomedical fields (Mungall *et al.* 2010). TOP, which is designed as a thesaurus focused on the definitions of concepts, cannot be considered as an ontology *per se*, because so far it only provides suggestions for very simple relationships between the concepts (synonyms, broader, narrower and related terms). However, TOP is intended to serve as the basis to develop a domain ontology for plant ecology (a prototype, PLATON – PLAnT characteristics ONtology for ecology – is under development using the W3C's OWL format: Laporte 2011). To do so, and in agreement with approaches promoting the use of existing ontologies (Pinto & Martins 2004; and see recommendations of the Planteome project: www.planteome.org), TOP (i) reuses concepts from existing ontologies whenever this is possible and (ii) will be mapped to the higher level ontology OBOE ('Extensible Observation Ontology': Madin *et al.* 2007), which has been designed to capture the semantics of generic ecological observations and measurements. This will require adapting the simple EQ model used here to the OBOE framework, in which observations and measurements are the central concepts linking an entity to a quality (see Kattge *et al.* 2011b for an application of this framework in the context of the TRY data base). This development of TOP towards a more expressive domain ontology, will provide functional ecologists with a semantic framework enabling scientists to produce new knowledge sets from large information systems (Laporte, Mougenot & Garnier 2012).

TOP constitutes a step toward solving the problem of data heterogeneity across thematic, organizational, spatial and temporal scales inherent to biodiversity and ecological data. As a first proof of concept, it is already in use in the context of the TRY data base of plant traits (cf. https://www.try-db.org). By providing well-defined and harmonized concepts, it has also the potential to improve communication and data interoperability beyond academic science, with domains including citizen science, teaching activities, as well as environmental management (cf. Herrando-Pérez, Brook & Bradshaw 2014).

Authors' contributions

E.G., U.S., M.-A.L., J.K. and I.M. designed the study and were involved in the main stages of the thesaurus development. B.L. is in charge of the maintenance of the web site hosting the thesaurus. All authors contributed to the choice of terms and definitions. E.G., U.S., M.-A.L., J.K. and S.K. wrote the manuscript with contributions from all authors, who gave final approval for publication.

Acknowledgements

This work was partially funded by the following grants: GDR 2574 ('Traits') of CNRS (France); the TraitNet Research Coordination Network (US National Science Foundation Grant No. 0639161); the SONet effort (US NSF Grant No. 0753144); the project 'Towards a unifying system of plant trait definition and measurement' (the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig/Deutsche Forschungsgemeinschaft via Grant No. 5017 0649_#16); the German Federation for Biological Data - GFBio (Deutsche Forschungsgemeinschaft via Grant No. KA 3323/2-2); FONCyT, CONICET, Universidad Nacional de Córdoba and Inter-American Institute for Global Change Research (CRN 2015 and SGP-CRA2015, supported by NS grants GEO-0452325 and GEO-1138881), which supported the collaboration with the Plant Trait Handbook. Marie-Claude Quidoz, from the 'Système d'Information en Ecologie' platform at CEFE made possible preliminary work on the ThesauForm tool on the web.

Data accessibility

All data from the Thesaurus Of Plant characteristics are freely available on the web under the URL http://www.top-thesaurus.org.

References

Bannister, P. (2007) Godley review: a touch of frost? Cold hardiness of plants in the southern hemisphere. *New Zealand Journal of Botany*, **45**, 1–33.

Cooper, L., Walls, R.L., Elser, J. *et al.* (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant and Cell Physiology*, **54**, e1.

Cornelissen, J.H.C., Lavorel, S., Garnier, E. *et al.* (2003) A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, **51**, 335–380.

Deans, A.R., Lewis, S.E., Huala, E. *et al.* (2015) Finding our way through phenotypes. *PLoS Biology*, **13**, e1002033.

Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulissen, D. (1992) Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica*, **18**, 1–258.

Enquist, B.J., Norberg, J., Bonser, S.P., Violle, C., Webb, C.T. & Savage, V.M. (2015) Scaling from traits to ecosystems: developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories. *Advances in Ecological Research*, **52**, 249–318.

Fitter, A.H. & Peat, H.J. (1994) The ecological flora database. *Journal of Ecology*, **82**, 415–425.

Garnier, E. & Navas, M.-L. (2012) A trait-based approach to comparative functional plant ecology: concepts, methods and applications for agroecology. A review. *Agronomy for Sustainable Development*, **32**, 365–399.

Garnier, E., Navas, M.-L. & Grigulis, K. (2016) *Plant Functional Diversity – Organism Traits, Community Structure, and Ecosystem Properties*. Oxford University Press, Oxford, UK.

Gruber, T.R. (1995) Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies*, **43**, 907–928.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S. & Porter, J.H. (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*, **11**, 156–162.

Heidorn, P.B. (2008) Shedding light on the dark data in the long tail of science. *Library Trends*, **57**, 280–299.

Hendry, G.A.F. & Grime, J.P. (1993) *Methods in Comparative Plant Ecology*. Chapman & Hall, London, UK.

Herrando-Pérez, S., Brook, B.W. & Bradshaw, C.J.A. (2014) Ecology needs a convention of nomenclature. *BioScience*, **64**, 311–321.

Jaiswal, P., Ware, D., Ni, J. *et al.* (2002) Gramene: development and integration of trait and gene ontologies for rice. *Comparative and Functional Genomics*, **3**, 132–136.

Jones, M.B., Schildhauer, M.P., Reichman, O.J. & Bowers, S. (2006) The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution and Systematics*, **37**, 519–544.

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

Kattge, J., Ogle, K., Bönisch, G. *et al.* (2011b) A generic structure for plant trait databases. *Methods in Ecology and Evolution*, **2**, 202–213.

Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G. & Hooker, G. (2009) Data-intensive science: a new paradigm for biodiversity studies. *BioScience*, **59**, 613–620.

Kleyer, M., Bekker, R.M., Knevel, I.C. *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., Kühn, I. & Durka, W. (2002) BIOLFLOR – Eine Datenbank zu biologisch-ökologischen Merkmalen zur Flora von Deutschland. *Schriftenreihe für Vegetationskunde*, **38**, 1–333.

Knevel, I.C., Bekker, R.M., Kunzmann, D., Stadler, M. & Thompson, K. (2005) *The LEDA Traitbase. Collecting and Measuring Standards of Life-History Traits of the Northern European Flora*. University of Groningen, Groningen, the Netherlands.

Laporte, M.-A. (2011) *Définition de standards de données relatifs aux traits fonctionnels des végétaux pour l'étude de la biodiversité*. PhD, Université Montpellier II – Sciences et Techniques du Languedoc, Montpellier, France.

Laporte, M.-A., Mougenot, I. & Garnier, E. (2012) ThesauForm – Traits: a web based collaborative tool to develop a thesaurus for plant functional diversity research. *Ecological Informatics*, **11**, 34–44.

Laporte, M.A., Mougenot, I., Garnier, E., Stahl, U., Maicher, L. & Kattge, J. (2014) A semantic web faceted search system for facilitating building of biodiversity and ecosystems services. *Data Integration in the Life Sciences, DILS 2014* (eds H. Galhardas & E. Rahm), pp. 50–57. Springer Int Publishing Ag, Cham, Switzerland.

Lavorel, S., Díaz, S., Cornelissen, J.H.C. *et al.* (2007) Plant functional types: are we getting any closer to the Holy Grail? *Terrestrial Ecosystems in a Changing World* (eds J. Canadell, D. Pataki & L. Pitelka), pp. 149–164. Springer-Verlag, Berlin, Germany.

Madin, J.S., 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 and Evolution*, **23**, 159–168.

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

McIntyre, S., Lavorel, S., Landsberg, J. & Forbes, T.D.A. (1999) Disturbance response in vegetation – towards a global perspective on functional traits. *Journal of Vegetation Science*, **10**, 621–630.

Michener, W.K. (2006) Meta-information concepts for ecological data management. *Ecological Informatics*, **1**, 3–7.

Michener, W.K. & Brunt, J.W. (2000) *Ecological Data – Design, Management and Processing. Methods in Ecology*. Blackwell, Oxford, UK.

Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, **27**, 85–93.

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. & Stafford, S.G. (1997) Nongeospatial metadata for the ecological sciences. *Ecological Applications*, **7**, 330–342.

Miles, A. & Bechhofer, S. (eds) (2009) SKOS simple knowledge organization system reference. W3C Recommendation 18 August 2009. http://www.w3.org/TR/skos-reference/.

Mungall, C.J., Gkoutos, G.V., Smith, C.L., Haendel, M.A., Lewis, S.E. & Ashburner, M. (2010) Integrating phenotype ontologies across multiple species. *Genome Biology*, **11**, R2.

Parr, C.S., Schulz, K., Hammock, J., Wilson, N., Leary, P., Rice, J. & Corrigan, R.J. Jr (2016) TraitBank: practical semantics for organism attribute data. *Semantic Web Journal*, **7**, 6.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E. *et al.* (2013) New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal Botany*, **61**, 167–234.

Pey, B., Laporte, M.-A., Nahmani, J. *et al.* (2014) A thesaurus for soil invertebrate trait-based approaches. *PLoS ONE*, **9**, e108985.

Pinto, H.S. & Martins, J.P. (2004) Ontologies: how can they be built? *Knowledge and Information Systems*, **6**, 441–464.

Raunkiaer, C. (1934) *The Life Forms of Plants and Statistical Plant Geography* (English edn). Oxford University Press, Oxford, UK.

Reichman, O.J., Jones, M.B. & Schildhauer, M.P. (2011) Challenges and opportunities of open data in ecology. *Science*, **331**, 703–705.

Rijgersberg, H., van Assem, M. & Top, J. (2013) Ontology of units of measure and related concepts. *Semantic Web*, **4**, 3–13.

Sack, L., Cornwell, W.K., Santiago, L.S., Barbour, M.M., Choat, B., Evans, J.R., Munns, R. & Nicotra, A. (2010) A unique web resource for physiology, ecology and the environmental sciences: PrometheusWiki. *Functional Plant Biology*, **37**, 687–693.

Underwood, A.J., Chapman, M.G. & Crowe, T.P. (2004) Identifying and understanding ecological preferences for habitat or prey. *Journal of Experimental Marine Biology and Ecology*, **300**, 161–187.

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.

Walls, R.L., Athreya, B., Cooper, L. *et al.* (2012) Ontologies as integrative tools for plant science. *American Journal of Botany*, **99**, 1263–1275.

Walls, R.L., Deck, J., Guralnick, R. *et al.* (2014) Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PLoS ONE*, **9**, e89606.

Received 8 June 2016; accepted 26 October 2016
Handling Editor: Peter Vesk

Supporting Information

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

Appendix S1. Types of characteristics.

Appendix S2. Searching TOP.

Appendix S3. Curation of TOP.