

FORUM

Plant names in vegetation databases – a neglected source of bias

Florian Jansen & Jürgen Dengler

Abstract

Problem: The increasing availability of large vegetation databases holds great potential in ecological research and biodiversity informatics. However, inconsistent application of plant names compromises the usefulness of these databases. This problem has been acknowledged in recent years, and solutions have been proposed, such as the concept of “potential taxa” or “taxon views”. Unfortunately, awareness of the problem remains low among vegetation scientists.

Methods: We demonstrate how misleading interpretations caused by inconsistent use of plant names might occur through the course of vegetation analysis, from relevés upward through databases, and then to the final analyses. We discuss how these problems might be minimized.

Results: We highlight the importance of taxonomic reference lists for standardizing plant names and outline standards they should fulfill to be useful for vegetation databases. Additionally, we present the R package *vegdata*, which is designed to solve name-related problems that arise when analysing vegetation databases.

Conclusions: We conclude that by giving more consideration to the appropriate application of plant names, vegetation scientists might enhance the reliability of analyses obtained from large vegetation databases.

Keywords: Biodiversity informatics; Checklist; Plant nomenclature; Potential taxon; R package; Species; Synonym; Taxonomic interpretation; Taxon view; *Vegdata*.

Nomenclature: As long as not stated otherwise the nomenclature follows Jansen & Dengler (2008).

Jansen, F. (corresponding author, jansen@uni-greifswald.de; URL <http://www.botanik.uni-greifswald.de/jansen.html>): Institute of Botany and Landscape Ecology, University of Greifswald, Grimmer Straße 88, 17487 Greifswald, DE.

Dengler, J. (dengler@botanik.uni-hamburg.de): Biodiversity, Evolution and Ecology of Plants, Biocentre Klein Flottbek and Botanical Garden, University of Hamburg, Ohnhorststraße 18, 22609 Hamburg, DE.

Introduction

Vegetation databases that contain species co-occurrence data have been compiled all over the globe. It is estimated that in Europe alone nearly two million phytosociological relevés are stored electronically (Schaminée et al. 2009), and similar databases are emerging outside Europe (Ewald 2001). This legacy of plot data spanning 100 years is a powerful tool for vegetation science (Ewald 2003). These databases open new avenues for analysis, from classical synthetic classification to predictive mapping and tests of fundamental ecological hypotheses regarding functional traits, assembly rules and biodiversity patterns (Dengler et al. 2008). They potentially offer many options for analysing patterns and processes of global change caused by anthropogenic climate warming, land-use changes and biotic invasions (Schaminée et al. 2009).

Despite the immense potential of vegetation databases, applications beyond consistent large-scale vegetation classification (e.g. Schaminée et al. 1995; Berg et al. 2004; Chytrý 2007) are still limited (e.g. Chytrý et al. 2008). This discrepancy can probably be attributed to a lack of awareness about the high potential of this resource but also to methodological problems resulting from the use of data that originate from numerous and heterogeneous sources (Ewald 2003). Some of the major problems have been addressed recently. At least partial solutions have been suggested concerning, e.g. geographic bias of relevés (e.g. Knollová et al. 2005), non-random sampling (e.g. Botta-Dukát et al. 2007; Roleček et al. 2007), varying completeness of relevés (e.g. Chytrý 2001) or varying plot sizes and their effect on the perception of differences in floristic composition (e.g. Otýpková & Chytrý 2006; Dengler et al. 2009).

By contrast, little acknowledgement has been given to problems associated with the use of plant names in vegetation databases. Actually, consistent application of “research object” names is an indispensable step in any analysis of biotic data. This is particularly true when combining plot data from different sources. In fact, the procedure of normalizing plant name usage can become one of the most

time-consuming processes among all the methods. The problems we aim to address in this contribution arise from the fact that plant taxonomy is a dynamic discipline (Stuessy 2009). A consequence of this constant taxonomic flux and individual idiosyncrasy of taxonomic understanding is that the thousands of floras and checklists in use worldwide are seldom congruent in their taxonomy.

On the one hand, the same taxon name may refer to a different entity (taxonomic homonyms, misapplied names or different delimitations of a certain taxon). On the other hand, different names might be applied to the same entity (illegitimate names, homotypic or heterotypic synonyms). Plant taxonomists first addressed the consequences of such complicated and dynamic relationships between different names and the entities they are applied to. To reflect this complexity adequately, a new concept, named “potential taxon” (Berendsohn 1995), “taxon view” (Zhong et al. 1996) or “taxonym” (Koperski et al. 2000), has been developed, as well as data models (Berendsohn et al. 2003; Kennedy et al. 2006; Thau & Ludäscher 2007; Franz & Peet 2009) and database concepts (Berendsohn 1997). The implementation of a similar concept for vegetation databases has only recently started (Vegbank: <http://www.vegbank.org/vegdocs/design/planttaxaoverview.html>; VegetWeb: Ewald et al. 2006). However, we will show that these implementations can only partly solve the prevailing problems, and we should acknowledge that e.g. in Europe at least 60% of electronically available relevés are stored without a taxon view link (Schaminée et al. 2009).

Focusing on vegetation databases, the aim of this paper is to: show how inconsistent application of plant names can lead to biased analyses; propose ways of minimizing such shortcomings; and demonstrate how our ideas can be put into practice.

When Inconsistent Use of Plant Names Cause Problems

Problems caused by inconsistent use of plant names arise in floristic or vegetation databases as a result of the accumulation of data from hundreds or thousands of researchers (with their regional and individual idiosyncrasies). For example, Schaminée et al. (2009) state that in order to establish a joint European vegetation database (SynBioSys Europe), 30 national species lists with 300 000 names (or rather: taxon views) have to be united. The effects on analyses are likely to be profound, but the methods and results in published vegetation studies will not allow the reader to detect any artifice, or make it

possible to trace the problem back to its source in the vegetation database. Below, we highlight these difficulties using three arbitrarily chosen analyses of floristic and vegetation databases, in which effects of inconsistent use of plant names are apparent.

Using names that are interpreted differently in their extent by various authors is probably the most frequent source of problems in floristically well-studied regions. For instance, Benkert et al. (1996), referring to the taxonomic concept of Bäßler et al. (1996), presented distribution maps of the *Festuca ovina* aggregate and its micro-species in E Germany. According to these maps, *F. ovina* and *F. brevipila* would be equally frequent in lowland habitats. The latter species, however, is much more characteristic of this region than the former (e.g. Dengler 1994). The mapping error likely did not result from misidentifications but rather from different delimitations of “*Festuca ovina*” among the surveyors who contributed data to the atlas of Benkert et al. (1996). For example, Ascherson (1864) accepts only one species, *F. ovina*, with several infraspecific taxa (including *F. ovina* subsp. *duriuscula sensu* Ascherson = *F. brevipila* of modern floras), this taxon view corresponds to *F. ovina* agg. in most modern classifications. When compiling data from hundreds of field researchers, one has to account for the fact that at least these three different wide concepts of the name *F. ovina* are still regularly used in Germany (see Fig. 1). Thus, the assignment of all records with field name “*F. ovina*” to the *F. ovina* map in the atlas might not have been correct based on the taxon view adopted there.

Mahecha & Schmidlein (2008) analysed species distribution patterns, based on floristic grid mapping data from Germany (FlorKart), with multidimensional ordination methods. The authors state that a significant proportion of the patterns they found can be attributed to biases caused by different sampling intensities, as well as to regional differences in taxonomic concepts. They highlight the distribution of “*Tripleurospermum perforatum*” versus “*T. maritimum*” in the FlorKart database, which is highly correlated to biogeographic patterns. The first species seemingly occurs in nearly all grid cells of W Germany, Berlin and Saxony, but is absent from the rest of the former German Democratic Republic (GDR). The second seems to be highly frequent in Thuringia, scattered in the rest of the former GDR, and nearly absent from W Germany. Actually, *T. maritimum sensu* Wisskirchen & Haeupler 1998 is a coastal species that does not occur in Thuringia at all, while *T. perforatum sensu* Wisskirchen & Haeupler 1998 is very common in all German federal states (Jäger & Werner 2005).

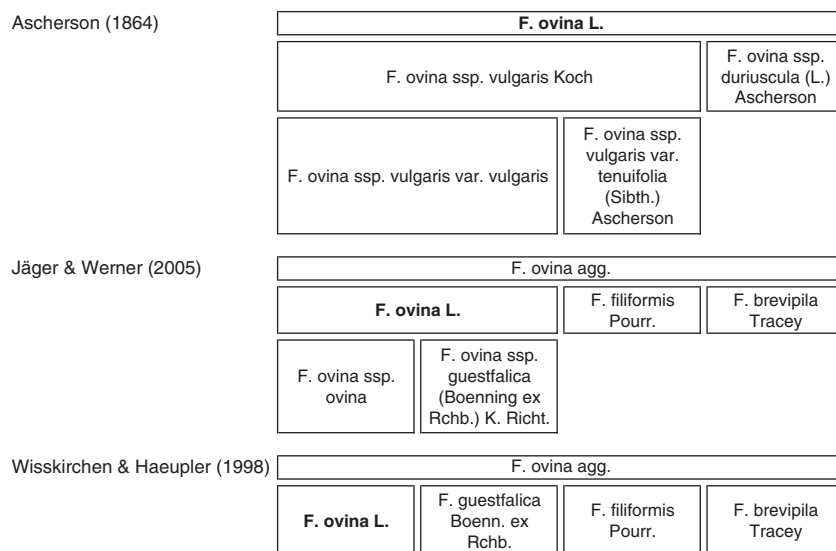


Fig. 1. Comparison of different taxon views as exemplified using *Festuca ovina* L. All sources refer correctly to the same name, author and type specimen of *Festuca ovina*, but represent different views of the circumscription of this taxon. All names can be mapped unambiguously into a single reference list, when referring not only to the author but also to the source of taxonomic interpretation (flora, commented checklist).

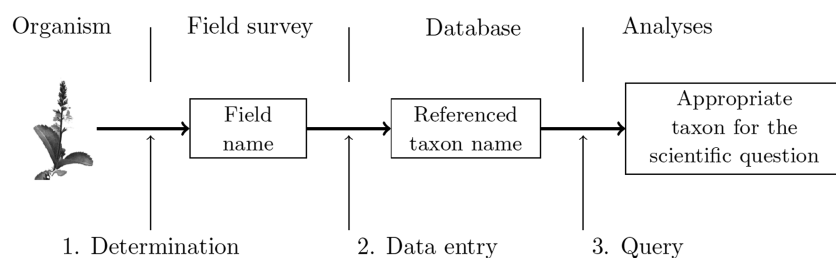


Fig. 2. Necessary steps for taxonomic interpretation. The presented R package *vegdata* can only solve problems occurring at step number three (see text). Additionally, it draws attention to possible mistakes during data entry.

Thus, we witness two problems. Similar to *Festuca ovina*, the name *T. maritimum* is frequently applied in a sense deviating from FlorKart, i.e. including *T. perforatum* (*sensu* Wisskirchen & Haeupler 1998) as *T. maritimum* subsp. *inodorum*. Further, the near complete “absence” of both taxa from NE Germany is caused merely by the fact that they have been mapped at the aggregate level in that area (Benkert et al. 1996).

In our third example, Wamelink et al. (2005) modelled species response curves against pH for 556 taxa. In several cases, both infraspecific taxa and the corresponding species have been modelled separately, with deviating results for the mean response (e.g. *Erodium cicutarium* subsp. *cutarium*: pH = 7.4; *Erodium cicutarium*: pH = 7.0). Modelling species in addition to subspecies can be reasonable when more than one subspecies is involved, but would require including all the data of the sub-

ordinate taxa at the higher level. Further, we can assume that in their study area each of the species occurred with only one subspecies (i.e. they were “regionally monotypic”). Thus, while being two valid taxa of different rank, the biological “content” of the species and the subspecies was identical. The authors seemingly modelled the “names” in the database, neglecting their hierarchical relationships.

To exemplify the extent of the named problems, we provide a (non-exhaustive) list of critical names from the German flora in Appendix S1.

Steps of Taxonomic Interpretation

Three steps of taxonomic interpretation are essential when moving from the real-world object to the results of scientific research (Fig. 2). In the fol-

lowing, we discuss the difficulties and necessary methods in each step.

Field survey and species determination

The first step of taxonomic interpretation occurs in the field, when recording relevé data (Fig. 2). Many studies have demonstrated that even experienced surveyors can typically overlook or misidentify a fraction of the plants present (e.g. Tüxen 1972; Scott & Hallam 2003; Archaux et al. 2006; Vittoz & Guisan 2007). The problems arising at this point are not within the focus of this article, but shortfalls at this first step can hardly be re-addressed at a later stage of database entry or preparation for analysis. In this context, it appears that students in vegetation ecology are increasingly less trained in handling taxonomy-related issues and making proper use of floras or checklists. The taxonomic education at universities should thus be strengthened to match the demands for high quality primary data.

Surveyors should document their references for determination and naming for every single taxon, i.e. it is important to disclose the author of the taxonomic concept (who is not identical to the author of the taxon name) when the determination or naming deviates from the identification guide that is otherwise followed in the study. In many cases (particularly in botanically less well studied regions of the Earth), it is essential to collect voucher specimens of critical plant taxa and to quote the herbarium where these have been deposited. Additionally, we recommend further explanation of the uncertainty of determination by specifying at which taxonomic level the uncertainty is located (cf. *Festuca ovina*, *F. cf. ovina* agg. or *F. cf. ovina*) or which is the next higher taxon known with certainty. Further, it would be useful to indicate the degree of uncertainty on a rough ordinal scale (“certain”, “with some uncertainty”, “uncertain, but best guess”).

Data entry

The second step is the data entry into an electronic list or database (Fig. 2). Widely used vegetation database programs such as TURBOVEG (Hennekens & Schaminée 2001) do not provide a field for the entry of information on the preciseness of plant identification as default. Accordingly, information is often lost, be it by replacing “*Festuca cf. ovina*” with *F. ovina*, by assigning it to some superior taxon, or by omitting it altogether. Instead,

we suggest retaining information on lack of certainty and leaving decisions on how to deal with such entries to the step involving preparation for analysis (see Appendix S2).

More troublesome are problems caused by different meanings of the same name. The problem cannot be avoided by publishing the taxon names together with their authority. *Festuca ovina*, for example, bears the same correct author citation “L.” (for C. Linnaeus) irrespective of whether a narrow or a wide taxon view is applied (see Fig. 1). Accordingly, the taxonomic concept must be disclosed, whether by citing the flora used or by describing the taxonomic concept. However, there are only a few systems, e.g. Vegbank (<http://www.vegbank.org/vegdocs/design/planttaxaoverview.html>) and VegetWeb (Ewald et al. 2006, <http://www.planto.de/OekoArt/ModellLog.php>), that allow the referencing of each plant name to a source, but even these databases actually currently work with only one preferred taxonomic view. BIOTABase (http://www.biota-africa.de/biotabase_ba.php) is another example of a database system designed to map the relationships between field names and accepted names in a very flexible way, together with links to voucher specimens. This software has been designed to account for the absence of checklists useful for standardizing the use of names of African plants, but its approach is useful for other regions as well.

While we strongly support the application of the “taxon view” approach in vegetation databases, i.e. storing both the original plant name and a link to a reference that defines its content, we are sceptical as to whether this alone would solve the “nomenclatural confusion” for such databases. In many published sources of vegetation relevés, no explicit references to floras or checklists are made, or the general declaration is not reliable for all species. Further, names are often used that do not exist in the source (“*Taraxacum officinale*” is probably one of the most frequently occurring names in vegetation relevés worldwide, despite the fact that many recent floras reject such a taxon), or names are used in an aberrant mode from the source (e.g. *Festuca ovina* in the sense of *F. ovina* agg.).

Thus, we suggest that a taxonomic interpretation of names, together with documentation of the decision, should take place instead of an automatic assignment of names. While often the assignment to a higher taxonomic level is required, in other cases, *a-posteriori* application of a subordinate name might be sensible. For example, *Silene latifolia* has four accepted subspecies in Europe (Tutin et al. 1993), of which only subspp. *alba* occurs in Germany

(Wisskirchen & Haeupler 1998). Thus, researchers in Germany often do not note the subspecies because they consider this information superfluous. Thus, in the data entry step, it is both reasonable and useful to re-assign any German field record of *S. latifolia* to *S. latifolia* subsp. *alba*. When combining databases at a supranational level, the infra-specific classification becomes relevant, and the German data are already correctly assigned to subspecies. These examples show that it is important to have the digitized data “verified” by an experienced botanist before they are added to a larger database.

Expert knowledge is also required when combining datasets based on different taxonomic reference lists, as is typical for international collaborations. The correct assignment of names and concepts can be a very time-consuming process, but is essential for credible results. In the software package *vegdata*, the function *tv.compRefl* is included, which compares taxon numbers and/or taxon names of different TURBOVEG reference lists as a starting point (see Appendix S3).

Preparation for analyses

For the third step, the stored data have to be prepared for the needs of the scientific analyses (Fig. 2). If the interpretations suggested for steps 1 and 2 remain unresolved, they have to be dealt with now – which obviously will be more complicated and error-prone because the necessary information about taxon views is usually not available. In any case, the outcome should be the correct fusion of synonyms and homonyms into accepted names of the chosen taxonomic reference. If alternative taxon views are available in the vegetation database, this would be the point to select one. To ensure sound and consistent results, four further decisions are necessary prior to the analyses, depending on the type of study.

- (1) A choice has to be made as to the use of *uncertain determinations* (often indicated with the abbreviation “cf.” in field names), with two options: acceptance as a confident determination or aggregation at next highest taxonomic level (then see 4).
- (2) *Nested taxa* may make sense in some studies and not in others. For example, one could model separate response curves along a gradient for a species and its subspecies, with the ecological amplitude of the latter probably being narrower. By contrast, counting a species in addition to its subordinate taxa is unacceptable when species richness is analysed. In both cases, however, all data of

subordinate taxa have to be included in the analysis of the higher-rank taxon.

- (3) Also, for the *taxonomic level* of the analyses, there are two options: one may choose (i) the “terminal” taxa (i.e. the lowest-ranked accepted taxa with available data, which in one case may be a species and in another a subspecies or variety), or (ii) a uniform taxonomic rank (e.g. species).

- (4) Further, one has to decide what to do with taxon records available only at *coarser taxonomic levels* than desired (e.g. genus level): either combining the more precise data to the higher level (see 2) or excluding the coarse-scale data from the analyses.

For none of these four decisions is there a general right or wrong method. Instead, we face a trade-off between different solutions, each of which necessarily introduces additional uncertainty to the data. If the introduced uncertainty is too big, the exclusion of the whole relevé from the analyses might be the better choice.

In order to apply the principles presented in our paper to real datasets and to automate the process of taxonomic interpretation, the first author has written scripts for the statistical environment R (R Development Core Team, Vienna, Austria). These are capable of solving problems occurring at the last step, but also assists with checking results of the second (and first) step. The scripts are included in the package *vegdata* (version 0.2, <http://cran.r-project.org/package=vegdata>), whose present version supports data access from TURBOVEG. In Appendix S2 we present an example session with a commentary.

Taxonomic Reference Lists – A Necessary Prerequisite

As outlined above, reference to taxonomic lists is needed during all three interpretative steps (Fig. 2). These lists could be (i) floras, which define the meaning of plant names by use of keys and descriptions, (ii) checklists (providing reference to species belonging to the current flora of a region) or (iii) specialized reference lists for survey data. They may simply present a preferred taxon view, potentially accompanied by synonyms (e.g. Ehrendorfer 1973), but in a more useful variant, they additionally outline the links between the preferred and other taxon views (see Wisskirchen & Haeupler 1998; Koperski et al. 2000). Such lists are the “backbones” of all databases containing information related to plant taxa (e.g. museum collections, distribut-

ion, taxonomy, vegetation, traits) and for the connection of different types of such databases. However, floras and checklists are typically developed according to other goals, and so they seldom fully meet the needs of survey databases.

In order to be useful for taxonomic interpretation in the framework of vegetation databases, a taxonomic reference list for survey data should fulfill the following requirements:

(1) *Use of taxon views* to link the present taxonomic interpretation to the views of major floras, other reference works (e.g. Red Lists) and previous editions of the reference list.

(2) *Unambiguous recording of different taxon views.* All widely applied names used in the past should find a direct counterpart in a new reference list, instead of referring to several names *pro parte* or only partially referring to one higher-rank taxon. For example, if *Tripleurospermum maritimum* in the sense of earlier floras is split into *T. maritimum* and *T. perforatum* in the reference list, then *T. maritimum* of the older sources must be included in the new reference list (e.g. as *T. maritimum* agg.) and all records must be mapped to it.

(3) *Inclusion of hybrids, non-naturalized neophytes and frequently cultivated plants.* Most checklists purposefully exclude these three groups from listing, since in statistical evaluations of floras, such taxa normally should not be counted. Nevertheless, taxonomic reference lists for survey data should provide reliable nomenclatural data for all taxa occurring in the vegetation.

(4) *Hierarchical levels.* All taxa should be assigned to hierarchical levels unambiguously to make automatic data checks and taxonomic refinement possible and to allow analyses at different taxonomic resolutions.

(5) *Information on regionally monotypic taxa.* When a species is present in a region with only one infraspecific taxon, it is necessary to list which infraspecific taxon is present, even if it is the nominal taxon (e.g. *Silene latifolia* subsp. *alba*).

While points 1-5 apply to printed lists, additional considerations become relevant when implementing reference lists as an electronic tool within the framework of vegetation databases (e.g. *GermanSL*, see Jansen & Dengler 2008):

(6) *Joint list for all "plant" taxa.* The use of vegetation databases requires that reference lists of different "plant" taxa (vascular plants, bryophytes, lichens, macroscopic algae, macroscopic cyanobac-

teria), which are usually covered in separate printed lists, are combined in one electronic list.

(7) *Continuous update and versioning.* Unlike print versions, the electronic format allows frequent updates to accommodate newly recorded plants and new taxonomic views. This is necessary and highly useful, but sequential versioning is necessary to unambiguously link to previously used taxonomies and nomenclature.

(8) *Comprehensive documentation.* The electronic reference list should either completely match the printed version, or all deviations must be disclosed unambiguously.

(9) *Unique and unambiguous names* with clear reference to the represented taxon view. To be able to convey the information from the original source with an unambiguous name, we urge the use of abbreviations such as "s.str." or "s.l." for differently wide species concepts and "auct. non" for misapplications in those, but only in those cases. Of course, the source of the applied taxon view ("secundum") has to be referenced in additional fields of the list.

(10) *Warning about critical taxa:* It would be useful if database programs provide warnings during data entry or merging of different databases when "critical" taxon names are involved that can be misapplied (i.e. when there are variants of the same taxonym with and without "s.l.", "s.str.", "agg." or "auct. non" or if monotypic species occur), and require a specific decision/confirmation of the researcher for such assignments.

Conclusions

Within the rapidly evolving field of biodiversity informatics (see Canhos et al. 2004; Jones et al. 2006), vegetation databases can contribute as a comprehensive and informative data source, as they contain not only many millions of records of individual species but also combine these with information on species composition, vegetation structure and environmental conditions in a spatially and temporally explicit manner. Tapping the full potential of vegetation databases currently remains seriously hindered by unappreciated nomenclatural problems that arise when combining data from different sources. In solving these problems, the "taxon view" approach developed for organismic taxonomy (e.g. Berendsohn 1995; Zhong et al. 1996; Koperski et al. 2000) is a welcome contribution. However, vegetation scientists should be aware that the situation for their data is even more

complicated because field records normally are not connected to voucher specimens that would allow a later correction of the identification. While reference lists are a central tool for the correct use of plant names in vegetation databases, their utility is often limited because their layout does not cover the requirements of the users. In conclusion, vegetation scientists have to accept that their data will always contain some uncertainty, but with a taxonomic reference list layout, as proposed in this contribution, and appropriate tools like the R package *vegdata* shown in Appendices S2 and S3, they should be able to avoid flawed results that would ensue from inconsistent use of plant names. Finally, also editors and reviewers of journals should put more emphasis on taxonomic accuracy in ecological and biodiversity articles and allow for – even long if need be – taxon lists in research articles.

Acknowledgements. We thank Erwin Bergmeier, Curtis Björk, Zoltán Botta-Dukát, Manfred Finckh, Michael Manthey and Gerhard Muche for useful comments on former versions of the manuscript. Curtis Björk also polished our English usage.

References

- Archaux, F., Gosselin, F., Bergès, L. & Chevalier, R. 2006. Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *Journal of Vegetation Science* 17: 299–306.
- Ascherson, P. 1864. *Flora der Provinz Brandenburg, der Altmark und des Herzogthums Magdeburg*. Hirschwald, Berlin, DE.
- Bäbler, M., Jäger, E. & Werner, K. (eds.). 1996. *Exkursionsflora von Deutschland. Band 2: Gefäßpflanzen: Grundband*. 16th ed. Fischer, Jena, DE.
- Benkert, D., Fukarek, F. & Korsch, H. (eds.) 1996. *Verbreitungsatlas der Farn- und Blütenpflanzen Ostdeutschlands*. Fischer, Jena, DE.
- Berendsohn, W.G. 1995. The concept of “potential taxa” in databases. *Taxon* 44: 207–212.
- Berendsohn, W.G. 1997. A taxonomic information model for botanical databases: the IOPI Model. *Taxon* 46: 283–309.
- Berendsohn, W.G., Döring, M., Geoffroy, M., Glück, K., Güntsch, A., Hahn, A., Kusber, W.-H., Li, J., Röpert, D. & Specht, F. 2003. The Berlin Model: a concept-based taxonomic information model. In: Berendsohn, W.G. (ed.) *MoReTax – handling factual information linked to taxonomic concepts in biology [Schriftenreihe für Vegetationskunde no. 39]*. pp. 15–42. Federal Agency for Nature Conservation, Bonn, DE.
- Berg, C., Dengler, J., Abdank, A. & Isermann, M. (eds.). 2004. *Die Pflanzengesellschaften Mecklenburg-Vorpommerns und ihre Gefährdung – Textband*. Weissdorn, Jena, DE.
- Botta-Dukát, Z., Kovács-Láng, E., Rédei, T., Kertész, M. & Garadnai, J. 2007. Statistical and biological consequences of preferential sampling in phytosociology: theoretical considerations and a case study. *Folia Geobotanica* 42: 141–152.
- Canhos, V.P., Souza, S., Giovanni, R. & Canhos, D.A.L. 2004. Global biodiversity informatics: setting the scene for a “new world” of ecological modelling. *Biodiversity Informatics* 1: 1–13.
- Chytrý, M. 2001. Phytosociological data give biased estimates of species richness. *Journal of Vegetation Science* 12: 439–444.
- Chytrý, M. (ed.). 2007. *Vegetation of the Czech Republic – 1. Grassland and heathland vegetation [in Czech, with English summary]*. Academia, Praha, CZ.
- Chytrý, M., Maskell, L.C., Pino, J., Pyšek, P., Vilà, M., Font, X. & Smart, S.M. 2008. Habitat invasions by alien plants: a quantitative comparison among Mediterranean, subcontinental and oceanic regions of Europe. *Journal of Applied Ecology* 45: 448–458.
- Dengler, J. 1994. Flora und Vegetation von Trockenrasen und verwandten Gesellschaften im Biosphärenreservat Schorfheide-Chorin. *Gleditschia* 22: 179–321.
- Dengler, J., Chytrý, M. & Ewald, J. 2008. Phytosociology. In: Jørgensen, S.E. & Fath, B.D. (eds.) *Encyclopedia of ecology*. pp. 2767–2779. Elsevier, Oxford, UK.
- Dengler, J., Löbel, S. & Dolnik, C. 2009. Species constancy depends on plot size – a problem for vegetation classification and how it can be solved. *Journal of Vegetation Science* 20: 754–766.
- Ehrendorfer, F. (ed.) 1973. *Liste der Gefäßpflanzen Mitteleuropas*. 2nd ed. Fischer, Stuttgart, DE.
- Ewald, J. 2001. Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. *Berichte der Reinhold-Tüxen-Gesellschaft* 13: 53–69.
- Ewald, J. 2003. A critique for phytosociology. *Journal of Vegetation Science* 14: 291–296.
- Ewald, J., May, R. & Kleikamp, M. 2006. VegetWeb – die pflanzensoziologische online-Datenbank unter www.floraWeb.de. In: Berg, C., Bergmeier, E., Hövelmann, T. & Ristow, M. (eds.) *Ein Netzwerk für botanischen Naturschutz – Neue Herausforderungen für die Botanikerinnen und Botaniker Deutschlands [BfN-Skripten no. 178]*. pp. 127–131. Bundesamt für Naturschutz, Bonn, DE.
- Franz, N.M. & Peet, R.K. 2009. Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity* 7: 5–20.
- Hennekens, S. & Schaminée, J. 2001. TURBOVEG, a comprehensive data base management system for vegetation data. *Journal of Vegetation Science* 12: 589–591.
- Jäger, E.J. & Werner, K. (eds.) 2005. *Exkursionsflora von Deutschland. Band 4: Gefäßpflanzen: Kritischer Band*. 10th ed. Spektrum Akademischer Verlag, München, DE.

- Jansen, F. & Dengler, J. 2008. GermanSL – eine universelle taxonomische Referenzliste für Vegetationsdatenbanken. *Tuexenia* 28: 239–253.
- 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.
- Kennedy, J., Hyam, R., Kukla, R. & Paterson, T. 2006. A standard data model representation for taxonomic information. *OMICS* 10: 220–230.
- Knollová, I., Chytrý, M., Tichý, L. & Hájek, O. 2005. Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *Journal of Vegetation Science* 16: 479–486.
- Koperski, M., Sauer, M., Braun, W. & Gradstein, S.R. 2000. *Referenzliste der Moose Deutschlands – Dokumentation unterschiedlicher taxonomischer Auffassungen [Schriftenreihe für Vegetationskunde no. 34]*. Bundesamt für Naturschutz, Bonn, DE.
- Mahecha, M.D. & Schmidlein, S. 2008. Revealing biogeographical patterns by nonlinear ordinations and derived anisotropic spatial filters. *Global Ecology and Biogeography* 17: 284–296.
- Otápková, Z. & Chytrý, M. 2006. Effects of plot size on the ordination of vegetation samples. *Journal of Vegetation Science* 17: 465–472.
- Roleček, J., Chytrý, M., Hájek, M., Lvončík, S. & Tichý, L. 2007. Sampling design in large-scale vegetation studies: do not sacrifice ecological thinking to statistical purism!. *Folia Geobotanica* 42: 199–208.
- Schaminée, J.H.J., Stortelder, A.H.F. & Weesthoff, V. (eds.) 1995. *De Vegetatie van Nederland – Deel 1. Inleiding tot de plantensociologie – grondslagen, methoden en toepassingen*. Opulus, Uppsala, SE.
- Schaminée, J.H.J., Hennekens, S.M., Chytrý, M. & Rodwell, J.S. 2009. Vegetation-plot data and databases in Europe: an overview. *Preslia* 81: 173–185.
- Scott, W.A. & Hallam, C.J. 2003. Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology* 165: 101–115.
- Stuessy, T.F. 2009. *Plant taxonomy: the systematic evaluation of comparative data*. 2nd ed. Columbia University Press, New York, NY, US.
- Thau, D. & Ludäscher, B. 2007. Reasoning about taxonomies in first-order logic. *Ecological Informatics* 2: 195–209.
- Tüxen, R. 1972. Kritische Bemerkungen zur Interpretation pflanzensoziologischer Tabellen. In: Tüxen, R. (ed.) *Grundfragen und Methoden in der Pflanzensoziologie [Bericht über das Internationale Symposium der Internationalen Vereinigung für Vegetationskunde 1970]*. pp. 168–182. Junk, Den Haag, NL.
- Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M. & Webb, D.A. (eds.) 1993. *Flora Europaea – Volume 1: Psilotaceae to Platanaceae*. 2nd ed. Cambridge University Press, Cambridge, UK.
- Vittoz, P. & Guisan, A. 2007. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science* 18: 413–422.
- Wamelink, G.W.W., Goedhart, P.W., Van Dobben, H.F. & Berendse, F. 2005. Plant species as predictors of soil pH: replacing expert judgement with measurements. *Journal of Vegetation Science* 16: 461–470.
- Wisskirchen, R. & Haeupler, H. 1998. *Standardliste der Farn- und Blütenpflanzen Deutschlands*. Ulmer, Stuttgart, DE.
- Zhong, Y., Jung, S., Pramanik, S. & Beaman, J.H. 1996. Data model and comparison and query methods for interacting classifications in a taxonomic database. *Taxon* 45: 223–241.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. List of “critical” plant names in the German flora, generated from the reference list *GermanSL*. Critical are the taxon names that might be mixed with equal names from other content, either homonyms or taxa with a smaller or wider extent regarding other taxon views.

Appendix S2. An exemplary session of data access and taxonomic interpretation from a TURBOVEG database with the R package *vegdata*, together with a commentary.

Appendix S3. An exemplary session from merging two reference lists with the R package *vegdata*, together with a commentary.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author of the article.

Received 30 November 2009;

Accepted 25 June 2010.

Co-ordinating Editor: Dr. Angelika Schwabe-Kratochwil.