Data management for the Braun-Blanquet project and the European Vegetation Archive

Borja Jiménez-Alfaro[1], Stephan Hennekens[2], Milan Chytrý[3] and the partners of the Braun-Blanquet project[4] and the European Vegetation Archive[5]

[1]*Masaryk University, Brno, Czech Republic, borja@sci.muni.cz;*
[2]*Alterra, Wageningen, The Netherlands, stephan. hennekens@wur.nl;*
[3]*Masaryk University, Brno, Czech Republic, chytry@sci.muni.cz;*
[4]*http://www.sci.muni.cz/botany/vegsci/braun_blanquet.php?lang=en*
[5]*http://euroveg.org/eva-database*

Introduction

The Braun-Blanquet project and the European Vegetation Archive are among the first initiatives for analyzing comprehensive datasets of vegetation plots in Europe (Jiménez-Alfaro et al. 2013). Both initiatives are based on the compilation of vegetation data from different collaborators, including national and regional databases and additional data from individual researchers or research groups or the literature. The management of this information is complex since it derives from heterogeneous sources and many different research contexts.

Here we report the conceptual management plan developed for merging European databases and for creating taxonomically consistent outputs to be used for vegetation analyses. The main aim is to develop an archive of data sets which can be regularly updated, allowing to create comprehensive matrices of species x plots, and ensuring that the databases are compatible in terms of species taxonomy and header data.

Storing data

The data sets are managed separately in Turboveg 2, a software program widely used for storing vegetation data in Europe (Hennekens & Schaminée 2001). Our general procedure is to preserve the original structure of the databases in order to facilitate regular updates from data providers.

Databases provided by partners of the Braun-Blanquet project or the European Vegetation Archive are in most cases linked to one of the species lists available for Turboveg 2, although in some cases they are linked to adhoc lists created by one or more authors for specific projects. As a general rule, we suggest data providers to use one of approximately 30 most commonly used European national or regional checklists. Accordingly, new digitized data are linked to these lists or to the general European checklist for Turboveg which is based on *Flora Europaea* (Tutin et al. 1993).

Header data are also very heterogeneous, and only a few fields (e.g. plot size, total cover and altitude) are regularly assigned to the plots in the databases. For the specific purposes of the Braun-Blanquet project (i.e. the characterization of phytosociological alliances), we prioritized the standardization of only three fields: plot size, geographical coordinates and vegetation or habitat type. However, a more ambitious system of header data harmonization will be created for the European Vegetation Archive, which is expected to provide data for many different purposes.

Combining data

We are using a prototype of Turboveg 3 (Figure 1) to combine species and header data from the original databases that are regularly managed in Turboveg 2. A copy of each of these databases is imported into Turboveg 3 from a single repository that is shared in GoogleDrive by the data managers. The general settings of Turboveg 3 are then fixed to link any version of the original databases. Thus further update of a given database with the same structure will be automatically integrated into the system.

Figure 1. General view of the main panel of Turboveg 3 prototype (version January 2014).

Figure 2. Cross-link species system of SynBioSys integrated in Turboveg 3 prototype (version January 2014).

The most important issue for combining the databases is to crosslink the various species checklists. We followed the general procedure developed for SynBioSys Europe (Schaminée et al. 2007) to create a crosslink between taxon concepts of different species checklists (Figure 2). On the one hand, species names from different checklists that fit at 100% are linked automatically and identified by the same alphanumeric code. On the other hand, species that are not matched must be linked manually to harmonize taxon concepts. This process is dynamic and can be continuously reviewed by data contributors under the supervision of a number of taxonomical authorities selected among regional experts. At the moment, more than 80% of the species included in 30 European checklists have been taxonomically harmonized, although more effort is still necessary to create formal guidelines for the harmonization of taxon concepts in SynBioSys Europe and Turboveg 3.

Under this system, we are able to perform queries in Turboveg 3 based on the presence or cover of a given species that is systematically checked in more than 40 individual databases. This allows us to create outputs in form of species x plot matrices including the associated header data for each plot (when existing). These outputs can be then used for performing analyses based on species composition (e.g. ordination or classification) or the properties of vegetation (e.g. distribution patterns of plots assigned to the same community type).

Further steps

Under the proposed data management plan, new functionalities of Turboveg 3 are being developed, and a more detailed procedure for managing European databases will be developed in the year 2014. Among the main priorities for the integration of vegetation databases into the Braun-Blanquet project, the European Vegetation Archive or any other initiative, we highlight the following:

- Quality control of the original datasets
- Feedback with data providers for improvement of header data
- Involvement of new databases from underrepresented regions
- Continuous updating of species crosswalks in SynBioSys Europe
- New functionalities for exporting output matrices and associated data in Turboveg 3
- Project-specific analyses at continental scale

References

Hennekens, S.M. & Schaminée, J.H.J. 2001. TURBOVEG, a comprehensive data base management system for vegetation data. *Journal of Vegetation Science* 12: 589–591.

Jiménez-Alfaro B. et al. (2013) Unifying and analyzing vegetation-plot databases in Europe: the European Vegetation Archive (EVA) and the Braun-Blanquet project. In: Walker, D.A., Breen, A.L., Raynolds, M.K. & Walker, M.D. (Ed). Arctic Vegetation Archive (AVA) Workshop, Krakow, Poland, April 14-16, 2013. *CAFF Proceedings Report #10*. Akureyri, Iceland.

Schaminée, J.H.J., Hennekens, S.M. & Ozinga, W.A. 2007. Use of the ecological information system SynBioSys for the analysis of large datasets. *Journal of Vegetation Science* 18: 463–470.

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. 1993. Flora Europaea, vol 1–5. Cambridge University Press, Cambridge, UK.